

Universidad Nacional de Misiones

Facultad de Ciencias Exactas, Químicas y Naturales

Tesis de Maestría en Tecnologías de la Información

**“Metodología para el desarrollo de *Datamarts* en
Universidades Nacionales del Consorcio SIU.
Caso de Estudio: Área Graduados de la FCEQyN”**

Autor: Lic. María Daniela Rolón

Director: Dr. Eduardo Zamudio

Co Director: Mgter. J. Germán Andrés Pautsch

Año 2021

Dedicatorias

A mi compañero de vida, por su amor incondicional.

A mis pequeñas hijas Delfina, Sofia y Abril, por sus abrazos y sonrisas regaladas para seguir adelante.

A mis padres, por haberme dado la oportunidad de estudiar y estar presentes en cada etapa de mi vida.

A mis hermanas, por acompañarme en cada desafío.

Para los que intentan encontrar el “para qué” de las cosas, dejando de lado los “por qué”

Resumen

La Universidad Nacional de Misiones (UNaM) ha implementado el módulo de consulta gerencial SIU-Wichi que utiliza Data Marts para el procesamiento de sus datos y la consulta de información. Los Data Marts que existen actualmente, proporcionados por el SIU, son poco representativos y no responden a todos los requerimientos propios de las instituciones académicas. En este trabajo final de maestría se propone el diseño e implementación de una metodología para el desarrollo de Data Marts, en universidades nacionales que utilizan como fuente de datos los sistemas provistos por el Consorcio SIU, que permitan cubrir las necesidades de información requeridas por las autoridades e interesados de una institución, con la participación activa de los mismos en cada una de sus etapas. Como caso de estudio y validación se toma el Área de Graduados de la Facultad de Ciencias Exactas, Químicas y Naturales de la UNaM. Se elaboró una metodología que integra las contribuciones de autores como Inmon, Kimball y Bernabeu, junto a técnicas y herramientas para la gestión de requisitos basadas en las buenas prácticas del modelo CMMI v1.3. Esta implementación ha permitido el modelado de un conjunto de Data Marts con indicadores claves para la gestión educativa que contribuyeron al proceso de toma de decisión en base a los requerimientos y políticas definidas en la UNaM.

Palabras claves: *data warehouses, data marts, OLAP, metodología, universidades nacionales, SIU*

Abstract

The National University of Misiones (UNaM) has implemented the management consultation module that uses Data Marts for the processing of its data and the access of information. The existing Data Marts, provided by the SIU, are unrepresentative and do not respond to all the own requirements of the academic institutions. In this final master's degree paper, it is proposed the design and implementation of a methodology for the development of Data Marts in National Universities using the systems provided by the SIU Consortium as a data source. Its purpose is to meet the information needs required by the authorities of an institution, with their active participation and the management of the requirements laid down. As a case study and validation, the Graduate Area of the Faculty of Exact, Chemical and Natural Sciences (FCEQyN) of the UNaM is taken. A methodology is developed that integrates the contributions of authors such as Inmon, Kimball and Bernabeu, along with techniques and tools for the management of requirements based on the best practices of the CMMI v1.3 model. This implementation has allowed the modeling of a set of Datamarts with key indicators for educational management, which contributed to the decision – making process based on the requirements and policies defined in the UNaM.

Keywords: data warehouses, datamarts, OLAP, methodology, national universities, SIU.

Agradecimientos

A mi Director y Co Director de Tesis por brindarme sus conocimientos y guiarme durante todo el proceso de investigación.

A las autoridades de la Facultad de Ciencias Exactas, Químicas y Naturales por permitirme realizar el presente trabajo.

Al personal de la Dirección de Enseñanza de la FCEQyN, principalmente a su directora, por el acompañamiento brindado.

Al coordinador de la Oficina de Graduados de la FCEQyN por su buena predisposición y participación durante el desarrollo del presente trabajo.

Al equipo técnico de la Dirección Área de Informática y Comunicaciones de la FCEQyN por el apoyo de siempre.

A todos los docentes de la Maestría en Tecnologías de la Información por contribuir a nuestra formación profesional.

Índice

1.	Introducción	15
1.1	<u>Descripción del problema y justificación</u>	15
1.2	<u>Objetivos</u>	16
1.3	<u>Elección del caso</u>	17
1.4	<u>Plan metodológico</u>	18
1.5	<u>Área de conocimiento disciplinar específico</u>	18
2.	Marco Teórico	22
2.1	Sistemas de Soporte para la Toma de Decisión	22
2.1.1	Inteligencia de Negocios	22
2.1.2	Proceso de Inteligencia de Negocios	23
2.1.3	Beneficios de la Inteligencia de Negocios	24
2.2	Data Warehouse	25
2.2.1	Arquitectura de un Data Warehouse	26
2.2.2	Data Mart	28
2.2.3	Dimensiones	29
2.2.4	Tabla de Hechos	30
2.2.5	Esquema de Modelado Multidimensional	30
2.2.6	Operaciones	32
2.2.7	Tipos de implementación de Data Warehouses	33
2.2.8	Modelos para el desarrollo de Data Warehouses	34
2.2.8.1	Top-down (Inmon)	34
2.2.8.2	Bottom-up (Kimball)	35
2.2.8.3	HEFESTO	39
2.3	Ingeniería de Requerimientos	42
2.3.1	Evolución de los Requisitos	43
2.3.2	Proceso de Ingeniería de Requisitos	43
2.3.3	Modelos de Ingeniería de Requisitos	45
2.4	Antecedentes	48
3	Situación Actual	52
3.1	Módulo de información gerencial: SIU-Wichi	52

3.2 Proceso utilizado por el SIU para el desarrollo de Data Marts	53
3.3 Herramientas utilizadas por el SIU para el desarrollo de Data Marts	56
3.4 Arquitectura actual utilizada por el SIU para el desarrollo de Data Marts	57
3.5 Evaluación de la situación actual	58
4. Metodología propuesta para el desarrollo de Data Marts en Universidades Nacionales	62
4.1 Etapas de la metodología	62
4.1.1 E1: Entendimiento del Negocio	64
4.1.2 E2: Relevamiento y Análisis	71
4.1.3 E3: Diseño y Modelado Multidimensional	75
4.1.4 E4: Procesos ETL	79
4.1.5 E5: Validación del modelo final	83
5. Validación de la metodología en el caso de estudio propuesto	86
5.1 La Facultad de Ciencias Exactas, Químicas y Naturales	86
5.1.1 La Oficina de Vinculación con los Graduados	87
5.1.2 Módulo de Gestión de Encuestas: SIU-Kolla	87
5.1.3 Problemas identificados	89
5.2 Circuito administrativo propuesto para la completitud de la Encuesta al Recién Graduado	89
5.2.1 Sitio web para Graduados de la FCEQyN	91
5.3 Implementación de la Metodología propuesta	92
5.3.1 E1: Entendimiento del Negocio	92
5.3.2 E2: Relevamiento y Análisis	107
5.3.3 E3: Diseño y Modelado Multidimensional	122
5.3.4 E4: Procesos ETL	127
5.3.5 E5: Validación del modelo final	148
6. Resultados	152
7. Discusión	203
8. Conclusiones	206
8.1 Principales contribuciones del trabajo	206
8.2 Futuras líneas de investigación	208
Bibliografía	210
Anexo 1	217
Anexo 2	221
Anexo 3	222

Anexo 4	231
Glosario de acrónimos	233

Índice de Figuras

Figura 1: Fases del proceso de Inteligencia de Negocios [Elaboración propia]	24
Figura 2: Proceso de Extracción, Transformación y Carga de datos. [Elaboración propia]	27
Figura 3: Arquitectura de un almacén de datos [Elaboración propia]	28
Figura 4: Modelo multidimensional [Elaboración propia].	29
Figura 5: Esquema Estrella [Elaboración propia].....	31
Figura 6: Esquema Copo de Nieve [Elaboración propia].	31
Figura 7: Esquema Constelación [Elaboración propia].	32
Figura 8: Modelo ROLAP y MOLAP [18].....	34
Figura 9: Diagrama del Ciclo de vida de Kimball [13]	39
Figura 10: Etapas de la metodología Hefesto [6]	41
Figura 11: Actividades de la Ingeniería de Requisitos [24].....	44
Figura 12: Actividades de la Ingeniería de Requisitos frente a la evolución del universo de discurso [27].....	44
Figura 13: Procesos utilizados por el SIU para el desarrollo de Data Warehouses [Elaboración propia]	56
Figura 14: Arquitectura utilizada por el SIU para el desarrollo de Data Marts [Elaboración propia].	57
Figura 15: Mapa de implementaciones del SIU-Wichi en el territorio de la República Argentina [46].....	58
Figura 16: Modelo Conceptual preliminar [6].....	68
Figura 17: Actividades de la Etapa 1 [Elaboración propia].....	70
Figura 18: Modelo Conceptual ampliado [3].....	73
Figura 19: Actividades de la Etapa 2 [Elaboración propia].....	75
Figura 20: Diseño de una tabla de dimensiones [3].....	76
Figura 21: Diseño de una tabla de hechos [3].....	77
Figura 22: Diseño modelo físico [3].....	78
Figura 23: Actividades de la Etapa 3 [Elaboración propia]	79
Figura 24: Arquitectura ETL / DM propuesta [Elaboración propia]	80
Figura 25: Actividades de la Etapa 4 [Elaboración propia]	82
Figura 26: Actividades de la Etapa 5 [Elaboración propia].....	84
Figura 27: Comprobante de completitud de la Encuesta.	90
Figura 28: Circuito administrativo propuesto para la aplicación de la Encuesta al Recién Graduado. [Elaboración propia]	91
Figura 29: Modelo Conceptual preliminar: TA01-Situación laboral de los graduados por carrera según su distribución geográfica [Elaboración propia]	98
Figura 30: Modelo Conceptual preliminar: TA02- Valoración académica de los graduados por carrera [Elaboración propia]	101
Figura 31: Modelo Conceptual preliminar: TA03-Necesidad de capacitación de los graduados por carrera según su distribución geográfica [Elaboración propia].....	103
Figura 32: Estructura de la base de datos SIU-Kolla [61]	108
Figura 33: Módulo ubicación geográfica SIU-Kolla [61]	109
Figura 34: Estructura de la base de datos SIU-Guarani [62]	110
Figura 35: Modelo Conceptual ampliado: TA01-Situación laboral de los graduados por carrera según su distribución geográfica [Elaboración propia]	118
Figura 36: Modelo Conceptual ampliado: TA02- Valoración académica de los graduados por carrera [Elaboración propia]	120
Figura 37: Modelo Físico: TA01-Situación laboral de los graduados por carrera según su distribución geográfica [Elaboración propia]	126
Figura 38: Modelo Físico: TA02- Valoración académica de los graduados por carrera [Elaboración propia].....	126
Figura 39: Arquitectura propuesta y utilizada para la carga de los Data Marts [Elaboración propia].....	127
Figura 40: Datos de generación del reporte desde el SIU-Kolla	130

Figura 41: Exportación del reporte generado desde el SIU-Kolla.....	130
Figura 42: Listado de reportes Generados en SIU-Kolla.....	130
Figura 43: Implementación de una Primary Key en la tabla dsa.recien_graduado.....	133
Figura 44: Proceso ETL dsa_recien_graduado.....	133
Figura 45: Mapeo de campos en la tabla dsa.recien_graduado	134
Figura 46: Datos cargados en la Encuesta al Recién Graduado referidos a Título, Año de Ingreso, Ciudad, Provincia, País de procedencia	134
Figura 47: Datos cargados en la Encuesta al Recién Graduado referidos a fecha de nacimiento.....	136
Figura 48: Datos cargados en la Encuesta al Recién Graduado referidos a la ciudad de residencia.	137
Figura 49: Datos cargados en la Encuesta al Recién Graduado referidos a la provincia y país de residencia.....	137
Figura 50: Imputación de valores en bi_loc_proc	138
Figura 51: Imputación de valores en bi_loc_res.....	138
Figura 52: Transformación en PDI sobre migración desde Informix a PostgreSQL.....	140
Figura 53: Importación de datos. Conexión a Informix (SIU-Guarani)	141
Figura 54: Correspondencia entre los atributos de Informix y PostgreSQL.....	142
Figura 55: Flujo de trabajo en PDI, carga de datasets	143
Figura 56: Transformación mapeo de valores con PDI	143
Figura 57: Configuración del objeto mapeo de valores en PDI.....	144
Figura 58: Configuración del objeto d_situacion_laboral (Tabla- Salida)	145
Figura 59: Configuración del objeto Actualizar.....	146
Figura 60: Modelo multidimensional Situación laboral	148
Figura 61: Modelo multidimensional Valoración académica.....	148
Figura 62: Visualización de la herramienta JPivot	149
Figura 63: Etapas de la metodología propuesta [Elaboración propia].....	153
Figura 64: TA01 - Las diez carreras con mayor número de graduados que residen en Misiones.	161
Figura 65: TA01 - Ranking general de carreras cuyos graduados residen en Posadas, Misiones.	162
Figura 66: TA01 - Situación laboral de los graduados de la carrera Ingeniería Química.....	163
Figura 67: TA01 - Graduados de la carrera Ingeniería Química que no trabajan discriminado por año, sexo y promedio de edad.....	164
Figura 68: TA01 - Porcentaje de participación de los países extranjeros donde residen los graduados FCEQyN	165
Figura 69: TA01-Graduados de la carrera de Farmacia que trabajan en relación a su profesión	166
Figura 70: TA01-Carreras con mayor inserción laboral en Misiones.....	167
Figura 71: TA01-Gráficamente las diez carreras con mayor inserción laboral en Misiones	167
Figura 72: TA01-Cantidad de graduados de Profesorado en Matemática y Farmacia insertos en el mercado local	168
Figura 73: TA01-Grado de satisfacción laboral de las carreras de Ingeniería Química e Ingeniería en Alimentos	169
Figura 74: TA01-Registro sobre la carga horaria laboral informada por los graduados.....	170
Figura 75: TA01-Carreras que registraron una carga horaria laboral semanal de “Mas de 31 y hasta 45”	171
Figura 76: TA02 -Valoración académica de los graduados de la FCEQyN	172
Figura 77: TA02 -Valoración académica de los graduados de la FCEQyN por cohorte	173
Figura 78: TA02-Valoración de los graduados para la elección de la carrera estudiada	174
Figura 79: TA02-Valoración de los graduados para la elección de la carrera estudiada por cohorte.....	175
Figura 80: TA02-Ranking descendente por carrera sobre la duración real de egreso en años.	176
Figura 81: TA02-Duración teórica y duración real de egreso en años, por carrera.	177
Figura 82: TA02-Desvío de la duración real en años por carrera.....	178

Figura 83: TA02-Representación gráfica del desvío de la duración real del plan de estudio por carrera en años	179
Figura 84: TA02-Valoración sobre la duración y contenido de los planes de estudio por carrera	180
Figura 85: TA02-Aportes valorados por los graduados sean brindados por la Facultad/Universidad.....	181
Figura 86: TA02-Cantidad de graduados por carrera que desean recibir ‘Cursos, seminarios de actualización profesional’	182
Figura 87: Tablero situación laboral de los graduados por carrera.....	184
Figura 88: Tablero valoración académica de los graduados por carrera.....	185
Figura 89: Distribución geográfica de los graduados en la Argentina.....	186
Figura 90: Distribución geográfica de los graduados en Misiones.....	187
Figura 91: Cantidad de graduados por Carrera residentes en Posadas	188
Figura 92: Duración del proyecto [Elaboración propia].	193
Figura 93: Lista de tareas en Redmine FCEQyN	193
Figura 94: Detalle de un requerimiento en Redmine FCEQyN.....	194
Figura 95: Encuesta de validación de la metodología propuesta.	196
Figura 96: Encuesta de validación de la metodología propuesta (cont.)	197
Figura 97: Resultados de la encuesta de validación de la metodología propuesta (Pregunta 1).....	198
Figura 98: Resultados de la encuesta de validación de la metodología propuesta (Pregunta 2).....	198
Figura 99: Resultados de la encuesta de validación de la metodología propuesta (Pregunta 3).....	199
Figura 100: Resultados de la encuesta de validación de la metodología propuesta (Pregunta 4).....	199
Figura 101: Resultados de la encuesta de validación de la metodología propuesta (Pregunta 6).....	200
Figura 102: Resultados de la encuesta de validación de la metodología propuesta (Pregunta 5).....	200
Figura 103: Resultados de la encuesta de validación de la metodología propuesta (Pregunta 7).....	201
Figura 104: Presupuesto del servidor adquirido por la FCEQyN.	221
Figura 105: Consulta SQL para el <i>dataset</i> A.....	231
Figura 106: Consulta SQL para el <i>dataset</i> B.	231
Figura 107: Consulta SQL para el <i>dataset</i> C.....	232
Figura 108: Consulta SQL para el <i>dataset</i> D.....	232
Figura 109: Consulta SQL para el <i>dataset</i> F.	232

Índice de Tablas

Tabla 1: Tabla 1: Beneficios tangibles, intangibles y estratégicos [Elaboración propia]	25
Tabla 2: Tabla 2: Modelos de Ingeniería de Requisitos [Elaboración propia]	47
Tabla 3: Definición de roles y responsabilidades.	65
Tabla 4: Definición de roles intervinientes en cada etapa.	66
Tabla 5: Documento de Especificación de Requerimientos de Usuario (DERU)	67
Tabla 6: Formato propuesto para definir los temas analíticos a desarrollar.	68
Tabla 7: Diccionario de Datos (DD).....	69
Tabla 8: Identificación de las Fuentes de Datos (FUDA).....	72
Tabla 9: Conformación de indicadores (CI)	73
Tabla 10: Definición de Tabla de Dimensiones (DTD).....	77
Tabla 11: Definición de Tabla de Hechos (DTH)	78
Tabla 12: TA01-Situación laboral de los graduados por carrera según su distribución geográfica.....	94
Tabla 13: TA02- Valoración académica de los graduados por carrera.....	94
Tabla 14: TA03-Necesidad de capacitación del graduado por carrera según su distribución geográfica.....	95
Tabla 15: Diccionario de Datos TA01.....	96
Tabla 16: Diccionario de Datos TA02.....	99
Tabla 17: Diccionario de Datos TA03.....	102
Tabla 18: Requerimientos mínimos de hardware [54].....	104
Tabla 19: Configuración de hardware servidores configurados	105
Tabla 20: Accesibilidad y usabilidad de las bases de datos seleccionadas.....	107
Tabla 21: FUDA TA01-Situación laboral de los graduados por carrera según su distribución geográfica [Elaboración propia]	110
Tabla 22: FUDA TA02- Valoración académica de los graduados por carrera [Elaboración propia]	112
Tabla 23: FUDA TA03-Necesidad de capacitación de los graduados por carrera según su distribución geográfica [Elaboración propia]	113
Tabla 24: Conformación de indicadores TA01	115
Tabla 25: Conformación de indicadores TA02	116
Tabla 26: Conformación de indicadores TA03	116
Tabla 27: Descripción de atributos por perspectiva (TA01).....	118
Tabla 28: Descripción de atributos por perspectiva (TA02).....	120
Tabla 29: DTD d_carrera.....	123
Tabla 30: DTH ft_situacion_laboral.....	123
Tabla 31: DTH ft_valoracion_academica.....	124
Tabla 32: Descripción de los datos de los datasets A, B, C, D y F.....	128
Tabla 33: Descripción de los datos del dataset E	131
Tabla 34: Plantillas diseñadas para la gestión de requisitos	156
Tabla 35: Atributos editables del SIU-Kolla	157
Tabla 36: Repositorio de dimensiones.....	158
Tabla 37: Consultas derivadas surgidas de cada tema analítico planteado.....	159
Tabla 38: Validación de los usuarios.....	183
Tabla 39: Formato inicial de la plantilla DERU	189
Tabla 40: Plantilla de perspectivas e indicadores	189
Tabla 41: Plantilla sobre el nivel de granularidad.	190
Tabla 42: Graduados que completaron la encuesta por colación SIU-Kolla	191

Capítulo 1

Introducción

1. Introducción

Las Universidades Nacionales en la República Argentina utilizan en general, los sistemas transaccionales y de gestión desarrollados por el Consorcio SIU [1], en sus distintas áreas operacionales y de gestión como son la académica (SIU-Guaraní), recursos humanos (SIU-Mapuche), financiera (SIU-Pilaga), patrimonial (SIU-Diaguita), entre otros.

El sistema SIU-Wichi, a diferencia de los otros sistemas, es un módulo de información gerencial y es utilizado en el ámbito universitario como herramienta de Inteligencia de Negocios.

La Universidad Nacional de Misiones (UNaM) ha implementado en el año 2013 éste módulo como herramienta de soporte para la toma de decisión, utilizando como insumo el conjunto de Data Marts provistos por el SIU.

En este contexto, y sabiendo la importancia de contar con información que aporte al proceso de toma de decisión en un ambiente universitario, resulta indispensable contar con Data Marts que cubran las necesidades de información requeridas por las autoridades de una institución.

Este trabajo final de maestría propone diseñar e implementar una metodología para el desarrollo de Data Marts en Universidades Nacionales, que aporten al proceso de toma de decisión.

Como caso de estudio y validación se toma el área de Graduados de la Facultad de Ciencias Exactas, Químicas y Naturales (FCEQyN) de la UNaM.

1.1 Descripción del problema y justificación

Las Universidades Nacionales en nuestro país utilizan en general sistemas de Procesamiento de Transacciones en Línea (OLTP) desarrollados por el Consorcio SIU, que acompañan la gestión en el ámbito operacional universitario. Entre ellos podemos mencionar el módulo SIU-Guaraní (Sistema de Gestión Académica), el módulo SIU-Pilaga (Sistema Económico, Presupuestario, Financiero y Contable), el módulo SIU-Mapuche (Sistema de Recursos Humanos), el módulo SIU-Diaguita (Sistema de Compras, Contrataciones y Patrimonio), el módulo SIU-Kolla (Sistema de Gestión de Encuesta), entre otros.

Los sistemas tradicionales de gestión suelen realizar tareas repetitivas muy bien estructuradas que implican transacciones cortas y actualizaciones. Sin embargo, estos sistemas transaccionales no siempre son apropiados para dar soporte al proceso de toma de decisión que requieren los mandos medios y altos de una institución.

Para poder analizar el caudal de datos generados por los sistemas fuentes y, así lograr una visión más integral de los mismos por parte de las gerencias de las distintas áreas, el Consorcio SIU desarrolló el módulo de consultas gerenciales SIU-Wichi con herramientas de Data Warehouse como soporte para la toma de decisión. El mismo utiliza un conjunto de Data Marts para el procesamiento de sus datos y la consulta de información.

Los Data Marts que existen actualmente en el ámbito de las universidades, proporcionados por el SIU, son poco representativos y no responden a todos los requerimientos propios de las instituciones académicas. Por otro lado, adaptar los Data Marts existentes a las necesidades propias de una institución se vuelve complejo debido a que no se encuentra una guía oficial de referencia para personalizar uno existente; o desarrollar uno nuevo.

A pesar que el Consorcio SIU cuenta con un proceso de desarrollo para diseñar un Data Warehouse y, que las universidades nacionales adheridas al mismo tienen la posibilidad de diseñar sus propios Data Marts, se observa la ausencia de una adecuada metodología para el desarrollo de los mismos.

Esto conlleva a la dificultad de garantizar el cumplimiento de algunos requisitos fundamentales de los Data Marts como ser: flexibilidad, escalabilidad, integración de múltiples y heterogéneas fuentes de datos, coherencia de los datos y perspectiva histórica de los mismos, entre otros.

Como propuesta a esta carencia en la formalización de los procesos necesarios para el diseño y construcción de Data Marts, en este trabajo final de maestría se ofrece una metodología para el desarrollo de los mismos, que resulte adecuada para el contexto de Universidades Nacionales que utilicen herramientas provistas por el Consorcio SIU.

1.2 Objetivos

El objetivo general del presente trabajo consiste en elaborar una metodología para el desarrollo de Data Marts en Universidades Nacionales adheridas al Consorcio SIU.

Los objetivos particulares son:

- Relevar metodologías existentes para el desarrollo de Data Marts, y los procesos utilizados actualmente para el desarrollo de Data Marts en el ámbito del Consorcio SIU.
- Analizar las características de las metodologías relevadas, identificando procesos, roles, recursos y estrategias para la evaluación de desempeño de sus actividades.

- Proponer una metodología para el desarrollo de Data Marts utilizando la información analizada sobre los procesos actuales y las metodologías relevadas.
- Desarrollar e implementar un conjunto de Data Marts utilizando al área de Graduados de la Facultad de Ciencias Exactas, Químicas y Naturales como caso de estudio para la evaluación de la metodología propuesta.
- Analizar los resultados obtenidos en la experiencia adquirida a partir del caso de estudio y aplicar ajustes sobre la metodología propuesta.

1.3 Elección del caso

Por requerimientos establecidos por la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU) [2] en el marco de la Autoevaluación Institucional de la Universidad Nacional de Misiones (UNaM), en el año 2019 la Facultad de Ciencias Exactas, Químicas y Naturales (FCEQyN) aprueba mediante la Resolución Nro.: 008/19 el Plan de Desarrollo Institucional (PDI) 2018 – 2026 [3].

Este Plan de Desarrollo Institucional, tiene sus fundamentos en la construcción colectiva que traduce la misión y la visión de la UNaM en metas y objetivos planificados para un período de tiempo determinado.

El mismo se encuentra organizado en ejes estratégicos con distintas líneas de acción para alcanzar los objetivos propuestos, el cual engloba todas las dimensiones del quehacer universitario. Una de ellas es el área de los Graduados.

Si bien en la FCEQyN existieron espacios relacionados a los graduados, recién en el año 2018, adquirió mayor relevancia contar con información asociada a los mismos y se comenzó a recabar datos a través de la Encuesta al Recién Graduado implementada mediante el sistema SIU-Kolla.

En el año 2019 se crea en la FCEQyN la “Oficina de Vinculación con los Graduados (OVIG)” mediante la Resolución Nro.: 040/19 [4], para cubrir los requerimientos de la CONEAU y acompañar las políticas propuestas en el PDI de la FCEQyN y de la UNaM [5] referidas a los graduados.

Con objeto de evaluar la metodología propuesta se opta por desarrollar un producto software basado en un conjunto de Data Marts correspondientes al área de Graduados de la FCEQyN de la UNaM, que facilite el proceso de toma de decisiones y aporte al cumplimiento de los objetivos específicos definidos en el PDI.

La implementación de la metodología para el caso de estudio implica un diseño de tipo experimental en el que se aplicará la metodología propuesta en el desarrollo de un conjunto de Data Marts.

Actualmente, mediante el sistema SIU-Kolla, es posible obtener datos de los graduados y utilizarlos para conseguir información estadística. Sin embargo, la implementación de un conjunto de Data Marts específico para esta área, prevé una herramienta de acceso a los datos en un formato adecuado para generar información valiosa para la toma de decisiones en los niveles gerenciales.

1.4 Plan metodológico

La investigación asociada a este trabajo es del tipo “Aplicada”.

Dicha investigación comenzará con el relevamiento de las metodologías existentes para el desarrollo de Data Marts, y los procesos utilizados actualmente para el desarrollo de Data Marts en el ámbito del Consorcio SIU.

Luego se realizará un análisis de las características de las distintas metodologías, identificando al menos, procesos, roles, recursos y estrategias para la evaluación de desempeño de sus actividades.

En base a las metodologías relevadas y la información analizada sobre los procesos actuales, se propondrá una metodología para el desarrollo de Data Marts adecuada para el contexto de las Universidades Nacionales adheridas al Consorcio SIU.

Esta nueva metodología propuesta será evaluada utilizando como caso de estudio el Área de Graduados de la FCEQyN de la UNaM.

Finalmente, los resultados obtenidos de la implementación del caso de estudio mediante la metodología propuesta, serán analizados, procesados y validados para la elaboración del informe final.

1.5 Área de conocimiento disciplinar específico

Principalmente los temas centrales que se abordarán en el presente trabajo final de maestría son: Metodología de desarrollo de Software, Bases de Datos, Data Marts y Data Warehouses, por lo que este trabajo se enmarca dentro del área de conocimiento de la ingeniería del software

referenciado en SWEBoK¹, documento creado por la *Software Engineering Coordinating Committee* y promovido por la IEEE².

1.6 Contexto del presente trabajo

El presente trabajo ha sido desarrollado para la obtención del título de “Magister en Tecnologías de la Información”. El mismo se desarrolló dentro del ámbito de la FCEQyN y cuenta con el aval institucional para la implementación del mismo.

El caso de estudio escogido se encuentra dentro del PDI de la FCEQyN (UNaM).

La Dirección Área de Informática y Comunicaciones (DAIyC) de la mencionada Facultad cuenta con el equipamiento y personal calificado para dar soporte a este proyecto.

1.7 Estructura del presente trabajo

Este documento se encuentra dividido en ocho capítulos.

En el capítulo 2 se brinda el marco teórico utilizado como base para el desarrollo del presente Trabajo Final de Maestría. En principio se introduce el concepto sobre los Sistemas de Apoyo a la toma de Decisión junto a la definición sobre Inteligencia de Negocios. Luego, se presenta la definición de Data Warehouse, seguido a la definición de sus elementos, operaciones, tipos de implementación y los modelos de referencia más relevantes para el desarrollo de Data Warehouses. A continuación, se realiza una descripción de la Ingeniería de Requisitos, las actividades presentes en el proceso y modelos relacionados a la misma. Finalmente un resumen de antecedentes relacionados a la integración de las metodologías existentes

El capítulo 3 describe el módulo de información gerencial SIU-Wichi, el proceso de desarrollo de Data Marts utilizado por el SIU³, las herramientas informáticas que utiliza y la arquitectura que se aplica para la construcción de los mismos. Finalmente se realiza un resumen de la situación actual en el ámbito de la UNaM.

En el capítulo 4 se describe la metodología propuesta para el desarrollo de Data Marts, sus etapas y actividades.

El capítulo 5 describe el alcance de los sistemas de información implementados en la FCEQyN, la evolución del espacio de los graduados en el ámbito institucional junto a la

¹SWEBoK: *Software Engineering Body of Knowledge* - Cuerpo del Conocimiento de la Ingeniería del Software.

² IEEE: *Institute of Electrical and Electronics Engineers* – Instituto de Ingenieros Eléctricos y Electrónicos

³ SIU: Sistema de Información Universitaria

implementación del módulo de encuestas SIU-Kolla, la readecuación del circuito administrativo para la implementación de la encuesta al recién graduado en la institución y la generación de un espacio web para los graduados. Luego se describen los problemas identificados relacionados al caso de estudio que se analiza y finalmente se realiza la implementación de la metodología propuesta.

Los capítulos 6 y 7 describen los resultados y la discusión respectivamente, sobre las experimentaciones realizadas, y finalmente en el capítulo 8 las principales conclusiones con relación a los objetivos que han sido establecidos, contribuciones realizadas y las futuras líneas de investigación que pueden llevarse a cabo.

Capítulo 2

Marco Teórico

2. Marco Teórico

En este capítulo se presenta el marco teórico el cual describe los conceptos de Sistemas de Soporte para la Toma de Decisión, Data Warehouse, metodologías de desarrollo de Data Marts / Data Warehouse, el proceso de Ingeniería de Requisitos y sus principales modelos, y finalmente, algunos antecedentes vinculados con la integración de las metodologías existentes en ambientes académicos - universitarios.

2.1 Sistemas de Soporte para la Toma de Decisión

Los sistemas de apoyo para la toma de decisiones tuvieron su mayor auge en los años ochenta. Según Trujillo *et al.* [6], estos sistemas se caracterizaron por presentar los datos de manera multidimensional.

Entre ellos se encuentran los Sistemas de Soporte para la Toma de Decisión (DSS - Decision Support Systems) que ofrecen datos detallados y están orientados a ejecutivos de la escala intermedia de la empresa, y los Sistemas de Información para Ejecutivos (EIS - Executive Information Systems) que ofrecen datos más consolidados y están orientados a ejecutivos de escala superior. Estos sistemas están enfocados en extraer conocimiento de datos históricos.

2.1.1 Inteligencia de Negocios

El término Inteligencia de Negocios (BI -Business Intelligence) se remonta al año 1958 cuando Luhn [7] expresaba en un artículo que una comunicación eficiente era la llave para poder avanzar en todos los campos del esfuerzo humano. En ese artículo, propuso un sistema de inteligencia de negocios que aceptaba la información en su formato original, y organizaba la misma mediante la automatización de procesos de datos, de manera tal que los usuarios finales pudieran acceder a la información lo más directamente posible.

En 1989 Howard Dresner [8], un analítico del *Gartner Group*⁴ retomó las ideas de Luhn sobre la Inteligencia de Negocios y reformuló la importancia de estos sistemas como soporte a las decisiones.

Según Bernabeu [9] el concepto de Inteligencia de Negocios integra, por un lado el almacenamiento y por el otro el procesamiento de grandes cantidades de datos, cuyo objetivo principal es transformar esos datos en conocimiento y en decisiones en tiempo real.

2.1.2 Proceso de Inteligencia de Negocios

Comprender cómo una organización puede generar conocimiento a partir del análisis de un conjunto de datos, se describe a continuación a través del proceso de Inteligencia de Negocios basado en cinco fases: [9]

FASE 1 - Dirigir y Planear: en esta fase inicial se relevan los requerimientos específicos de información y las necesidades de los usuarios que participan en las distintas actividades diarias del proyecto.

FASE 2 – Recolección de Información: una vez identificados los requerimientos específicos, se determinan las posibles fuentes de datos tanto internas como externas, que permitan dar respuesta a las necesidades planteadas en la fase anterior.

FASE 3 – Procesamiento de Datos: En esta fase se integran los datos obtenidos en la fase anterior mediante procesos de extracción, transformación y carga (ETL - Extract, Transform and Load), que luego son cargados en un Data Warehouse para su posterior análisis.

FASE 4 – Análisis y Producción: En esta fase se procede a analizar los datos extraídos e integrados del Data Warehouse, utilizando herramientas y técnicas propias de inteligencia de negocios. Como resultado de esta fase se obtienen las respuestas a las necesidades de información mediante la creación de reportes, indicadores de rendimiento, cuadros de mando, gráficos, etc.

FASE 5 – Difusión: En esta fase final se entrega al usuario las herramientas necesarias para que pueda explorar los datos de manera sencilla e intuitiva.

En la Figura 1 se observan las fases del proceso de inteligencia de negocio descripto.

⁴ Gartner Inc. empresa consultora y de investigación de las tecnologías de la información con sede en Stamford, Connecticut, Estados Unidos

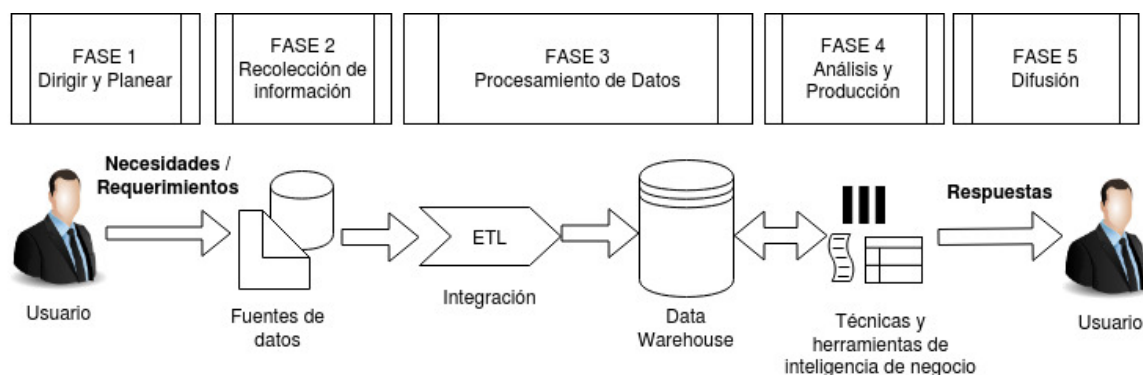


Figura 1: Fases del proceso de Inteligencia de Negocios [Elaboración propia]

2.1.3 Beneficios de la Inteligencia de Negocios

Entre los beneficios más importantes que otorga la Inteligencia de Negocios a las organizaciones se encuentran: [9]

- Reduce el tiempo necesario para recoger toda la información relevante sobre un tema particular.
- Permite actualizar la información de manera automatizada, gracias a la implementación de procesos predefinidos.
- Facilita el análisis de los datos proporcionando herramientas que ayudan a realizar comparaciones y tomar decisiones.
- Genera reportes e informes de manera dinámica.
- Facilita el acceso y análisis de los indicadores de éxito.
- Los usuarios pueden consultar y analizar los datos de manera sencilla e intuitiva.
- Permite estimar el comportamiento futuro de la organización.

Ramos [10], en otro enfoque, clasifica a los beneficios que se logran con la obtención de información mediante el uso de inteligencia de negocios, en “tangibles”, “intangibles” y “estratégicos”.

En la Tabla 1 se describe cada una de los beneficios:

Tabla 1: Beneficios tangibles, intangibles y estratégicos [Elaboración propia]

Beneficios	Descripción
Tangibles	Enfocado a las diferentes actividades del negocio: bajar costos, aumentar ingresos y reducir tiempos.
Intangibles	Generación de información confiable y oportuna: favorecer la toma de decisión.
Estratégicos	Facilitar la formulación de nuevas estrategias de negocio con la información obtenida.

2.2 Data Warehouse

Los Data Warehouses, son almacenes de datos que contienen datos históricos y son utilizados por los sistemas de soporte para la toma de decisiones tangibles, intangibles y estratégicas.

Bill Inmon [11] define un Data Warehouse como una colección de datos orientados por tema, integrados, variables en el tiempo y no volátiles que se emplean como apoyo a las decisiones estratégicas.

Se puede decir que un Data Warehouse es una colección de datos, obtenidos a partir de los datos transaccionales y estructurados de las distintas fuentes, de manera tal que se puedan realizar consultas y analizar dicha información [12]. Comúnmente se dice que los Data Warehouses son fuentes secundarias de información pues no generan datos por sí mismos, sino que son alimentados desde sistemas existentes en la organización o desde datos externos.

Los usuarios de un Data Warehouse tienen sólo permisos de lectura sobre el repositorio de datos.

Por otro lado, Chaudhuri, *et al* [13], remarca que un Data Warehouse es un proceso de construcción, y no un producto, que utiliza técnicas para consolidar y administrar datos de variadas fuentes, eliminando gran cantidad de datos inútiles o no deseados, con el objetivo de responder consultas a los administradores de las organizaciones, y tomar decisiones de una forma que antes no era posible.

2.2.1 Arquitectura de un Data Warehouse

Según Chaudhuri, *et al* [13], la arquitectura de un Data Warehouse se estructura en cinco partes principales:

- ✓ **Fuentes de datos:** Las fuentes de datos operacionales son aquellas con las que se puede llegar a disponer, tanto internas como externas a la organización. Entre ellas se encuentran:
 - **Fuentes internas** (OLTP - OnLine Transaction Processing): Compuesta por los distintos sistemas transaccionales que operan en la organización, en las cuales se reflejan las transacciones diarias.
 - **Sistemas de archivos:** Archivos tales como planillas de cálculo o documentos de texto.
 - **Fuentes externas:** Son datos obtenidos fuera de la organización. Estas fuentes pueden ser públicas o privadas.
- ✓ **Metadatos:** Los metadatos describen o dan información de otros datos que existen dentro del Data Warehouse. Brindan información de estructura, localización y definición de los mismos. [9]
- ✓ **Procesos de Extracción, Transformación y Carga (ETL):** Los procesos de ETL son los encargados de extraer los datos de las distintas fuentes de datos operacionales, realizar las transformaciones necesarias y cargarlos en el Data Warehouse.
 - **Extracción:** refiere a los procesos que se encargan de conectarse a las distintas fuentes de datos para extraer los datos con los que se poblará el Data Warehouse.
 - **Transformación:** refiere a la limpieza y/o transformación que deben recibir los datos provenientes de diversas fuentes de datos para eliminar los posibles errores que estos puedan tener previo al proceso de carga en el Data Warehouse. Esta limpieza y transformación refiere a la integración y estandarización de los datos, unificación de los estándares y codificación de los mismos, cálculo de datos derivados, creación de claves, obtención de agregados y eliminación de datos duplicados.
 - **Carga:** una vez que los datos han sido extraídos de las fuentes de datos, transformados y limpiados, se los debe cargar en el Data Warehouse en las correspondientes estructuras de datos. La carga inicial puede consumir mucho tiempo, debido a que se mueven grandes volúmenes de los mismos.

Con respecto al mantenimiento del Data Warehouse, el mismo está determinado por su granularidad y los requisitos de los usuarios, entre ellos la determinación de un horario, la frecuencia de actualización, etc.

En la Figura 2 se puede observar el proceso extracción, transformación y carga completo.

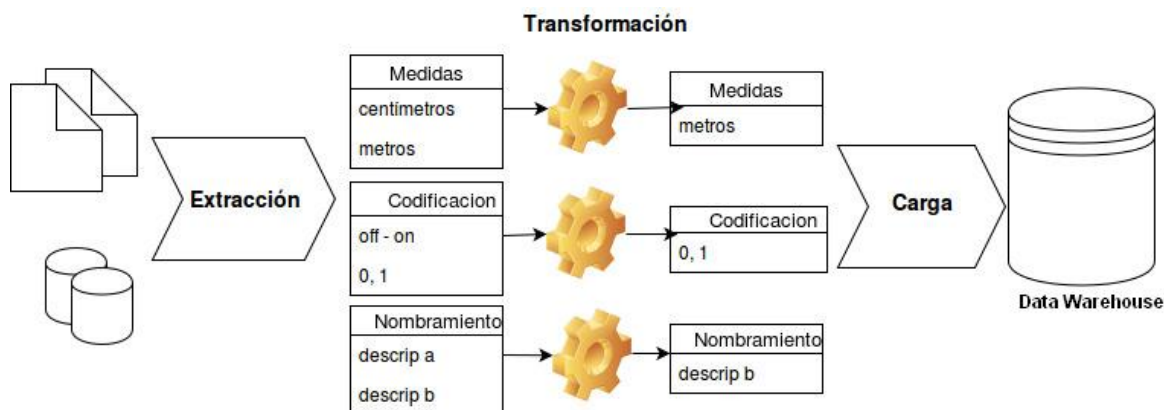


Figura 2: Proceso de Extracción, Transformación y Carga de datos. [Elaboración propia]

Los procesos ETL son ejecutados en un Área de almacenamiento de Datos (DSA – Design Stage Area), el cual permite realizar las transformaciones y la integración de las distintas fuentes de datos, sin interrumpir el normal funcionamiento de las bases de datos transaccionales ni del Data Warehouse.

- ✓ **Data Warehouse:** Almacena los datos obtenidos de las distintas fuentes de datos. Los mismos se encuentran transformados mediante procesos de extracción, transformación y carga y, orientado hacia la información relevante de la organización. Los Data Marts son subconjuntos del Data Warehouse orientados a un tema en particular, como ser un departamento o una sección dentro de la organización.
- ✓ **Herramientas para el Procesamiento Analítico en Línea (OLAP- On-Line Analytical Processing),** presentan al usuario una visión multidimensional de los datos, para cada actividad que es objeto de análisis. Estas herramientas generan la consulta y se las envía al usuario final del sistema.

En la Figura 3 se muestra la arquitectura básica de un Data Warehouse.

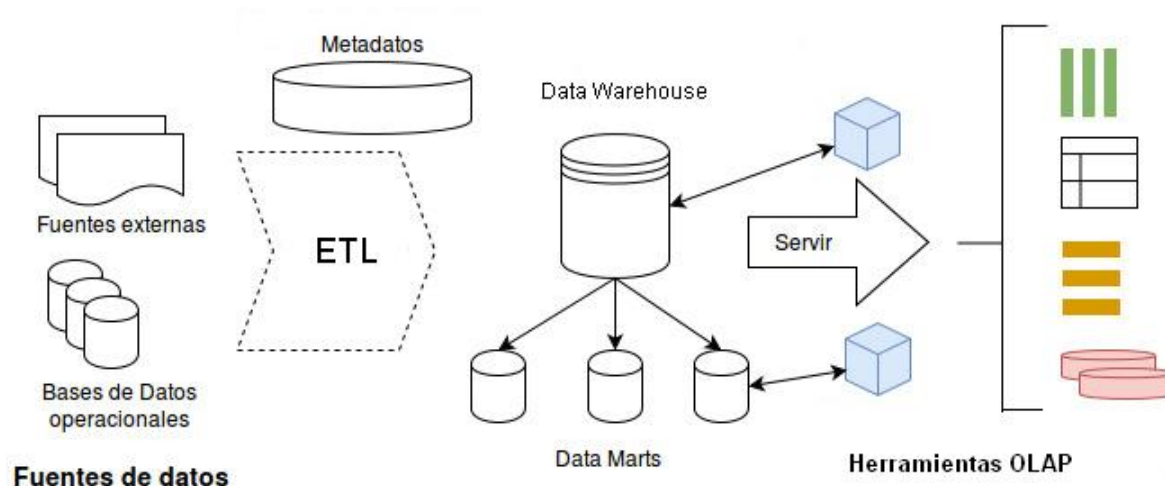


Figura 3: Arquitectura de un almacén de datos [Elaboración propia]

2.2.2 Data Mart

Un Data Mart es un almacén de datos orientado a un área o departamento específico de la organización. Responde a un determinado análisis, función o necesidad, de una población de usuarios específica. [12]

Los Data Marts permiten navegar entre las distintas fuentes de datos, agruparlas, filtrarlas y realizar comparaciones para, de este modo, ofrecer al usuario obtener información que sea de utilidad y aporte al proceso de toma de decisiones. El análisis de dichos datos se caracteriza por el hallazgo de nuevo conocimiento. [14]

Date [15] describe a un Data Mart como un almacén de datos que contiene información específica, orientada e integrada a un tema, que da apoyo a un área determinada de análisis de negocios.

Inmon [11] por otro lado, lo describe como la implementación de un Data Warehouse restringido a un área funcional, en el cual la implementación de un Data Mart permite brindar soluciones a requerimientos específicos sobre un tema dentro de la organización.

La visualización de la información contenida en un Data Mart se realiza mediante sistemas OLAP. Estos sistemas son aplicaciones que generan información táctica y estratégica que sirven a la organización como soporte para la toma de decisiones.

A diferencia de los sistemas transaccionales, que utilizan base de datos relacionales u otros archivos, los sistemas OLAP logran su máximo rendimiento y flexibilidad trabajando sobre un Data Mart.

Como se muestra en la Figura 4, los sistemas OLAP presentan al usuario un esquema o modelo multidimensional llamado también “Cubo”, en el cual se pueden realizar consultas seleccionando atributos sobre el tema en particular que se trate. La aplicación OLAP se encarga de generar la consulta y enviar los resultados al usuario para su análisis.



Figura 4: Modelo multidimensional [Elaboración propia].

El modelo multidimensional convierte los atributos y tuplas, en una matriz de n dimensiones.

La estructura multidimensional consta de una tabla de hechos y tablas de dimensiones. En la siguiente sección se describe cada uno de estos conceptos.

2.2.3 Dimensiones

Las dimensiones representan los aspectos de interés que se desea analizar dentro del contexto de negocios de la organización. Cada dimensión contiene una serie de atributos, por los cuales se pueden agrupar, comparar o filtrar la información.

Las dimensiones definen la manera que están organizados los datos lógicamente. Las mismas permiten a los usuarios filtrar y manipular la información almacenada en las tablas de hechos cuando se realizan consultas en un almacén de datos. [9]

Las tablas de dimensiones contienen datos cualitativos.

Cabe aclarar, que en las distintas metodologías existentes para el desarrollo de un almacén de datos ([9], [16]), la terminología utilizada para describir una dimensión también puede ser encontrada como perspectiva.

2.2.4 Tabla de Hechos

A diferencia de las dimensiones, las tablas de hechos contienen medidas (datos cuantitativos), que son los puntos de interés utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones. También las tablas de hechos están conformadas por las claves de cada una de las tablas de dimensión relacionadas.

Los hechos son valores numéricos y representan lo que se desea analizar y/o medir concretamente, pudiendo ser filtrados, agrupados y explorados a través de las condiciones definidas en las tablas de dimensiones. Permiten consultar información relativa a las actividades diarias de la organización. [9]

Cabe aclarar, que en las distintas metodologías existentes para el desarrollo de un almacén de datos ([9], [16]), la terminología utilizada para describir una medida también puede ser encontrada como indicador.

2.2.5 Esquema de Modelado Multidimensional

Para abordar el modelado de un almacén de datos se puede optar por distintas formas de relacionar la tabla de hecho con las dimensiones. [9]

En un esquema multidimensional un hecho representa una actividad que es objeto de análisis y las dimensiones lo que caracterizan dicha actividad. [17]

La información relevante de un hecho se encuentra representada por un conjunto de indicadores (atributos del hecho o medidas) que lo contienen. Así mismo, la información descriptiva de una dimensión se encuentra representada a través de los atributos de la misma.

Entre los atributos de una dimensión se pueden definir jerarquías. Una jerarquía representa una relación lógica entre dos o más atributos. [9]

El uso de jerarquías permite analizar los datos desde su nivel más general al más detallado y viceversa.

Entre los distintos tipos de modelado de un almacén de datos se encuentran:

- **Esquema en Estrella:** El esquema en estrella consta de una sola tabla de hechos central y de varias dimensiones relacionadas a esta. Este tipo de esquema, optimiza los tiempos de respuestas ante las consultas de los usuarios (Figura 5)

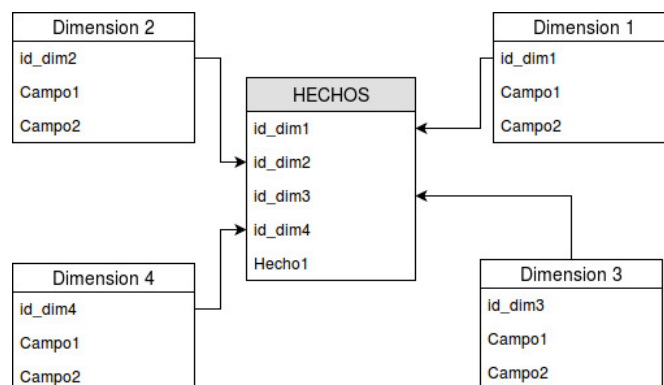


Figura 5: Esquema Estrella [Elaboración propia].

- **Esquema Copo de Nieve:** El esquema copo de nieve representa una extensión del modelo en estrella, en el cual cada tabla de dimensiones representa exactamente un nivel en una jerarquía; es decir, este tipo de esquema puede tener varias dimensiones y cada dimensión puede tener varios niveles de jerarquía (Por ejemplo: Año, Mes, Día).

En la Figura 6 se muestra un esquema de copo de nieve con cuatro dimensiones (Dimensión 1, Dimensión 2, Dimensión 3, Dimensión 4), dos de ellas con dos niveles de jerarquía (Dimensión 4 - Dimensión 4a y Dimensión 3 - Dimensión 3a):

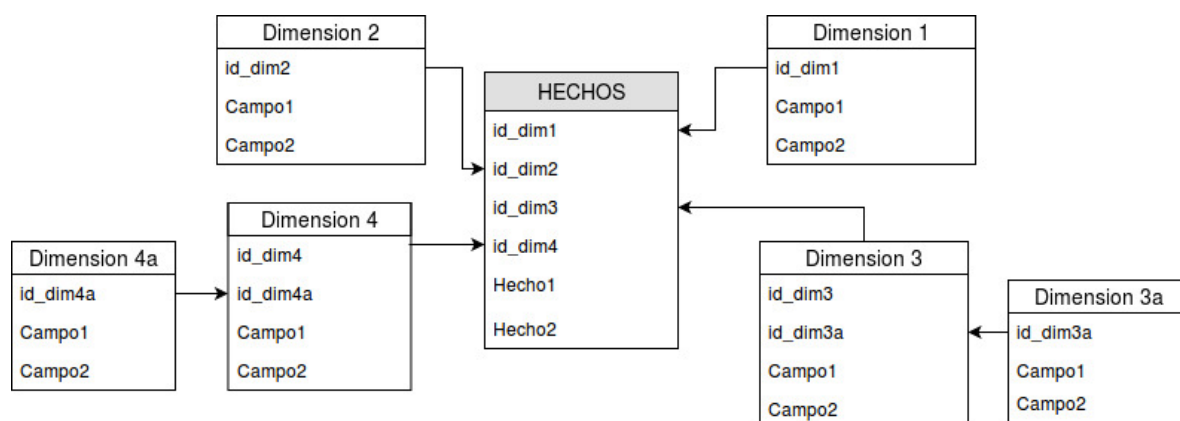


Figura 6: Esquema Copo de Nieve [Elaboración propia].

- **Esquema Constelación:** Este modelo propone un conjunto de esquemas en estrella, con una tabla de hechos principal y otras tablas de hechos auxiliares que pueden ser calculadas de la principal. Las tablas de hechos se mantienen en el centro pudiendo estar relacionadas a cualquiera de las dimensiones.

Este esquema al proponer más de una tabla de hechos, permite el análisis de más de un tema de interés y la reutilización de dimensiones. Como el nivel de complejidad de este tipo de esquema es alto, no todas las herramientas de consulta y análisis permiten su implementación (Figura 7).

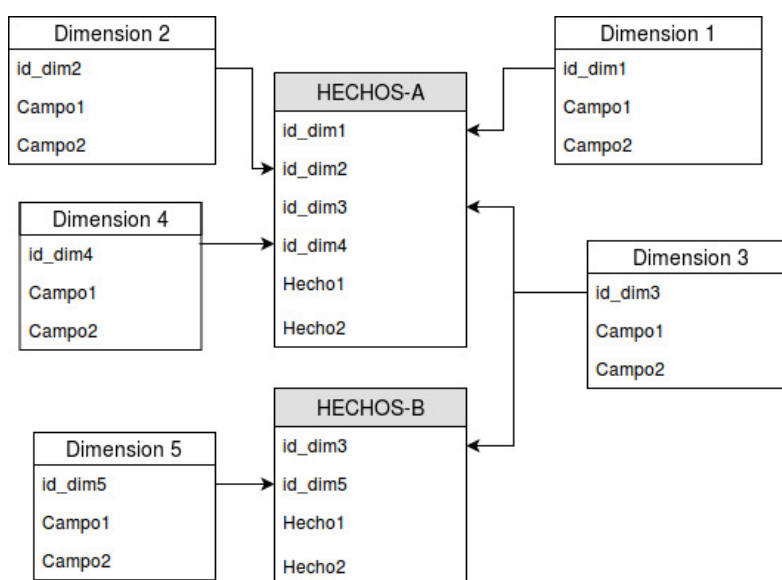


Figura 7: Esquema Constelación [Elaboración propia].

2.2.6 Operaciones

Las operaciones que se pueden realizar sobre modelos multidimensionales mediante herramientas OLAP, permiten a los usuarios explorar los datos sin necesidad de conocer la estructura interna del mismo.

Estas operaciones son las siguientes: *Drill-down*, *Roll-up*, *Drill-across*, *Roll-across*, *Pivot*, *Page*, y *Drill-through*. [9]

A continuación se explican cada uno de ellas:

- ***Drill-down*:** permite explorar los datos con un mayor nivel de detalle o criterio de agregación, accediendo a diferentes jerarquías definidas en el cubo. Esta operación permite ir de lo general a lo específico.

- **Roll-up:** permite observar los datos con un menor nivel de detalle, accediendo a una jerarquía superior en el cubo. Esta operación permite ir de lo específico a lo general.
- **Drill-across:** funciona de forma similar a la operación *Drill-down*, con la diferencia que *Drill-across* no se realiza sobre una jerarquía, sino que va de lo general a lo específico agregando atributos a su criterio de análisis.
- **Roll-across:** funciona de forma similar a la operación *Roll-up*, con la diferencia que *Roll-across* no se realiza sobre una jerarquía, sino que va de lo específico a lo general quitando atributos de la consulta, eliminando de esta manera criterios de análisis.
- **Pivot:** permite seleccionar la forma de visualizar los atributos y medidas, de manera tal de poder analizar la información de diferentes perspectivas.
- **Drill-through:** permite analizar los datos que componen una medida sumariada, en su mayor nivel de detalle.

2.2.7 Tipos de implementación de Data Warehouses

Las vistas multidimensionales pueden ser construidas de distintas formas. Entre ellas se mencionan las siguientes:

- **ROLAP (Procesamiento Analítico en Línea Relacional):** se implementa sobre tecnología relacional, pero cuenta con algunas facilidades para mejorar el rendimiento, como ser, utiliza índice de mapa de bits e índices *join*.

Los cubos multidimensionales se generan al momento de realizar las consultas, haciendo el proceso transparente para el usuario.

Este tipo de implementación puede ser aplicado a cualquiera de los esquemas mencionados en secciones anteriores: estrella, copo de nieve y constelación [16].

En los sistemas ROLAP, un almacén de datos puede ser soportado por un sistema de gestión de base de datos relacional.

- **MOLAP (Procesamiento Analítico en Línea Multidimensional):** Tiene una estructura de *arrays* multidimensionales. Los cubos son generados y almacenados antes de ser consultados, por lo tanto el análisis se realiza sobre los datos agregados y las métricas o los indicadores pre calculados.

El objetivo de los sistemas MOLAP es almacenar los datos físicamente en estructuras multidimensionales de manera que la representación externa y la representación interna coincidan. [9]

- **HOLAP (Procesamiento Analítico en línea Híbrido):** Es una combinación de las técnicas ROLAP y MOLAP. Los cubos frecuentemente consultados son generados y almacenados. Cualquier otra consulta debe generarse en tiempo de ejecución.

En la Figura 8 se pueden observar los modelos ROLAP y MOLAP respectivamente.

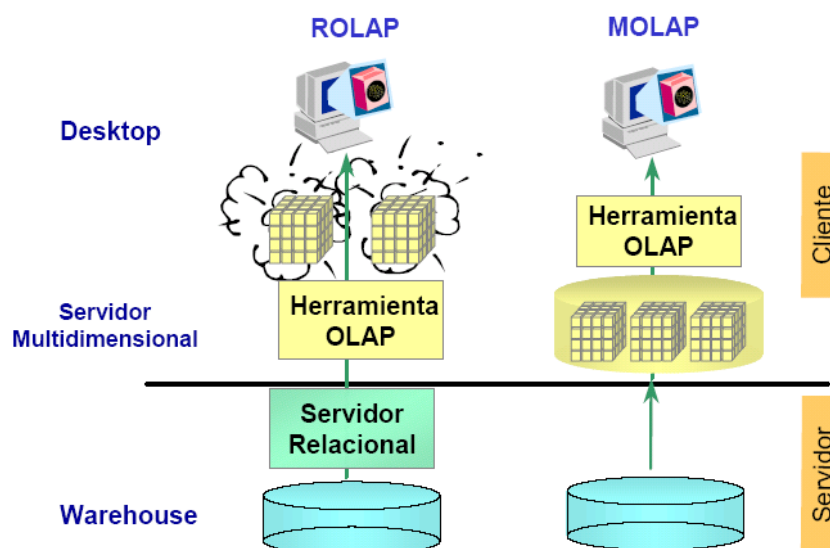


Figura 8: Modelo ROLAP y MOLAP [18]

2.2.8 Modelos para el desarrollo de Data Warehouses

En esta sección se realiza una descripción detallada de los modelos más difundidos en la actualidad para el desarrollo de Data Warehouses.

2.2.8.1 Top-down (Inmon)

El modelo propuesto por Bill Inmon [11] para desarrollar un Data Warehouse, es un modelo iterativo y presenta un enfoque top-down.

Bill Inmon [11] considera que un Data Warehouse debe cumplir con las siguientes características:

- **Orientado a un tema:** los datos deben estar organizados para que los distintos eventos puedan acceder con mayor facilidad.

- **Integrado:** las diferentes fuentes de datos deben estar integradas en una única base de datos de manera coherente.
- **Variable en el tiempo:** todos los datos pertenecen a un período de tiempo determinado.
- **No volátil:** los datos, una vez almacenados, no se modifican ni se eliminan. Los mismos se convierten en información de solo lectura.

Este modelo se basa en un esquema donde trabaja primero con los datos, se los integra para ser probados y se los programa para luego analizar los resultados.

Un Data Warehouse es considerado un gran almacén de datos donde se almacenan todos los datos de la organización. El mismo está compuesto por pequeños Data Marts temáticos que son consultados mediante herramientas de análisis y explotación.

2.2.8.2 Bottom-up (Kimball)

Ralph Kimball en 1996 propone una metodología para la construcción de un Data Warehouse denominada Ciclo de Vida Dimensional del Negocio (*Business Dimensional Lifecycle*). La misma engloba cuatro principios fundamentales [16]:

- **Centrarse en el negocio:** identificación de los requerimientos del negocio y el análisis de los mismos.
- **Construir una infraestructura de información adecuada:** diseño de una base de información única e integrada, que sea fácil de usar y de alto rendimiento. Debe contener los requerimientos de negocio identificados en la empresa.
- **Realizar entregas en incrementos significativos:** crear el almacén de datos en incrementos entregables en plazos cortos de tiempo (6 a 12 meses).
- **Ofrecer una solución completa a los usuarios de negocios:** proporcionar a los usuarios un almacén de datos consistente, bien diseñado, de calidad y accesible con herramientas de consulta, aplicaciones para informes, capacitación, soporte y documentación.

Las distintas etapas del modelo de Kimball se pueden agrupar en tres secciones: “**tecnología**”, donde se diseña la arquitectura técnica y se realiza la selección de productos a implementar, “**datos**” donde se diseña e implementa el modelo dimensional y se diseñan los procesos de extracción, transformación y carga; y por último “**aplicaciones de BI**” cuyas tareas refieren al diseño y desarrollo de las aplicaciones de negocios para los usuarios finales.

A continuación se describen cada una de las etapas de la metodología:

- Planificación del proyecto:

En esta etapa se determina el propósito del proyecto, sus objetivos y el alcance del mismo, los principales riesgos y una aproximación inicial a las necesidades de información.

Un proyecto, según Kimball [16] refiere a una iteración simple del ciclo de vida desde el inicio hasta su implementación.

- Definición de requerimientos del negocio:

La definición de requerimientos, se basa en la realización de entrevistas al personal del negocio y personal técnicos. Además, se debe relevar la documentación sobre la terminología y la estrategia interna de la organización. [16]

El personal que se recomienda entrevistar durante esta tarea son:

- **Directivos:** son los responsables de tomar decisiones estratégicas en la organización.
- **Administradores intermedios:** son los responsables de explorar las alternativas estratégicas y aplicar las decisiones.
- **Personal de sistemas:** técnicos informáticos de la organización que administran los sistemas y los recursos existentes en el sector.
- **Otros:** personal que se entrevista por razones políticas.

En la sección del nivel superior llamado “**tecnología**” se diseña la arquitectura técnica y se realiza la selección de productos a implementar. A continuación se describen cada una de las etapas:

- Diseño de la arquitectura técnica:

Para establecer el diseño de la arquitectura técnica del Data Warehouse, es necesario considerar los siguientes factores: requerimientos del negocio, el actual entorno técnico y las instrucciones técnicas estratégicas planificadas. [16]

-Selección de productos e instalación:

En base al diseño de la arquitectura técnica, se deben definir diversos componentes específicos de estructura, entre ellos la plataforma de hardware, el sistema de gestión de base de

datos, las herramientas de *data staging* o de acceso de datos, para ser seleccionados, procesados, limpiados, unificados, evaluados y probados, y así asegurar la correcta integración de los mismos dentro del entorno del almacén de datos. [16]

En la sección del nivel intermedio llamado “**datos**” se diseña e implementa el modelo dimensional y se definen los procesos de extracción, transformación y carga. A continuación se describe cada una de las etapas:

- Modelado dimensional:

El modelado dimensional de Kimball es un proceso dinámico e iterativo. El mismo consiste en cuatro pasos:

- **Elegir el proceso de negocio:** elección del área a modelar.
- **Establecer el nivel de granularidad:** definición del nivel de detalle del área a modelar, dependiendo de los requerimientos del negocio y lo que se encuentre disponible entre los datos actuales.
- **Elegir las dimensiones:** definición de las dimensiones que contendrá el modelo en base a las discusiones del equipo, el nivel de granularidad definido y los procesos de negocios determinados.
- **Identificar medidas y tablas de hechos:** identificación de las medidas que surgen de los procesos de negocios. Una medida es un atributo de una tabla que se desea analizar, sumando o agrupando sus datos a través de las dimensiones. Las medidas se encuentran en tablas denominadas tablas de hechos (fact tables).

- Diseño físico:

El diseño físico de la base de datos refiere a la definición de las estructuras físicas necesarias para soportar el diseño lógico de la misma. Los elementos principales de este proceso incluyen la definición de los estándares de nomenclatura y la configuración del entorno de la base de datos. [16]

- Diseño e implementación del subsistema de extracción, transformación y carga (ETL):

A través de los procesos de extracción, transformación y carga se alimenta el Data Warehouse. Estos procesos extraen los datos de los sistemas de origen de los datos, aplican diferentes reglas para aumentar la calidad y la consistencia de los mismos, consolida la información

proveniente de los distintos sistemas, y finalmente cargan la información en el Data Warehouse en un formato adecuado para la utilización por parte de las herramientas de análisis. [16]

- Implementación:

La implementación representa la convergencia de los tres grupos definidos como tecnología, datos y aplicaciones de usuarios finales. Además existen otros factores extras que aseguran el correcto funcionamiento de todas las partes, entre ellos se encuentran la capacitación, el soporte técnico, la comunicación y las estrategias de retroalimentación. [16]

- Mantenimiento y crecimiento del Data Warehouse

- **Mantenimiento:** el mantenimiento del Data Warehouse debe enfocarse en los usuarios de negocio, definiendo políticas que permitan gestionar de manera adecuada las operaciones del Data Warehouse y establecer un flujo de retroalimentación constante.
- **Crecimiento:** el crecimiento y la evolución del Data Warehouse debe ser realizado, de forma iterativa utilizando el ciclo de vida propuesto.

En la sección del nivel inferior llamado “**aplicaciones de BI**”⁵ se encuentran las tareas correspondientes al diseño y desarrollo de aplicaciones para los usuarios finales. [16]

- Especificación y desarrollo de aplicaciones de BI:

Las especificaciones y el desarrollo de aplicaciones de inteligencia de negocios proporcionan a los usuarios una manera más estructurada de acceder a los datos de un almacén. Estas aplicaciones incluyen una amplia variedad de tipos de informes y herramientas de análisis para su desarrollo. Kimball divide a estos informes en dos categorías:

- **Informes estándar:** son informes simples de formato predefinidos y parámetros de consulta fijos.
- **Aplicaciones analíticas:** son más complejas que los informes estándares, estas aplicaciones pueden incluir algoritmos y modelos de minería de datos.

⁵BI: Inteligencia de negocios (*Business Intelligence*)

En la Figura 9 se observan las distintas tareas del ciclo de vida del modelo de Kimball.

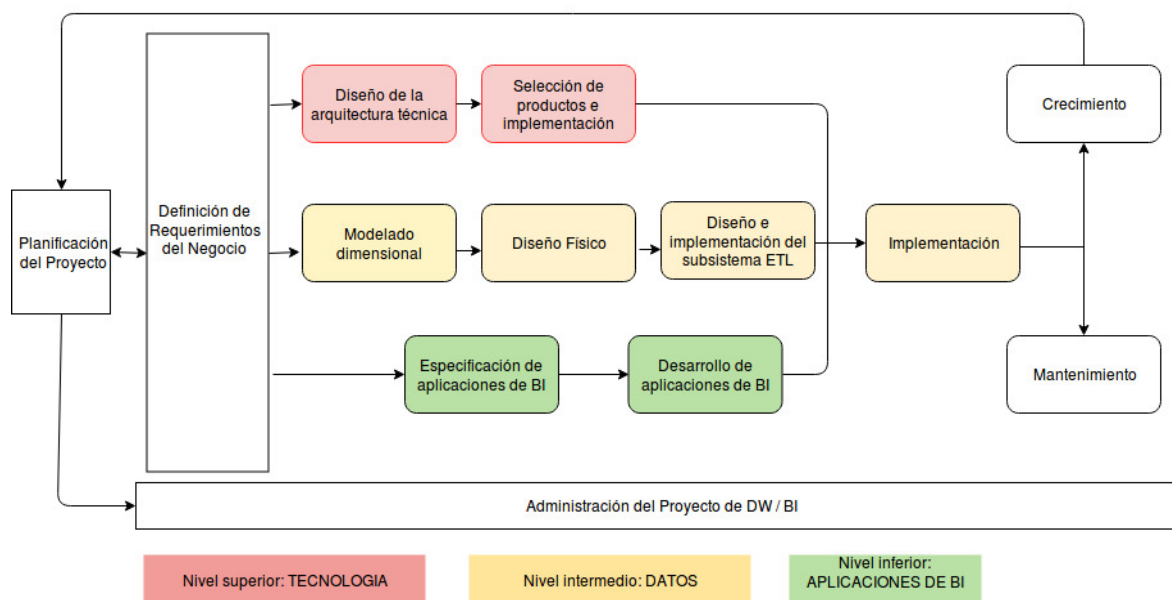


Figura 9: Diagrama del Ciclo de vida de Kimball [13]

2.2.8.3 HEFESTO

HEFESTO [9] es una metodología creada por el Ing. Bernabeu. La misma consta de cuatro etapas principales: Análisis de requerimientos, Análisis de los OLTP, Modelo lógico del Data Warehouse y Procesos de extracción, transformación y carga.

Estas etapas pueden ser utilizadas en cualquier ciclo de vida que no requiera fases extensas de requerimientos y análisis, con el fin de entregar una implementación que cumpla con una parte de las necesidades proporcionadas por el usuario.

A continuación se describen cada etapa de esta metodología.

ETAPA 1 - Análisis de Requerimientos

En esta etapa se identifican y recolectan los requerimientos de información de los usuarios con el fin de entender los objetivos de la organización, mediante distintas técnicas y herramientas de elicitación, como ser: entrevistas, encuestas, cuestionarios, diagramas de flujo, diccionario de datos y la observación. Como resultado se obtienen una serie de preguntas que se deben analizar con el propósito de establecer cuáles serán los indicadores y perspectivas que se utilizarán para la construcción del almacén de datos.

Por último se realiza un modelo conceptual donde se puede visualizar el resultado obtenido en esta primer etapa.

ETAPA 2 - Análisis de los OLTP

En esta etapa de la metodología se analizan las distintas fuentes de datos existentes, en base a los resultados obtenidos de la etapa anterior. Se determinan y definen como se van a calcular los indicadores, para luego establecer las correspondencias entre el modelo conceptual y las fuentes de referencia. Luego, se identifican y establecen los campos que se incluirán en cada perspectiva.

Por último, se amplía el modelo conceptual con la información obtenida en este paso.

ETAPA 3 – Modelo lógico del DW

En esta etapa se define el modelo lógico de la estructura del almacén de datos, en base al modelo conceptual obtenido en la etapa anterior. Luego se confeccionan las tablas de dimensiones y las tablas de hechos, para posteriormente efectuar las uniones entre los mismos.

ETAPA 4 – Integración de Datos

En esta etapa se elaboran y se aplican los procesos de extracción, transformación y carga. Esta compleja actividad se basa en la extracción de los datos de las diferentes fuentes de datos, la transformación de los mismos y la carga de estos en el almacén de datos diseñado en la etapa anterior.

Además, mediante procesos de extracción, transformación y carga, se definen las políticas y estrategias para la carga inicial del Data Warehouse y su respectiva actualización.

En la Figura 10 se puede observar las actividades que se realizan en cada etapa.



Figura 10: Etapas de la metodología Hefesto [6]

Según Bernabeu [9], esta metodología cuenta con las siguientes características:

- Los objetivos y resultados esperados en cada fase se los puede distinguir y comprender.
- La estructura propuesta en la metodología es adaptable ante los cambios en el negocio.
- Reduce la resistencia al cambio por parte de los usuarios, por su participación en cada etapa.
- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de las estructuras físicas que contengan el Data Warehouse y de su respectiva distribución.
- Los resultados obtenidos en cada etapa constituyen las entradas de la etapa siguiente.
- Se aplica tanto para Data Warehouse como para Data Mart.

2.3 Ingeniería de Requerimientos

La especificación de requerimientos, o la gestión de requisitos de usuarios, son una de las fases más críticas en el desarrollo y producción de proyectos de software.

Según Brooks [19], la parte más difícil de construir de un sistema software es decidir qué construir; esta parte del trabajo afecta negativamente al sistema final si se realiza de manera incorrecta.

La ingeniería de requerimientos es el proceso de recopilar, analizar y verificar las necesidades del cliente para un sistema de software. El objetivo de una ingeniería de requerimientos es lograr entregar una especificación de requerimientos de software completa y correcta.

Boehm [20] define a la ingeniería de requerimientos como una disciplina para desarrollar una especificación completa, consistente y no ambigua, la cual servirá como base para acuerdos comunes entre las partes involucradas y en donde se describen las funciones que realizará el sistema.

Leite [21], por su parte, considera a la ingeniería de requerimientos como un proceso mediante el cual se intercambian diferentes puntos de vistas, para recopilar y modelar lo que el sistema va a hacer.

Este proceso utiliza una combinación de métodos, herramientas y actores, cuyo producto es un modelo del cual se genera un documento de requerimientos. Estos deben expresarse de forma clara y estructurada de manera que puedan ser entendidos tanto por expertos como por los usuarios quienes deberán participar en la validación.

Según la Real Academia Española [22], un Requerimiento es un: *“Aviso, manifestación o pregunta que se hace.”*

Según la IEEE⁶ [23], un requerimiento es una *“Condición o capacidad requerida por el usuario para resolver un problema o alcanzar un objetivo”*.

En muchos casos es difícil especificar los requerimientos de un problema, con la mayor calidad posible en una etapa temprana del proceso de desarrollo. Por ello, la ingeniería de requerimientos proporciona el mecanismo apropiado para entender lo que desea el usuario, analizar las necesidades, evaluar la factibilidad, brindar una solución razonable, validar la especificación y administrar los requerimientos a medida que se transforman en un sistema funcional [24].

El proceso de la ingeniería de requerimientos es un trabajo cooperativo que involucra a usuarios, clientes, ingenieros de software, consultores, expertos, entre otros. Mediante este proceso

⁶ IEEE: Instituto de ingenieros eléctricos y electrónicos (*Software Engineering Body of Knowledge*)

se debe lograr reducir la distancia entre el usuario, que es quien tiene las necesidades y requiere soluciones a su problema, pero no posee los conocimientos para especificar los requisitos, y el ingeniero de software que sí tiene esos conocimientos técnicos pero desconoce las necesidades y los problemas del entorno del usuario.

2.3.1 Evolución de los Requisitos

Los cambios en los requisitos ocurren durante todo el ciclo de vida del software, por lo que merecen una consideración especial durante el proceso de desarrollo.

Según menciona Leite [25], “...*el cambio es una propiedad intrínseca al software*”, por lo que la administración de dichos cambios merece una consideración especial en el proceso de desarrollo del software.

Obtener un resultado exitoso en un proyecto de software depende fuertemente de resolver el problema correcto [26], es decir encontrar el conjunto de soluciones adecuadas en el universo del discurso⁷.

Desde la perspectiva de la ingeniería de requerimientos esta actividad se denomina “Gestión de requisitos”. Todas las tareas asociadas a la gestión de requisitos se realizan durante todo el ciclo de vida del proyecto, el seguimiento y cambio de los mismos, además del control de versiones y control de cambios.

2.3.2 Proceso de Ingeniería de Requisitos

El proceso de ingeniería de requisitos, según Leite [27], es el proceso de definición de requisitos el cual debe ser elicitado, modelado y analizado. El objetivo de este proceso es obtener una comprensión acabada del problema de nuestros clientes para establecer la solución más adecuada del problema.

El proceso de definición de requisitos es iterativo e involucra las siguientes actividades: Elicitación, Modelado, Análisis y Gestión (Figura 11).

⁷ Universo de discurso: todo el contexto en el cual el software se desarrolla e incluye todas las fuentes de información y toda la gente relacionada con el software.

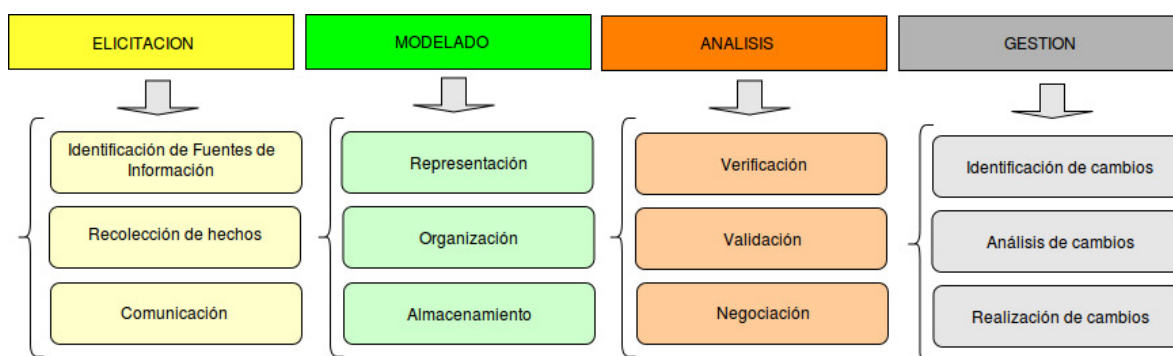


Figura 11: Actividades de la Ingeniería de Requisitos [24]

Estas actividades van interactuando entre sí a medida que se va adquiriendo conocimiento. La gestión de requisitos involucra las tres primeras actividades a lo largo del tiempo, interactuando entre las mismas por retroalimentación.

En la Figura 12 se puede observar la iteración entre estas actividades por retroalimentación junto con la evolución del universo del discurso a lo largo del tiempo. El proceso de definición de requisitos continúa a lo largo de todo el ciclo de vida del software.

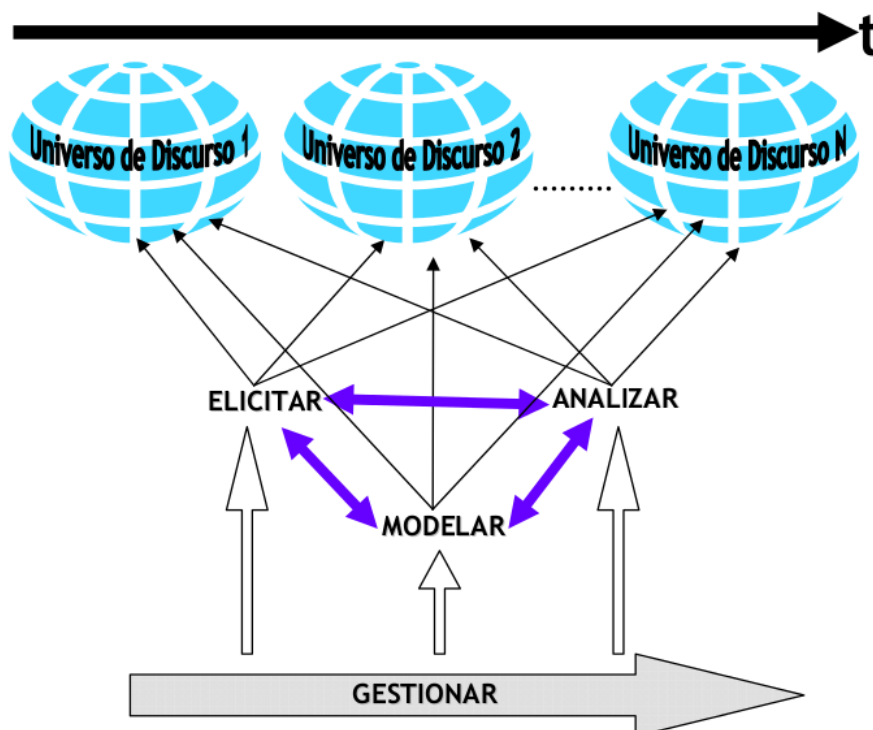


Figura 12: Actividades de la Ingeniería de Requisitos frente a la evolución del universo de discurso [27]

A continuación se describen cada una de las actividades:

Elicitación: es la fase de adquisición de conocimiento. Se identifican y se seleccionan fuentes de información, técnicas de elicitación y se procede a la recolección de hechos propiamente dicho. En esta fase es muy importante definir una buena comunicación entre los involucrados. El analista debe lograr homogeneizar el vocabulario, centrándolo en el léxico del propio cliente [21].

Existen diferentes técnicas de elicitación, entre las que se encuentran las entrevistas, los cuestionarios, tormenta de ideas (brainstorming), casos de uso, entre otros.

Modelado: es la fase donde se representan, se organizan y se registran los hechos recolectados durante la elicitación. Existen diferentes modelos de diseño para modelar requisitos que se basan en el lenguaje natural, tales como los Casos de Uso [26], los Escenarios cuya estructura está compuesta por sentencias declarativas (objetivo, contexto, recursos actores y episodios) y por un conjunto de sentencias con un lenguaje muy simple que hace posible la descripción operativa de comportamiento (Leite *et al* [25], Carroll [28]), el Léxico Extendido del Lenguaje compuesto por símbolos que identifican el lenguaje de la aplicación [29], entre otros.

Análisis: es la fase donde se verifican y validan los hechos modelados. Se identifican y resuelven requisitos ambiguos, contradictorios, omitidos, no posibles, erróneos; aplicando distintas técnicas de verificación y validación. Esta etapa también, involucra la negociación entre las partes involucradas: intercambio de información, asignación de prioridades a los requisitos, la resolución de conflictos, etc.

Gestión de Requisitos: es la fase donde se identifican, analizan y realizan los cambios necesarios provenientes de nuevos requisitos o modificaciones a los ya existentes. Esta actividad se encarga de ir registrando los cambios en los mismos, para luego poder verificarlos y validarlos.

En esta fase es importante la rastreabilidad de los requisitos y el control del versionado, debido a la evolución de los mismos.

2.3.3 Modelos de Ingeniería de Requisitos

Existen diferentes modelos de procesos para la Ingeniería de Requisitos. Entre ellos se encuentran el modelo espiral [30] [31], basado en el modelo espiral de Boehm [32]; el modelo de Polh [33]; el modelo SWEBoK [34]; el modelo de Leite [27]; el modelo CMMI v1.3 [35]; entre otros.

El modelo espiral [32] tiene como actividades principales la Elicitación, Análisis / Validación, y Negociación de Requisitos. Este modelo asume que existe una cuarta actividad, que es la Gestión de Requisitos, que se realiza durante todo el proceso y se encarga de gestionar la obtención de requisitos y los cambios de los mismos durante el ciclo de vida del software.

El modelo de Polh [33] es un modelo iterativo compuesto de cuatro actividades. Estas actividades pueden realizarse en cualquier orden, pero el modelo asume una secuencia donde los requisitos son primero elicitados, luego negociados e integrados al resto de la documentación, y finalmente validados y verificados, para asegurar una secuencia lógica de las necesidades de los clientes.

El modelo SWEBoK [34] es un proyecto en conjunto entre IEEE [23] y ACM⁸ para producir un cuerpo del conocimiento sobre ingeniería del software [34].

Leite [27] en su modelo, propone las siguientes actividades: Elicitación, Modelado y Análisis, que ayudan al correcto entendimiento del problema antes del acuerdo de una solución final.

Finalmente el modelo CMMI v1.3 [35] donde, dentro de su área específica de gestión de requisitos, permite gestionar todos los requisitos generados o recibidos por el proyecto. Entre las prácticas de este modelo se encuentran: Comprender los requisitos, Obtener el compromiso sobre los requisitos, Gestionar los cambios a los requisitos, Mantener la trazabilidad bidireccional de los requisitos y Asegurar el lineamiento entre el trabajo del proyecto y los requisitos.

A continuación se muestra un resumen de los distintos modelos de ingeniería de requisitos y sus principales actividades (Tabla 2).

⁸ ACM: Asociación de Maquinaria Computacional (Association for Computing Machinery)

Tabla 2: Modelos de Ingeniería de Requisitos [Elaboración propia]

MODELOS DE INGENIERIA DE REQUISITOS	AÑO	ACTIVIDADES DEL MODELO
Modelo de Leite [27]	1994	• Elicitación • Modelado • Análisis
Modelo de Polh [33]	1997	• Elicitación • Negociación • Especificación y documentación • Validación y verificación
Modelo espiral de Boehm [32]	1997	• Elicitación • Análisis y validación • Negociación de requisitos • Gestión de requisitos
Modelo SWEBoK ⁹ [34]	1999	• Elicitación • Análisis y negociación de requisitos • Documentación de requisitos • Validación de requisitos
Modelo CMMI v1.3 [35]	2010	• Comprender los requisitos • Obtener el compromiso sobre los requisitos • Gestionar los cambios a los requisitos • Mantener la trazabilidad bidireccional de los requisitos • Asegurar el lineamiento entre el trabajo del proyecto y los requisitos

⁹ SWEBoK: Guía del Cuerpo de Conocimientos de Ingeniería de Software (*Software Engineering Body of Knowledge*)

2.4 Antecedentes

La revisión de antecedentes estuvo centrada en la evaluación de otras metodologías, relacionadas a la problemática abordada en el trabajo.

Entre algunos trabajos realizados se encuentran:

➤ Torres Romo [36] presenta una investigación sobre las metodologías consideradas referentes sobre el desarrollo de un Data Warehouse (Inmon [11], Kimball [16] y HEFESTO [9]), desarrollando una comparativa entre las mismas. Esta comparativa está orientada al análisis de las dimensiones: enfoque, diseño, arquitectura, integración de los datos y mantenimiento, entre otras.

En su informe, el autor no propone cambios en ninguna de sus etapas y concluye que la metodología HEFESTO es la que mejor se adapta a cualquier etapa de desarrollo de un Data Warehouse, por ser una metodología híbrida.

➤ Por otro lado, Rojas Muñoz y Saquicela Galarza [37] presentan una investigación sobre el desarrollo de una nueva arquitectura para el desarrollo de un Data Warehouse para la Universidad de Cuenca de Ecuador, donde analizan la utilización conjunta de las metodologías de Inmon [11], Kimball [16] y HEFESTO [9] combinadas con metodologías ágiles como SCRUM [38].

Como conclusión, los autores exponen que la metodología HEFESTO [9] combinada con la metodología de desarrollo SCRUM [38] brindan la posibilidad de mejorar la planificación y el avance del trabajo desarrollado.

➤ Otra investigación presentada por Leonard Brizuela y Castro Blanco [39], releva otra metodología existente para el desarrollo de almacenes de datos, entre ellas DWEP¹⁰, RWM¹¹ y Trujillo [40], además de la de Kimball [16] y HEFESTO [9]. Analizan sus principales características y mencionan que la metodología que se elija debe demostrar por un lado, “ventajas” en el desarrollo de un almacén de datos y por otro lado, motivar a los usuarios en el uso del mismo.

Los autores concluyen que la integración de las metodologías HEFESTO [9] y DWEP son las que mejor se adaptan para cumplir con los requisitos de los usuarios y, mejorar la calidad y el uso del Data Warehouse.

➤ Castillo-Rojas, Medina Q. y Fariñas M. [41] proponen en su trabajo de investigación, el desarrollo de una metodología para el diseño de un Data Warehouse en la

¹⁰ DWDEP: Proceso de Ingeniería de un Almacén de Datos (*Data Warehouse Engineering Process*)

¹¹ RWM: Metodología de Almacenamiento Rápido (*Rapid Warehousing Methodology*)

Universidad Chilena donde incorporan al proceso de desarrollo indicadores de productividad académica.

Esta propuesta metodológica integra diversos enfoques, técnicas y metodologías como las de Kimball [16] y HEFESTO [9]. A la integración de ambas propuestas, agregaron una fase de validación de indicadores de productividad académica que favorecen la efectividad y el éxito del desarrollo del Data Warehouse.

En su informe, los autores proponen una reingeniería de las etapas de las metodologías relevadas, e incorporan como contribución a su investigación una fase de análisis de los indicadores de productividad y otra fase de validación de los mismos.

En resumen, a partir de las investigaciones relevadas se puede inferir que la mayoría concuerdan con los modelos tradicionales existentes y mencionados en el presente trabajo para el desarrollo de un almacén de datos: Kimball [16], Inmon [11] y HEFESTO [9].

Por otro lado, cabe destacar que la mayoría de las investigaciones se han realizado comparativas y combinaciones entre estos modelos, siendo escasas aquellas investigaciones que realizaron una reestructuración de las etapas incorporando nuevos métodos y herramientas para optimizar su uso y aplicabilidad.

A diferencia de los trabajos descriptos relacionados a la actual investigación, el aporte del presente Trabajo Final de Maestría está basado en la incorporación de etapas que involucran a la gestión de requisitos durante todo el proceso de desarrollo de un Data Warehouse.

Con referencia a la gestión de requisitos asociado al trabajo colaborativo que desarrolla el Consorcio SIU, Cuadrado y Oliveros [42] realizaron un estudio sobre la importancia de aplicar el método *Grounded Theory* (Teoría Fundada) al proceso actual de gestión de requerimientos, de manera tal que se puedan analizar los datos obtenidos en escenarios no cubiertos por las metodologías tradicionales, como es el caso de los *stakeholders* (partes interesadas) distribuidos.

Sin embargo, la gestión de requerimientos a la que apunta dicha investigación se aplica para el desarrollo de los distintos sistemas de gestión que brinda el Consorcio SIU y no a la gestión de requisitos necesarios para el desarrollo de Data Marts.

Capítulo 3

Situación Actual

3 Situación Actual

En este capítulo se describe el módulo de información gerencial SIU-Wichi, el proceso de desarrollo de Data Marts utilizado por el SIU, las herramientas informáticas que utiliza y la arquitectura que se aplica para la construcción de los mismos. Finalmente se realiza un resumen de la situación actual en el ámbito de la UNaM.

3.1 Módulo de información gerencial: SIU-Wichi

Desde el año 1996 el SIU desarrolla soluciones informáticas que brindan servicios tecnológicos al sistema universitario nacional argentino y otros organismos de gobierno [43].

Como respuesta a las necesidades de las distintas instituciones educativas que conforman el SIU, éste desarrolló numerosos sistemas transaccionales que sirvieron de soporte en la gestión diaria de una institución en sus distintos aspectos operacionales.

Entre ellos se destacan el Módulo Económico, Presupuestario, Financiero y Contable (SIU-Pilagá), Módulo de Gestión Académica (SIU-Guaraní), Módulo de Recursos Humanos (SIU-Mapuche), Módulo de Compras, Contrataciones y Patrimonio (SIU-Diaguíta), Módulo de Gestión de Encuestas (SIU-Kolla), Módulo de Estadísticas de Alumnos (SIU-Araucano), Módulo de Prestaciones Generales de Bibliotecas (SIU-Bibliotecas), Módulo de Facturación y Cobranzas (SIU-Sanavirón/Quilmes), entre otros [1].

Una vez cubierta la demanda básica de las instituciones en sus distintas áreas de gestión, las mismas comenzaron a solicitar otras prestaciones como el desarrollo de sistemas de gestión que aporten al proceso de toma de decisiones por parte de las autoridades.

En el año 2003, el SIU comenzó a trabajar en el desarrollo y la implementación de herramientas de Inteligencia de Negocio en el ámbito universitario.

En el año 2008 desarrollaron los primeros módulos que tuvieron como objetivo abarcar la mayoría de los datos generados a través de los sistemas transaccionales mencionados anteriormente, permitiendo una rápida inserción, comprensión y aceptación entre los usuarios de los distintos sectores.

A pesar de ello, el SIU detectó que para las necesidades particulares de los usuarios resultaba más sencillo trabajar con datos parciales y no con la totalidad de los mismos. Es por ello que impulsó, como alternativa, comenzar a trabajar con subconjuntos de datos mediante la construcción de Data Marts [43].

Surge entonces el módulo de información gerencial SIU-Wichi como herramienta de soporte para la toma de decisión en las instituciones universitarias. El mismo permitió visualizar y analizar, de manera integrada los datos provenientes de los sistemas transaccionales, buscando aportar al proceso de toma de decisiones por parte de las gerencias de las distintas áreas: académicas, administración y planificación.

El módulo SIU-Wichi utiliza un conjunto de Data Marts para el procesamiento de sus datos y la generación de información, permitiendo el análisis multidimensional de los mismos.

Los beneficios que aporta esta herramienta para los mandos medios y altos son numerosos. Suministra información consolidada, permite construir indicadores y evaluar tendencias, ofrece imágenes integradas, posibilita acceder a los datos de manera fácil, rápido y en un lenguaje que puede ser interpretado fácilmente por el usuario final.

En la UNaM el sistema SIU-Wichi se encuentra implementado desde el año 2013 mediante la Resolución Consejo Superior Nro.:1349/13 [5]. El mismo se encuentra en funcionamiento con los siguientes sistemas transaccionales de gestión: SIU-Pilaga, SIU-Mapuche, SIU-Diaguita y SIU-Guaraní.

En el año 2015, mediante la Resolución del Consejo Superior Nro.: 035/15 [5] se implementan en la UNaM los circuitos administrativos necesarios para recopilar información de estos sistemas de cada unidad académica, e incorporarlos al sistema SIU-Wichi.

3.2 Proceso utilizado por el SIU para el desarrollo de Data Marts

En la actualidad, el SIU no presenta documentación publicada sobre la metodología utilizada para el desarrollo de productos Data Warehouse y en particular, en lo referente al desarrollo de Data Marts. Es por ello que surge la necesidad de relevar información referida al desarrollo de estas actividades por parte del mismo.

Para ello se realizó una búsqueda exhaustiva de información en la web, se realizaron consultas vía correo electrónico a representantes técnicos del SIU (coordinador técnico y personal técnico del equipo SIU-Wichi) y en el foro de la comunidad sobre cuáles eran las actividades que realizaban para el desarrollo de soluciones OLAP.

Entre las fuentes de datos consultadas se pueden mencionar:

- Boletines electrónicos informativos del SIU: existen publicados un total de 118 números desde el año 2005 hasta el mes de Octubre del 2020 [44].
- Publicaciones realizadas por el equipo SIU-Wichi [43].

- El portal de servicios de la Comunidad SIU: en el cual se encontró bibliografía relacionada a Inteligencia de Negocios, metodologías de desarrollo de Data Warehouses e información relacionada, que proporcionan al usuario el conocimiento necesario para comprender la potencialidad de la herramienta y su administración mediante una interface web [1]. Toda la documentación recomendada por el SIU se la puede encontrar en formato blog, tesis, libros, notas y videos.
- Sitio oficial del foro de la Comunidad SIU [45].

En base a la documentación relevada en las distintas fuentes de datos y principalmente, los aportes realizados por el equipo técnico que conforma el SIU-Wichi, se pudieron identificar los procesos involucrados en el desarrollo de soluciones de Data Warehouse por parte del SIU.

A continuación se describen cada uno de los procesos:

1. Relevamiento y Análisis

El primer paso del proceso se realiza con los referentes de cada proyecto, a nivel institución educativa y analistas del Consorcio SIU (SIU-Guaraní, SIU-Mapuche, SIU-Pilagá, etc.) que son quienes conocen los datos y el negocio. En este paso se realiza la definición de medidas y dimensiones del negocio que se quiere modelar. Por ejemplo: académico, recursos humanos, presupuesto, infraestructura edilicia, etc.

Posteriormente, se genera una primera versión del cubo que se valida con los usuarios de dichos sistemas de diferentes formas de acuerdo a cada circunstancia y al negocio que haya modelado.

Lo más habitual es que dicha validación lo hagan en los talleres que se dictan a lo largo del año y/o también en reuniones puntuales con algunos usuarios.

Básicamente, para relevar el interés existente en lo que refiere a los temas académicos y graduados han utilizado encuestas como herramientas no informáticas de recolección de datos.

2. Diseño de la estructura de la Base de Datos

De acuerdo a las medidas y dimensiones relevadas en el paso anterior, se definen la/s tabla/s de hechos y las tablas de dimensiones respectivamente, utilizando un esquema estrella.

En esta instancia se determina la granularidad que tendrá/n la/s tabla/s de hechos.

3. Procesos ETL

Básicamente en este paso se desarrollan las consultas y los procesos que permiten extraer los datos desde los sistemas de gestión (SIU-Guaraní, SIU-Mapuche, SIU-Pilagá, SIU-Querandíes, etc.). Este proceso se realiza generando archivos ‘txt’ que luego son importados a esquemas de base de datos temporales (tmp), DSA y finalmente al Data Warehouse.

El desarrollo de los procesos de extracción, transformación y carga se realiza habitualmente con la herramienta Pentaho Data Integration.

4. Desarrollo del modelo multidimensional

En este proceso se realiza el diseño y desarrollo del cubo. En el caso de Pentaho Data Integration, es un archivo ‘.xml’ que contiene las dimensiones y las medidas definidas apuntando a la estructura de base de datos definida en el proceso anterior.

5. Testeo

Durante el testeo se controla el funcionamiento de los procesos. Se cargan los datos al Data Warehouse y se validan los mismos con los datos del sistema de gestión origen.

En este proceso siempre surgen correcciones a realizar en la estructura de la base de datos, en los procesos de extracción, transformación y carga diseñados o en el modelo (archivo xml).

6. Documentación

En este proceso se genera la documentación técnica y funcional para la wiki de la comunidad del sistema de información universitaria, por versión.

La misma puede ser encontrada en el portal de la Comunidad SIU [46].

7. Versionado

Una vez cumplimentados los procesos anteriores, y realizadas todas las correcciones necesarias, se libera la versión actualizando la documentación correspondiente a la misma.

En la Figura 13 se observa cada uno de los procesos definidos anteriormente:

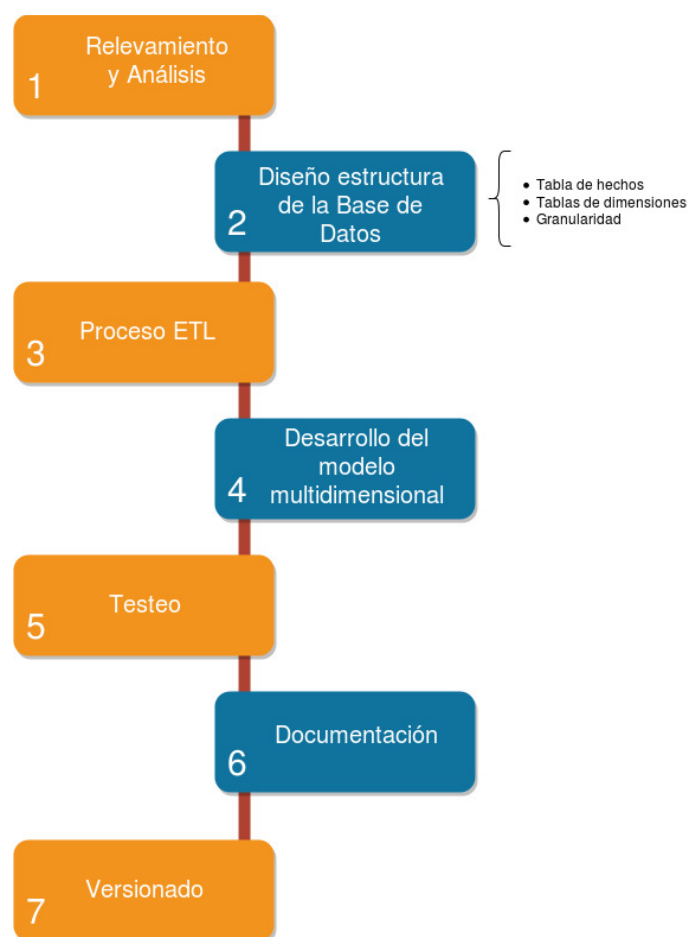


Figura 13: Procesos utilizados por el SIU para el desarrollo de Data Warehouses [Elaboración propia]

3.3 Herramientas utilizadas por el SIU para el desarrollo de Data Marts

Las herramientas informáticas utilizadas por el SIU para el desarrollo de Data Marts son las siguientes: Pentaho Data Integration, Schema Workbench, Report Designer de Pentaho Suite, motor de base de datos PostgreSQL¹², gestor de ticket de la herramienta Track¹³ y como gestor de requerimientos y solicitudes que realizan las universidades el Gestor de Solicitudes¹⁴ (GDS).

Como elaboración propia, el SIU cuenta con un instructivo de “Estándares de desarrollos para el SIU-Wichi” [46], donde se especifican y definen los estándares para el diseño y esquemas de bases de datos SIU-Wichi según: el sistema o fuente de datos a utilizar, el nombre de tablas y

¹² PostgreSQL (<https://www.postgresql.org/>)

¹³ Track (<https://colab.siu.edu.ar/trac/>)

¹⁴ GDS (<https://herramientas.comunidad.siu.edu.ar/>)

atributos de las tablas, registro para valores nulos o vacío/sin referencia en la dimensión, dimensiones comunes, nombres de esquemas y cubos en Pentaho Data Integration, nombre de dimensiones, niveles y medidas, descripciones en dimensiones, extracción de datos y los procesos de extracción, transformación y carga.

El módulo de consultas gerenciales SIU-Wichi utiliza como lenguaje de programación Java y como motor de base de datos PostgreSQL. El mismo está desarrollado sobre Pentaho y los cubos pueden ser analizados con la herramienta Saiku (consola de usuario de Pentaho).

3.4 Arquitectura actual utilizada por el SIU para el desarrollo de Data Marts

La arquitectura utilizada por el SIU para el desarrollo de Data Marts se la describe a continuación: como sistema fuente de origen utiliza los sistemas transaccionales de gestión como ser el SIU-Mapuche (Módulo de Recursos Humanos), SIU-Pilaga (Módulo Económico, Presupuestario, Financiero y Contable), SIU-Guaraní (Módulo de Gestión Académica), entre otros. La integración de los mismos se realiza de manera aislada y no integrada.

Luego mediante un conjunto de procesos ETL se extraen los datos de los sistemas fuentes, y se los transforma en un área temporal de almacenamiento. Luego, los datos transformados son cargados en Data Marts para que sean consultados y analizados por los usuarios mediante tableros y reportes OLAP.

En la Figura 14 se muestran los principales componentes de la arquitectura utilizada por el SIU para el desarrollo de Data Marts.

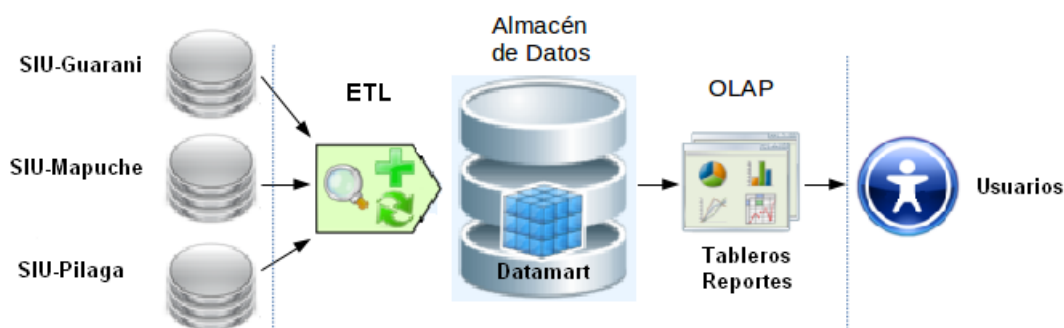


Figura 14: Arquitectura utilizada por el SIU para el desarrollo de Data Marts [Elaboración propia].

3.5 Evaluación de la situación actual

El módulo de información gerencial SIU-Wichi se encuentra implementado en numerosas Universidades Nacionales de nuestro país. En la Figura 15 se observa el mapa de implementaciones publicado en el sitio oficial del SIU [46].



Figura 15: Mapa de implementaciones del SIU-Wichi en el territorio de la República Argentina [46]

A pesar de ello, existen necesidades de información en distintas instituciones del país que no son cubiertas totalmente por el diseño actual que presentan los Data Marts incorporados al SIU-Wichi. Esto, debido a la falta de participación activa de los expertos en los distintos procesos de negocios que se modelan.

Esta situación también se ve reflejada en la validación de los Data Marts resultantes. Situación que deriva y resulta en lo antes expuesto, donde el Data Mart desarrollado no satisface muchas veces las necesidades inicialmente planteadas. Como ejemplo de ello, surge la necesidad de contar con el rango etario de los graduados de una institución a la hora de analizar los datos

relacionados a estos; dicha demanda es una solicitud realizada por numerosas universidades que no es contemplada actualmente, en ninguno de los Data Marts desarrollados.

Otro ejemplo de lo antes mencionado, se ve reflejado en la necesidad de identificar de manera rápida e intuitiva la ubicación geográfica de los alumnos al momento de graduarse y la situación laboral que poseen, de manera tal, de contribuir con el “Programa de inserción laboral profesional” [47] en desarrollo en la UNaM y articulado con las oficinas de graduados de cada Unidad Académica. Este programa, tiene como objetivo promover la incorporación y participación de los mismos en áreas de trabajo vinculadas con el sector privado y el gobierno provincial.

Cabe mencionar que, adaptar los Data Marts que provee el SIU a las necesidades propias de cada institución se vuelve problemático dado que no se encuentra una guía oficial de referencia para personalizar uno existente o en su defecto desarrollar uno nuevo.

Otra situación detectada se encuentra relacionada directamente al análisis de los datos llevada a cabo de manera aislada y no integrada, limitando el cruzamiento de los datos provenientes de distintos sistemas fuentes. A pesar que el sistema SIU-Kolla recaba mucha información relacionada al graduado, la información específica sobre su rendimiento académico solo se lo encuentra en el sistema SIU-Guaraní.

Por otro lado, en lo que refiere a la definición y gestión de los requisitos, no se llegan a cubrir aspectos necesarios para una adecuada determinación del origen, comprensión, registración, implementación y gestión de cambios, asociado a los mismos.

Particularmente, la UNaM cuenta con la implementación del modulo gerencial SIU-Wichi desde hace varios años, pero la inexistencia de un procedimiento formal de generación de conocimiento provoca que los sistemas instalados se utilicen solo para la carga de datos sin incluir el procesamiento de los mismos y la derivación hacia procesos de toma de decisión. Esto genera que se realicen esfuerzos en forma aislada en los distintos ámbitos de la universidad, siendo que existen herramientas disponibles para el abordaje de esta problemática.

A pesar que el SIU cuenta con un proceso de desarrollo para diseñar Data Warehouses, y que las universidades nacionales adheridas al consorcio SIU tienen la posibilidad de desarrollar sus propios Data Marts, se observa la ausencia de una adecuada metodología en el desarrollo de los mismos.

Esto dificulta poder cumplir y garantizar algunos requisitos fundamentales de los Data Marts como ser la flexibilidad, escalabilidad, integración de múltiples y heterogéneas fuentes de datos, coherencia de los datos, integración y perspectiva histórica de los mismos, entre otros.

Capítulo 4

Metodología propuesta para el desarrollo de Data Marts en Universidades Nacionales

4. Metodología propuesta para el desarrollo de Data Marts en Universidades Nacionales

En este capítulo se describe la metodología propuesta para el desarrollo de Data Marts, sus etapas y actividades. También se describe una estrategia dedicada a mantener la trazabilidad de los requisitos y el mantenimiento de los mismos.

4.1 Etapas de la metodología

Como propuesta metodológica se presenta un marco de trabajo para el desarrollo de Data Marts el cual consiste en un conjunto de etapas compuestas por actividades, definición de roles y responsabilidades; y un conjunto de técnicas y herramientas para dar soporte a los temas de análisis planteados.

Las etapas y actividades, se ejecutan de manera secuencial admitiendo posibles repeticiones correctivas si fuera necesario, incrementando el entendimiento y el resultado del producto final.

La propuesta metodológica se desarrolla en base a las buenas prácticas aplicadas en el modelo CMMI¹⁵ v1.3 para la gestión de requisitos.

Estas buenas prácticas ayudan a definir y comprender los requisitos, obtener el compromiso sobre los mismos, gestionar sus cambios, mantener la trazabilidad bidireccional y la validación de resultados haciendo participar en todo momento al experto en la temática, y a la persona que ha solicitado el tema de análisis.

Todos estos atributos de calidad, aplicados mediante la incorporación del modelo CMMI, dieron a la metodología propuesta la posibilidad de mantener la trazabilidad de los requisitos a lo largo de toda la metodología. Este aspecto importante, permite evaluar el impacto de los cambios de los requerimientos definidos sobre las actividades planificadas y el producto final.

Mantener una trazabilidad bidireccional de los requisitos ayuda a determinar si todos los requisitos fuentes han sido tratados en su totalidad hasta su nivel más bajo, y viceversa.

Para mantener dicha trazabilidad se sugiere utilizar un sistema de gestión de proyectos que permita a los usuarios crear un registro de sus tareas, asignar roles, gestionar los cambios y realizar un seguimiento de las mismas con respecto al tiempo.

¹⁵ CMMI: *Capability Maturity Model Integration*

Otro aspecto a considerar durante el desarrollo de la metodología es asegurar el lineamiento entre el producto final y los requisitos planteados. El mismo hace referencia a la práctica de encontrar inconsistencias entre los requisitos definidos, el tema analítico planteado y el producto final que se encuentra en desarrollo; planteando e iniciando acciones correctivas para resolverlas.

Esta actividad se aplica durante todo el desarrollo de la metodología y se gestiona a través de herramientas informáticas aplicadas a gestión de proyectos.

Para el desarrollo de la metodología propuesta se combinaron, integraron y redefinieron etapas en común que presentan los modelos de Kimball y HEFESTO para el desarrollo de Data Marts; y se agregaron nuevas etapas de control y aseguramiento de calidad.

El proceso definido en el modelo CMMI v1.3, permite contar con una base para la planificación, realización y mejoras de las tareas y actividades dentro del proyecto.

Por otro lado, se definen roles y responsabilidades que se van a desempeñar en las distintas etapas y actividades que se desarrollan en la metodología propuesta.

La metodología propuesta consta de cinco etapas:

- Etapa 1: Entendimiento del Negocio
- Etapa 2: Relevamiento y Análisis
- Etapa 3: Diseño y modelado multidimensional
- Etapa 4: Procesos ETL
- Etapa 5: Validación del modelo final

Cada etapa tiene definido un objetivo, entradas, actividades, usuarios participantes y salidas.

Del modelo propuesto por Kimball [16] se consideraron las siguientes etapas: “Planificación”, “Requerimientos del negocio”, “Modelado dimensional”, “Diseño físico” y “Diseño e implementación del subsistema ETL”.

Por otro lado, del modelo HEFESTO [9] se tuvieron en cuenta sus cuatro etapas principales: “Análisis de los requerimientos”, “Análisis de los OLTP”, “Modelo lógico del DW” e “Integración de datos”.

En la metodología propuesta, la primera etapa denominada **Entendimiento del Negocio** se fusionaron y reestructuraron las etapas de “Planificación” y “Requerimientos del Negocio” del modelo de Kimball y la etapa de “Análisis de los Requisitos” del modelo HEFESTO, con el agregado de la definición de los roles y las responsabilidades del equipo de trabajo, y la determinación de recursos disponibles.

La segunda etapa denominada **Relevamiento y Análisis** se encuentra basada principalmente en la etapa de “Análisis de los OLTP” del modelo HEFESTO. En esta etapa se

incorporó una actividad asociada al estudio de factibilidad y aceptación sobre los requerimientos planteados por parte del experto y los recursos disponibles, determinante para la continuidad del proyecto.

La tercer etapa de la metodología propuesta se denomina **Diseño y modelado dimensional** y es una propuesta basada en la integración y combinación de las etapas “Modelo dimensional” y “Diseño físico” del modelo de Kimball con la etapa “Modelo lógico del DW” del modelo HEFESTO.

La cuarta etapa denominada **Procesos ETL** resulta de la integración de la etapa “Diseño e implementación del subsistema ETL” del modelo de Kimball con la etapa de “Integración” de datos del modelo de HEFESTO.

Por último, se propone una quinta etapa no contemplada en los modelos relevados, denominada **Validación del modelo final** donde se validan los datos modelados, y su representación en el Data Mart desarrollado.

Durante el desarrollo de las distintas etapas de la metodología se van generando plantillas y documentación que aporta a la gestión de los requisitos permitiendo poner énfasis en el seguimiento y cambio de los mismos, además del control de versiones y control de cambios.

Como objetivo, esta metodología propone desarrollar Data Marts que contribuyan fehacientemente a la toma de decisión en las instituciones educativas, en base a los requerimientos y políticas definidas en las mismas.

La metodología propuesta utiliza el enfoque *botton-up* del modelo de Kimball, lo que significa que se construyen Data Marts para dar solución a los requerimientos de la institución en base a la problemática planteada.

4.1.1 E1: Entendimiento del Negocio

El objetivo de esta etapa es analizar las necesidades de negocio a partir de una problemática planteada, identificando quiénes serán los roles involucrados, el alcance del estudio, los recursos disponibles con los que cuenta la institución, la factibilidad operativa, económica y técnica, los temas analíticos que se abordarán y la planificación de las distintas actividades de desarrollo.

Entrada:

- ✓ Necesidad planteada por el usuario.

Actividades:

✓ E1.A1- Roles y responsabilidades

Identificar y definir quiénes serán los usuarios involucrados en el desarrollo del caso de estudio a analizar, en base a los roles y las responsabilidades definidas a continuación (Tabla 3):

Tabla 3: Definición de roles y responsabilidades.

Rol	Responsabilidades
Responsable Institución	- Entiende de las necesidades de información a nivel estratégico sobre las políticas y gestión de la institución. - Provee la visión y misión de la institución.
Experto / Usuario	- Manifiesta las necesidades de información de un área en particular. - Ayuda a obtener y validar las necesidades de información clave para el desarrollo del almacén de datos (Data Mart). - Interactúa con los demás roles para asegurar el cumplimiento de los objetivos del desarrollo.
Analista	- Define el alcance de la solución. - Selecciona y aplica técnicas de elicitación de requisitos. - Valida los requisitos relevados entre los miembros del equipo y aclara dudas que surjan. - Diseña el modelo conceptual. - Elabora la documentación con la información relevada en cada una de las etapas.

En base a los roles y las responsabilidades definidas anteriormente, se recomienda considerar la elección del equipo de trabajo en relación a los cargos y/u ocupaciones existentes en las universidades nacionales. Como por ejemplo: Responsable institucional: decano / vice decano de la Unidad Académica; Experto/Usuario: coordinador / director / responsable de proyectos / secretario / áreas / departamentos / programas; Analista: director / jefe / encargado del área de informática.

Los roles propuestos en esta actividad se encuentran establecidos en el organigrama institucional de cada casa de estudios.

Una vez identificados los roles y responsabilidades en la institución, conformar el equipo de trabajo cuya función recae en el responsable institucional.

De acuerdo a cada una de las Etapas definidas en la metodología propuesta los roles intervinientes en cada una de ellas se especifican en el siguiente cuadro (Tabla 4):

Tabla 4: Definición de roles intervinientes en cada etapa.

Etapas	Roles intervinientes
E1: Entendimiento del Negocio	Responsable institucional Experto / Usuario Analista
E2: Relevamiento y Análisis	Responsable institucional Experto / Usuario Analista
E3: Diseño y modelado multidimensional	Analista
E4: Procesos ETL	Analista
E5: Verificación y Validación del modelo final	Responsable institucional Experto / Usuario Analista

✓ E1.A2- Elicitación de Requerimientos

Descubrimiento de los requerimientos de usuarios

En esta actividad se descubren los requerimientos de los usuarios, de manera tal de comprender las necesidades de información de la institución e identificar los conceptos que sean relevantes para el análisis que permitan estudiar la información desde diferentes perspectivas.

El analista y el experto relevan las necesidades y los requisitos de información de la institución, con el acompañamiento del responsable institucional.

Para llevar a cabo esta actividad se eligen y seleccionan técnicas (grupales, tradicionales, etc.) y herramientas de educación (entrevistas, cuestionarios, reuniones, *Brainstorming*¹⁶, etc.) [48].

El analista genera un Documento de Especificación de Requerimientos de Usuario (DERU) por cada entrevista realizada.

Este documento corresponde a una planilla diseñada para reconocer y entender las necesidades del experto/usuario. En la Tabla 5 se puede observar el diseño de la plantilla que contiene: un número identificador del documento, el nombre del analista que realiza el relevamiento, la fecha, la institución involucrada, el experto consultado, el cargo que ocupa el experto y el detalle del relevamiento realizado.

¹⁶ *Brainstorming*: Conocido también como “Tormenta de ideas”, es una técnica grupal de educación que favorece la aparición de nuevas ideas sobre un problema o tema concreto.

Tabla 5: Documento de Especificación de Requerimientos de Usuario (DERU)

Documento de Especificación de Requerimientos de Usuario	
DERU Nro.:	[número identificador]
Analista	[nombre del analista]
Fecha	[fecha de relevamiento]
Institución	[institución involucrada]
Experto consultado	[nombre del experto]
Cargo	[cargo que ocupa el experto]
Información relevada:	[detalle del relevamiento]

✓ E1.A3- Definición del Alcance

En esta actividad se define el alcance del proyecto especificando el área a abordar, los temas a tratar, los sectores que estarán involucrados y la institución que intervendrá en el desarrollo del mismo.

✓ E1.A4- Definición de Temas Analíticos

En esta actividad el analista junto al experto definen los Temas Analíticos (TA) que serán desarrollados y analizados en base a los requerimientos y necesidades de información relevados durante la actividad E1.A2.

Cabe aclarar que un tema analítico es una expresión sobre una necesidad de información que incluye variables de análisis relevantes, las cuales permiten estudiar la información desde diferentes perspectivas.

La planilla diseñada para definir un tema analítico puede observarse en la Tabla 6., en la cual se define un identificador junto a una descripción del tema analítico a analizar, el número del Documento de Especificación de Requerimientos de Usuario asociado, el rol del usuario/experto quien solicita el tema a analizar (Cómo), el objetivo de esta solicitud (Quiero) y cuál es el beneficio de lo solicitado (Para).

Tabla 6: Formato propuesto para definir los temas analíticos a desarrollar.

TEMA ANALITICO	
[identificador del tema analítico]	[descripción del tema analítico]
DERU Nro.:	[DERU asociado]
Cómo	[rol del usuario]
Quiero	[objetivo]
Para	[beneficio]

✓ E1.A5- Modelado preliminar

En esta actividad, el analista procede a conceptualizar los requisitos relevados en cada tema analítico definido en la actividad anterior y representarlos en un Modelo Conceptual para un mejor entendimiento de los mismos.

En la Figura 16 se observa el diseño del modelo conceptual en el cual, del lado izquierdo se ubican las perspectivas y del lado derecho los indicadores posibles a analizar.

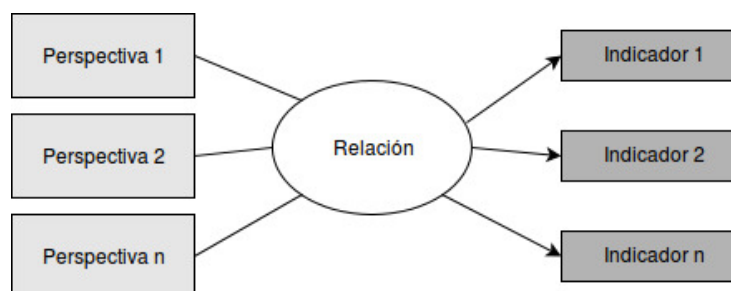


Figura 16: Modelo Conceptual preliminar [6]

La relación, a través de la cual se unen las diferentes perspectivas y se obtiene como resultado los indicadores requeridos por los usuarios, constituye el tema analítico seleccionado para el análisis.

Junto al diseño del modelo conceptual, se comienza a elaborar un Diccionario de Datos (DD) con las características lógicas de los distintos conceptos relevados, atributos que caracterizan dicho concepto, significado, contenido y posible origen de los datos (sistema fuente).

Durante el desarrollo de esta actividad, el analista descubre e identifica las perspectivas e indicadores de análisis que intervendrán, en base al Diccionario de Datos y al tema analítico que se está analizando.

En la Tabla 7 se puede observar el diseño del Diccionario de Datos propuesto.

Tabla 7: Diccionario de Datos (DD)

Diccionario de Datos			
DERU Nro.	[DERU asociado]	Tema analítico: [descripción del tema analítico]	
Fecha	[fecha de relevamiento]	Analista: [nombre del analista]	
Atributos	Significado	Contenido	Sistema Fuente
[atributo]	[descripción]	[contenido]	[fuente de dato]

✓ **E1.A6- Determinación de recursos**

La determinación de recursos para el estudio de factibilidad que se propone está basado en tres áreas específicas: técnica, económica y operacional.

- Factibilidad técnica: El analista evalúa y determina si los recursos técnicos actuales pueden satisfacer las necesidades planteadas. Dentro de los recursos técnicos a analizar se encuentra: equipamiento, accesibilidad y disponibilidad de los datos.
- Factibilidad económica: Los recursos económicos a ser considerados se encuentran basados principalmente en el costo estimado del software y el costo estimado del hardware, requeridos para el desarrollo del presente proyecto.
- Factibilidad operacional: La factibilidad operacional involucra el recurso humano necesario para llevar adelante el proyecto. El mismo, puede estar basado en la estructura organizacional de la institución.

Cabe aclarar, que para el estudio de la factibilidad económica, los recursos económicos pueden no ser evaluados en caso que ya se encuentre implementado un DW en la institución.

✓ **E1.A7- Planificación**

En esta actividad se realiza la planificación inicial de cada una de las etapas en base al alcance definido, los recursos disponibles y las necesidades de información planteadas.

Salidas:

- ✓ Temas analíticos generados como resultado del trabajo con el experto.
- ✓ Modelo Conceptual preliminar.
- ✓ Diccionario de Datos.

En la Figura 17 se observan las actividades de la Etapa 1 junto a las plantillas generadas en cada una de ellas.

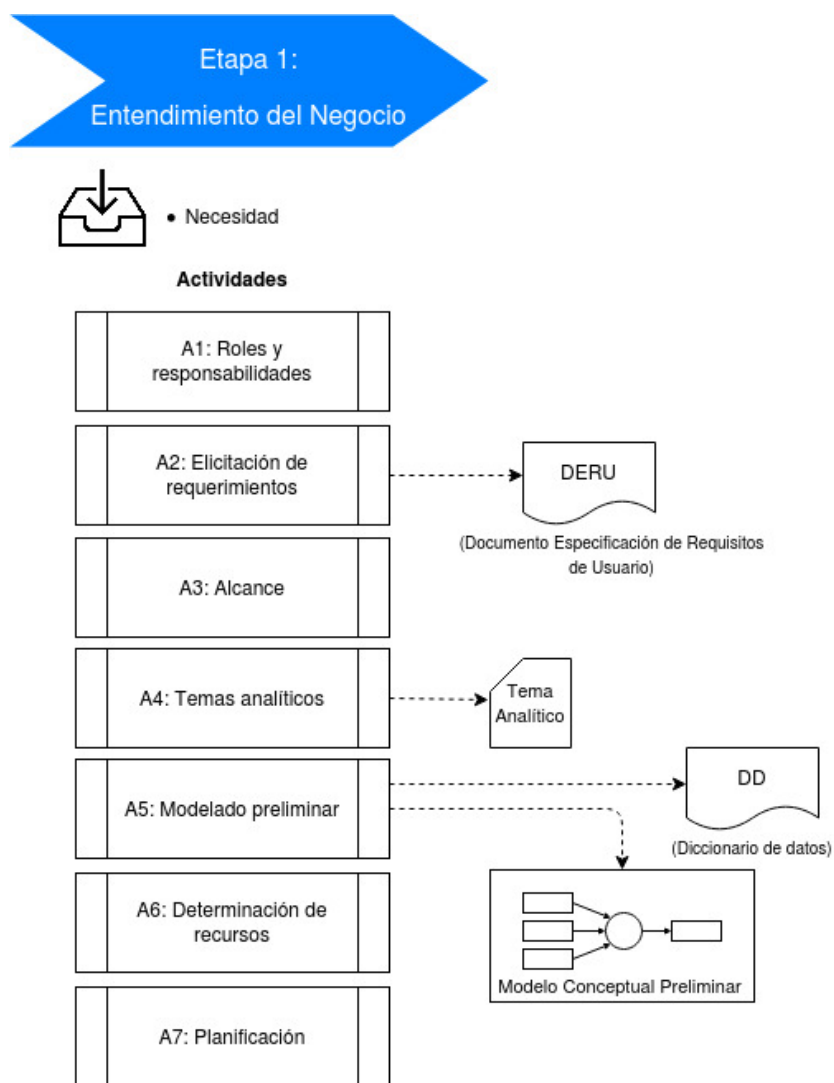


Figura 17: Actividades de la Etapa 1 [Elaboración propia]

4.1.2 E2: Relevamiento y Análisis

El objetivo de esta etapa es relevar y analizar con mayor profundidad las distintas fuentes de datos tanto internas como externas a utilizar en base al/los temas analíticos definidos en la etapa anterior. Determinar cómo serán calculados los indicadores y cómo se establecerán las respectivas correspondencias entre el modelo conceptual preliminar y las fuentes de datos existentes.

Luego se confecciona un modelo conceptual ampliado donde se puede visualizar el resultado obtenido en esta etapa.

Finalmente, se evalúa la factibilidad y se obtiene la aceptación del modelo por parte del experto.

Entrada:

- ✓ Temas analíticos generados como resultado del trabajo con el experto.
- ✓ Modelo Conceptual preliminar.
- ✓ Diccionario de Datos.

Actividades:

- ✓ E2.A1- **Análisis de los datos**

En esta actividad, el analista examina el diccionario y las fuentes de datos disponibles, que contengan la información requerida, de manera tal que se puedan identificar las correspondencias entre el modelo conceptual preliminar y las fuentes de datos.

- ✓ E2. A2 - **Mapeo al modelo conceptual**

En esta actividad el analista genera una plantilla de Identificación de las Fuentes de Datos (FUDA) donde se ubican las distintas fuentes de datos disponibles para cada perspectiva identificada.

El diseño de esta plantilla permite contener e identificar las distintas fuentes de datos de donde extraer la correspondencia con las perspectivas relevadas. En la Tabla 8 se puede observar el diseño de la plantilla donde se especifica un número identificador del tema analítico a analizar, el nombre del analista interviniente, el nombre y versión de la fuente de datos de donde extraer la información, tablas, atributos, perspectivas, atributos de la perspectiva y los indicadores.

Tabla 8: Identificación de las Fuentes de Datos (FUDA)

Identificación de las Fuentes de Datos						
Tema analítico	[tema analítico a analizar]					
Analista	[nombre del analista]					
Fuente de Datos	Versión	Tabla	Atributo	Perspectiva	Atributo Perspectiva	Indicador
<i>Fuente de Datos 1</i>	<i>Versión xx</i>	<i>Tabla 1</i>	<i>Atributo a</i>	<i>Perspectiva A</i>	<i>Atributo b</i>	<i>Indicador a</i>

Se sugiere para la selección de las fuentes de datos, considerar la factibilidad de las siguientes características:

- Disponibilidad: acceso a la fuente de datos.
- Frecuencia de actualización: frecuencia de actualización de los datos.
- Formato de presentación.

✓ E2.A3 - Cálculo de los indicadores

En esta actividad el analista procede a explicitar la manera que se van a calcular los indicadores, definiendo para cada uno de ellos los siguientes conceptos:

Indicador: nombre del indicador.

Hecho/s que lo componen más su fórmula de cálculo.

Ejemplo: Hecho 1 + Hecho 2

Función de sumarización que se utiliza para su agregación.

Ejemplo: SUM, COUNT, AVG, etc.

Descripción: descripción del indicador.

En la Tabla 9 se observa el diseño de la plantilla donde se constituyen los indicadores del tema analítico asociado.

Tabla 9: Conformación de indicadores (CI)

Conformación de Indicadores			
Tema analítico	[tema analítico a analizar]	Fecha	[fecha de relevamiento]
Indicador		Fórmula	Descripción
[nombre del indicador]		[fórmula de calculo]	[descripción del indicador]

✓ E2.A4-Nivel de granularidad

Una vez establecidas las correspondencias con las fuentes de datos, se debe especificar el nivel de granularidad que se desea obtener seleccionando los campos que contendrá cada perspectiva. A través de los mismos se examinarán y filtrarán los indicadores, influyendo en el tipo de análisis que se podrá realizar.

✓ E2.A5-Perspectiva “Tiempo”

En esta actividad se define el nivel de granularidad de la perspectiva Tiempo. Esta definición determina la manera en la que se agruparán o sumarán los datos.

Ejemplo: año, semestre, cuatrimestre, trimestre, mes, quincena, semana, día.

✓ E2.A6- Modelo Conceptual ampliado

Terminada la actividad anterior, se procede a la elaboración del Modelo Conceptual ampliado, colocando debajo de cada perspectiva los campos identificados como relevantes y debajo de cada indicador su fórmula correspondiente (Figura 18).

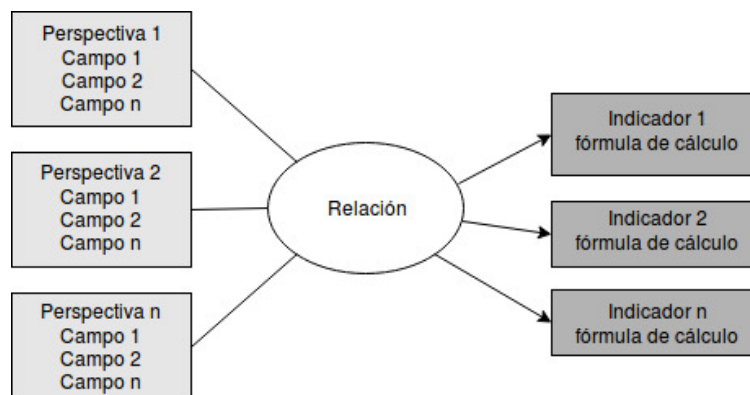


Figura 18: Modelo Conceptual ampliado [3]

✓ E2.A7- Factibilidad y aceptación del modelo

En esta actividad se procede a verificar la factibilidad del modelo conceptual diseñado en la actividad anterior con el experto. El objetivo de esta actividad es asegurar que los datos, con los que se cuentan y proporcionan al experto, cubran sus expectativas de información.

Si la verificación no resulta positiva, y las necesidades de información del experto no pueden ser cubiertas por falta de adecuación en los datos o, por inexistencia de los mismos, el analista elabora un informe técnico con sugerencias y modificaciones a ser aplicados en los sistemas fuentes existentes, o recomendaciones para la implementación de uno nuevo. El mismo es elevado al experto y al responsable institucional para su análisis.

El objetivo de estas sugerencias es aportar a la institución mejoras que ayuden a contar con atributos apropiados que permitan relevar la información necesaria a través de un método, como así también recomendar la implementación de nuevos circuitos institucionales que permitan recabar los datos faltantes.

Estas recomendaciones tienen su fundamento en el análisis realizado en las actividades anteriores sobre: la estructura de los sistemas fuentes, los procedimientos de carga y actualización de los mismos, la configuración de los campos y su obligatoriedad de carga.

Salidas:

- ✓ Identificación de las Fuentes de Datos.
- ✓ Modelo Conceptual ampliado.
- ✓ Informe técnico (opcional).

En la Figura 19 se observan las actividades de la Etapa 2 junto a las plantillas generadas en cada una de ellas.

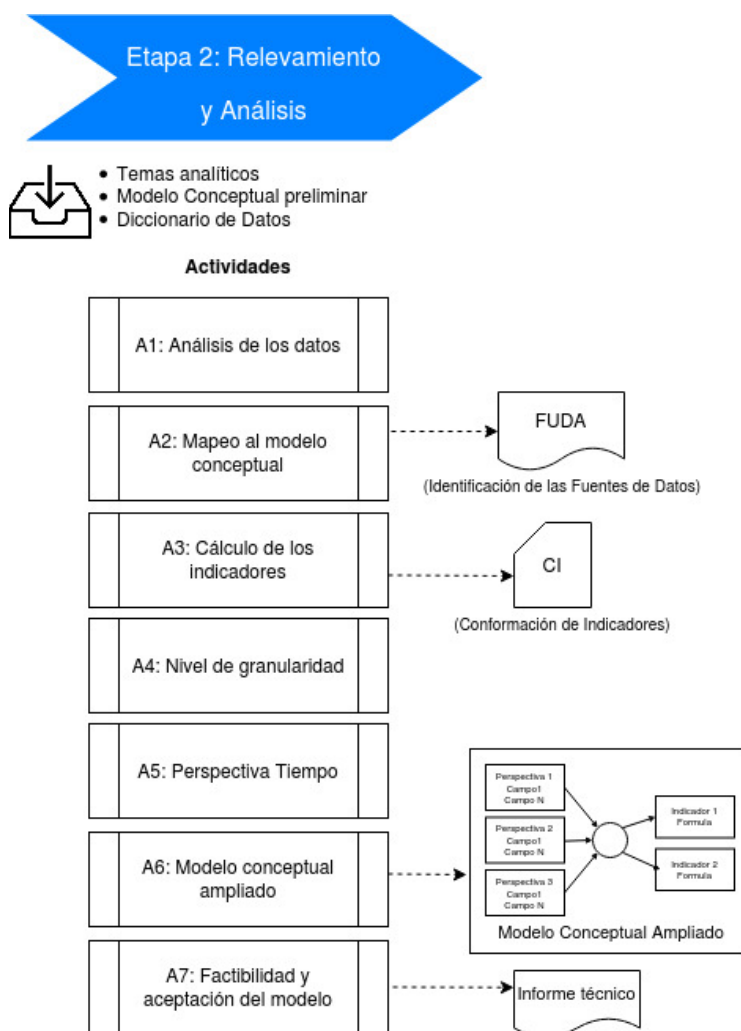


Figura 19: Actividades de la Etapa 2 [Elaboración propia]

4.1.3 E3: Diseño y Modelado Multidimensional

El objetivo de esta etapa es diseñar el modelo lógico del Data Mart. Definir el tipo de esquema que se va a utilizar, las tablas de dimensiones, la tabla de hechos y las uniones que habrá entre las mismas. Finalmente diseñar el modelo físico que tendrá el Data Mart.

Entrada:

- ✓ Identificación de las Fuentes de Datos.
- ✓ Modelo Conceptual ampliado.

Actividades:

✓ E3.A1- Esquema de modelado

En esta actividad, el analista realiza la selección del esquema de modelado multidimensional que se utilizará para elaborar el almacén de datos.

Esta selección debe ser la que mejor se adapte a los requerimientos y necesidades relevados en la etapa anterior.

✓ E3.A2- Diseño tabla de dimensiones

En esta actividad, el analista constituye las tablas de dimensiones. Se diseña una por cada perspectiva definida en el Modelo Conceptual ampliado.

En el modelo lógico cada tabla de dimensión debe ser identificada con un nombre único. Sus campos, deben ser nombrados de manera intuitiva y contemplar un campo adicional que represente su clave principal (Figura 20).

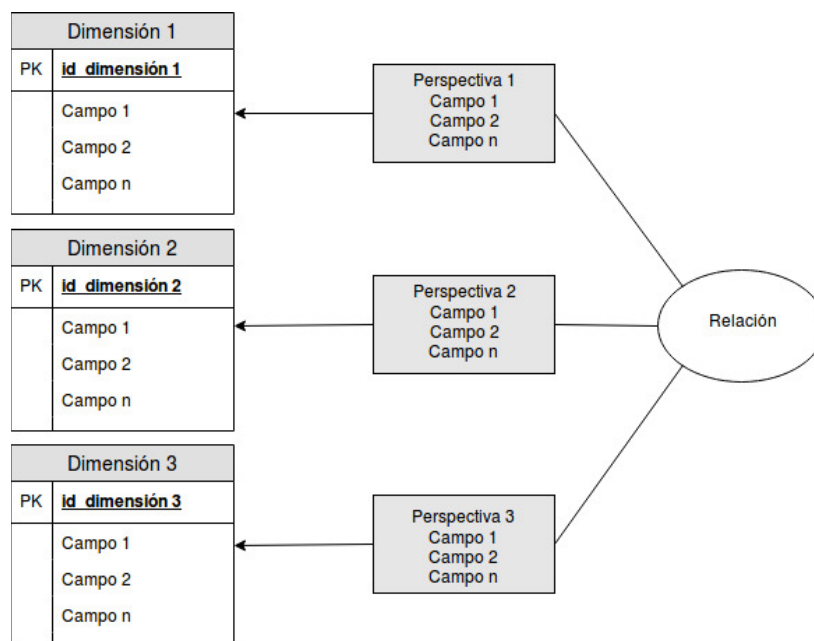


Figura 20: Diseño de una tabla de dimensiones [3]

Se elabora el documento de Definición de Tablas de Dimensiones (DTD), el cual corresponde a una planilla diseñada por el analista para contener los distintos campos existentes de una tabla de dimensiones y su correspondencia con dicha tabla y atributo de origen.

En la Tabla 10 se puede observar el diseño de la plantilla donde se especifica un número identificador del documento que corresponde al tema analítico a analizar, el nombre de la

perspectiva proporcionado en el modelo conceptual, el nombre de la dimensión, los campos de la dimensión y su tipo, la tabla de origen asociada y su atributo.

Tabla 10: Definición de Tabla de Dimensiones (DTD)

Definición de Tablas de Dimensiones			
Tema analítico	[tema analítico a analizar]		
Perspectiva	[nombre de la perspectiva del modelo conceptual]		
Dimensión	[nombre de la dimensión]		
Campo Dimensión	Tipo Campo	Tabla origen	Atributo origen
<i>Campo 1</i>	<i>Integer, Character, Char,....</i>	<i>Tabla 1</i>	<i>Atributo 1</i>
...			
<i>Campo n</i>			

✓ E3.A3- Diseño tabla de hechos

En esta actividad, el analista constituye la tabla de hechos que contendrá los hechos a través de los cuales se construirán los indicadores de las distintas dimensiones relacionadas.

En el modelo lógico, cada tabla de hecho debe ser identificada con un nombre único que represente la información a analizar. La clave principal estará compuesta por las claves primarias de cada tabla de dimensión relacionada y, tendrá tantos campos de hechos (medidas) como indicadores se hayan definido en el modelo conceptual ampliado (Figura 21).

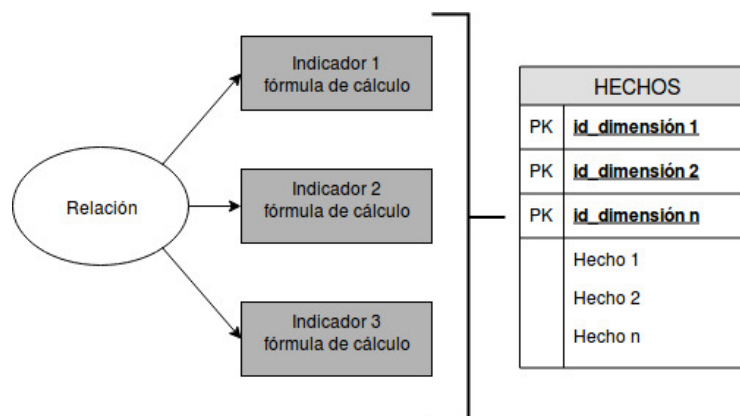


Figura 21: Diseño de una tabla de hechos [3]

El analista elabora una plantilla denominada Definición de Tablas de Hechos (DTH), para contener el detalle de la tabla de hechos en base al Modelo Conceptual ampliado modelado en la etapa anterior.

En la Tabla 11 se puede observar el diseño de la plantilla donde se especifica un número identificador del documento que corresponde al tema analítico a analizar, el nombre de la tabla de hechos, la clave principal de cada dimensión involucrada, una descripción de la misma, la medida utilizada y la manera que se calcula cada una, en base a los indicadores identificados.

Tabla 11: Definición de Tabla de Hechos (DTH)

Definición de Tabla de Hechos			
Tema analítico	[tema analítico a analizar]		
Hecho	[nombre tabla de hechos]		
Campo Dimensión	Descripción	Medida	Calculo
<i>id dimesion 1</i>	[descripción campo dimensión]	[medida utilizada]	[fórmula de cálculo]
...			
<i>id dimesion n</i>			

✓ E3.A4- Modelo Físico

Definidas las tablas de dimensiones y las tablas de hechos, se realizan las uniones correspondientes entre ellos.

Una vez realizadas las uniones, el analista diseña la estructura física necesaria para soportar el diseño lógico del almacén de datos (Figura 22).

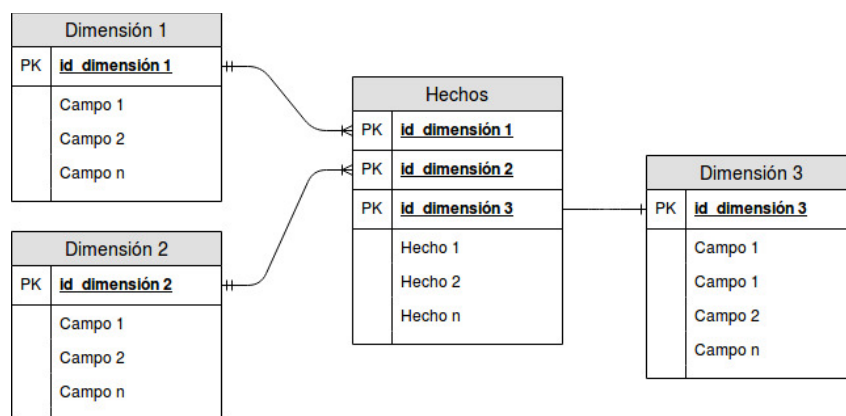


Figura 22: Diseño modelo físico [3]

Salida:

- ✓ Modelo físico de datos.

En la Figura 23 se observan las actividades de la Etapa 3 junto a las plantillas generadas en cada una de ellas.

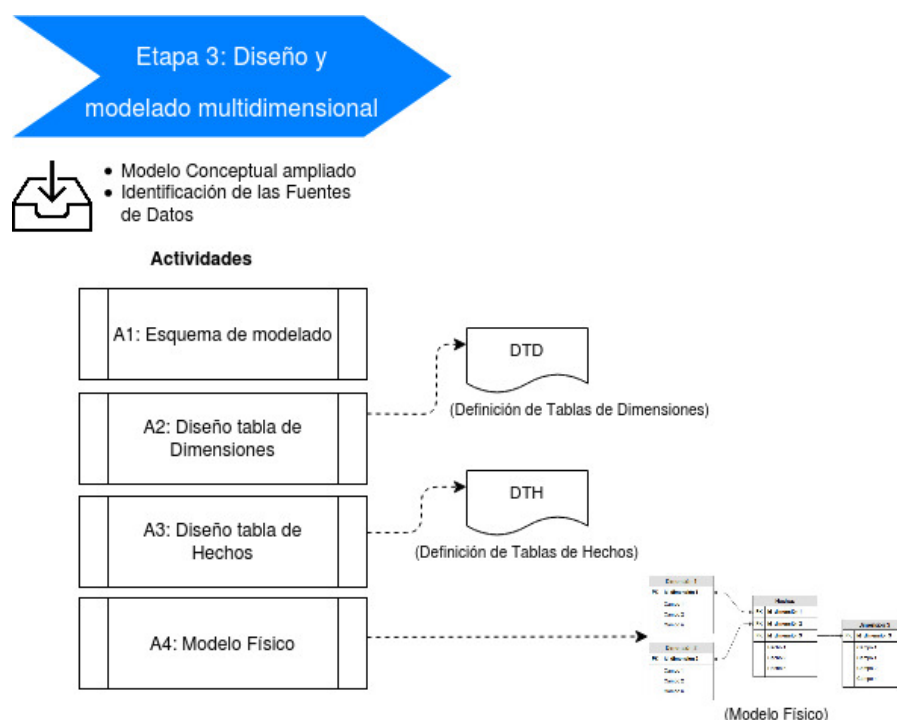


Figura 23: Actividades de la Etapa 3 [Elaboración propia]

4.1.4 E4: Procesos ETL

El objetivo de esta etapa es realizar la integración de los datos provenientes de las distintas fuentes de datos y realizar la carga a las estructuras de los Data Marts mediante procesos ETL.

En la Figura 24 se observa la arquitectura propuesta para el tratamiento de los datos, en el cual se toman los datos provenientes de los distintos sistemas fuentes (BD1, BD2) y se cargan en un área de almacenamiento temporal (DSA) sin ningún tipo de transformación. Para realizar la carga a la estructura correspondiente al Data Mart se obtienen los datos del DSA, se ejecutan un conjunto de procesos de extracción, transformación y limpieza, y se cargan los datos en las tablas correspondientes al modelo físico definido en la etapa anterior: Data Mart (DM).

Por último, se exploran los datos a través de herramientas de consultas, como ser reportes, análisis OLAP, Data Mining, etc.

En cada uno de estos procesos se aplican técnicas de verificación que garantizan una correcta selección de los datos a analizar, que los mismos sean correctos y válidos para el análisis, y que se carguen y actualicen correctamente.

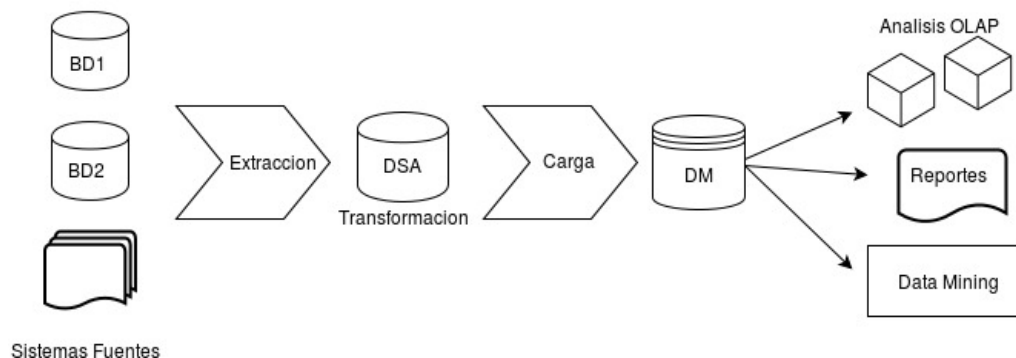


Figura 24: Arquitectura ETL / DM propuesta [Elaboración propia]

Entrada:

- ✓ Modelo físico de datos.
- ✓ Sistemas fuentes.

Actividades:

✓ E4.A1- Elección de la herramienta de software

En esta actividad se elige la herramienta de software a utilizar para crear los procesos ETL de carga inicial y de actualización de los datos.

✓ E4.A2- Fuentes de datos

En esta actividad se definen y establecen las fuentes de datos que contienen la información necesaria para el análisis.

✓ E4.A3- Selección de datos

En esta actividad se selecciona el conjunto de datos a ser utilizados para el análisis en base a: periodo de tiempo, volumen de datos, disponibilidad, etc.

La extracción de los datos, de las distintas fuentes de datos, se realiza mediante la generación de consultas SQL (Structured Query Language).

✓ E4.A4- **Área de almacenamiento temporal**

En esta actividad se crea un área de almacenamiento temporal (DSA) para almacenar temporalmente los datos obtenidos de las distintas fuentes de datos.

En caso que exista un DW implementado, para llevar a cabo el proceso de actualización, se crea un área de almacenamiento temporario (TEMP) para el almacenamiento de los datos de las distintas fuentes de datos sin procesar.

✓ E4.A5- **Tratamiento de los datos**

En esta actividad se crean procesos que serán los encargados de capturar, limpiar, depurar, filtrar, validar y cargar los datos al Data Mart correspondiente. La elaboración de los procesos ETL pueden llevar un 70% del esfuerzo en un proyecto de Inteligencia de Negocios.
[16]

Para ello considerar lo siguiente:

- Tratamiento de campos editables: homogeneizar el contenido de los campos configurados como editables, de manera que puedan ser procesados e interpretados por el usuario.
- Tratamiento de los datos nulos: insertar valores por defecto o estimar el impacto de los campos nulos dentro del contexto de análisis.
- Tratamiento de los datos duplicados: eliminar los datos considerados duplicados.
- Generación de claves subrogadas: generar claves subrogadas como identificador de cada tabla para obtener un mejor rendimiento en las consultas.
- Formato de los datos: modificar sintácticamente el contenido del campo sin alterar su significado.

✓ E4.A6- **Carga Inicial y Actualización**

Una vez que los datos son extraídos, limpiados y transformados, se los incorpora al Data Mart mediante un proceso de carga.

La actualización de los datos almacenados en el Data Mart, se realiza de acuerdo a las políticas de actualización definidas con el experto y el responsable de la institución.

✓ E4.A7- **Modelo multidimensional**

En esta actividad se crea el modelo multidimensional basado en el modelo lógico diseñado en la Etapa 3 (E3: Diseño y modelado multidimensional).

Salida:

- ✓ Un conjunto de procesos ETL que pueblan el Data Mart.
- ✓ Políticas de actualización del Data Mart.
- ✓ El Data Mart.

En la Figura 25 se observan las actividades de la Etapa 4.

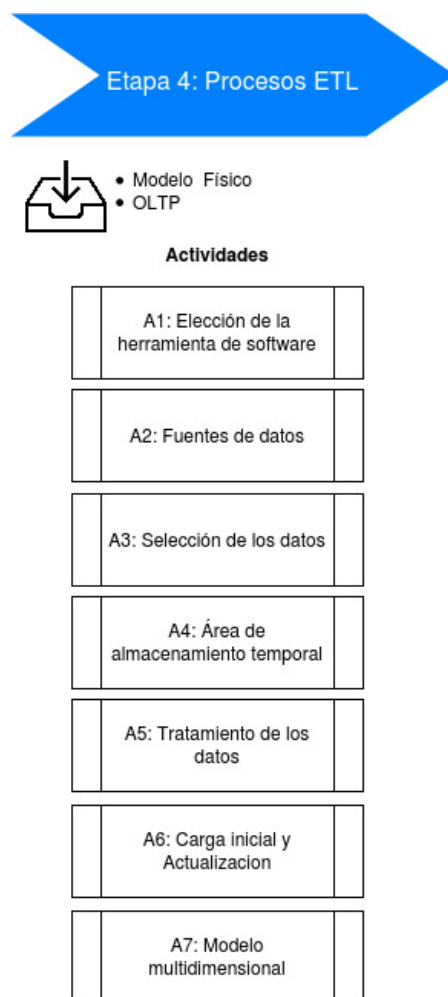


Figura 25: Actividades de la Etapa 4 [Elaboración propia]

4.1.5 E5: Validación del modelo final

El objetivo de esta fase es verificar y validar los datos modelados, y su representación en el Data Mart desarrollado.

La verificación y validación de los datos es llevada a cabo por el experto y el responsable institucional, en conjunto con el analista.

Entrada:

- ✓ El Data Mart.

Actividades:

✓ E5.A1- Verificación del Data Mart

Verificar si el Data Mart cumple con el beneficio planteado en la E1.A4, mediante la ejecución del modelo desarrollado, generando reportes que respondan al interrogante planteado.

✓ E5.A2- Validación del Data Mart

Validar los resultados obtenidos del Data Mart según el grado de satisfacción obtenida por los usuarios, ante el cumplimiento de los requerimientos planteados para cada caso.

Salida:

- ✓ Aceptación del Modelo Multidimensional (Data Mart)

En la Figura 26 se observan las actividades de la Etapa 5.

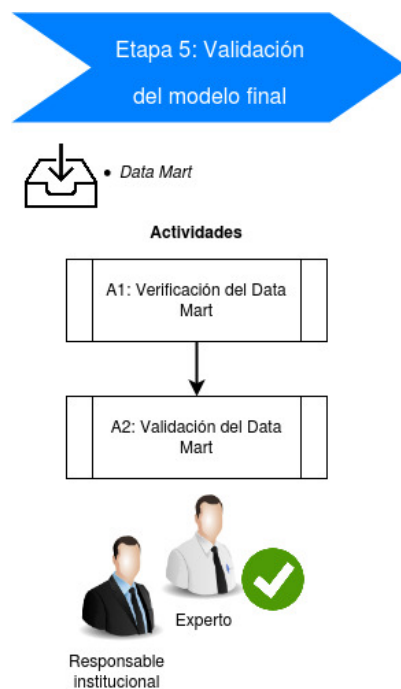


Figura 26: Actividades de la Etapa 5 [Elaboración propia]

Capítulo 5

Validación de la metodología en el caso de estudio propuesto

5. Validación de la metodología en el caso de estudio propuesto.

En este capítulo se describe el alcance de los sistemas de información implementados en la FCEQyN, la evolución del espacio de los graduados en el ámbito institucional junto a la implementación del módulo de encuestas SIU-Kolla, la readecuación del circuito administrativo para la implementación de la encuesta al recién graduado en la institución y la generación de un espacio web para los graduados. Luego se describen los problemas identificados relacionados al caso de estudio que se está analizando y finalmente se realiza la implementación de la metodología propuesta.

5.1 La Facultad de Ciencias Exactas, Químicas y Naturales

La FCEQyN de la UNaM tiene implementado numerosos sistemas informáticos que contribuyen a mejorar la gestión institucional.

Entre los sistemas informáticos implementados se encuentran: SIU-Guaraní (Módulo de Gestión Académica), SIU-Mapuche (Módulo de Recursos Humanos), SIU-Pilaga (Módulo Económico, Presupuestario, Financiero y Contable), SIU-Diaguíta (Módulo de Compras, Contrataciones y Patrimonio), SIU-Kolla (Módulo de Encuestas), entre otros.

Si bien existen estos sistemas transaccionales satisfacen las necesidades básicas operativas, los mismos no tienen una dinámica que se adapten a la realidad cambiante de la universidad.

El conocimiento, es un activo fundamental para la institución. El mismo produce ventajas competitivas, permitiendo optimizar el crecimiento y generar valor agregado a la misma.

En muchos casos ante una solicitud realizada desde el nivel gerencial de la unidad académica (Decanato, Secretarías u oficinas operativas) se busca una solución operativa; pero la gestión del conocimiento no llega a lograrse de manera eficiente porque los sistemas actuales no permiten el cruzamiento de datos ni la integración de los mismos.

Actualmente, en las distintas unidades académicas de la UNaM se realiza la gestión de la información sobre distintos sistemas transaccionales instalados en las mismas, pero no se utiliza esta gestión de información con un método de gestión de conocimiento que permita orientar de mejor manera el recurso humano y tecnológico existente dentro de la institución.

5.1.1 La Oficina de Vinculación con los Graduados

En el año 2011, la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU) a partir de la Resolución Nro.: 382-11 [2], incorpora políticas de seguimiento a los graduados en los procesos de acreditación de las carreras de grado y postgrado, la cual considera una pieza fundamental para el mejoramiento de las mismas.

En el año 2012, ante la necesidad surgida de acudir a un sistema de seguimiento, análisis y gestión de los graduados, la FCEQyN crea el “Programa de Seguimiento y Asistencia al Graduado” (Disposición Nro.: 828/12 [49]), para ofrecer alternativas de formación y/o articulación con la Universidad, registro de empleos y demandas laborales, asistencia sobre becas e inserción en el medio y asistencia sobre incumbencias profesionales. Dicho programa estuvo activo hasta el año 2019, donde la FCEQyN crea la “Oficina de Vinculación con los Graduados (OVIG)” mediante la Resolución Nro.: 040/19 [50] dependiente de la Secretaría de Extensión y Vinculación Tecnológica.

La OVIG tiene como finalidad promover actividades que permitan elevar los estándares de calidad de la institución, cubrir los requerimientos de la CONEAU y acompañar las políticas propuestas en el Plan de Desarrollo Institucional (PDI) de la FCEQyN [3] y de la UNaM [47] referidas a los graduados.

A pesar de que los programas generados para la asistencia y seguimiento de graduados comenzaron a surgir desde el año 2012, en la actualidad no se cuentan con procesos formales de comunicación con el Graduado que permitan conocer sobre su situación actual profesional, académica, económico-social y personal.

El seguimiento a los graduados es todavía, junto a otros temas, materia pendiente tanto a nivel FCEQyN como a nivel UNaM.

5.1.2 Módulo de Gestión de Encuestas: SIU-Kolla

El módulo de gestión de encuestas SIU-Kolla es un sistema provisto por el SIU que proporciona tres tipos de encuestas relacionadas al seguimiento de los Graduados de una Unidad Académica.

Las mismas se describen a continuación:

- **Encuesta al Recién Graduado:** Esta encuesta está dirigida al alumno que egresa de una carrera universitaria. Recaba información sobre la relación del graduado con la universidad, su plan de estudios y su situación laboral.

- **Encuesta al año de graduado:** Esta encuesta está dirigida al alumno que se graduó hace un año. Recaba información más detallada sobre la situación laboral del mismo, la relación de trabajo con los estudios cursados y el interés que tiene por seguir capacitándose (estudios de grado, posgrado o capacitación continua).
- **Encuesta a los 5 años de graduado:** Esta encuesta está dirigida al alumno que se graduó hace cinco años o más. Recaba información sobre su trayectoria laboral actual y pasada, otros estudios de grado y posgrado que haya realizado y su relación con la comunidad y la universidad.

En la FCEQyN, la implantación del sistema SIU-Kolla para el seguimiento y asistencia a los graduados se comenzó a utilizar recientemente. Por este motivo, la Encuesta al Recién Graduado es la única encuesta cuyo proceso de implementación fue aplicada hasta el momento.

Dicha encuesta es enviada vía correo electrónico al alumno que haya rendido su última materia y se haya acercado al Departamento de Títulos, Diplomas y Legalizaciones de la Dirección de Enseñanza de la facultad a realizar su trámite de egreso. El graduado recibe vía correo electrónico, el enlace de la encuesta y los datos de acceso a la misma.

En el SIU-Kolla, los grupos de encuestados correspondiente a los graduados se los configura de acuerdo a las Colaciones registradas en el sistema SIU-Guaraní. En dicho sistema, se configura una Colación por año calendario y se incluyen a los graduados que egresaron en ese año.

Las encuestas son disparadas semanalmente, a todos los graduados cuya completitud de la misma no se encuentre registrada en el sistema SIU-Kolla.

Los datos recabados a través de la Encuesta al Recién Graduado se encuentran divididos en trece bloques de preguntas. Cada uno de estos bloques contiene preguntas relacionadas al tema asociado al mismo.

Entre los bloques presentes se encuentran:

1. Bloque de Preguntas: DATOS PERSONALES
2. Bloque de Preguntas: RESIDENCIA
3. Bloque de Preguntas: PROCEDENCIA
4. Bloque de Preguntas: HÁBITOS
5. Bloque de Preguntas: NIVEL DE ESTUDIO ALCANZADO POR ... (Padre, Madre, Pareja)
6. Bloque de Preguntas: DATOS LABORALES
7. Bloque de Preguntas: TRABAJO ACTUAL
8. Bloque de Preguntas: SINO TRABAJA
9. Bloque de Preguntas: EN CASO DE ESTAR BUSCANDO TRABAJO

10. Bloque de Preguntas: EN CASO DE ESTAR BUSCANDO TRABAJO
11. Bloque de Preguntas: RELACIÓN CON LA UNIVERSIDAD
12. Bloque de Preguntas: SU RELACIÓN EN LA UNIVERSIDAD CON RELACIÓN A (Alumnado, Biblioteca, Laboratorios)
13. Bloque de Preguntas: OTROS ESTUDIOS

El modelo de la encuesta se encuentra disponible en el ANEXO 1.

5.1.3 Problemas identificados

El primer problema detectado en la FCEQyN es la carencia de datos existentes en relación a los graduados. Los datos recabados son escasos y de baja calidad, debido a que no se cuenta con un circuito administrativo que permita la actualización permanente de los mismos a través del tiempo.

A pesar que existen encuestas para los graduados generadas a través del sistema SIU Kolla, las mismas al no ser obligatorias, no son respondidas por los mismos. En su mayoría por falta de interés y desmotivación por no conocer cuál es la finalidad de las mismas.

La implementación de un circuito administrativo que permita el seguimiento a los graduados, es todavía materia pendiente en todo el ámbito de la UNaM.

El sistema SIU-Kolla permite recabar una gran cantidad de datos en cada una de las encuestas los cuales pueden ser muy valiosos, pero resultan muy difíciles de analizar debido a la configuración de carga que presentan los mismos. Esta dificultad se observa principalmente, en los campos configurados como editables dentro de las encuestas.

Para que los datos recabados sirvan como base para el análisis de la problemática, es necesario someterlos a un proceso previo de limpieza, transformación e integración para asegurar su calidad y consistencia de los mismos.

5.2 Circuito administrativo propuesto para la completitud de la Encuesta al Recién Graduado

Para garantizar la completitud de la Encuesta al Recién Graduado se sugiere la modificación del circuito administrativo llevado a cabo en el Departamento de Títulos, Diplomas y Legalizaciones de la Dirección de Enseñanza de la FCEQyN, incorporando como un nuevo

requisito para tramitar el título la presentación del comprobante de completitud de la encuesta por parte del graduado.

El circuito administrativo sugerido e implementado para la completitud de las encuestas por parte de los recién graduados es el siguiente:

- 1) La Dirección de Enseñanza, luego de cargar la aprobación de la última materia del alumno y marcar la completitud del plan de estudios, dispara la “Encuesta al Recién Graduado” para el alumno en cuestión.
- 2) El alumno debe responder obligatoriamente la misma, para obtener un comprobante de completitud de la encuesta (Figura 27) que deberá presentar en el Departamento de Títulos, Diplomas y Legalizaciones de la Dirección de Enseñanza para retirar su título. Es importante la aplicación de esta restricción para contar con datos actualizados.

Comprobante de encuesta respondida		
Encuesta al Recién Graduado		
Apellido y nombres	Toledo, Facundo Manuel	
Tipo de documento	Documento Nacional de Identidad	
Número de documento	99999999	
Fecha	07/06/2019 08:28	
Código de recuperación	1171779912	
Código de verificación	d9d9e1974f286ca70cfd5a6bec7fc610c1751e4	

Figura 27: Comprobante de completitud de la Encuesta.

De esta manera la información es relevada, actualizada y almacenada en la base de dato del sistema SIU-Kolla, pudiendo la misma ser utilizada para contribuir a la construcción de Data Marts específicos que aporten al proceso de toma de decisión por parte de los mandos medios y altos de la Unidad Académica.

El circuito administrativo propuesto se puede observar en la Figura 28:

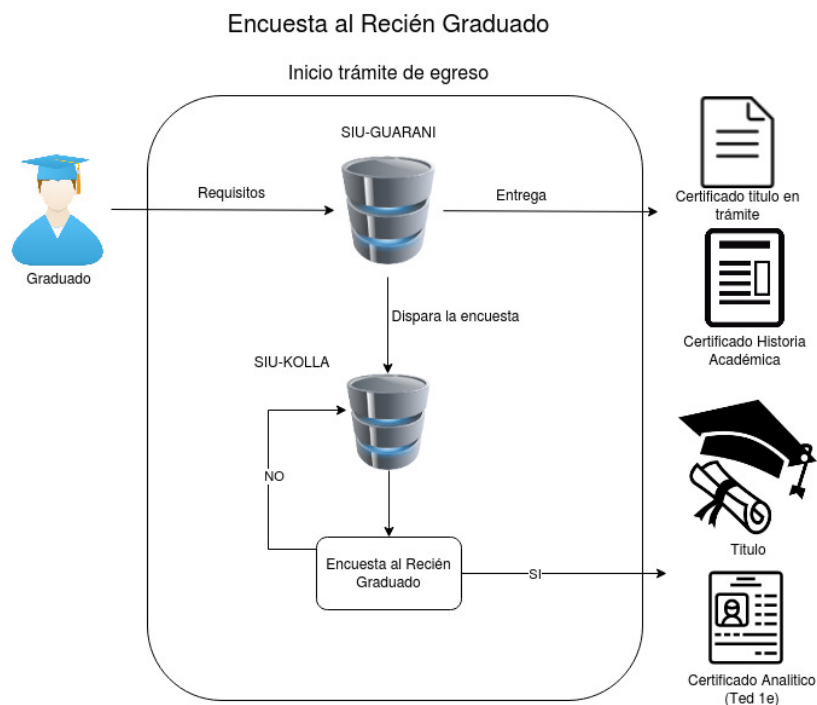


Figura 28: Circuito administrativo propuesto para la aplicación de la Encuesta al Recién Graduado.
[Elaboración propia]

5.2.1 Sitio web para Graduados de la FCEQyN

Lograr institucionalizar y visibilizar las medidas que se van adoptando para la construcción de información sobre los graduados, impulsó la construcción de un espacio web institucional para los mismos (<https://www.fceqyn.unam.edu.ar/graduados/>).

La primera propuesta generada en dicho espacio fue la invitación a participar de las encuestas, comunicando el objetivo y los pasos a seguir para poder acceder a las mismas y responderlas.

Luego, con el uso del mismo, se incorporó mayor información de interés y asesoramiento a los graduados sobre las distintas actividades que se generan desde la Unidad Académica.

Actualmente el sitio web proporciona información sobre:

- Oficina de Vinculación con el Graduado: Sitio web de presentación y contacto con la Oficina de Vinculación con el Graduado.
- Trámite de títulos: Sitio web con los requisitos para iniciar los trámites de título.
- SIU-Kolla: Acceso al sistema SIU-Kolla para la completitud de las encuestas.

- Registro Público de Graduados Universitarios: Sitio web que facilita la búsqueda de títulos oficiales de graduados en universidades argentinas.

5.3 Implementación de la Metodología propuesta

En esta sección se realiza la implementación de la metodología propuesta tomando como caso de estudio el área de Graduados de la FCEQyN de la UNaM.

La incorporación de las buenas prácticas propuestas por el modelo CMMI, incluyendo la trazabilidad, se lleva a cabo mediante la documentación generada en cada etapa y la utilización de herramientas de gestión de proyectos Redmine y GitHub. Este último utilizado hacia final del proyecto, como repositorio con toda la documentación, procesos desarrollados y las distintas instancias de prueba, producción e implementación.

Las actividades sobre el desarrollo del proyecto se registraron en el Redmine de la FCEQyN, dentro del proyecto Business Intelligence (BI) creado para tal fin. La registración de cada una de ellas genera un número de ticket, el cual fue asociado a cada tarea realizada.

En los siguientes apartados se describe el caso práctico desarrollado, detallando cada una de las Etapas (E) definidas en la metodología.

5.3.1 E1: Entendimiento del Negocio

E1.A1 – Roles y Responsabilidades

Teniendo en cuenta los diferentes roles y responsabilidades definidas en la metodología propuesta, se identificaron los integrantes del equipo de trabajo en base a la Estructura Orgánica Funcional Administrativa-Técnica de la FCEQyN aprobada por Disposición Nro.: 1620/09 [51].

- **Responsable Institución:** Decano y Vice Decano de la FCEQyN.
- **Experto / Usuario:** Coordinador de la Oficina de Vinculación con el Graduado y el Director de la carrera Licenciatura en Sistemas de Información de la FCEQyN.
- **Analista:** Directora Área de Informática y Comunicaciones de la FCEQyN.

E1.A2 – Elicitación de Requerimientos

Las herramientas utilizadas para la elicitación de requerimientos fueron entrevistas abiertas, reuniones y cuestionarios. La documentación relacionada a las mismas se encuentran registradas en

el Redmine de FCEQyN (Tickets Nro.: 290, 291, 292,293) y también pueden descargarse desde el siguiente enlace: <https://github.com/mdarolon/BBII-FCEQyN>

Las entrevistas estuvieron centradas en recopilar información acerca de las necesidades y requerimientos de información demandados actualmente en la institución para la gestión de los graduados, las herramientas tecnológicas utilizadas para recabar información sobre los mismos y las utilizadas para la comunicación con estos.

Otro mecanismo de comunicación que predominó fueron las reuniones con personal de la Dirección de Enseñanza y Secretaría Académica, para la redefinición e implementación del circuito administrativo de solicitud de título por parte de un alumno y el mecanismo de envíos de las encuestas a los graduados. Por otro lado, se realizaron numerosas reuniones, entre el analista, los expertos y el responsable institucional, en las cuales se conformó el equipo de trabajo, se consensuaron y unificaron conceptos, y se definieron criterios de abordaje de la información para alcanzar soluciones acordes a sus necesidades.

Como canales de comunicación se utilizaron: correos electrónicos, servicio de mensajería y videoconferencias (Contexto ASPO – Aislamiento Social Preventivo y Obligatorio [52]).

La selección de los expertos a entrevistar estuvo enfocada en el cargo ocupacional de los mismos, sus funciones y trabajos asociados a los graduados, las necesidades de información planteadas surgidas durante su desempeño ocupacional, y en el caso de uno de ellos sus conocimientos y experiencia sobre Inteligencia de Negocios.

En cuanto a la documentación relevada, se obtuvo la Resolución Nro.: 040/19 [50] a través de la cual se crea la Oficina de Vinculación con los Graduados dependiente de la Secretaría de Extensión y Vinculación Tecnológica de la FCEQyN y en donde se describen los fundamentos y objetivos de la misma, el PDI de la FCEQyN [3] y de la UNaM [47].

E1.A3 – Definición del Alcance

El presente proyecto de investigación se desarrolló dentro del ámbito de la FCEQyN de la UNaM. El tema a abordar refiere la generación de una base de conocimiento sobre los Graduados de ésta Unidad Académica.

Los sectores intervinientes para el desarrollo del mismo fueron la Oficina de Vinculación con los Graduados perteneciente a la Secretaría de Extensión y Vinculación Tecnológica, la Secretaría Académica, la Dirección de Enseñanza, el Decanato y la Dirección Área de Informática y Comunicaciones.

E1.A4 – Definición de Temas Analíticos (TA)

En base al descubrimiento de las necesidades de los expertos consultados durante el proceso de elicitación de requerimientos, se definieron los temas analíticos a analizar. Los mismos se encuentran registrados en el Redmine de la FCEQyN (Tickets 294, 320 y 345).

En las Tablas siguientes (Tabla 12, Tabla 13 y Tabla 14) se describen cada uno de ellos:

Tabla 12: TA01-Situación laboral de los graduados por carrera según su distribución geográfica.

TEMA ANALITICO	
TA01	Situación laboral de los graduados por carrera según su distribución geográfica
DERU Nro.:	1 (Tickets 290)
Cómo	Coordinador de la Oficina de Graduados
Quiero	Conocer la situación laboral de los graduados y su distribución geográfica.
Para	Generar espacios laborales y poder articularlos con distintos sectores socio-productivos e industriales de la región.

Tabla 13: TA02- Valoración académica de los graduados por carrera

TEMA ANALITICO	
TA02	Valoración académica de los graduados por carrera.
DERU Nro.:	2 - 3 (Tickets 291, 292)
Cómo	Coordinador de la Oficina de Graduados
Quiero	Conocer la valoración de los graduados sobre la formación académica recibida.
Para	Alimentar cambios, mejoras y actualizaciones en los planes de estudios.

Tabla 14: TA03-Necesidad de capacitación del graduado por carrera según su distribución geográfica

TEMA ANALITICO	
TA03	Necesidad de capacitación del graduado por carrera según su distribución geográfica.
DERU Nro.:	4 (Ticket 293)
Cómo	Director de la carrera Licenciatura en Sistemas de Información.
Quiero	Conocer por región, las necesidades de capacitación de los graduados por carrera.
Para	Poder orientar las actividades de capacitación, extensión y vinculación de la carrera, a cubrir dicha demanda.

E1.A5 – Modelado preliminar

Para el desarrollo de esta actividad, es importante la participación del analista y el conocimiento que posee éste sobre el manejo de los datos y de los sistemas transaccionales que se utilizan en la institución, dado que la información que se necesita debe estar soportada de alguna manera en algún OLTP o en algún contenedor digital para poder elaborar el Data Mart.

En base a los temas analíticos definidos durante la etapa anterior de la metodología y considerando que el SIU-Wichi no ofrece un cubo específico que aborde las temáticas planteadas, se comienza a desarrollar el modelado conceptual de los mismos.

- TA01- ‘Situación laboral de los graduados por carrera según su distribución geográfica’ (Tabla 12).

Para comenzar a organizar y representar de manera preliminar, los requerimientos sobre el dominio particular a analizar, se realiza una transcripción del TA01 definido y se extraen los conceptos que pueden aportar información para el análisis.

En cuanto a la distribución geográfica, se analiza el contexto y se expone al experto las variantes existentes relacionadas a este concepto. La distribución geográfica puede ser modelada según la ubicación de residencia o de procedencia del graduado, donde lugar de residencia refiere al lugar donde reside el graduado actualmente y lugar de procedencia, de donde proviene el mismo.

Por otro lado, para una mejor interpretación, se pone a consideración evaluar ésta distribución mediante el dato de la región/departamento donde se encuentra localizada la ciudad de

procedencia o residencia del graduado, para que los datos tengan una representación que se ajuste a la realidad.

De la misma manera, se debate con el experto sobre el concepto de situación laboral; si de dicho concepto solo desea conocer la situación laboral actual del graduado, siendo que además en los OLTP se registran otros aspectos de la experiencia laboral del mismo, como ser la cantidad de horas que trabaja, si el trabajo tiene relación con la carrera estudiada, cual es la satisfacción que siente con su trabajo, la entidad donde trabaja, entre otros.

Por otro lado, se considera la posibilidad de incorporar datos personales como: estado civil, tipo de convivencia, género, entre otros.

Una vez identificados los conceptos a abordar, el analista analiza la documentación existente de las distintas fuentes de datos de donde extraer la información y se comienza a confeccionar un diccionario de datos propio, que aporte valor al tema de análisis planteado.

En el diccionario de datos se describen los atributos que se relacionan a los conceptos que se fueron relevando y que pudieran ser considerados para el diseño del Data Mart. De dichos atributos se describe su significado, contenido y la posible fuente de datos de donde extraer dicho valor.

Los conceptos relevados e incorporados al diccionario de datos son los siguientes (Tabla 15):

Tabla 15: Diccionario de Datos TA01

DERU Nro.	1	Tema analítico: TA01-Situación laboral de los graduados por carrera según su distribución geográfica	
Fecha	Abril 2020	Analista: Lic. María Daniela Rolón	
Atributo	Descripción	Contenido	Sistema Fuente
Titulo	Título universitario obtenido por el graduado.	Bioquímico, Farmacéutico, Ingeniero Químico, etc.	SIU-Kolla / SIU-Guarani
Carrera	Carrera en la que se graduó el alumno.	Bioquímica, Farmacia, Ingeniería Química, etc.	SIU-Guarani
Procedencia →Ciudad →Provincia →País	Ciudad, provincia y país de procedencia del graduado.	Posadas, Misiones, Argentina	SIU-Kolla / SIU-Guarani
Residencia →Ciudad →Provincia	Ciudad, provincia y país de residencia del graduado.	Posadas, Misiones, Argentina	SIU-Kolla / SIU-Guarani

→País			
Situación actual	Situación laboral del graduado.	No trabaja / Trabaja en relación con la profesión / Trabaja, pero no tiene relación con la profesión / No responde.	SIU-Kolla
Rango horas semanales	Rango de horas semanales que trabaja el graduado.	Hasta 15 / 16- 30 / 31-45 / Más de 45 / No responde.	SIU-Kolla
Satisfacción laboral	Satisfacción por el trabajo que realiza.	Muy satisfecho / Medianamente satisfecho / Poco satisfecho / No responde.	SIU-Kolla
Tipo entidad laboral	Tipo de entidad donde trabaja el graduado.	Privada / Pública / Ambas / No responde	SIU-Kolla
Sexo	Género del graduado	Masculino / Femenino / No responde	SIU-Kolla / SIU-Guarani
Estado civil	Estado civil del graduado	Unido de hecho / Casado / Soltero / Divorciado / Unido civilmente / No responde.	SIU-Kolla
Grupo de convivencia	Grupo de convivencia del graduado	Con familia parental / Solo / Con familia propia / Otra / No responde.	SIU-Kolla
Rango etario	Grupo de edades según los siguientes rangos	Menos a 16 / 17-20 / 21-24 / 25-29 / 30-34 / 35-39 / 40-49 / 50-59 / Mayor a 60	SIU-Kolla / SIU-Guarani
Tiempo	Fecha de finalización de la encuesta	23/04/2019,04/10/2019, etc.	SIU-Kolla
Cantidad de graduados	Cantidad de graduados que completaron la encuesta (*)	1,2,3,5,6, etc.	SIU-Kolla / SIU-Guarani
Promedio edad	Promedio de edad de los graduados (*)	25, 33, 28, etc.	SIU-Kolla / SIU-Guarani

(*) Indicador

El diccionario de datos asociado al TA01 se registra en el Redmine de FCEQyN (Ticket 297) y se encuentra disponible en el siguiente enlace: <https://github.com/mdarolon/BBII-FCEQyN>.

Finalmente, se diseña un modelo conceptual preliminar que permite determinar el alcance del tema analítico analizado en base a las distintas perspectivas e indicadores identificados hasta el momento.

En la Figura 29 se representa el modelo conceptual preliminar, en el cual las perspectivas se encuentran ubicadas a la izquierda del modelo y los indicadores a la derecha. El círculo central representa el tema analítico que ha sido definido.

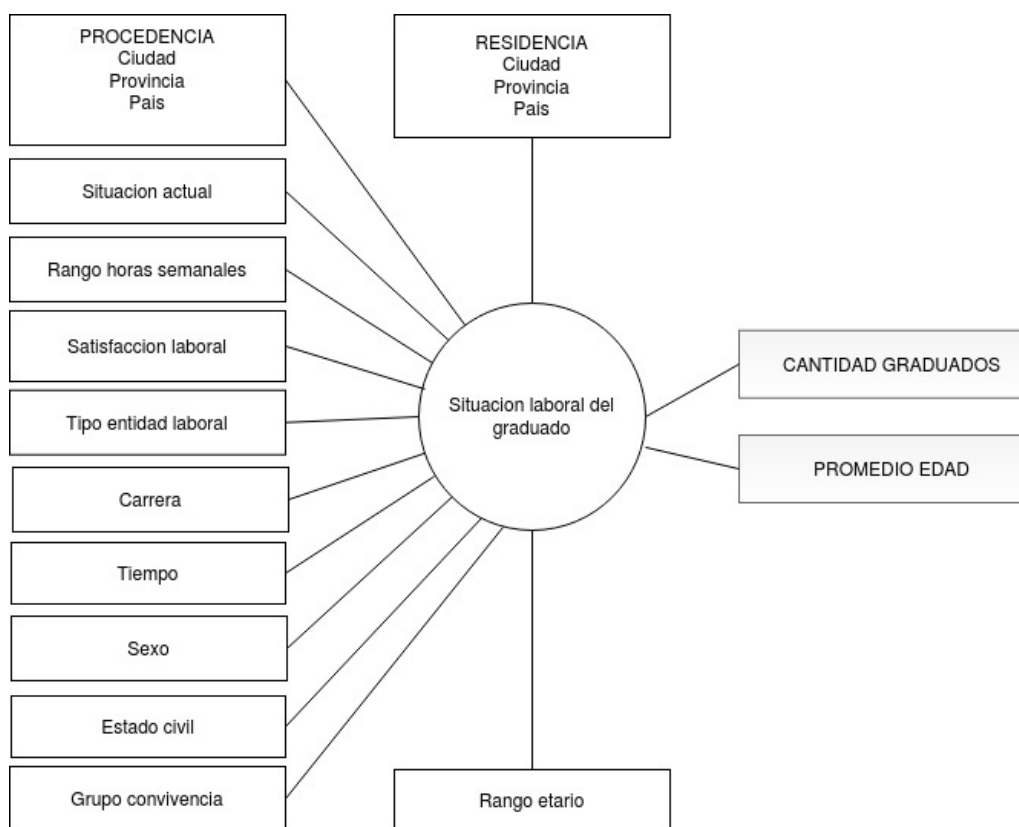


Figura 29: Modelo Conceptual preliminar: TA01-Situación laboral de los graduados por carrera según su distribución geográfica [Elaboración propia]

Como puede apreciarse en el Modelo Conceptual preliminar, se han incorporado nuevos contenidos para el análisis como alternativas de solución propuesta.

- TA02 -‘Valoración académica de los graduados por carrera’ (Tabla 13).

En cuanto a la valoración académica se analiza el contexto, y se expone al experto las variantes existentes relacionadas a este concepto. Dichas variantes se encuentran relacionadas directamente con la opinión y evaluación del graduado sobre la formación académica recibida relacionada a su carrera.

Esta evaluación se basa en conceptos relacionados a la elección de la carrera, su plan de estudios, contenidos, duración, nivel de exigencia y la elección de estudiar en la FCEQyN (UNaM).

Por otro lado, para poder promover cambios, mejoras y actualizaciones a los planes de estudios de las distintas carreras, el experto considera importante contar con la información

relacionada al tiempo que les lleva a los graduados terminar sus estudios, teniendo en cuenta la duración real de cada plan de estudio.

Una vez identificados los conceptos se comienza a confeccionar un diccionario de datos propio, que aporte al proceso de negocio que se está diseñando.

Los conceptos relevados e incorporados al diccionario de datos son los siguientes (Tabla 16):

Tabla 16: Diccionario de Datos TA02

DERU Nro.	2-3	Tema analítico: TA02: Valoración académica de los graduados por carrera.	
Fecha	Julio 2020	Analista: Lic. María Daniela Rolón	
Atributo	Descripción	Contenido	Sistema Fuente
Título	Título universitario obtenido por el graduado.	Bioquímico, Farmacéutico, Ingeniero Químico, etc.	SIU-Kolla / SIU-Guarani
Carrera	Carrera en la que se graduó el alumno.	Bioquímica, Farmacia, Ingeniería Química, etc.	SIU-Guarani
Selección carrera	Motivos por los cuales eligió su carrera universitaria	Desarrollo Profesional / Deseos de ser útil a la sociedad / Interés por el avance científico y tecnológico / No Responde / Otros / Posibilidades de trabajo / Prestigio social / Sugerencia del núcleo familiar / Una clara vocación	SIU-Kolla
Selección universidad	Motivos por los cuales eligió estudiar en esta Universidad	Cercanía con el lugar de residencia o de trabajo / Costo de la matrícula / Información en medios de difusión / No Responde / Otros / Prestigio académico / Recomendación de estudiantes / Recomendación de familiares o amigos / Recomendación de graduados de la universidad / Única institución que dictaba la carrera	SIU-Kolla
Interés contenidos	Interés sobre los contenidos estudiados durante la carrera.	De acuerdo / En desacuerdo / Ni acuerdo ni en desacuerdo / No Responde / Totalmente de acuerdo	SIU-Kolla

Expectativas contenido	Cumplimiento de expectativas sobre los contenidos estudiados durante su carrera.	De acuerdo / En desacuerdo / Ni acuerdo ni en desacuerdo / No Responde / Totalmente de acuerdo / Totalmente en desacuerdo	SIU-Kolla
Nivel exigencia	Nivel de exigencia de la carrera.	Ni elevado ni bajo / Bajo / Elevado / Totalmente elevado / No responde	SIU-Kolla
Sexo	Género del graduado	Masculino / Femenino / No responde.	SIU-Kolla / SIU-Guarani
Plan corto	El plan de estudios debería ser más corto	Si / No / No responde.	SIU-Kolla
Plan generalista	El plan de estudios debería ser más generalista.	Si / No / No responde.	SIU-Kolla
Plan técnico	El plan de estudios debería tener un perfil más técnico	Si / No / No responde.	SIU-Kolla
Valoración universidad	Aspectos de la Universidad que el graduado más valora con relación a la carrera que estudió.	El clima que se vive en la universidad / El equipo de profesores / El interés de la universidad por la inserción de sus egresados / El nivel académico / Las relaciones de compañerismo / Los planes de estudio / No Responde /Otros	SIU-Kolla
Tiempo	Fecha de finalización de la encuesta.	23/04/2019,04/10/2019, etc.	SIU-Kolla
Universidad sociedad	Percepción personal del reconocimiento de la sociedad a esta Universidad	Insatisfactoria / Ni satisfactoria ni insatisfactoria / No Responde / Satisfactoria / Totalmente satisfactoria	SIU-Kolla
Aportes universidad	Otros aportes que se desea brinde la Universidad	Actividades culturales / Carreras de posgrado / Cursos, seminarios de actualización profesional / No Responde / Otros / Posibilidad de trabajo en docencia / Posibilidad de trabajo en investigación.	SIU-Kolla
Cantidad de graduados	Cantidad de graduados que completaron la encuesta (*)	1,2,3,5,6, etc.	SIU-Kolla / SIU-Guarani
Duración real carrera	Tiempo que le tomo al graduado de una Carrera completar el plan de estudio (*)	4, 5,3, etc.	SIU-Guarani
desvío duración	Es la diferencia entre la		SIU-Guarani

real	duración real de la carrera y la duración teórica de la misma (*)		
------	---	--	--

(*) Indicador

El diccionario de datos asociado al TA02 se registra en el Redmine de FCEQyN (Ticket 381) y en el siguiente enlace: <https://github.com/mdarolon/BBII-FCEQyN/DD>.

Finalmente, se diseña un modelo conceptual preliminar que permite determinar el alcance del tema analítico analizado en base a las distintas perspectivas e indicadores identificados hasta el momento.

En la Figura 30 se observa el modelo conceptual preliminar, en el cual las perspectivas se encuentran ubicadas a la izquierda del modelo y los indicadores a la derecha. El círculo central representa el tema analítico que se ha definido.

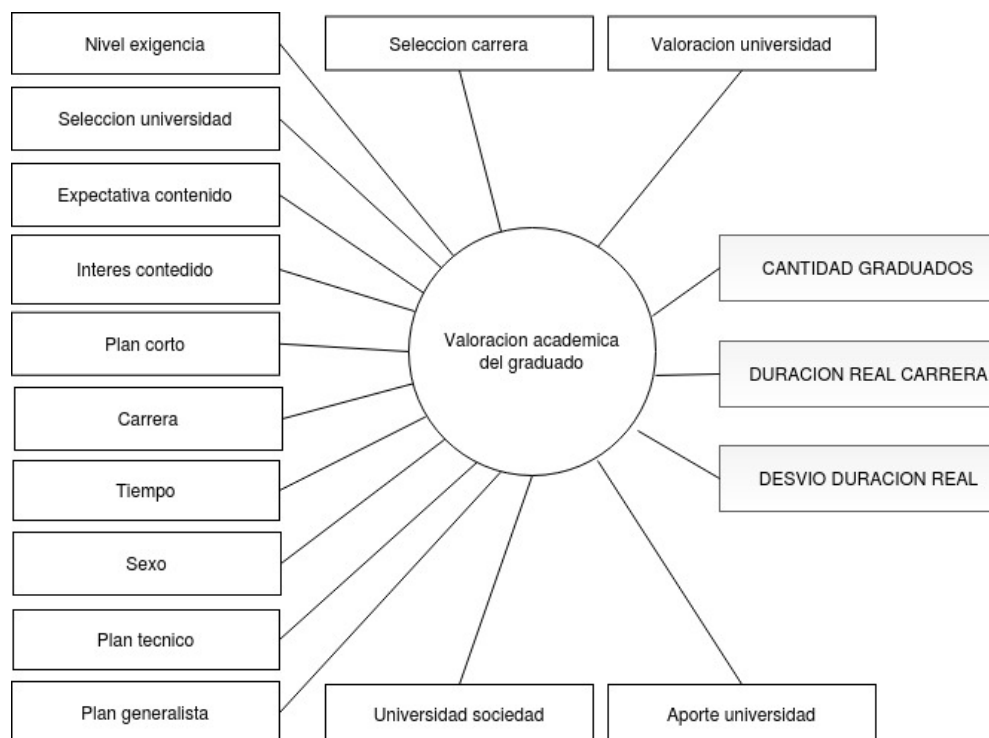


Figura 30: Modelo Conceptual preliminar: TA02- Valoración académica de los graduados por carrera [Elaboración propia]

- TA03- “Necesidad de capacitación del graduado por carrera según su distribución geográfica” (Tabla 14)

En cuanto a las necesidades de capacitación de los graduados se analiza el contexto, y se comunica al experto las pocas variantes existentes relacionadas a este concepto. Dichas variantes se

encuentran relacionadas directamente con el interés de los graduados en realizar capacitaciones y con su distribución geográfica, pero otros datos relacionados con el tipo de formación profesional desean recibir, turno preferido, áreas de formación, tipo de dictado, entre otros, no estaban presentes en los sistemas fuentes implementados actualmente.

Se comienza a confeccionar el diccionario de datos, en base a los conceptos existentes y a distinguir aquellos relevantes para el análisis. En la Tabla 17 se listan cada uno de ellos.

Tabla 17: Diccionario de Datos TA03

DERU Nro.	4	Tema analítico: TA03- Necesidad de capacitación de los graduados por carrera según su distribución geográfica	
Fecha	Mayo 2020	Analista: Lic. María Daniela Rolón	
Atributo	Descripción	Contenido	Sistema Fuente
Título	Título universitario obtenido por el graduado.	Bioquímico, Farmacéutico, Ingeniero Químico, etc.	SIU-Kolla / SIU-Guarani
Carrera	Carrera en la que se graduó el alumno.	Bioquímica, Farmacia, Ingeniería Química, etc.	SIU-Guarani
Procedencia →Ciudad →Provincia →País	Ciudad, provincia y país de procedencia del graduado.	Posadas, Misiones, Argentina	SIU-Kolla / SIU-Guarani
Residencia →Ciudad →Provincia →País	Ciudad, provincia y país de residencia del graduado.	Posadas, Misiones, Argentina	SIU-Kolla / SIU-Guarani
Formación continua	Interés en cursos de formación continua	No responde / Si / No.	SIU-Kolla
Sexo	Género del graduado	Masculino / Femenino / No responde	SIU-Kolla / SIU-Guarani
Tiempo	Fecha de finalización de la encuesta	23/04/2019,04/10/2019, etc.	SIU-Kolla
Área de formación	Área de formación profesional	Básicas / Aplicadas	--
Herramientas	Herramientas específicas relacionadas con su profesión.	--	--
Idioma	Idioma en el cual quiera formarse.	Inglés / Portugués / Otro	--
Tipo de	Tipo de actividad de	Taller / Seminario / Curso	--

Actividad	actualización y capacitación desea realizar.		
Turno preferido	Turno proferido para realizar la capacitación.	Mañana / Tarde / Noche	--
Costeo del curso	Cómo costea el mismo.	Beca / Recursos Propios / Otro	--
Tipo de dictado	Tipo de dictado del curso.	Presencial / Virtual / Ambos	--
Cantidad de graduados	Cantidad de graduados que completaron la encuesta (*)	1,2,3,5,6, etc.	SIU-Kolla / SIU-Guarani

(*) Indicador

El diccionario de datos asociado al TA03 se registra en el Redmine de FCEQyN (Ticket 411) y en el siguiente enlace: <https://github.com/mdarolon/BBII-FCEQyN>.

En base a los conceptos relevados, se diseña un modelo conceptual preliminar (Figura 31), donde se deja de manifiesto con color rojo aquellas perspectivas cuya fuente de origen no pudo ser determinada.

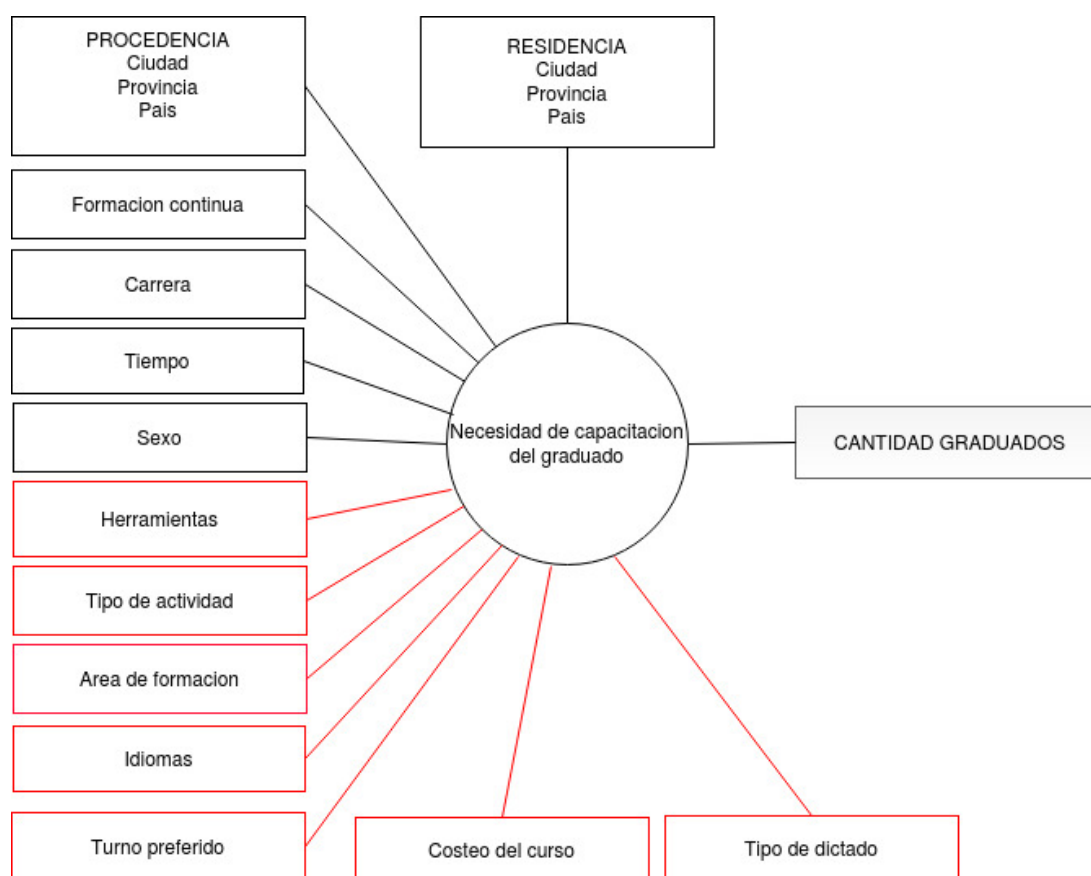


Figura 31: Modelo Conceptual preliminar: TA03-Necesidad de capacitación de los graduados por carrera según su distribución geográfica [Elaboración propia]

E1.A6 – Determinación de los recursos

La FCEQyN cuenta con la Dirección Área de Informática y Comunicaciones (DAIyC) encargada de mantener, administrar y gestionar los distintos sistemas implementados en la institución.

Esto permite una interacción mucho más directa a la hora de abordar la problemática debido a la accesibilidad de los datos y al entendimiento y tratamiento de los mismos.

La DAIyC recibe periódicamente solicitudes de datos, de estadísticas y de información por parte de Docentes, No Docentes, Secretarios y Autoridades, basadas en necesidades propias de la institución (Departamentos, Secretarías, Direcciones, Cátedras). En su mayoría, dichas solicitudes involucran datos sensibles de los alumnos, por lo que son atendidas bajo autorización de la autoridad de turno y dentro del marco de la Ley 25.326 (*habeas data*) [53], garantizando el derecho a la intimidad de las personas. Las mismas son gestionadas mediante la herramienta informática de gestión de proyectos Redmine¹⁷ implementado en esta Dirección.

Como recurso humano se cuenta con el personal informático idóneo para el desarrollo del presente trabajo y autoridades de la Facultad que acompañarán con el aval institucional al caso de estudio que se está analizando.

La presente investigación se lleva a cabo en la DAIyC de la FCEQyN de la UNaM, donde se encuentran implementados los sistemas SIU-Kolla y SIU-Guarani.

A continuación se detalla el estudio de factibilidad realizado, en base a tres aspectos específicos: técnico, económico y operacional.

➤ Factibilidad técnica:

Según la última “Guía Técnica de SIU-Wichi 6.7.0” [54] confeccionada por el SIU, las características mínimas requeridas de un servidor de Pentaho son las siguientes (Tabla 18):

Tabla 18: Requerimientos mínimos de hardware [54]

Característica	Servidor mínimo recomendado para realizar una instalación de prueba	Servidor recomendado para una instalación de Producción
Servidores	Único servidor para instalar base de datos y aplicación Pentaho	• Un servidor de base de datos • Un servidor para la aplicación Pentaho
RAM	4 Gb	8 Gb cada servidor

¹⁷ Redmine (<https://www.redmine.org/>).

Disco	40 Gb	• Base de datos: 80 Gb • Pentaho: 30 Gb
Núcleos	2 núcleos	4 núcleos cada servidor

Para el desarrollo e implementación del presente trabajo se ha instalado y configurado un servidor de producción en el datacenter de la FCEQyN, bajo un entorno de virtualización de servidores de código abierto Proxmox VE; y un servidor de desarrollo local en una Notebook particular.

Las características de cada servidor se observa en la Tabla 19:

Tabla 19: Configuración de hardware servidores configurados

Característica	Servidor de prueba	Servidor de producción
Servidores	Único servidor para instalar base de datos y aplicación Pentaho	Único servidor para instalar base de datos y aplicación Pentaho
RAM	8 Gb	8 Gb
Disco	1 Gb	80 Gb
Núcleos	1 núcleo	1 núcleo

En cuanto al software los requerimientos mínimos que recomienda el SIU [54] para la instalación de un servidor de Pentaho son los siguientes:

- **Sistema Operativo:** Linux - Ubuntu 12.04 o superior / Debian 6.0 o superior
- **Servidor Pentaho:** BI-Server 6.1 (Edición comunidad)
- **Motor Base de Datos:** PostgreSQL 9.3 o superior
- **Java:** JRE versión 1.7

Teniendo en cuenta las recomendaciones propuestas por el SIU, para el desarrollo del presente trabajo se instalaron las siguientes herramientas:

- **Sistema Operativo:** Debian GNU/Linux 10
- **Servidor Pentaho:** BI-Server 8.0 (Edición comunidad)
 - Pentaho Data Integration
 - Schema Workbench
 - User Console
 - Analyzer
 - Dashboard Designer
- **Motor Base de Datos:** PostgreSQL 9.2

- pgAdmin III en su versión 1.16.1
- **Java:** JRE versión 8
- Librerías para Informix (ifxjdbc.jar, ifxtools.jar, ifxlang.jar, ifxsupp.jar)
- Informix Server Versión 12.10.FC6IE
- SQL Power Architect Community Edition
- OpenVPN 2.0
- Redmine versión 3.2.0
- GitHub versión free
- SQL Power Architect Community Edition

Con referencia a los datos, la accesibilidad y disponibilidad de los mismos se encuentran avaladas institucionalmente mediante CUDAP: FCEQYN_NOTA-S01:0002558/2020.

Por lo tanto, se puede decir que el proyecto es técnicamente factible para la institución.

➤ **Factibilidad económica:**

Con referencia a la inversión de hardware para llevar adelante la presente investigación, se han utilizado recursos disponibles en la DAlyC de la FCEQyN. No obstante, la facultad recientemente ha adquirido un servidor para uso institucional, cuyas características técnicas cumplen con los requerimientos recomendados por el SIU para la implementación de un servidor de producción de Pentaho, cuyo costo estimado del mismo se encuentra detallado en el ANEXO 2 (Redmine de la FCEQyN - Ticket 387).

En cuanto al software se utilizó software open source sin costo de inversión.

➤ **Factibilidad operacional:**

Usuarios Operativos: Los usuarios operativos serán:

- Coordinador de la Oficina de Graduados de la FCEQyN
- Director de la carrera Licenciatura en Sistema de Información de la FCEQyN
- Vice Decano de la FCEQyN

Usuario Técnico (Analista):

- Directora Área de Informática y Comunicaciones de la FCEQyN

Definidos los usuarios, se puede decir que el proyecto es operacionalmente factible para la institución.

E1.A7 – Planificación

Como esta investigación involucra una metodología que se encuentra en proceso de desarrollo, los tiempos necesarios para su aplicación se irán relevando y este tema se abordará posteriormente en el Capítulo 6 de Resultados determinando los tiempos de duración de cada etapa.

5.3.2 E2: Relevamiento y Análisis

E2.A1 – Análisis de los Datos

Para el análisis de los datos se examinan los diccionarios de datos y los Modelos Conceptuales provenientes de la etapa anterior (E1.A5 – **Modelado preliminar**).

Para el caso de estudio elegido, se analizó la accesibilidad y usabilidad de las bases de datos de los sistemas SIU-Kolla y SIU-Guaraní implementados en la FCEQyN (Tabla 20), como así también sus estructuras.

Tabla 20: Accesibilidad y usabilidad de las bases de datos seleccionadas

Fuentes de datos	Disponibilidad	Frecuencia de actualización	Tipo de sistema
SIU-Guaraní	Alta	Diaria	OLTP
SIU-Kolla	Alta	Mensual	OLTP

El sistema SIU-Kolla presenta una estructura de base de datos modelada mediante una serie de tablas, que concentran la información almacenada a nivel bloques, preguntas y respuestas. Dichas tablas conforman el detalle del formulario habilitado (encuesta) que se está respondiendo.

Este formulario resulta en un nivel más de abstracción, donde su composición es de uno o más bloques, y donde cada una de estos puede habilitarse una o más veces.

Desde que un formulario ya tiene una habilitación asociada, los alumnos habilitados pueden responder la encuesta o no. Para cada encuestado que responde, se guarda la información principal y si está terminado o no (`sge_respondido_formulario`), y el encuestado mencionado (`sge_respondido_encuestado`).

El almacenamiento de las encuestas está modelado mediante una tabla que indica a que detalle del formulario habilitado responde, es decir a cual encuesta dentro del formulario responde (`sge_respondido_encuesta`), y otra tabla para almacenar las respuestas propiamente dichas (`sge_respondido_detalle`).

En la Figura 32 se observa la estructura de la base de datos del sistema SIU-Kolla correspondiente al modulo de completitud del formulario.

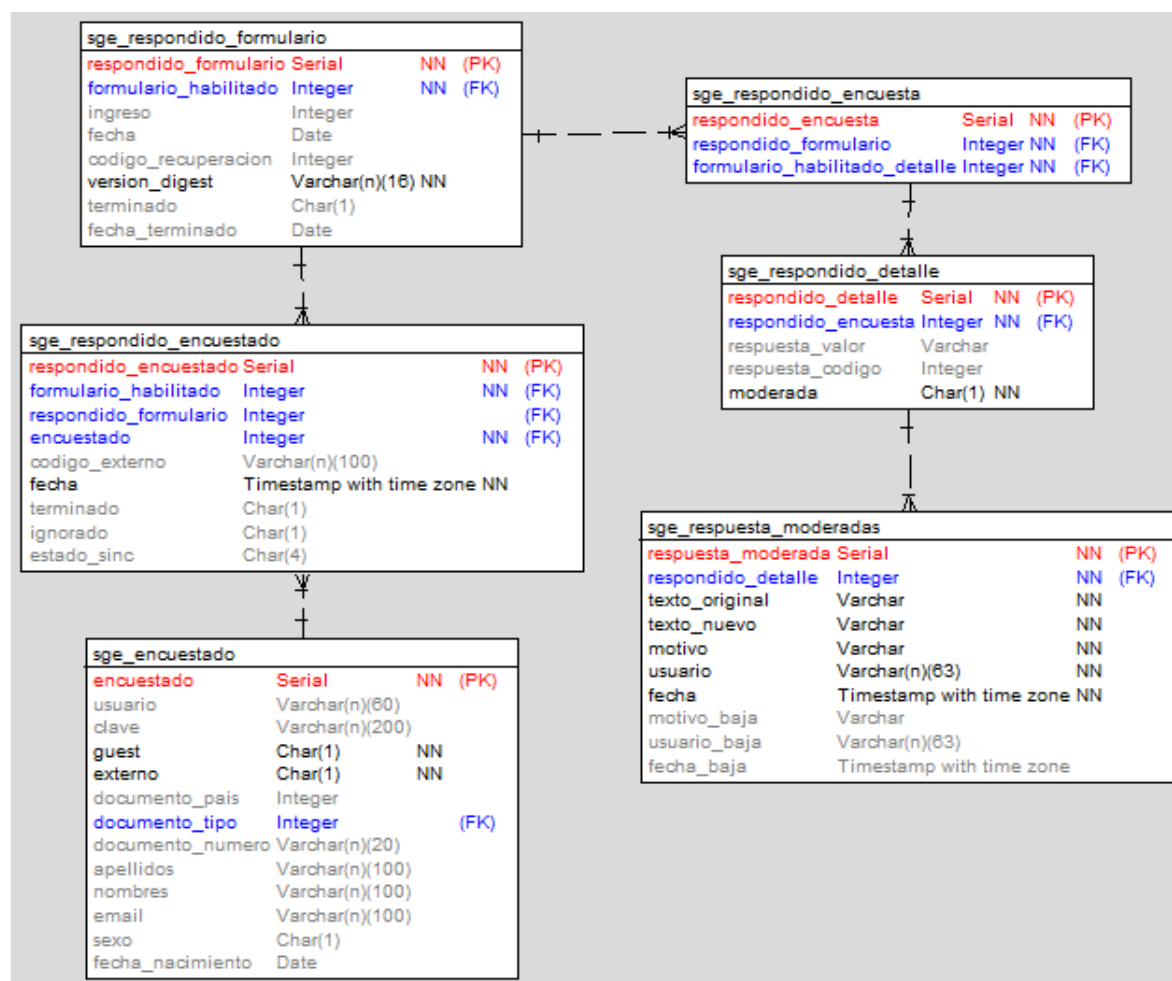


Figura 32: Estructura de la base de datos SIU-Kolla [61]

Ante esta particularidad detectada se decide crear una tabla (recien_graduado) dentro de un esquema de almacenamiento temporal, que contenga los resultados de las encuestas respondidas, extraídas a través de sus distintos reportes de gestión del sistema SIU-Kolla.

Con referencia al módulo denominado ubicación geográfica, el mismo es un módulo genérico en todos los sistemas desarrollados por el SIU. Por lo tanto, la información almacenada en el mismo se maneja de manera unívoca desde el SIU para poder mantener un catálogo unificado de localidades en todas sus distribuciones.

La estructura del módulo mencionado se observa en laFigura 33:

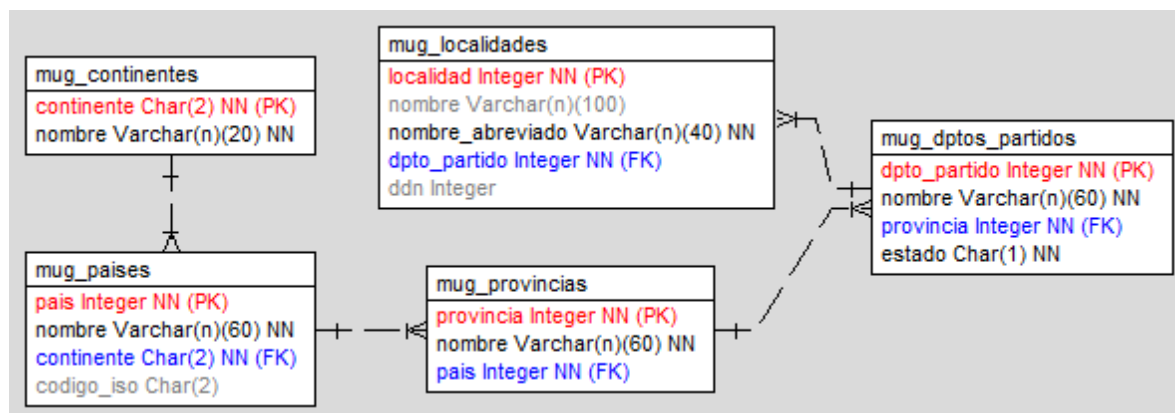


Figura 33: Módulo ubicación geográfica SIU-Kolla [61]

A pesar que desde el sistema SIU-Kolla se pueden extraer la totalidad de los datos para el modelado, se detecta que varios campos principales de la encuesta son preguntas abiertas que pueden ser contestadas por el graduado sin presentar ningún tipo de formato, categorización o control de carga.

Dichos campos corresponden a la fecha de nacimiento, año de ingreso a la carrera, título obtenido, localidad de procedencia y de residencia del graduado.

Por ello, se decide extraer dichos datos del sistema fuente SIU-Guaraní, donde los mismos son cargados por personal administrativo de la Dirección de Enseñanza en base a la documentación solicitada al momento de la inscripción a la carrera, evitando de esta manera inconsistencias en la integridad y correctitud de los datos.

La estructura de la base de datos analizada del sistema SIU-Guaraní se observa en la Figura 34 donde, toda la información relacionada al alumno (*sga_personas*, *sga_alumnos*) y al título obtenido (*sga_colaciones*, *sga_carreras*) se almacena en la tabla *sga_titulos_otorg*.

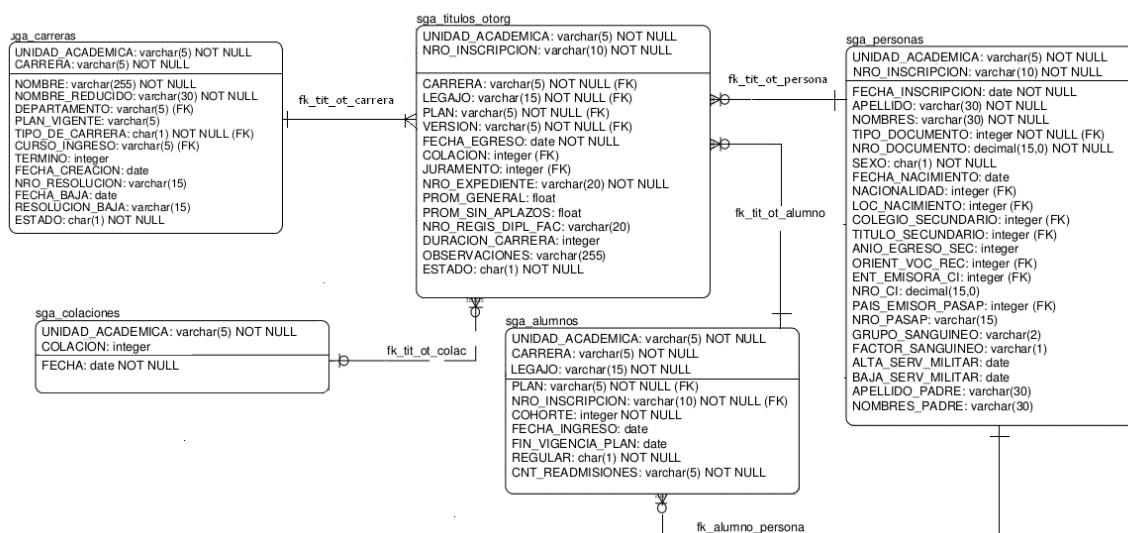


Figura 34: Estructura de la base de datos SIU-Guarani [62]

E2.A2- Mapeo al Modelo Conceptual

Una vez analizado los orígenes de los datos se establecieron las correspondencias entre las perspectivas identificadas en el Modelo Conceptual preliminar y las tablas de los sistemas fuentes involucradas para cada uno de los temas analíticos definidos.

- TA01- ‘Situación laboral de los graduados por carrera según su distribución geográfica’ (Tabla 21)

Tabla 21: FUDA TA01-Situación laboral de los graduados por carrera según su distribución geográfica [Elaboración propia]

Documento de Identificación de las Fuentes de Datos						
TA01	Situación laboral del graduado por carrera según su distribución geográfica					
Analista	Lic. María Daniela Rolón					
Base de Datos (BD)	Versión BD	Tabla	Atributo	Perspectiva	Atributo Perspectiva	Indicador
SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_localidades	nombre	Residencia	nom_ciudad_res	-

SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_dptos	nombre	Residencia	nom_dpto_res	-
SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_provincias	nombre	Residencia	nom_provincia_res	-
SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_paises	nombre	Residencia	nom_pais_res	-
SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_localidades	nombre	Procedencia	nom_ciudad_proc	-
SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_dptos	nombre	Procedencia	nom_dpto_proc	-
SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_provincias	nombre	Procedencia	nom_provincia_proc	-
SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_paises	nombre	Procedencia	nom_pais_proc	-
SIU-Guaraní	3.9.3	sga_carreras	nombre	Carrera	nom_carrera	-
SIU-Guaraní	3.9.3	sga_carreras	unidad_academica	Carrera	unidad_academica	-
SIU-Kolla	3.7.1	recien_graduado	cond_actual	Situación laboral	sit_laboral	-
SIU-Kolla	3.7.1	recien_graduado	cant_hs_semanal	Rango horas semanales	cant_hs_semanal	-
SIU-Kolla	3.7.1	recien_graduado	satisfecho_trabajo	Satisfacción laboral	satisfaccion_trabajo	-
SIU-Kolla	3.7.1	recien_graduado	entidad_trabaja	Tipo entidad laboral	entidad_trabaja	-
SIU-Kolla	3.7.1	recien_graduado	sexo	Sexo	sexo	-
SIU-Kolla	3.7.1	recien_graduado	estado_civil	Estado civil	estado_civil	-
SIU-Kolla	3.7.1	recien_graduado	grupo_conv	Grupo convivencia	grupo_conv	-
SIU-Guarani	2.09.3	sga_personas	fecha_nacimiento	Rango etario	fecha_nacimiento	-

SIU-Kolla	3.7.1	recien_graduado	fecha_fin_enc	Tiempo	-> fecha -> mes -> anio	-
SIU-Kolla	3.7.1	recien_graduado	nro_doc	-	-	cant_graduados
SIU-Kolla	3.7.1	recien_graduado	edad	-	-	promedio_edad

- TA02 -‘Valoración académica de los graduados por carrera’ (Tabla 22)

Tabla 22: FUDA TA02- Valoración académica de los graduados por carrera [Elaboración propia]

Documento de Identificación de las Fuentes de Datos						
Tema analítico	TA02- Valoración académica de los graduados por carrera.					
Analista	Lic. María Daniela Rolón					
Base de Datos (BD)	Versión BD	Tabla	Atributo	Perspectiva	Atributo Perspectiva	Indicador
SIU-Guaraní	3.9.3	sga_carreras	nombre	Carrera	nom_carrera	-
SIU-Guaraní	3.9.3	sga_carreras	unidad_academica	Carrera	ua	-
SIU-Kolla	3.7.1	recien_graduado	motivo_elec_carrera	Selección carrera	selec_carrera	
SIU-Kolla	3.7.1	recien_graduado	sexo	Sexo	sexo	-
SIU-Kolla	3.7.1	recien_graduado	motivo_estud_univer	Selección universidad	selec_univer	
SIU-Kolla	3.7.1	recien_graduado	cont_carrera_interes	Interés contenido	interes_cont	
SIU-Kolla	3.7.1	recien_graduado	cont_carrera_expect	Expectativa contenido	expec_cont	
SIU-Kolla	3.7.1	recien_graduado	nivel_exigencia_carrera	Nivel exigencia	nivel_exig	

SIU-Kolla	3.7.1	recien_graduado	plan_estudio_corto	Plan corto	plan_corto	-
SIU-Kolla	3.7.1	recien_graduado	plan_estudio_tecnico	Plan técnico	plan_tecnico	-
SIU-Kolla	3.7.1	recien_graduado	plan_estudio_gral	Plan generalista	plan_gen	-
SIU-Kolla	3.7.1	recien_graduado	aspectos_universitaria	Valoración universidad	valor_univ	-
SIU-Kolla	3.7.1	recien_graduado	reconom_universitaria	Universitaria sociedad	valor_univ_socie	-
SIU-Kolla	3.7.1	recien_graduado	brindar_universitaria	Aporte universidad	aporte_univ	-
SIU-Kolla	3.7.1	recien_graduado	fecha_fin_enc	Tiempo	-> fecha -> mes -> anio	
SIU-Guaraní	3.9.3	sga_planes	vigencia_minima	-	-	desvio_duracion_real
SIU-Kolla	3.7.1	recien_graduado	nro_doc	-	-	cant_graduados
SIU-Guaraní	3.9.3	sga_titulos_otor	fecha_egreso	-	-	duracion_real_carrera
SIU-Guaraní	3.9.3	sga_alumnos	fecha_ingreso	-	-	

- TA03- ‘Necesidad de capacitación de los graduados por carrera según su distribución geográfica’ (Tabla 23)

Tabla 23: FUDA TA03-Necesidad de capacitación de los graduados por carrera según su distribución geográfica
[Elaboración propia]

Documento de Identificación de las Fuentes de Datos	
TA03	Necesidad de capacitación de los graduados por carrera según su distribución geográfica.
Analista	Lic. María Daniela Rolón

Base de Datos (BD)	Versión BD	Tabla	Atributo	Perspectiva	Atributo Perspectiva	Indicador
SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_localidades	nombre	Residencia	nom_ciudad_res	-
SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_dptos	nombre	Residencia	nom_dpto_res	-
SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_provincias	nombre	Residencia	nom_provincia_res	-
SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_paises	nombre	Residencia	nom_pais_res	-
SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_localidades	nombre	Procedencia	nom_ciudad_proc	-
SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_dptos	nombre	Procedencia	nom_dpto_proc	-
SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_provincias	nombre	Procedencia	nom_provincia_proc	-
SIU-Kolla SIU-Guaraní	3.7.1 3.9.3	mug_paises	nombre	Procedencia	nom_pais_proc	-
SIU-Guaraní	3.9.3	sga_carreras	nombre	Carrera	nom_carrera	-
SIU-Guaraní	3.9.3	sga_carreras	unidad_academica	Carrera	unidad_academica	-
SIU-Kolla	3.7.1	recien_graduado	sexo	Sexo	sexo	-
SIU-Kolla	3.7.1	recien_graduado	fecha_fin_enc	Tiempo	-> fecha -> mes -> anio	-
--	--	--	--	Formación continua	--	-
--	--	--	--	Área de formación	--	-
--	--	--	--	Herramientas	--	-
--	--	--	--	Tipo de	--	-

				actividad		
--	--	--	--	Idiomas	--	-
--	--	--	--	Turno proferido	--	-
--	--	--	--	Costeo del curso	--	-
--	--	--	--	Tipo de dictado	--	
SIU-Kolla	3.7.1	recien_graduado	nro_doc	-	-	cant_grad uados

Los FUDA asociados a los temas analíticos TA01, TA02 y TA03 se registran en el Redmine de FCEQyN (Ticket 298, 390 y 412 respectivamente) y se encuentran disponibles en el siguiente enlace: <https://github.com/mdarolon/BBII-FCEQyN>.

E2.A3- Cálculo de los indicadores

A continuación se observa la conformación de indicadores correspondientes a los temas analíticos TA01 (Tabla 24), TA02 (Tabla 25) y TA03 (Tabla 26).

Tabla 24: Conformación de indicadores TA01

Conformación de Indicadores			
Tema analítico	TA01- Situación laboral de los graduados por carrera según su distribución geográfica.	Fecha	03/20
Indicador	Fórmula		Descripción
cant_graduados	COUNT(nro_docu) Donde: 'nro_docu' es el número de documento del graduado		Cantidad de graduados que respondieron las encuestas.
promedio_edad	AVG(edad) Donde: 'edad' corresponde a la edad del graduado (fecha_actual – fecha_nacimiento)		Promedio de edad de los graduados.

Tabla 25: Conformación de indicadores TA02

Conformación de Indicadores			
Tema analítico	TA02- Valoración académica de los graduados por carrera.	Fecha	07/20
Indicador	Fórmula		Descripción
cant_graduados	COUNT(nro_docu) Donde: ‘nro_docu’ es el número de documento del graduado		Cantidad de graduados que respondieron las encuestas.
duracion_real_carrera	(fecha_egreso – fecha_ingreso) Donde: ‘fecha_egreso’ corresponde a la fecha de egreso del alumno a la carrera. ‘fecha_ingreso’ corresponde a la fecha de ingreso del alumno a la carrera.		Tiempo que le tomó al graduado de una Carrera completar el plan de estudio.
desvio_duracion_real	(duración_real_carrera - duración_teórica_carrera) Donde: ‘duración_real_carrera’ corresponde al tiempo que le tomo al graduado de una Carrera completar el plan de estudio. ‘duracion_teórica_carrera’ corresponde a la duración teórica de la carrera definida en el plan de estudio.		Es la diferencia entre la duración real de la carrera y la duración teórica de la misma. (0= el alumno se graduó según lo aprobado en el plan de estudio, >1 = el alumno se graduó en mas años, <1= el alumno se graduó en menos años)

Tabla 26: Conformación de indicadores TA03

Conformación de Indicadores			
Tema analítico	TA03- Necesidad de capacitación de los graduados por carrera según su distribución geográfica.	Fecha	08/20
Indicador	Fórmula		Descripción
cant_graduados	COUNT(nro_docu) Donde: ‘nro_docu’ es el número de documento del graduado		Cantidad de graduados que respondieron las encuestas.

E2.A4- Nivel de granularidad

El nivel de granularidad definido se encuentra detallado en el FUDA definido para cada tema analítico, Tabla 21 y Tabla 22 respectivamente.

El mismo permite a los expertos consultar la información con un nivel de detalle conforme a sus necesidades.

Como ejemplo de ello, la granularidad definida en la perspectiva Residencia, es una granularidad con una jerarquía de tres niveles: País, Provincia, Departamento y Localidad.

Este nivel de granularidad permite a los expertos consultar los graduados que residen tanto en el extranjero como en nuestro país, como se encuentran distribuidos a nivel Provincias, a nivel Departamento y a nivel Localidad.

E2.A5- Perspectiva Tiempo

Para el caso de estudio analizado, la perspectiva Tiempo tendrá un nivel de granularidad definida con los siguientes campos:

- Fecha. Refiere a la fecha de completitud de la encuesta.
- Mes
- Año

E2.A6- Modelo Conceptual ampliado

Determinado el nivel de granularidad y realizado el cálculos de los indicadores se procedió a ampliar el Modelo Conceptual de cada tema analítico planteado, donde se ubicó debajo de cada perspectiva los atributos correspondientes a las fuentes de datos seleccionadas, y debajo de cada indicador su fórmula de cálculo.

En la Figura 35 se observa el Modelo Conceptual ampliado del TA01 y en la Tabla 27 se describe en general cada atributo seleccionado.

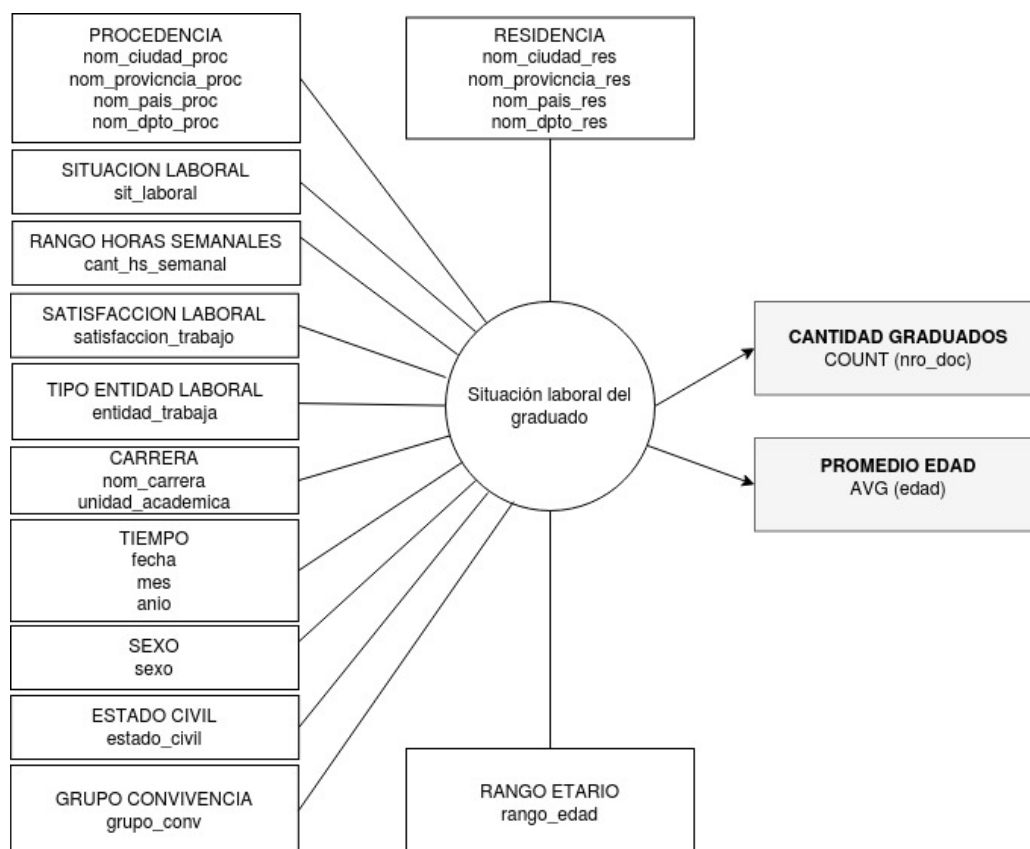


Figura 35: Modelo Conceptual ampliado: TA01-Situación laboral de los graduados por carrera según su distribución geográfica [Elaboración propia]

Donde:

Tabla 27: Descripción de atributos por perspectiva (TA01)

Perspectiva	Atributo	Descripción
Residencia	nom_ciudad_res	Ciudad de residencia del graduado.
	nom_provincia_res	Provincia de residencia del graduado.
	nom_pais_res	País de residencia del graduado.
	nom_dpto_res	Departamento de residencia del graduado.
Procedencia	nom_ciudad_proc	Ciudad de procedencia del graduado.
	nom_provincia_proc	Provincia de procedencia del graduado.
	nom_pais_proc	País de procedencia del graduado.
	nom_dpto_proc	Departamento de procedencia del graduado.
Situación laboral	sit_laboral	Situación laboral del graduado.

Rango semanales horas	cant_hs_semanal	Rango de horas semanales que trabaja el graduado.
Satisfacción laboral	satisfaccion_trabajo	Satisfacción por el trabajo que realiza.
Tipo entidad laboral	entidad_trabaja	Tipo de entidad donde trabaja el graduado.
Carrera	nom_carrera	Carrera en la que se graduó el alumno.
	unidad_academica	Unidad académica de la que se graduó el alumno.
Tiempo	fecha	Fecha de finalización de la encuesta.
	mes	Mes de finalización de la encuesta.
	anio	Año de finalización de la encuesta.
Sexo	sexo	Género del graduado.
Estado civil	estado_civil	Estado civil del graduado.
Grupo convivencia	grupo_conv	Grupo de convivencia del graduado.
Rango etario	rango_edad	Edad del graduado según rango etario.
Cantidad graduados (*)		Cantidad de graduados que respondieron las encuestas.
Promedio edad (*)		Promedio de edad de los graduados.
(*) Indicador		

En la Figura 36 se observa el Modelo Conceptual ampliado del TA02 y en la Tabla 28 se describe en general cada atributo seleccionado en el modelo.

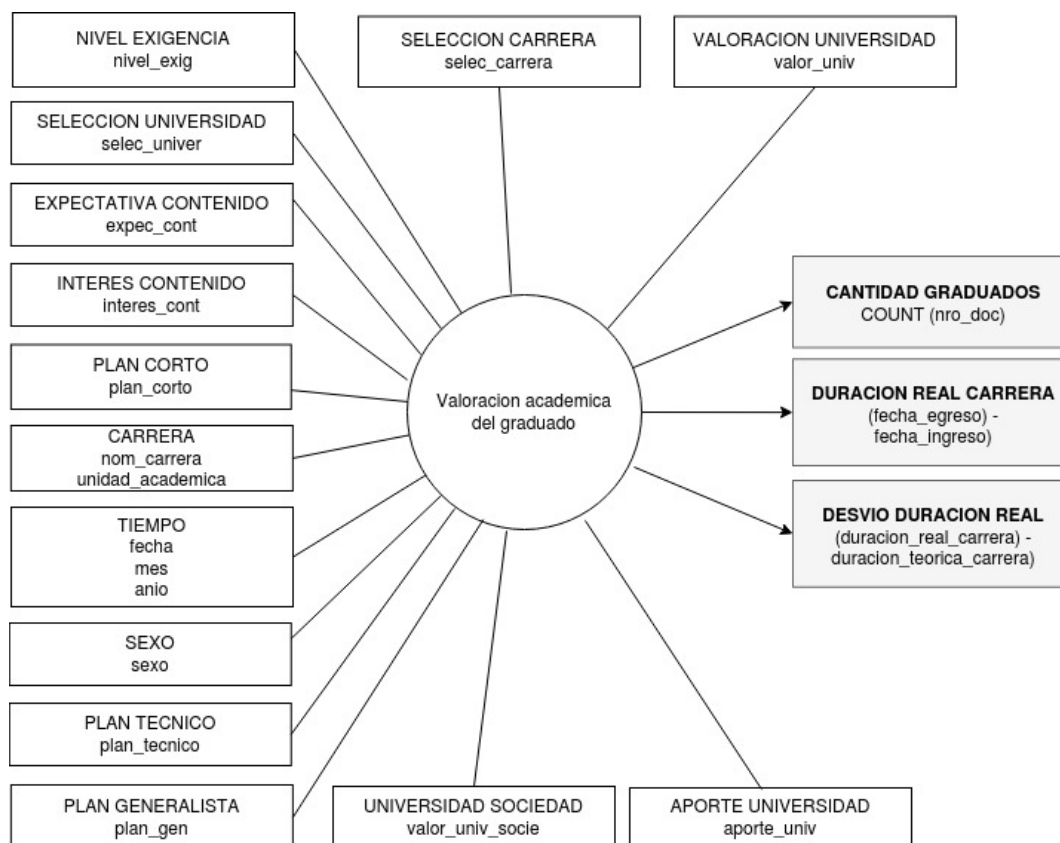


Figura 36: Modelo Conceptual ampliado: TA02- Valoración académica de los graduados por carrera
[Elaboración propia]

Donde:

Tabla 28: Descripción de atributos por perspectiva (TA02)

Perspectiva	Atributo	Descripción
Nivel exigencia	nivel_exig	Nivel de exigencia de la carrera.
Selección carrera	selec_carrera	Motivos por los cuales eligió su carrera universitaria.
Valoración universidad	valor_univ	Aspectos de la Universidad que el graduado más aprecia con relación a la carrera que estudió.
Selección universidad	selec_univer	Motivos por los cuales eligió estudiar en esta Universidad.
Expectativa contenido	expec_cont	Cumplimiento de expectativas sobre los contenidos estudiados durante su carrera.
Interés contenido	interes_cont	Interés sobre los contenidos estudiados durante la carrera.
Plan corto	plan_corto	El plan de estudios debería ser

		más corto.
Carrera	nom_carrera	Carrera en la que se graduó el alumno.
	unidad_academica	Unidad académica de la que se graduó el alumno.
Tiempo	fecha	Fecha de finalización de la encuesta.
	mes	Mes de finalización de la encuesta.
	anio	Año de finalización de la encuesta.
Sexo	sexo	Género del graduado.
Plan técnico	plan_tecnico	El plan de estudios debería tener un perfil más técnico.
Plan generalista	plan_gen	El plan de estudios debería ser más generalista.
Universidad sociedad	valor_univ_socie	Percepción personal del reconocimiento de la sociedad a esta Universidad.
Aporte universidad	aporte_univ	Otros aportes que se desea brinde la Universidad.
Cantidad graduados (*)		Cantidad de graduados que respondieron las encuestas.
Duración real carrera (*)		Tiempo que le tomo al graduado de una Carrera completar el plan de estudio.
Desvío duración real (*)		Es la diferencia entre la duración real de la carrera y la duración teórica de la misma.

(*) Indicador

E2.A7- Factibilidad y aceptación del modelo

Se realiza la verificación junto a los expertos de los modelos conceptuales diseñados, asociados al TA01 y TA02.

Durante el proceso se realiza una revisión de los requisitos planteados y se verifican que los mismos se encuentren representados en cada modelo.

Para el caso del TA01, el experto solicita la incorporación de una nueva variable referida al rango etario de los graduados, manifestando que es un requisito solicitado en las ofertas laborales que llegan a la Oficina de Vinculación con los Graduados. Esta solicitud fue considerada e incorporada al modelo, ya que se contaba con el dato requerido en los sistemas fuentes.

Con relación al TA02, el experto encuentra oportuno la incorporación de las medidas relacionadas a la duración de los graduados en la carrera. Esto aportaría al trabajo, a nivel académico, en los planes de estudios vigentes, contenidos y cantidad de asignaturas en los mismos, para mejorar y aumentar la tasa de egreso de la Facultad.

Con respecto al TA03, se informa al experto la falta de datos necesarios para cubrir la demanda de la institución y se procede a confeccionar un informe técnico al responsable de la institución, con alternativas y sugerencia de implementación y/o cambios en los circuitos administrativos o sistemas fuentes.

Una de las alternativas propuestas fue implementar la ‘Encuesta al año de graduado’ propuesta por el SIU. La misma contempla algunas de las variables solicitadas por el experto, como el fortalecimiento en la formación del graduado en áreas básicas, aplicadas, herramientas informáticas específicas relacionadas a su profesión e idiomas. Otra alternativa planteada, personalizar las encuestas existentes o generar una nueva encuesta (Ticket Nro.: 413).

Una vez validados con los expertos los modelos desarrollados, y obtenido la aceptación de los mismos mediante reuniones virtuales y videoconferencias (Contexto ASPO – Aislamiento Social Preventivo y Obligatorio [52]), se procede a actualizar el Modelo conceptual preliminar y el Modelo Conceptual ampliado asociados a cada tema analítico, de manera tal de garantizar la trazabilidad de los requisitos.

5.3.3 E3: Diseño y Modelado Multidimensional

Esta etapa involucra, teniendo como insumo los modelos conceptuales ampliados desarrollados en la etapa anterior, tareas de diseño de las dimensiones y tablas de hechos que luego se utilizan en la navegación, filtro y agrupamiento de la información definida para cada tema analítico planteado.

E3.A1- Esquema de modelado

El diseño lógico para cada tema analítico que se analiza se basa en un esquema estrella, donde cada tabla de dimensión tiene relación únicamente con una tabla de hechos y toda la información relacionada a cada dimensión se encuentra en una sola tabla.

Este tipo de esquema permite mayor eficiencia al momento de realizar las consultas.

E3.A2- -Diseño Tablas de Dimensiones

A continuación se procedió a construir las tablas de dimensiones para cada tema analítico.

Para la creación de las mismas se tuvieron en cuenta los estándares de desarrollo utilizados por el SIU para el sistema SIU-Wichi [55].

A manera de ejemplo, en la Tabla 29 se observa la conformación de la dimensión d_carrera, utilizada en ambos temas analíticos:

Tabla 29: DTD d_carrera

Definición de TABLAS de DIMENSIONES			
Tema analítico	TA01-Situación laboral del graduado por carrera según su distribución geográfica.		
Perspectiva	Carrera		
Dimensión	d_carrera		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
carrera_id	Integer	sga_carreras	Carrera
nom_carrera	Character varying	sga_carreras	Nombre
unidad_academica	Character varying	sga_carreras	Nombre

Los DTD asociados a los TA01 y al TA02 se encuentran detallados en el ANEXO 3, registrados en el Redmine de FCEQyN (Ticket 299 y 397 respectivamente), y compartidos en el siguiente enlace: <https://github.com/mdarolon/BBII-FCEQyN>.

E3.A3- -Diseño Tablas de Hechos

Para cada proceso de negocio modelado se generó una tabla de hechos que contiene los valores numéricos (medidas) y las claves de dimensiones asociadas al mismo.

Para la creación de las mismas se tuvo en cuenta los estándares de desarrollo utilizados por el SIU para el sistema SIU-Wichi.

En la Tabla 30 se observa la conformación de la tabla de hechos ft_situacion_laboral, utilizada en TA01.

Tabla 30: DTH ft_situacion_laboral

Definición de TABLAS de HECHOS			
Tema analítico	TA01-Situación laboral del graduado por carrera según su distribución geográfica.		
Hecho	ft_situacion_laboral		
Campo Dimensión	Dimensión	Medida	Calculo
carrera_id	d_carreras		
ciudades_id	d_residencia		
ciudadproc_id	d_procedencia		
sit_lab_id	d_situacion_laboral		

cant_hs_sem_id	d_cant_hs_sem		
sexo_id	d_sexo		
tiempo_id	d_tiempo		
estadocivil_id	d_estado_civil		
grupoconv_id	d_grupo_conv		
satisf_trabajo_id	d_satisf_trabajo		
entidad_id	d_entidad_trabaja		
edad_id	d_redad		
promedio_edad		Promedio	AVG(edad)
cant_graduados		Cantidad	COUNT(nro_docu)

En la Tabla 31 se observa la conformación de la tabla de hechos ft_valoracion_academica, utilizada en TA02.

Tabla 31: DTH ft_valoracion_academica

Definición de TABLAS de HECHOS			
Tema analítico	TA02- Valoración académica de los graduados por carrera.		
Hecho	ft_valoracion_academica		
Campo Dimensión	Dimensión	Medida	Calculo
carrera_id	d_carreras		
nivel_exig_id	d_nivel_exigencia		
selec_carrera_id	d_selec_carrera		
expec_cont_id	d_expec_cont		
interes_cont_id	d_interes_cont		
sexo_id	d_sexo		
tiempo_id	d_tiempo		
plan_corto_id	d_plan_corto		
plan_tecnico_id	d_plan_tecnico		
selec_carrera_id	d_selec_carrera		
valor_univ_id	d_valor_univ		

plan_gen_id	d_plan_gen		
valor_univ_socie_id	d_valor_univ_socie		
aporte_univ_id	d_aporte_univ		
duración_real_carrera		Promedio	fecha_egreso – fecha_ingreso
cant_graduados		Cantidad	COUNT(nro_docu)
desvio_duracion_real		Promedio	duración_real_carrera – duración_teorica_carrera

Los DTH asociados a los TA01 y al TA02 se encuentran registrados en el Redmine de FCEQyN (Ticket 300 y 398 respectivamente), y compartidos en el siguiente enlace: <https://github.com/mdarolon/BBII-FCEQyN>.

E3.A4- **Modelo Físico**

A partir de las dimensiones y las tablas de hechos definidas en la presente etapa se procede a realizar las uniones correspondientes entre las mismas para conformar los modelos de datos.

Para la confección del modelo de datos se utiliza la herramienta open source “SQL Power Architect Community Edition”.

En las Figura 37 y Figura 38 se observan los modelos de datos desarrollados.

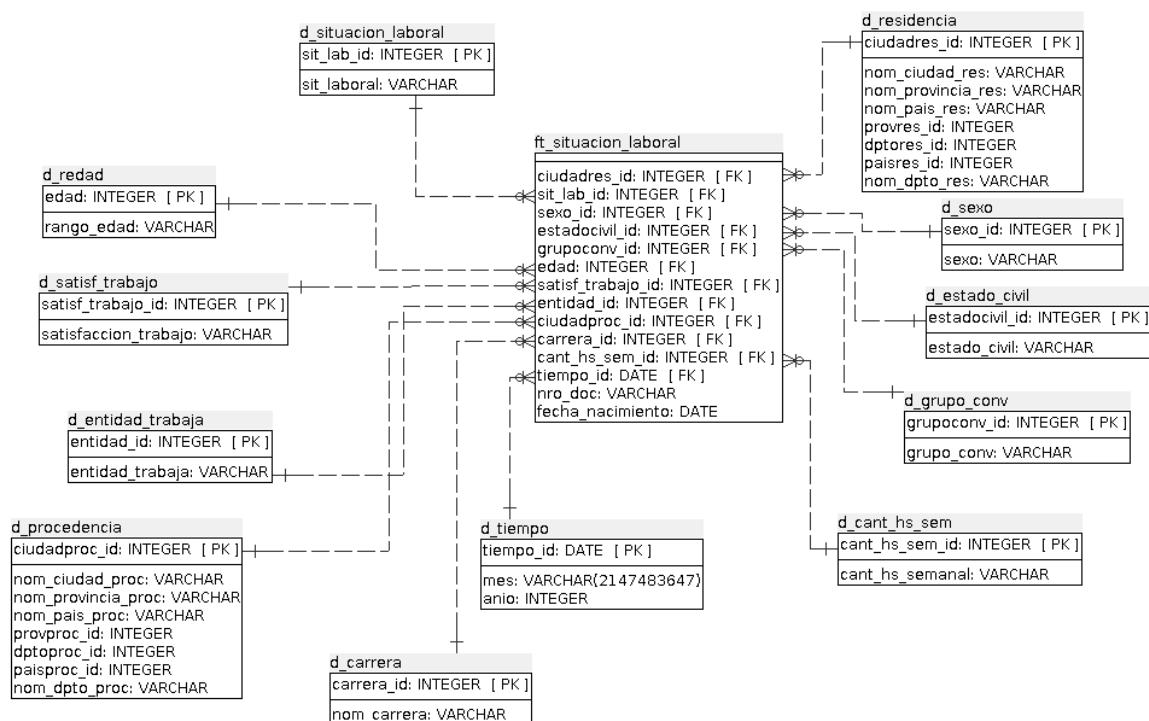


Figura 37: Modelo Físico: TA01-Situación laboral de los graduados por carrera según su distribución geográfica [Elaboración propia]

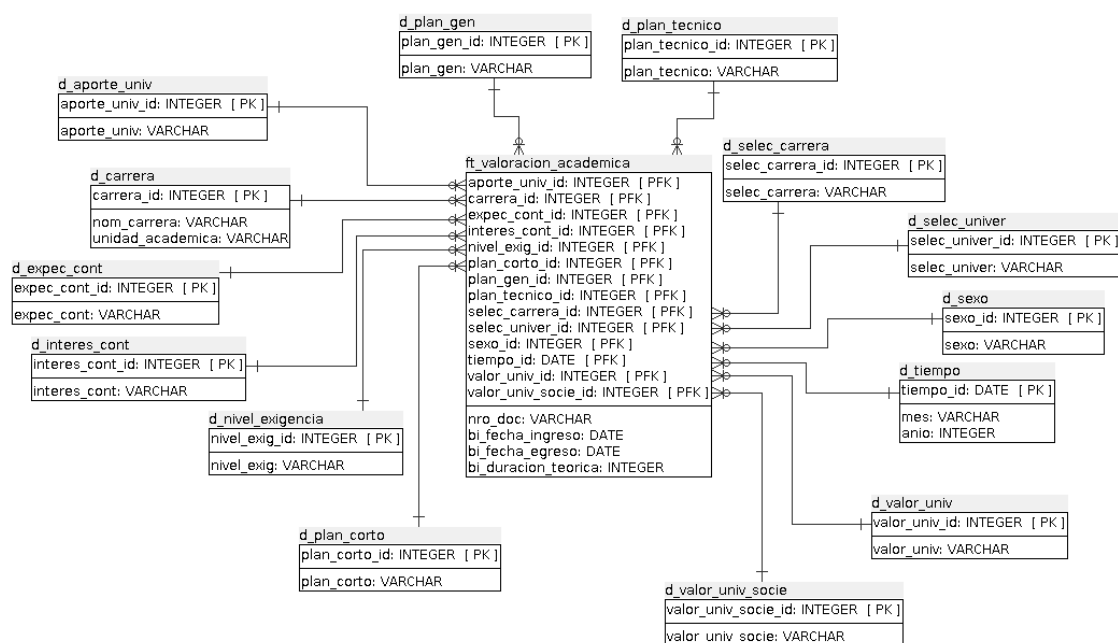


Figura 38: Modelo Físico: TA02- Valoración académica de los graduados por carrera [Elaboración propia]

5.3.4 E4: Procesos ETL

Para el tratamiento e integración de los datos, la arquitectura propuesta y utilizada se describe a continuación.

El conjunto de datos necesarios para el presente análisis (datasets) se los extrajo de los sistemas fuentes: SIU-Guaraní y SIU-Kolla.

Del sistema fuente SIU-Guaraní se extrajeron los datos de dos instalaciones distintas. Una instalada en la Sede Posadas y otra instalada en la Sede Apóstoles. Ambas correspondientes a la misma Unidad Académica (FCEQyN).

Con respecto al sistema SIU-Kolla, se extrajeron los datos de una única instalación implementada en la FCEQyN. En dicha instalación se encuentran contenidos los datos de ambas Sedes.

Mediante procesos de extracción y carga se almacenaron los datos en un área de almacenamiento temporal (DSA) sin ningún tipo de transformación. Luego mediante otro conjunto de procesos ETL, se tomaron los datos del DSA y se los depuró, transformó y limpió para su uso posterior.

Para la carga de las tablas correspondientes al Modelo Físico definido en la etapa anterior (Etapa 3), se ejecutó un nuevo conjunto de procesos ETL para extraer los datos transformados del DSA y poblar de esta manera el Data Mart.

Por último se exploraron los datos a través de las distintas herramientas de BI: análisis OLAP, reportes y dashboard (tableros de control).

En la Figura 39 se observa la arquitectura propuesta y utilizada en la presente investigación.

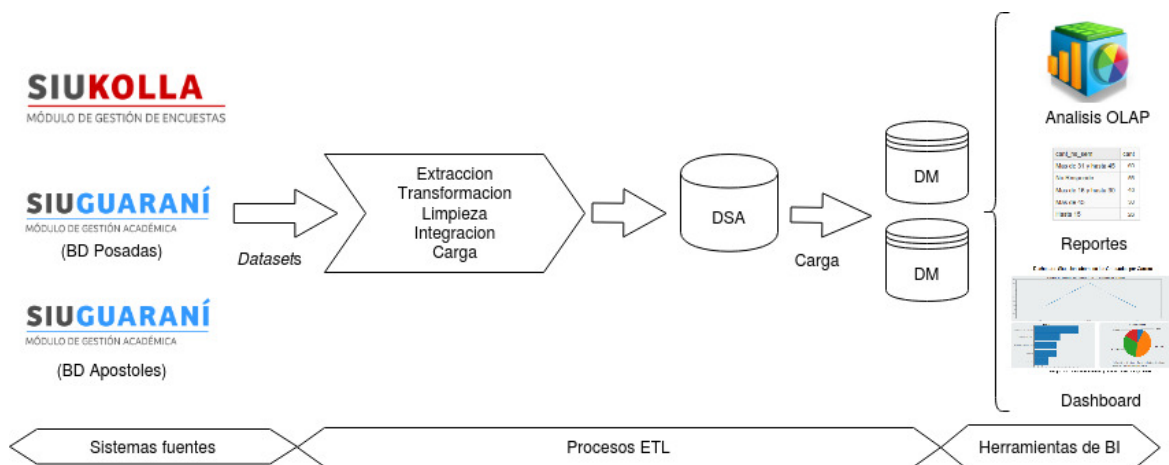


Figura 39: Arquitectura propuesta y utilizada para la carga de los Data Marts [Elaboración propia]

E4.A1-Elección de la herramienta de software

Para el desarrollo e implementación de la solución propuesta se utilizan herramientas de software libre.

Como servidor de BI Pentaho community bajo licencia GNU GPL, en su versión 8.0. [56], y como motor de base de datos open source PostgreSQL en su versión 9.2.

Ambas herramientas son utilizadas por el SIU para el desarrollo de sus distintas actividades de inteligencia de negocios, además de ser la plataforma en la cual ya se basan implementaciones en la UNaM como el SIU-Wichi.

E4.A2- Fuentes de datos

Las transacciones relacionadas a Alumnos y Graduados se registran en sistemas gestionados por personal de la Dirección de Enseñanza de la FCEQyN de la UNaM.

Para el caso de estudio, los sistemas transaccionales utilizados como fuentes de datos para el análisis fueron: el SIU-Guaraní (Sistema de Gestión Académica) y el SIU-Kolla (Sistema de Gestión de Encuestas) instalados e implementados en la FCEQyN.

E4.A3- Selección de datos

La selección de los datos se realizó recabando y analizando los datos correspondientes a la “Encuesta Recién Graduado” realizadas a los estudiantes de la FCEQyN que culminaron sus estudios en los años 2018, 2019 y en los meses de Enero, Febrero, Marzo, Abril, Mayo y Junio del año 2020; además de haber gestionado su trámite de título a través del sistema SIU-Guaraní.

El procesamiento de dichas encuestas incluye la oferta académica de pregrado, grado y postgrado de la FCEQyN.

El volumen de los datos analizados hasta Junio del año 2020 fue de 280 registros; de un total de 391 graduados que realizaron su trámite de título en el sistema SIU-Guaraní.

Los datos utilizados fueron extraídos de los sistemas fuentes mediante consultas SQL y cargados en el DSA sin ninguna transformación. Las consultas SQL generadas para la construcción de los conjuntos de datos (datasets A, B, C, D y F) pueden observarse en el ANEXO 4.

Los datos extraídos se los describen a continuación (Tabla 32):

Tabla 32: Descripción de los datos de los datasets A, B, C, D y F

Atributo	Descripción	OLTP	dataset
unidad_academica	Unidad Académica	SIU-Guaraní	A
carrera	Código de la carrera	SIU-Guaraní	A

nombre	Nombre de la carrera	SIU-Guaraní	A
fecha_egreso	Fecha de egreso del graduado	SIU-Guaraní	B
nro_inscripcion	Número de inscripción del graduado.	SIU-Guaraní	B
plan	Plan de estudios correspondiente a la carrera del graduado.	SIU-Guaraní	B
colación	Número de colación en la que participó el graduado.	SIU-Guaraní	B
fecha_ingreso	Fecha de ingreso del graduado a la carrera.	SIU-Guaraní	B
tipo_documento	Tipo de documento del graduado.	SIU-Guaraní	B
nro_documento	Número de documento del graduado.	SIU-Guaraní	B
fecha_nacimiento	Fecha nacimiento del graduado.	SIU-Guaraní	B
localidad	Código localidad	SIU-Guaraní	C
nombre	Nombre de la localidad	SIU-Guaraní	C
dpto_partido	Código del departamento	SIU-Guaraní	C
nombre	Nombre del departamento.	SIU-Guaraní	C
provincia	Código de la provincia	SIU-Guaraní	C
nombre	Nombre de la provincia.	SIU-Guaraní	C
país	Código del país	SIU-Guaraní	C
nombre	Nombre del país.	SIU-Guaraní	C
colación	Código de colación	SIU-Guaraní	D
descripción	Descripción de la colación	SIU-Guaraní	D
anio_colacion	Año de la colación	SIU-Guaraní	D
unidad_academica	Unidad Académica	SIU-Guaraní	F
carrera	Código de la carrera	SIU-Guaraní	F
plan	Plan de estudios correspondiente a la carrera.	SIU-Guaraní	F
vigencia_minima	Vigencia mínima en años del plan de estudios asociado a la carrera.	SIU-Guaraní	F

Datos provenientes del sistema SIU-Kolla:

Resultados de las encuestas: El dataset E correspondiente a las respuestas de la “Encuestas al Recién Graduado” se obtiene desde el perfil de gestión del sistema SIU-Kolla. Como se observa en la Figura 40, desde el menú “Emisión para Descarga”, opción “Datos de la Generación” se seleccionaron las opciones correspondientes a la generación de la encuesta a descargar.

Figura 40: Datos de generación del reporte desde el SIU-Kolla

Luego se exporta el archivo generado, como se observa en la Figura 41 y se lo descarga desde la opción “Listado de reportes Generados”, como se observa en la Figura 42.

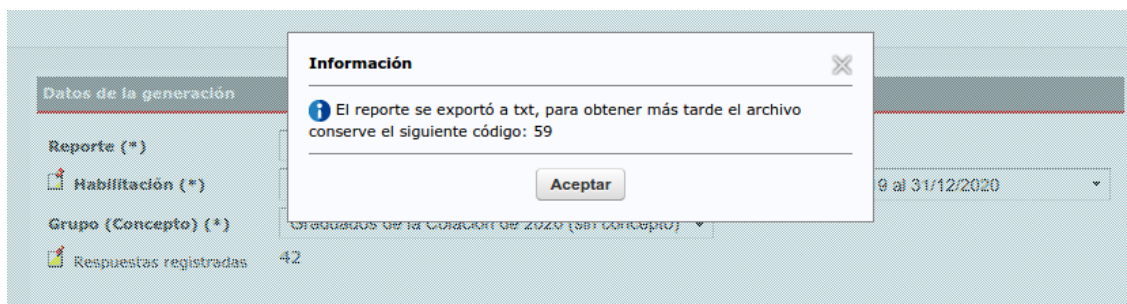


Figura 41: Exportación del reporte generado desde el SIU-Kolla

Listado de Reportes Generados											
Código	Fecha de generación	Habilitación	Formulario	Concepto	Tipo	Desde	Hasta	Inconclusas	Cod. Respuestas	Encuesta	Elemento
59		Encuesta al Recién Graduado. Colación 2020.	Encuesta al Recién Graduado		Resultados por Encuestado			NO	NO		
58		Encuesta al Recién Graduado. Colación 2020.	Encuesta al Recién Graduado		Resultados de habilitación por encuestado			NO	NO		

Figura 42: Listado de reportes Generados en SIU-Kolla

A continuación se identifican y recopilan del sistema SIU-Kolla, los atributos de interés necesarios para el análisis de los datos, relacionados a los temas analíticos seleccionados. Como se describen en la Tabla 33, los datos correspondientes al dataset E son los siguientes:

Tabla 33: Descripción de los datos del dataset E

Atributo	Descripción	OLTP
Título	Título universitario obtenido por el graduado.	SIU-Kolla
ciudad_res	Ciudad de residencia del graduado	SIU-Kolla
prov_res	Provincia de residencia correspondiente a la ciudad de residencia del graduado.	SIU-Kolla
país_res	País de residencia correspondiente a la provincia de residencia del graduado.	SIU-Kolla
ciudad_proc	Ciudad de procedencia del graduado	SIU-Kolla
prov_proc	Provincia de procedencia correspondiente a la ciudad de procedencia del graduado.	SIU-Kolla
país_proc	País de procedencia correspondiente a la provincia de procedencia del graduado.	SIU-Kolla
cond_actual	Condición laboral actual del graduado.	SIU-Kolla
cant_hs_semanal	Rango de horas semanales que trabaja el graduado.	SIU-Kolla
satisfecho_trabajo	Satisfacción por el trabajo que realiza.	SIU-Kolla
entidad_trabaja	Tipo de entidad donde trabaja el graduado.	SIU-Kolla
Sexo	Género del graduado	SIU-Kolla
estado_civil	Estado civil del graduado	SIU-Kolla
grupo_conv	Grupo de convivencia del graduado.	SIU-Kolla
fecha_nac	Fecha de nacimiento del graduado.	SIU-Kolla
fecha_fin_enc	Fecha de finalización de la encuesta.	SIU-Kolla
nro_doc	Número del documento nacional de identidad del graduado.	SIU-Kolla
motivo_elec_carrera	Motivo por el cual eligió la carrera.	SIU-Kolla
motivo_estud_univer	Motivo por el cual decidió estudiar en la universidad.	SIU-Kolla

cont_carrera_interes	Interés demostrado por el contenido de la carrera.	SIU-Kolla
cont_carrera_expec	Expectativas por el contenido de la carrera.	SIU-Kolla
nivel_exigencia_carrera	Nivel de exigencia de la carrera.	SIU-Kolla
plan_estudio_corto	El plan de estudios debería ser mas corto.	SIU-Kolla
plan_estudio_largo	El plan de estudios debería ser mas largo.	SIU-Kolla
plan_estudio_tecnico	El plan de estudios debería ser más técnico.	SIU-Kolla
plan_estudio_gral	El plan de estudios debería ser más generalista.	SIU-Kolla
volver_univer	Volver a estudiar a la universidad.	SIU-Kolla
aspectos_univer_aprecia	Aspecto de la universidad que el graduado aprecia.	SIU-Kolla
reconom_univer_socie	Reconocimiento de la universidad en la sociedad.	SIU-Kolla
brindar_univer	Que espera el graduado le brinde la universidad.	SIU-Kolla

E4.A4- Área de almacenamiento temporal

Se establece un área de almacenamiento temporal denominada DSA donde se almacenan temporalmente los datos obtenidos de las fuentes de datos SIU-Guaraní y SIU-Kolla.

Desde el DSA se ejecutan las tareas de extracción, transformación, limpieza y carga en las tablas correspondientes al Modelo Físico diseñado en la etapa E3.A4 – **Modelo Físico**. Esto permite la preparación e integración de los datos para garantizar que la información, que luego se cargue en el Data Mart, sea de calidad.

En PostgreSQL se crea una base de datos denominada ‘graduados’ y en ella, un área de almacenamiento temporal mediante la creación del esquema DSA.

E4.A5- Tratamiento de los datos

Para contener los datos provenientes del sistema SIU-Kolla, correspondientes a las respuestas de las encuestas, dentro del esquema DSA se crea una tabla cuya estructura soporta los datos de cada encuesta.

En dicha tabla (*dsa.recien_graduado*), se genera una clave primaria (Primary Key) para poder manipular los datos contenidos en la misma. En la Figura 43 se observa la creación de la misma.

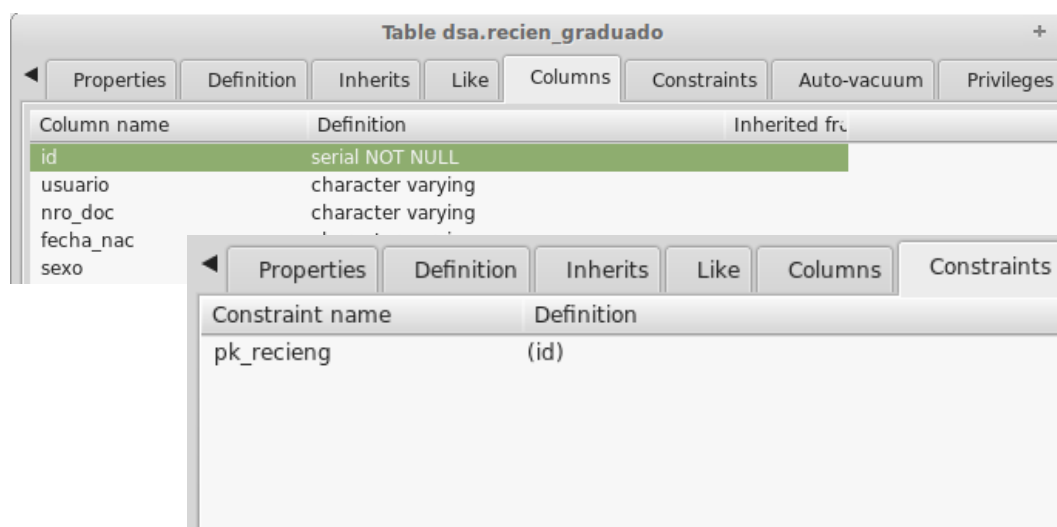


Figura 43: Implementación de una Primary Key en la tabla *dsa.recien_graduado*.

Para facilitar la carga de los metadatos de las encuestas en la tabla *dsa.recien_graduado*, con la herramienta Pentaho Data Integration, se crea una primera transformación que toma los datos contenidos en el archivo plano denominado *resultados.txt*, el cual contiene las respuestas de las encuestas, y los importa en la tabla *recién_graduado* dentro del esquema DSA (Figura 44).

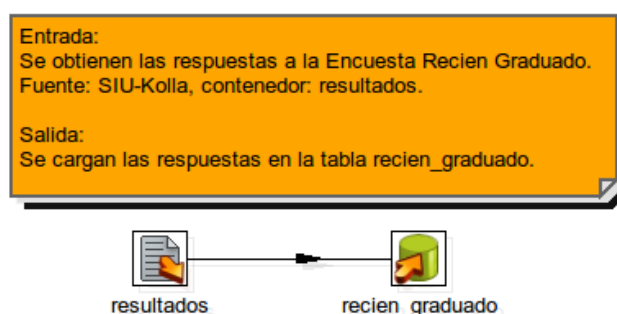


Figura 44: Proceso ETL *dsa_recien_graduado*

Previo a la importación de los datos en la tabla *dsa.recien_graduado*, se realiza un mapeo de campos. Dicho mapeo es realizado en el objeto *recién_graduado*, y el mismo puede observarse en la Figura 45.

Ingreso Mapeo

Campos de origen:	Campos destino:	Mapeos:
Respondida por	id	Usuario --> usuario
Apellido y Nombres	bi_edad	Sexo --> sexo
Tipo de documento	interes_formac_continua	Nacionalidad --> nacionalidad
Está interesado en cursos de formación continua	fecha_inicio_enc	Fecha de nacimiento --> fecha_nac
Fecha de inicio	fecha_fin_enc	Estado Civil --> estado_civil
Fecha fin		Cantidad de Hijos --> cant_hijos
		Correo Electrónico --> mail
		Grupo de Convivencia --> grupo_conv
		Título --> titulo
		Año de Ingreso --> anio_ingreso
		Ciudad --> ciudad_res
		Provincia --> prov_res
		País --> pais_res
		Ciudad_1 --> ciudad_proc

Figura 45: Mapeo de campos en la tabla *dsa.recien_graduado*

Durante la ejecución del proceso de mapeo de los campos, se detectaron numerosos errores en los tipos de datos, los cuales se analizaron y verificaron, llegando a la conclusión que el problema se encontraba en el tipo de preguntas abiertas y en los campos descriptivos. Esta particularidad imposibilita la categorización y comparación de los mismos.

Entre las mismas se identificaron los siguientes campos: “Título”, “Fecha de nacimiento”, “Año de ingreso”, “Ciudad_a”, “Provincia_a” y “País_a” correspondientes a la procedencia y “Ciudad”, “Provincia” y “País” correspondientes a la residencia del graduado.

Como se observa en la Figura 46, se ejemplifica lo expresado anteriormente. Un graduado en el campo “Año de Ingreso” ha cargado el valor ‘2009-2012’ y en el campo “Título” los dos títulos obtenidos en la unidad académica: ‘Licenciada en Genética y Profesora en Biología’; es decir, el último por el cual se le disparó la encuesta y el obtenido anteriormente.

Esto conlleva a una alteración de la información que se tiene que procesar, impactando directamente en la calidad de los datos.

Título name	Año de Ingreso name	Ciudad_a name	Provincia_a name	País_a name
Magister en Tecnologías de la Información	2016	Concepción	Misiones	Argentina
PROFESOR EN BIOLOGIA	2010	POSADAS	MISIONES	ARGENTINA
Profesor Universitario en Computación	2015	Apóstoles	Misiones	Argentina
Magister en Salud Publica y Enf.Transmisibles	2014	Oberá	Misiones	Argentina
Doctor en Ciencias Aplicadas	2014	Posadas	Misiones	Argentina
Licenciada en Genética y Profesora en Biología	2009-2012	Posadas	Misiones	Argentina
Porfesor Universitario en Computación	2015	Posadas	Misiones	Argentina
BIOQUIMICO	2012	Garupa	Misiones	Argentina
Doctor	2012	Posadas	Misiones	Argentina
Ingeniera Química	2006	Posadas	Misiones	Argentina
Profesor en Biología	2010	Posadas	Misiones	Argentina

Figura 46: Datos cargados en la Encuesta al Recién Graduado referidos a Título, Año de Ingreso, Ciudad, Provincia, País de procedencia

- Esta actividad involucra al tratamiento de campos editables que se generan como resultado de las respuestas abiertas que se encuentran definidas en la “Encuesta al Recién Graduado”.

Campo “Título”

En el atributo “título” se detecta que el nombre del título al cual accedió el graduado se encontraba escrito de maneras diferentes.

Como ejemplo de lo descripto anteriormente se encuentra el título “Doctor en Ciencias Aplicadas” y “Doctor”. Ambos refieren al mismo título de postgrado; pero al no estar escrito de la misma manera son considerados títulos diferentes, afectando directamente el procesamiento de los datos, distorsionando la información obtenida.

Para subsanar el inconveniente, se opta por extraer el dato de la carrera asociada al graduado directamente del sistema fuente SIU-Guaraní, para garantizar que la información extraída sea la institucionalmente correcta.

Se incorpora el driver JDBC (Java Database Connectivity) propio del motor de base de datos Informix a la herramienta Pentaho Data Integration, para poder establecer la comunicación entre Spoon y el sistema SIU-Guaraní.

Se crea una nueva columna en el `dsa.recien_graduado` con el prefijo `bi_` (`bi_carrera`) para realizar el mapeo correspondiente de la carrera.

Campo “Fecha de nacimiento”

El campo “`fecha_nac`” obtenido de la encuesta también fue identificado como un campo editable por lo que se ha encontrado disparidad a la hora de seleccionar el mismo y proporcionar un formato adecuado (*date*) para poder ser procesado y analizado.

Como puede observarse en la Figura 47, el campo “`fecha_nac`” presenta distintas formas de carga y simbologías, algunas con ‘/’, otras con ‘-’, otras con formato de texto, sin mencionar la longitud del dato.

fecha_nac character varying
27/07/1992
26/07/79
10/08/1992
23/11/1995
17-04-85
14/11/1995
13/06/1992
29 septiembre de 1991
05/07/2019
20/04/75
18/05/1991

Figura 47: Datos cargados en la Encuesta al Recién Graduado referidos a fecha de nacimiento.

Para subsanar el inconveniente, se opta por extraer este dato del sistema fuente SIU-Guaraní, ya que el mismo es cargado por personal del Departamento de Títulos, Diplomas y Legalizaciones de la Dirección de Enseñanza de la Facultad al momento de graduarse el alumno, garantizando que el dato sea verídico y se encuentre actualizado y cargado correctamente.

Por lo tanto, la fecha de nacimiento se extrae del sistema fuente SIU-Guaraní, y se importa a la tabla `dsa.graduado_colacion` del DSA mediante la transformación denominada `dsa_gradCola_ifx`.

Se crea una nueva columna en la tabla `dsa.recien_graduado` con el prefijo `bi_` (`bi_fecha_nac`) para realizar el mapeo correspondiente a la fecha de nacimiento.

Campos “Ciudad”, “Provincia” y “País”

Otros campos configurados como editables dentro de la encuesta, corresponde a los de la ciudad, provincia y país de procedencia y residencia del graduado.

Esto origina disparidad a la hora de seleccionar los mismos y darles una identificación única para su análisis.

Como puede observarse en la Figura 48, los datos referidos al atributo “ciudad_res” se encuentran nombrados de diferentes maneras. Ejemplo de ello: ‘Buenos Aires’, ‘CABA’ y ‘Ciudad Autónoma de Buenos Aires’ refieren a la misma ciudad pero como no poseen una misma descripción son considerados ciudades distintas afectando directamente en el procesamiento de los datos, distorsionando la información registrada en el sistema.

ciudad_res character varying	prov_res character varying	pais_res character varying
Apóstoles	Misiones	Argentina
Aristóbulo del Valle	Misiones	Argentina
Buenos Aires		Argentina
CABA	Buenos Aires	Argentina
Caleta Olivia	Santa Cruz	Argentina
Capioví	Misiones	Argentina
Capital Federal	Buenos Aires	Argentina
Ciudad Autónoma de Buenos Aires	Buenos Aires	Argentina
Col. Alicia Alta	Misiones	Argentina
Colonia Aurora	Misiones	Argentina

Figura 48: Datos cargados en la Encuesta al Recién Graduado referidos a la ciudad de residencia.

Otros ejemplos se pueden observar en la Figura 49, donde el campo que refiere a la provincia de residencia ‘Mendoza’ se encuentra escrito en diferentes formatos. La misma situación se observa con la provincia de residencia ‘Misiones’.

ciudad_res character varying	prov_res character varying	pais_res character varying
Ituzaingó	Corrientes	Argentina
Concepción del Uruguay	Entre Rios	Argentina
Hohenau	Itapua	Paraguay
San Salvador de Jujuy	Jujuy	Argentina
guaymallen	mendoza	argentina
Mendoza	Mendoza	Argentina
posadas	misiones	Argentina
posadas	misiones	argentna
posadas	misiones	argentina
posadas	misiones	argentina
Aristóbulo del Valle	Misiones	Argentina

Figura 49: Datos cargados en la Encuesta al Recién Graduado referidos a la provincia y país de residencia.

Ante esta situación se analiza como alternativa de solución, extraer estos datos del sistema fuente SIU-Guaraní, proveniente del trámite de título que realiza el alumno al momento de graduarse.

Sin embargo, durante el análisis de esta alternativa se detecta que los datos referidos a la procedencia de un graduado en el sistema fuente SIU-Guaraní, no se encuentra cargado para todos los registros consultados.

Se divisa que alrededor de un 14 % de los graduados no tienen el dato cargado o el mismo no se encuentra actualizado.

A pesar que para la realización del trámite de título, el sistema SIU-Guaraní contempla la carga de dichos campos, los relacionados a la procedencia no son obligatorios para la emisión del Certificado Analítico de Finalización de Carrera (Ted1.c). Esta situación, conlleva a que los datos referidos a la procedencia no se encuentren cargados ni actualizados, en su totalidad.

Por lo tanto este valor no contribuye a solucionar el problema surgido en el origen de los datos relacionados a la procedencia de un graduado. Por lo que se decide utilizar el catálogo de localidades propuesto por el SIU e implementado en todos los sistemas transaccionales provistos por éste.

Se crea una nueva transformación denominada `dsa_localidad_ifx` que extrae los datos correspondientes a ciudad, departamento, provincia y país del catálogo unificado de localidades desde el sistema SIU-Guaraní y se los importa en la tabla `dsa.localidad` del DSA.

Durante el proceso descrito, se opta por utilizar como clave primaria el *id* configurado en las propias tablas del sistema SIU-Guaraní, debido a que dichos códigos son compartidos en los distintos sistemas transaccionales provistos por el SIU.

Luego, en la tabla `dsa.recien_graduado` correspondiente a la base de datos ‘graduados’ PostgreSQL se crean dos nuevas columnas con el prefijo `bi_` (`bi_loc_res` y `bi_loc_proc`) para imputar los códigos de las localidades de residencia y procedencia respectivamente.

En la Figura 50 y la Figura 51 se observa la imputación de valores realizada.

Script SQL:

```
--QUITILIPÍ (CHACO)
UPDATE dsa.recien_graduado SET bi_loc_proc = '5415'
WHERE TRANSLATE(UPPER(TRIM(ciudad_proc)), 'áéíóúÃÊÍÓÚ', 'aeiouAEIOU') LIKE '%QUITILIPÍ%'
AND TRANSLATE(UPPER(TRIM(prov_proc)), 'áéíóúÃÊÍÓÚ', 'aeiouAEIOU') LIKE '%CHACO%';

--CONCEPCION DE LA SIERRA (MISIONES)
UPDATE dsa.recien_graduado SET bi_loc_proc = '4297'
WHERE TRANSLATE(UPPER(TRIM(ciudad_proc)), 'áéíóúÃÊÍÓÚ', 'aeiouAEIOU') LIKE '%CONCEPCION%'
AND TRANSLATE(UPPER(TRIM(prov_proc)), 'áéíóúÃÊÍÓÚ', 'aeiouAEIOU') LIKE '%MISIONES%';
```

Figura 50: Imputación de valores en `bi_loc_proc`

Script SQL:

```
--MENDOZA (MENDOZA)
UPDATE dsa.recien_graduado SET bi_loc_res = '13208'
WHERE TRANSLATE(UPPER(TRIM(ciudad_res)), 'áéíóúÃÊÍÓÚ', 'aeiouAEIOU') LIKE '%MENDOZA%'
AND TRANSLATE(UPPER(TRIM(prov_res)), 'áéíóúÃÊÍÓÚ', 'aeiouAEIOU') LIKE '%MENDOZA%';

--JARDIN AMERICA (MISIONES)
UPDATE dsa.recien_graduado SET bi_loc_res = '4122'
WHERE TRANSLATE(UPPER(TRIM(ciudad_res)), 'áéíóúÃÊÍÓÚ', 'aeiouAEIOU') LIKE '%JARDIN AMERICA%'
AND TRANSLATE(UPPER(TRIM(prov_res)), 'áéíóúÃÊÍÓÚ', 'aeiouAEIOU') LIKE '%MISIONES%';
```

Figura 51: Imputación de valores en `bi_loc_res`

- Esta actividad involucra al tratamiento de los datos faltantes que se generan como resultado de las respuestas no obligatorias que se encuentran definidas en la “Encuesta al Recién Graduado”

En la Encuesta al Recién Graduado existen campos definidos como **no obligatorios**, los cuales pueden ser cargados o no por el graduado. Esto promueve la aparición de datos faltantes o nulos.

Para el caso de estudio que se analiza, estos datos fueron considerados durante el proceso de análisis, de modo tal de evaluar el impacto que esto conlleva a la hora de tomar decisiones institucionales y en consecuencia, considerar la definición de lineamiento que ayuden a mejorar la calidad de los datos.

- Tratamiento de los datos duplicados

No se han detectado datos duplicados en la muestra de datos seleccionada para el análisis.

- Generación de claves subrogadas

Para la creación de las distintas dimensiones se generan y utilizan claves subrogadas de tipo numérica.

Para las dimensiones relacionadas a las carreras y localidades, tanto de residencia como de procedencia, se utilizan las claves provenientes de los sistemas fuentes para facilitar la integración de los datos con nuevas ofertas académicas de la UNaM.

- Especificación del formato de los datos

En el esquema DSA de la base de datos ‘graduados’ se crea una tabla con prefijo d_ por cada una de las dimensiones a transformar y utilizar. Los atributos con prefijo bi_ identifican los atributos transformados.

El valor del atributo “edad” del graduado es calculado mediante una función denominada dsa.f_edad (date) creada en PostgreSQL. El valor de dicho atributo se encuentra representado en el atributo “bi_edad” dentro del DSA.

La dimensión tiempo y la dimensión rango etario, a diferencia de las demás dimensiones, se cargan mediante funciones: dsa.spc_dim_tiempo(date, integer) y dsa.spc_dim_edad(integer, integer) respectivamente. Estas funciones se encuentran creadas y almacenadas en PostgreSQL.

Estas transformaciones permiten que los datos sean agrupados y filtrados correctamente brindando, de este modo, veracidad a la información resultante sobre los temas analíticos analizados.

Las transformaciones pueden descargarse desde el siguiente enlace: <https://github.com/mdarolon/BBII-FCEQyN>.

E4.A8- Carga inicial y Actualización

Esta actividad consiste en la población inicial de los Data Marts confeccionados en la etapa anterior.

Para la carga inicial, en primer lugar se definieron una serie de transformaciones, que luego se integraron a un flujo de trabajo llamado *carga_datasets*. Dichas transformaciones exportan los datos necesarios de la base de datos Informix (SIU-Guaraní) y de archivos planos correspondiente al sistema SIU-Kolla, y los importa como metadatos en la base de datos PostgreSQL creada para tal fin.

Como ejemplo, en la Figura 52 se observa una transformación en la cual se realiza la migración desde Informix a PostgreSQL.



Figura 52: Transformación en PDI sobre migración desde Informix a PostgreSQL

Donde, en el primer objeto *dsa_grad_cola_ifx* se realiza la importación de los graduados asociados a una colación dentro del sistema SIU-Guaraní. El detalle del objeto se puede observar en la Figura 53.

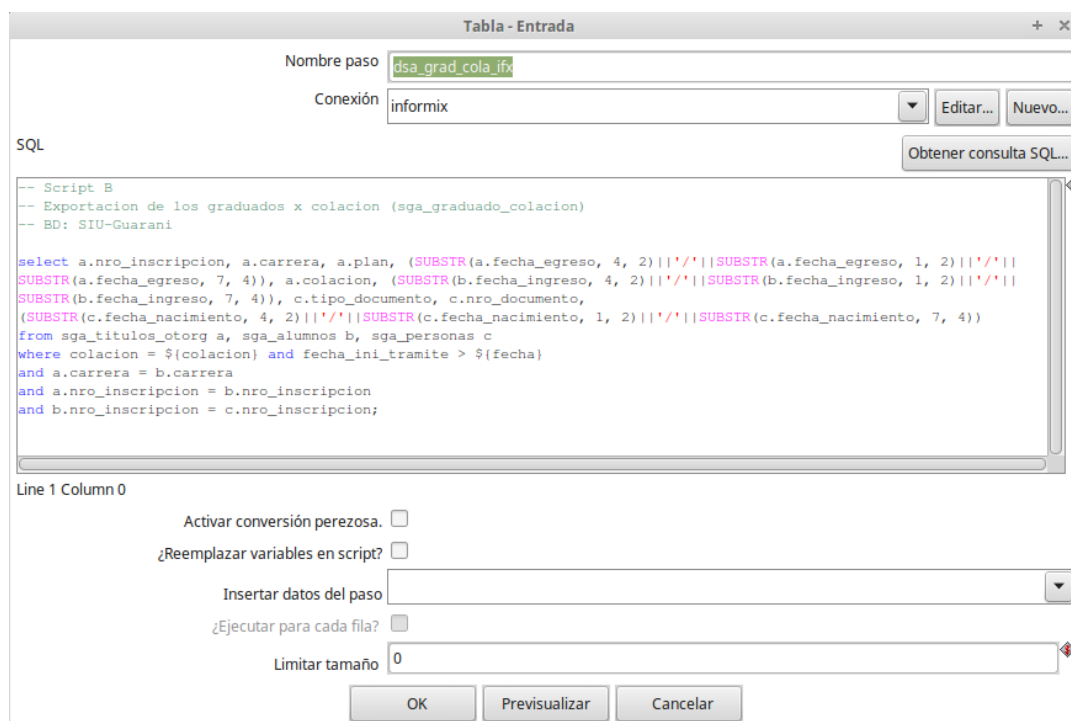


Figura 53: Importación de datos. Conexión a Informix (SIU-Guarani)

Como puede observarse en la sentencia SQL del script, la cláusula colación recibe los parámetros `${colacion}` y `${fecha}`. La utilización de parámetros embebidos en el SQL permite que estas transformaciones sean reutilizables.

Luego en el segundo objeto `dsa_graduado_colacion`, los datos obtenidos de la consulta son importados en la tabla `graduado_colacion` correspondiente a PostgreSQL, de manera que queden disponibles para su posterior procesamiento. La correspondencia entre los atributos obtenidos de la consulta SQL realizada al motor de Informix y los atributos correspondientes a la tabla del motor de PostgreSQL se observan en la Figura 54.

Tabla - Salida

Nombre de paso:

Conexión:

Esquema destino:

Tabla destino:

Tamaño transacción (commit):

Vaciar tabla: ☐

Ignorar errores de inserción: ☐

Specify database fields: ☒

Main options Database fields

Fields to insert:

#	Table field	Stream field
1	nro_inscripcion	nro_inscripcion
2	carrera	carrera
3	plan	plan
4	fecha_egreso	fecha_egreso
5	colacion	colacion
6	fecha_ingreso	fecha_ingreso
7	tipo_documento	tipo_documento
8	nro_documento	nro_documento
9	fecha_nacimient	fecha_nacimiento

Figura 54: Correspondencia entre los atributos de Informix y PostgreSQL

La carga inicial de las tablas correspondientes al DSA se realiza mediante el flujo de trabajo denominado *carga_datasets*. Al iniciarse (instancia STAR) ejecuta las distintas transformaciones que exportan los datos de los sistemas fuentes y los cargan en la base de datos ‘graduados’ de PostgreSQL. Si todo el proceso se ejecuta correctamente, se muestra un mensaje de Éxito y se envía un e-mail indicando que la carga fue exitosa. De lo contrario, si ocurre un error en alguna instancia de carga, primero genera una entrada en el registro log, luego se aborta el trabajo y finalmente se envía un e-mail indicando en qué punto se produjo el error.

En la Figura 55 se observa el proceso.

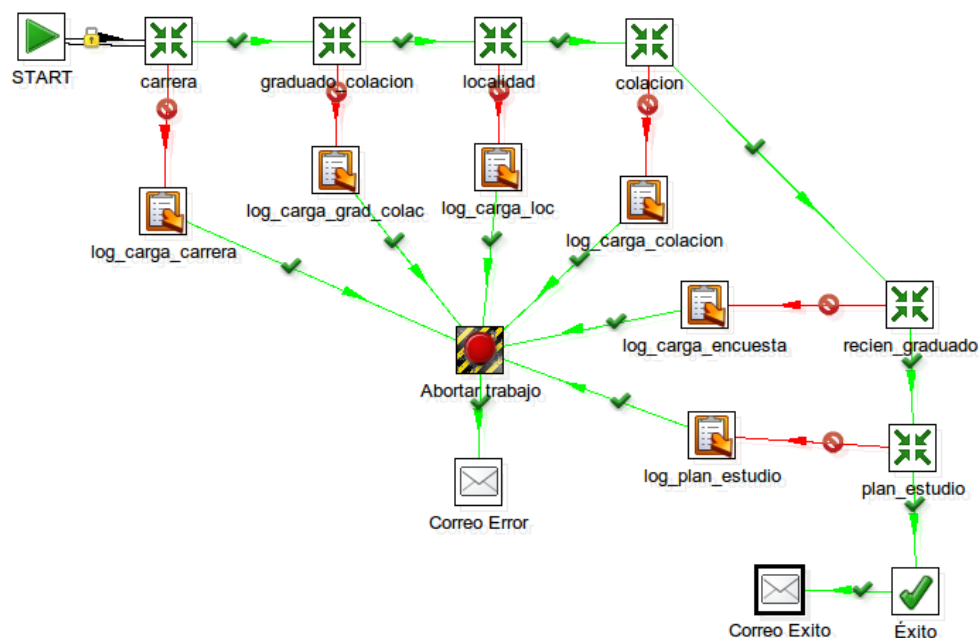


Figura 55: Flujo de trabajo en PDI, carga de datasets

En segunda instancia se definieron una nueva serie de transformaciones, integradas en flujos de trabajos diferentes llamados *job_mapeos* y *job_transformaciones*, las cuales toman los datos cargados en la base de datos ‘graduados’ de PostgreSQL, correspondientes al esquema DSA y realizan mapeos e imputación de valores, cálculos y transformaciones de acuerdo a las reglas establecidas para cada caso.

A continuación se describe una transformación diseñada (Figura 56).

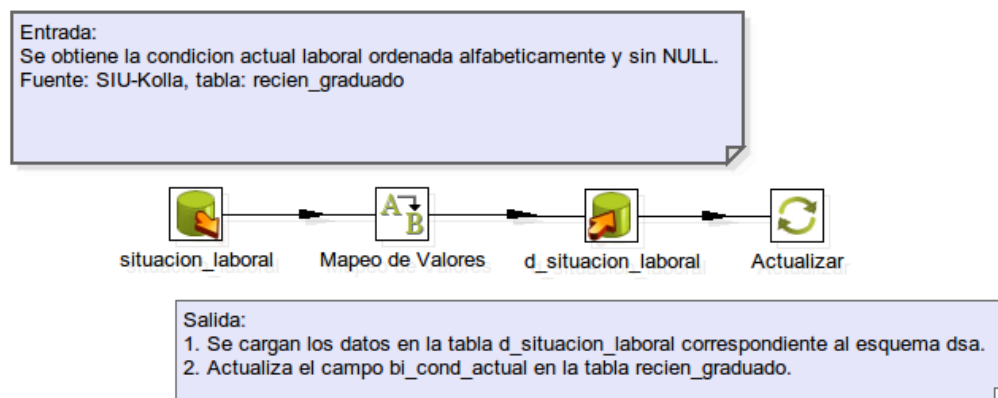


Figura 56: Transformación mapeo de valores con PDI

Donde, en el primer objeto *situación_laboral*, se obtiene un listado de los valores del campo *cond_actual* correspondiente a la tabla *dsa.recién_graduado* de la base de datos ‘graduados’ de PostgreSQL.

En la Figura 57 se observa la configuración del objeto *Mapeo de Valores* donde se evalúan los valores del campo origen *cond_actual* y se los mapea a un campo de destino llamado *aux_cond_actual*. En valores de campo se establecen las reglas de mapeo, donde el valor origen corresponde a los valores de la columna *cond_actual* y el valor destino es el nuevo valor que se obtendrá.

#	Valor origen	Valor destino
1	No Responde	1
2	No trabaja	2
3	Trabaja en relación con la profesión	3
4	Trabaja, pero no tiene relación con la profesión	4

Figura 57: Configuración del objeto mapeo de valores en PDI

El siguiente objeto *d_situacion_laboral* extrae los nuevos valores resultantes del mapeo y los importa en la tabla *dsa.d_situacion_laboral* de la base de datos ‘graduados’ de PostgreSQL, que será utilizada para la carga de la dimensión correspondiente. La configuración de dicho objeto se observa en la Figura 58.

Tabla - Salida

Nombre de paso: d_situacion_laboral

Conexión: dsaGraduados [Editar...] [Nuevo...]

Esquema destino: dsa [Examinar...]

Tabla destino: d_situacion_laboral [Examinar...]

Tamaño transacción (commit): 10

Vaciar tabla: ☒

Ignorar errores de inserción: ☐

Specify database fields: ☒

Main options Database fields

Fields to insert:

#	Table field	Stream field
1	sit_lab_id	aux_cond_actual
2	sit_laboral	cond_actual

[Get fields] [Enter field mapping]

[OK] [Cancelar] [SQL]

Figura 58: Configuración del objeto d_situacion_laboral (Tabla- Salida)

Finalmente, el objeto Actualizar busca en la tabla dsa.recien_graduado filas cuyo valor del campo cond_actual coincida con el valor del campo sit_laboral asignado a la tabla dsa.d_situacion_laboral; para cada coincidencia se actualiza el valor del campo bi_cond_actual con el valor del flujo de campo aux_cond_actual. La configuración del objeto se observa en la Figura 59.

Actualización

Nombre de paso: Actualizar

Conexión: dsaGraduados [Editar...] [Nuevo...]

Esquema destino: dsa [Examinar...]

Tabla destino: recien_graduado [Examinar...]

Tamaño transacción (commit): 100

Use batch updates? ☐

Skip lookup ☐

¿Ignorar fallo de búsqueda? ☐ Campo lógico (clave encontrada) []

La clave(s) para realizar la búsqueda del valor(es):

#	Campo de tabla	Comparador	Flujo de campo1
1	cond_actual	=	cond_actual

[Obtener campos]

Actualizar campos:

#	Campo de tabla	Flujo de campo
1	bi_cond_actual	aux_cond_actual

[Obtener campos actualización]

[OK] [Cancelar] [SQL]

Figura 59: Configuración del objeto Actualizar

Al igual que el flujo de trabajo anterior, los flujos de trabajos *job_mapeos* y *job_transformaciones* presentan el mismo tratamiento y comportamiento al iniciar su ejecución.

Para la carga de dimensiones y tabla de hechos se diseñaron dos flujos de trabajos diferentes, *carga_dimensiones* y *carga_ft* respectivamente, que ejecutan un conjunto de transformaciones las cuales extraen los datos transformados del DSA y los importan en las dimensiones y tablas de hechos correspondientes a los Data Marts.

Los flujos de trabajo como las transformaciones pueden descargarse desde el siguiente enlace: <https://github.com/mdarolon/BBII-FCEQyN>.

En cuanto a las políticas de actualización de los datos, las mismas definen los mecanismos que se utilizarán para insertar nuevos datos a los Data Marts o actualizar los existentes. En el presente caso de estudio las mismas se definieron conjuntamente con el experto y el responsable de la institución, y son las siguientes:

- La información referida a las encuesta se actualizará una vez por mes.
- Los datos de la tabla de dimensión *d_carrera* serán actualizados únicamente cuando exista una nueva oferta académica en la Unidad Académica.

- Los datos de las tablas de dimensiones d_residencia y d_procedencia serán actualizados cada vez que existan actualizaciones en el catálogo de localidades enviado por el SIU o sea necesario agregar una localidad inexistente.
- Los datos de la tabla de dimensión d_colacion serán actualizados una vez por año.
- Los datos de la tabla de dimensión d_tiempo serán actualizados al comienzo de cada año calendario.
- Los datos de las tablas de hechos se actualizarán de manera incremental teniendo en cuenta la fecha de la última actualización.
- Los datos de las restantes tablas de dimensiones permanecerán invariables, siempre que la estructura de la Encuesta al Recién Graduado no se modifique.

E4.A9- **Modelo multidimensional**

La definición del esquema multidimensional (egresados.xml) se realiza con la herramienta *Schema Workbench* de la suite de Pentaho. El mismo contiene los cubos desarrollados en base a los modelos lógicos diseñados en la Etapa 3- **Diseño y Modelado Multidimensional**.

Cada cubo se encuentra definido por la estructura de la tabla de hechos, las medidas, miembros calculados y las dimensiones.

Las dimensiones y sus jerarquías se crearon de manera general dentro del esquema y se utilizaron en cada cubo como dimensiones compartidas. Esto evitó, tener que definir varias veces la misma dimensión, pudiendo reutilizar elementos ya definidos en varios cubos.

Cada cubo contienen medidas (cant, prom_edad, duración_real, duración_teorica) y miembros calculados (desvio_duracion_real) que corresponden a los valores de análisis.

Las medidas se calcularon directamente con atributos de la base de datos y los miembros calculados mediante formulas que utilizan otras medidas. Por ejemplo, para calcular el desvío de la duración real, la fórmula es la siguiente: [Measures].[duracion_real]-[Measures].[duracion_teorica].

En la Figura 60 y Figura 61 se pueden visualizar las estructuras de los cubos definidos. El contenido del archivo XML¹⁸ generado puede ser descargado desde el siguiente enlace: <https://github.com/mdarolon/BBII-FCEQyN>.

¹⁸ XML: Lenguaje de Marcas Extensible (*eXtensibleMarkupLanguage*.)



Figura 60: Modelo multidimensional
Situación laboral



Figura 61: Modelo multidimensional
Valoración académica

5.3.5 E5: Validación del modelo final

✓ E5.A1- Verificación del Data Mart

En esta actividad se realiza la verificación y validación de los Data Marts desarrollados.

La **verificación** consiste en ejecutar y analizar los Data Marts desarrollados, respondiendo a los interrogantes planteados en la Etapa 1 (E1.A4 – **Definición de Temas Analíticos**).

El servidor de Pentaho cuenta con una herramienta para la navegación de los cubos llamada JPivot.

En cuanto a la usabilidad de esta herramienta, uno de los expertos consultados, sugirió incorporar alguna alternativa que sea más amigable e intuitiva para la navegabilidad y visualización de los Cubos debido a que JPivot presenta dureza en el manejo de los mismos, una interface poco amigable y requiere de una curva de aprendizaje mayor por parte de los usuarios.

Si bien JPivot cuenta con opciones para la representación gráficas de la información, la calidad percibida a la hora de mostrar los resultados es muy básica (Figura 62).

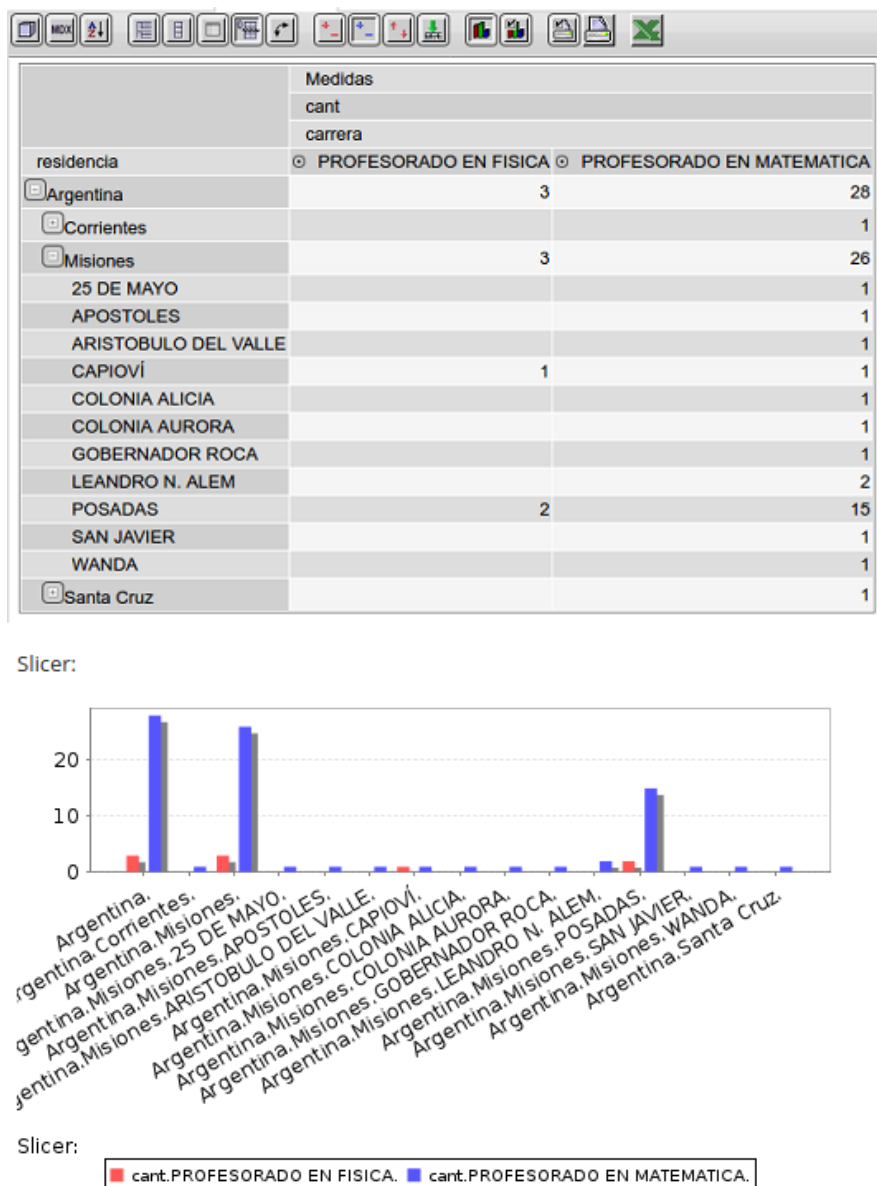


Figura 62: Visualización de la herramienta JPivot

Teniendo en cuenta las sugerencias hechas por el experto, se investigaron y analizaron alternativas que mejoraran la usabilidad de la herramienta y el análisis de los datos, resultando en la implementación del complemento Saiku Analytics de Pentaho.

La interface que brinda este complemento facilitó el manejo de las dimensiones y medidas, sin tener que acceder a la configuración del Data Mart. Por otro lado, la selección de las variables y

medidas a analizar mediante el sistema “drag & drop”¹⁹, permitió realizar el análisis de los datos en forma más fácil e intuitiva.

Debido a la imposibilidad de realizar reuniones presenciales, dentro del marco del Aislamiento Social Preventivo Obligatorio (ASPO) en todo el ámbito de la UNaM (Resolución Rectoral Nro.: 533/20 [57]) y hasta la finalización de la Emergencia Sanitaria y Epidemiológica decretada por el Gobierno Provincial hasta el 31 de Diciembre inclusive, sumado a los problemas de conectividad acontecidos en el domicilio particular de uno de los expertos, se realizó la grabación de una serie de videos tutoriales donde se propuso replicar la navegabilidad de los Data Mart por parte del experto buscando dar respuesta a los interrogantes planteados en los temas analíticos.

Para los demás expertos se puso a disposición una instancia de testeo, que brindó la posibilidad de acceder desde sus domicilios a la navegabilidad de los Data Marts mediante certificado VPN²⁰.

El análisis OLAP realizado sobre los Data Marts buscó responder los interrogantes planteados en cada tema de análisis y numerosas consultas derivadas asociadas a cada tema analítico, las cuales fueron incorporadas al análisis.

✓ E5.A2- Validación del Data Mart

Con respecto a la **validación** realizada por los expertos, se pudo obtener el grado de satisfacción alcanzado por los mismos en relación a los resultados obtenidos por cada consulta analizada.

En el Capítulo 6 de Resultados se muestran los resultados de la verificación y validación realizada por los expertos.

¹⁹ Drag and drop: Arrastrar y soltar con el mouse

²⁰ VPN: Red Privada Virtual (*Virtual Private Network*)

Capítulo 6

Resultados

6. Resultados

En este capítulo se presentan los principales resultados en base a los objetivos y el caso de estudio planteado.

El relevamiento realizado sobre los distintos modelos de desarrollo de Data Warehouses, permitió incorporar las contribuciones de autores como Kimball y Bernabeu (HEFESTO), a la propuesta metodológica desarrollada.

Como fue descripto en las secciones 2.2.8 y 2.3, se tomó como marco de referencia principal los modelos metodológicos mencionados, y se incorporaron técnicas y herramientas para la gestión de requisitos basadas en las buenas prácticas del modelo CMMI v1.3.

El análisis de dichos modelos permitió desarrollar tareas que se ajustaron a las necesidades de la FCEQyN, obteniendo como resultado una metodología que consta de cinco etapas, con actividades distribuidas estratégicamente en cada una de ellas.

En la Figura 63, se observa graficamente la metodología propuesta.

Metodología para el desarrollo de *Datamarts* en Universidades Nacionales del Consorcio SIU – Caso de Estudio: Área Graduados de la FCEQyN

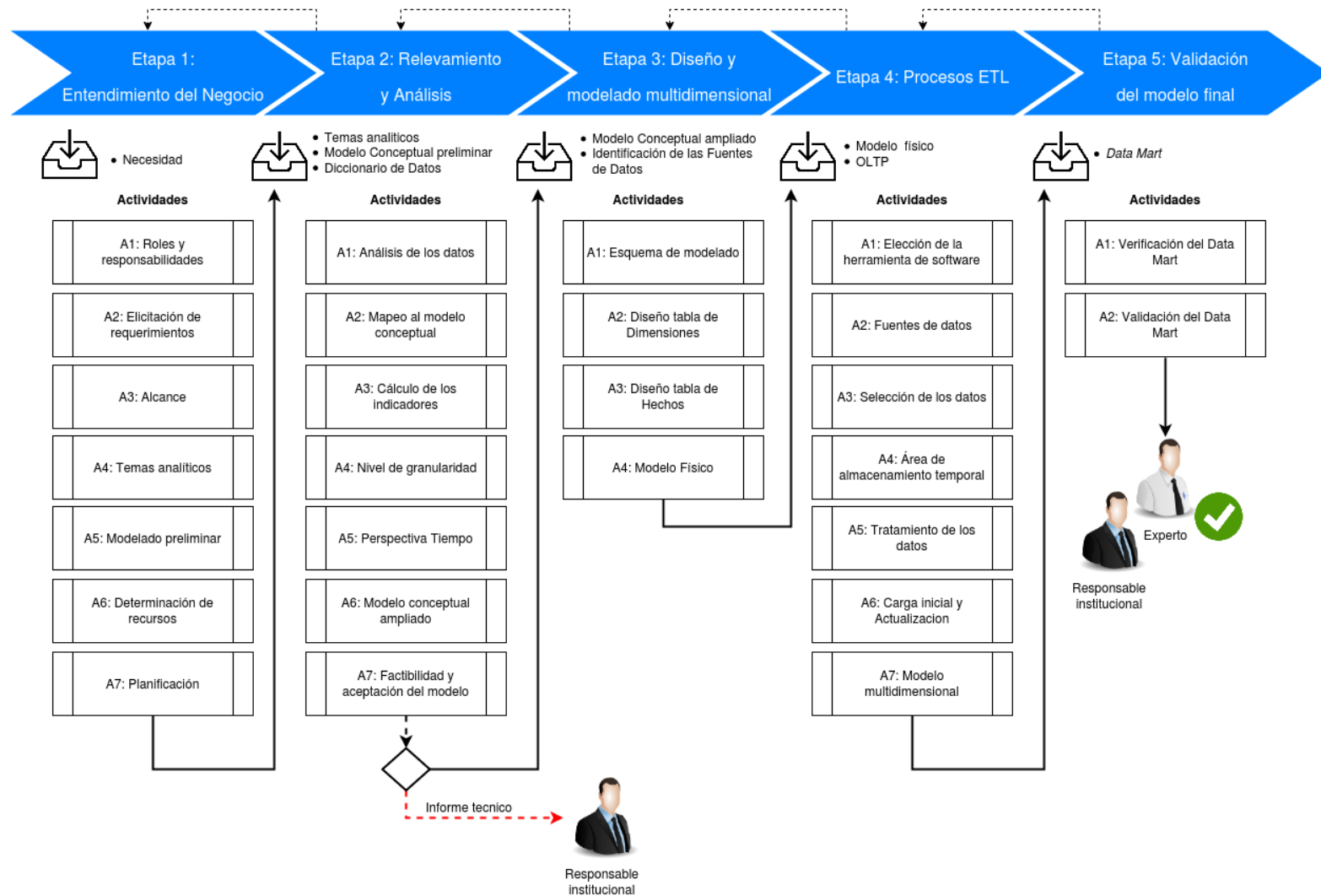


Figura 63: Etapas de la metodología propuesta [Elaboración propia]

La participación activa del experto en el desarrollo de las distintas actividades, permitió tener una mejor aproximación a la información necesaria sobre los temas de análisis planteados.

Las solicitudes de requerimientos e información en la FCEQyN se realizaban de manera no estructurada e informal mediante charlas, llamadas telefónicas o correos electrónicos. Si a posteriori se necesitaba profundizar en ese requerimiento, abarcar uno similar o sociabilizar la experiencia adquirida en el desarrollo de alguno de ellos, se perdía el seguimiento del mismo.

Con la metodología propuesta la FCEQyN cuenta con una estructura de cómo debe solicitarse un requerimiento en base a los estándares de CMMI y de SCRUM. Esta estructura incorporó la solicitud de un requerimiento mediante la generación de tickets en una herramienta de uso institucional (Redmine), acentuando el grado de formalización y compromiso de la solicitud

Aspectos relacionados a la trazabilidad fueron cubiertos mediante la gestión de los tickets generados, los cuales se le ha dado un tratamiento, agregado documentación o completado con tareas que han llevado a la resolución de los mismos. Esto facilitó la tarea tanto del analista como la del experto quién definió el requerimiento, como guía y seguimiento de lo solicitado.

Para comprender el requerimiento planteado se utilizaron diferentes técnicas y herramientas para la gestión de los mismos. Entre ellas, se aplicaron:

- Entrevistas: Se realizaron cuatro entrevistas en las cuales participaron el Vice Decano, el coordinador de la Oficina de Vinculación con los Graduados y el director de la carrera Licenciatura en Sistemas de Información. Las mismas tuvieron una duración promedio de dos horas y se realizaron durante el período de Mayo 2019 a Marzo 2020. Como resultado de las mismas se recabó información relevante relacionada a la gestión que se llevaba a cabo de los graduados en la FCEQyN y sobre las necesidades de información actuales en la institución referida a éstos.

Las mismas fueron plasmadas en plantillas diseñadas (DERU-Documento de Especificación de Requerimientos de Usuario) y personalizadas para facilitar la carga y el relevamiento de las necesidades de información manifestada por los expertos, permitiendo organizar la misma en base a la demanda planteada. Por otro lado, estas plantillas contribuyeron a identificar aquellos temas de análisis de mayor interés e importancia a ser abordados en una primera instancia de desarrollo.

Se definieron un total de tres Temas Analíticos (TA), los cuales fueron plasmados e individualizados en una nueva plantilla que sirvió de guía para identificar de manera concreta y resumida cada uno de ellos, determinar quién lo solicitó, cual es el objetivo de esa solicitud y cuál era el beneficio que se pretendía obtener.

Las entrevistas, cuestionario, documentación relevada y temas analíticos definidos se encuentran disponibles y pueden descargarse del siguiente enlace: <https://github.com/mdarolon/BBII-FCEQyN>

- Reuniones: Se realizaron numerosas reuniones de manera periódicas con la directora del Área de Enseñanza, Secretaría Académica, Vice Decano, coordinador de la Oficina de Vinculación con los Graduados y el director de la carrera Licenciatura en Sistemas de Información. Dichas reuniones tuvieron incidencia y relevancia a la hora de determinar acciones institucionales que permitieron acompañar el desarrollo del presente proyecto. Dichas acciones, resultaron en la definición del equipo de trabajo para la presente investigación, la modificación del circuito administrativo para la realización del trámite de título en el Área de Enseñanza, la implementación del sistema SIU-Kolla y la administración del mismo a nivel institucional.
- Diagramas conceptuales: Se diseñaron cinco modelos conceptuales que permitieron modelar la solución en base a la conceptualización de los requisitos, facilitando obtener una visión general del alcance de cada tema de análisis planteado. Por otro lado, los diagramas conceptuales aportaron al descubrimiento de nuevos conceptos a ser analizados.
- Diccionario de Datos (DD): Se confeccionaron un total de tres DD propios para el caso de estudio. Para la confección de los mismos se diseñó y estructuró una plantilla que asistió en la confección de cada uno de ellos, facilitando la incorporación de conceptos / atributos surgidos del análisis, junto a su significado y contenido. La implementación de esta plantilla permitió la reutilización de conceptos ya definidos para otros temas analíticos, agilizando los plazos de elaboración del siguiente DD.

Los DD elaborados se encuentran disponible en el siguiente enlace: <https://github.com/mdarolon/BBII-FCEQyN>

Por otro lado, para el desarrollo y documentación de cada fase de la metodología propuesta, se diseñaron nuevas guías que sirvieron como base para cada rol interviniente, facilitando la comprensión y síntesis del requerimiento.

Se generaron un total de siete plantillas durante toda la gestión del proyecto, utilizadas para la registración y gestión de los requisitos. Las mismas fueron modificándose en base a su utilización y funcionalidad.

Estas plantillas se especifican y describen en la Tabla 34:

Tabla 34: Plantillas diseñadas para la gestión de requisitos

Plantilla	Función	Etapas / Actividad
Documento de Especificación de Requerimientos de Usuario (DERU)	Plantilla diseñada para reconocer y entender las necesidades del experto/usuario	E1: Entendimiento del Negocio A2- Elicitación de Requerimientos
Tema Analítico (TA)	Plantilla diseñada para definir un tema analítico.	E1: Entendimiento del Negocio E1.A4- Definición de Temas analíticos
Diccionario de Datos (DD)	Plantilla diseñada para describir las características lógicas de los conceptos, atributos, significado, contenido y posible origen de los datos (sistema fuente).	E1: Entendimiento del Negocio E1.A5- Modelado preliminar
Identificación de las Fuentes de Datos (FUDA)	Plantilla diseñada para contener e identificar las distintas fuentes de datos.	E2: Relevamiento y Análisis E2. A2 - Mapeo al modelo conceptual
Conformación de Indicadores (CI)	Plantilla diseñada para detallar la conformación de indicadores.	E2: Relevamiento y Análisis E2.A3 - Cálculo de los indicadores
Definición de Tablas de Dimensiones (DTD)	Plantilla diseñada para describir los campos de una tabla de dimensiones y su correspondencia con la tabla y atributo de origen	E3: Diseño y modelado multidimensional E3.A2- Diseño tabla de dimensiones
Documento de Definición de Tabla de Hechos (DTH)	Plantilla diseñada para contener los campos de una tabla de hechos.	E3: Diseño y modelado multidimensional E3.A3- Diseño tabla de hechos

Como fuente de datos principal se utilizó el sistema SIU-Kolla implementado en la FCEQyN (versión 3.7.1). Con los cambios realizados en el circuito administrativo, incorporando como requisito obligatorio para el trámite de título la presentación del comprobante de completitud de la encuesta, este sistema comenzó a utilizarse de manera eficiente.

A medida que se avanzó con el relevamiento y análisis de los datos recabados sobre los graduados, se determinó que la calidad de los mismos no era apta para realizar Inteligencia de Negocio, principalmente por aquellos atributos configurados como campos editables. Se detectaron un total de nueve atributos, los cuales tuvieron que ser extraídos del sistema SIU-Guaraní (versión 3.9.3) para la determinación de dimensiones y conformación de indicadores.

En la Tabla 35 se describen cada uno de ellos.

Tabla 35: Atributos editables del SIU-Kolla

SIU-Kolla	SIU-Guaraní	
Atributo	Atributo	Tabla
Fecha de Nacimiento (Campo editable)	fecha_nacimiento	sga_personas
Año de Ingreso (Campo editable)	fecha_ingreso	sga_alumnos
Título (Campo editable)	Carrera	sga_titulos_otorg sga_carreras sga_colaciones
Ciudad (Campo editable) (*)	Nombre	mug_localidades mug_dpto_partidos
Provincia (Campo editable) (*)	Nombre	mug_provincias
País (Campo editable) (*)	Nombre	mug_paises
Ciudad (Campo editable) (**)	Nombre	mug_localidades mug_dpto_partidos
Provincia (Campo editable) (**)	Nombre	mug_provincias
País (Campo editable) (**)	Nombre	mug_paises

(*) Bloque Residencia – (**) Bloque Procedencia

Para la evaluación de la metodología propuesta se relevaron y analizaron un total de tres temas analíticos como se describe a continuación:

- TA01- Situación laboral de los graduados por carrera según su distribución geográfica.
- TA02- Valoración académica de los graduados por carrera.

- TA03- Necesidad de capacitación del graduado según su distribución geográfica.

En relación a los temas analíticos planteados anteriormente, se desarrollaron e implementaron un total de dos Data Marts:

- TA01: Situación laboral.
- TA02: Valoración académica.

Con referencia al TA03, se detectó en una etapa temprana de la metodología (E2.A7- **Factibilidad y aceptación del modelo**), que no se contaba con los datos necesarios para cubrir este requerimiento recomendando al experto generar una nueva encuesta o personalizar una existente para cubrir dicha demanda.

A partir del diseño e implementación de los distintos Data Marts, se generó un repositorio de veintidós dimensiones, pudiendo las mismas ser incorporadas en futuros temas analíticos. Estas dimensiones se encuentran detalladas en la Tabla 36:

Tabla 36: Repositorio de dimensiones

Dimensión	Descripción
d_carrera	Carrera de la cual se graduó el alumno.
d_residencia	Localidad de residencia del graduado.
d_procedencia	Localidad de procedencia del graduado.
d_satisf_trabajo	Nivel de satisfacción laboral del graduado.
d_situacion_laboral	Situación laboral del graduado.
d_cant_hs_sem	Cantidad de horas semanales que trabaja el graduado.
d_entidad_trabaja	Tipo de entidad donde trabaja el graduado.
d_tiempo	Fecha de completitud de la encuesta.
d_sexo	Sexo del graduado.
d_estado_civil	Estado civil del graduado.
d_grupo_conv	Grupo de convivencia del graduado.
d_redad	Edad del graduado.
d_nivel_exigencia	Nivel de exigencia de la carrera.
d_selec_carrera	Motivo por el cual seleccionó la carrera universitaria.
d_expec_cont	Expectativas del graduado por el contenido estudiado.
d_interes_cont	Interés del graduado por el contenido estudiado.

d_plan_corto	Duración del plan de estudios más corto.
d_valor_univ	Motivo por el cual eligió estudiar en la UNaM.
d_plan_tecnico	Plan de estudio más técnico.
d_plan_gen	Plan de estudio mas generalistas
d_valor_univ_socie	Reconocimiento de la sociedad a la UNaM.
d_aporte_univ	Aportes brindados por la Universidad

Todas estas dimensiones se encuentran disponibles en el siguiente enlace: <https://github.com/mdarolon/BBII-FCEQyN>

A modo de verificación de los TA desarrollados, se instaló e implementó un servidor de Pentaho en el Data Center de la FCEQyN.

Debido al ASPO, los expertos no estaban concurrendo a sus lugares de trabajo; por lo que fue necesario crear e instalar certificados VPN²¹ en las computadoras personales de los mismos para poder acceder al servidor de Pentaho desde sus domicilios particulares. Por otro lado, y debido a problemas de conectividad surgidos en varias oportunidades, se confeccionaron una serie videos que replicaban la navegabilidad de los Data Marts y que aportaron al proceso de verificación y validación de los mismos.

Se crearon un total de once videos para que puedan ser ejecutados por los expertos. El tamaño promedio de cada uno de ellos no superó los 6 MB y fueron puestos a disposición mediante la aplicación de YouTube.

En cuanto al nivel de granularidad aplicado, permitió satisfacer una mayor cantidad de consultas derivadas que surgieron durante la navegabilidad de los Data Marts. Con un total de nueve, estas consultas formaron parte del análisis y la validación de los resultados.

A continuación se listan las consultas derivadas surgidas de cada tema analítico planteado (Tabla 37):

Tabla 37: Consultas derivadas surgidas de cada tema analítico planteado

Data Mart	Tema analítico (TA)	Consulta derivada (CD)
Situación laboral	TA1: Situación laboral de los graduados por carrera según su distribución geográfica.	CD-1: Cantidad de graduados de una carrera que trabajan en relación a su profesión por tipo de entidad donde trabajan y lugar de residencia.

²¹ VPN: Red Privada Virtual (*Virtual Private Network*)

		CD-2: Cantidad de graduados por carrera con mayor inserción laboral en el mercado local.
		CD-3: Grado de satisfacción de los graduados que se encuentran trabajando por carrera.
		CD-4: Cantidad de graduados por carrera que cumplen una mayor carga horaria laboral.
Valoración académica	TA2: Valoración académica de los graduados por carrera.	CD-1: ¿Qué motivó al graduado a la elección de una carrera?
		CD-2: ¿Cual es la duración real de egreso de una carrera?
		CD-3: Graduados de que carrera termina sus estudios según la duración teórica del plan de estudios.
		CD-4: ¿Cómo debería ser el plan de estudio de una carrera según su duración y contenido?
		CD-5: ¿Qué aportes esperan los graduados de una carrera le brinde la universidad?

Para la navegación de los cubos se utilizó la herramienta Saiku Analytics de Pentaho, como fue sugerido por uno de los expertos en la Etapa 5 -**Validación del modelo final**. La misma facilitó la manipulación de las dimensiones y medidas, permitiendo realizar el análisis de los datos en forma más fácil e intuitiva.

A continuación se presenta el análisis realizado sobre los cubos desarrollados y el resultado obtenido por parte de los expertos.

Tema Analítico 1: Situación laboral de los graduados por carrera según su distribución geográfica.

Para realizar la consulta, se utilizaron las dimensiones disponibles: *carrera*, *sit_lab* y *residencia*, junto a la medida *cant*.

Se comenzó seleccionando la dimensión *residencia* y la medida *cant*, para conocer la distribución geográfica de los graduados en la Argentina.

Para obtener un mayor nivel de detalle, se utilizó la función (Include Level) para incorporar la jerarquía *provincia* y *ciudad* al análisis, y se realizó un desgranamiento por carrera. Para una mejor visualización de los datos se limitó el resultado para obtener las diez carreras con mayor número de graduados residentes en la ciudad de Posadas provincia de Misiones.

En la Figura 64, se observa el resultado de esta consulta.

carreara	cant
INGENIERIA QUIMICA	42
FARMACIA	37
LICENCIATURA EN GENETICA	35
PROFESORADO EN MATEMATICA	26
PROFESORADO EN BIOLOGIA	23
BIOQUIMICA	13
INGENIERIA EN ALIMENTOS	12
DOCTORADO EN CIENCIAS APLICADAS	11
PROFESORADO UNIVERSITARIO EN COMPUTACION	8
ANALISTA EN SISTEMAS DE COMPUTACION	7

Figura 64: TA01 - Las diez carreras con mayor número de graduados que residen en Misiones.

En la Figura 65 se pueden observar estos datos representados gráficamente, donde además se obtiene a simple vista un ranking general de todas las carreras.

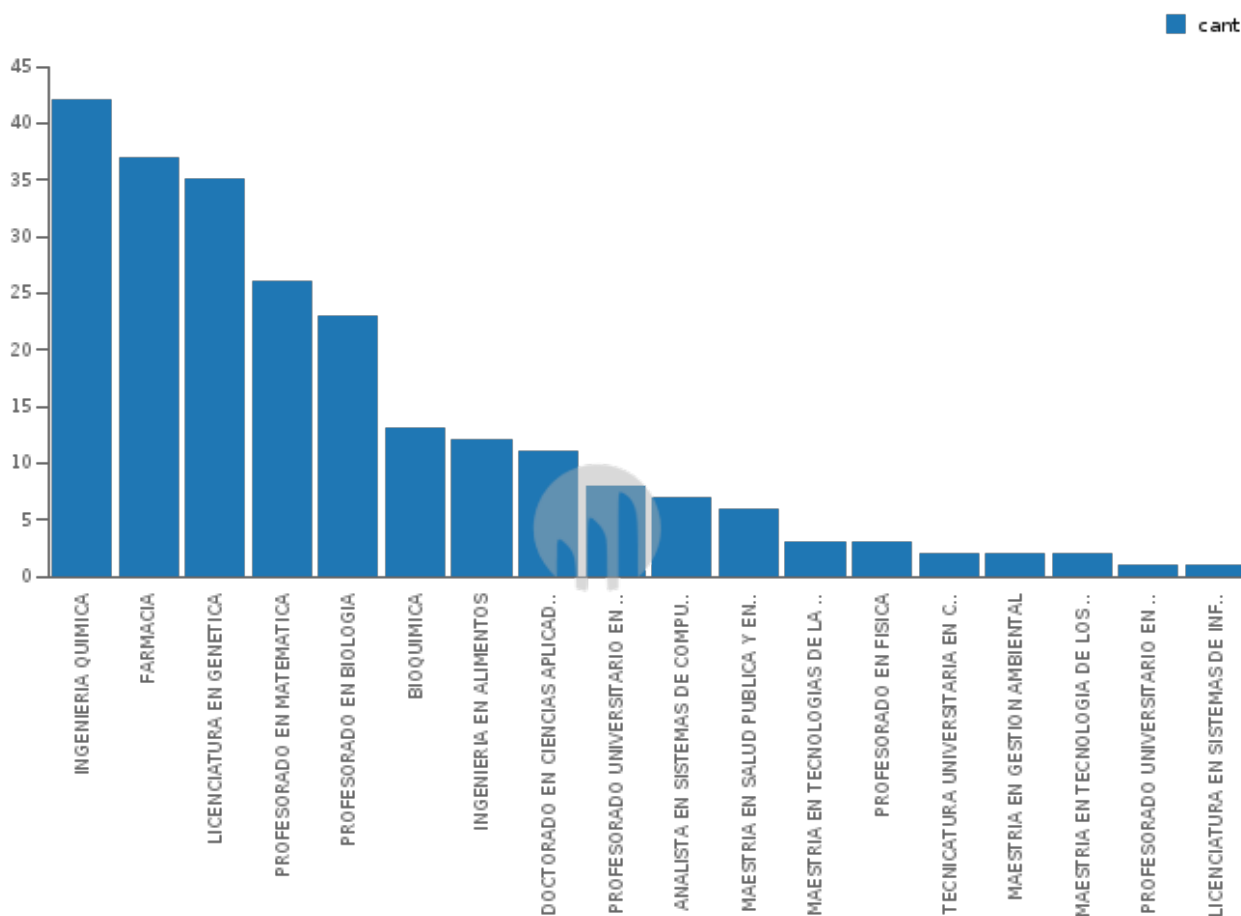


Figura 65: TA01 - Ranking general de carreras cuyos graduados residen en Posadas, Misiones.

De la carrera con mayor número de graduados (Ingeniería Química), se analizó la distribución de los mismos de acuerdo a su situación laboral. Para ello se agregó la dimensión *sit_lab* al análisis.

Como se observa en la Figura 66, el mayor porcentaje de los graduados (63,4%) manifestó **no** estar trabajando al momento de graduarse. De los que si se encontraban trabajando, un 31,7% manifestó estar trabajando en relación a su profesión y un 4,9% manifestó estar trabajando sin relación a su profesión.

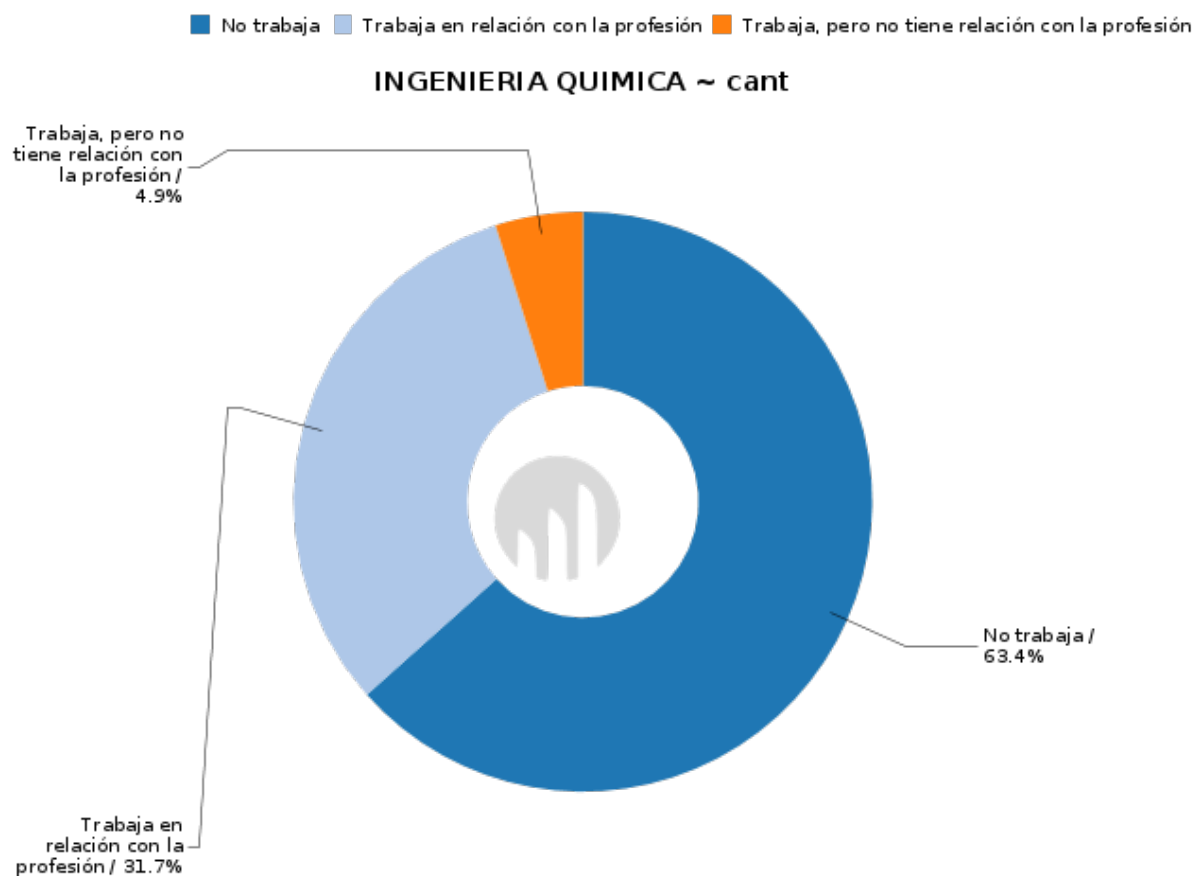


Figura 66: TA01 - Situación laboral de los graduados de la carrera Ingeniería Química.

Del porcentaje de graduados que no se encontraban trabajando, se lo aisló para realizar un análisis más detallado, incorporando las dimensiones *tiempo* y *sexo* junto a la medida *promedio_edad*, como características relevantes para descubrir posibles candidatos a cubrir la demanda laboral del mercado (Figura 67).

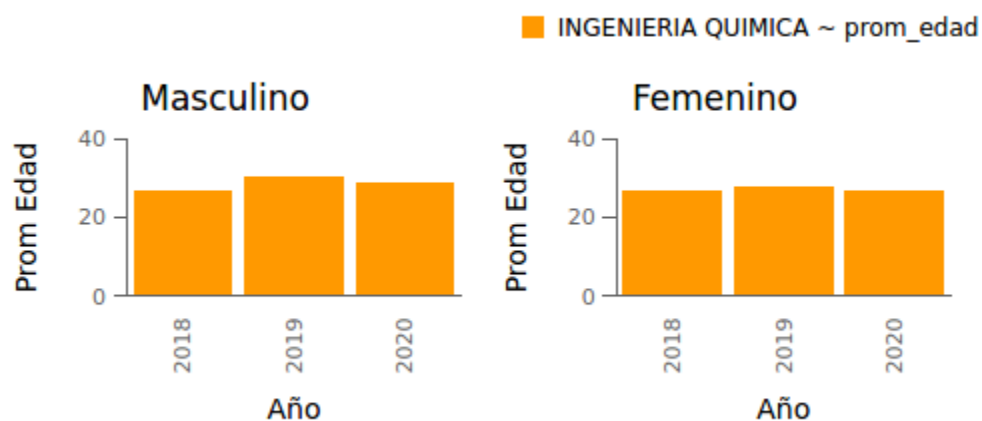


Figura 67: TA01 - Graduados de la carrera Ingeniería Química que no trabajan discriminado por año, sexo y promedio de edad

Como resultado de la consulta se obtuvo que el promedio de edad entre los graduados de género Masculino oscila entre los 28 y 29 años, mientras que los de género Femenino es de 27 años en los distintos años informados.

Otro aporte enriquecedor marcado por los expertos, es que también es posible obtener la distribución geográfica que tienen los graduados en el extranjero. En la Figura 68 se puede observar que Paraguay (50 %) es el segundo país donde reside la mayor cantidad de graduados de la FCEQyN.

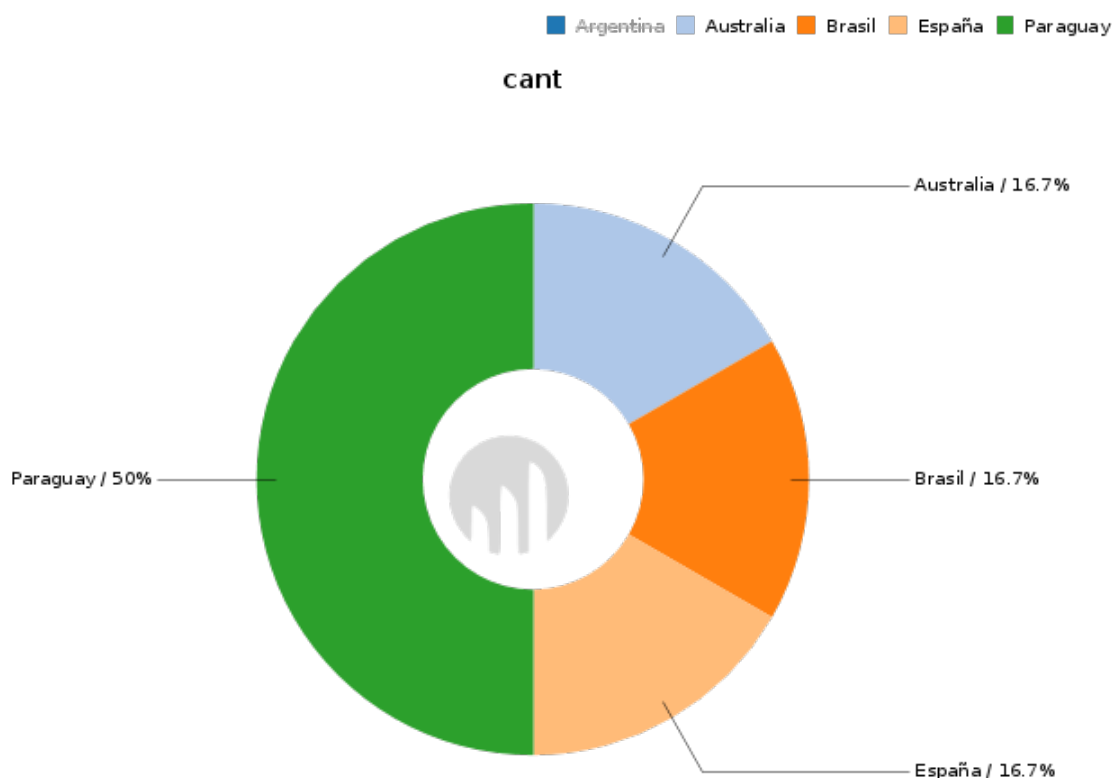


Figura 68: TA01 - Porcentaje de participación de los países extranjeros donde residen los graduados FCEQyN

Donde el porcentaje de participación de los graduados hace referencia a todas las carreras de la FCEQyN en un período de tiempo abarcado entre 2018 y 2020.

Tema Analítico 1: Situación laboral de los graduados por carrera según su distribución geográfica.

Consulta derivada 1: Cantidad de graduados de una carrera que trabajan en relación a su profesión, tipo de entidad donde trabajan y lugar de residencia.

Para realizar la consulta se utilizaron las dimensiones *carrera*, *situación laboral*, *residencia* y *tipo entidad*. Los atributos país, provincia, ciudad y carrera se agregaron como filas para tener una mejor visualización de los datos.

De la dimensión *situación laboral* se tildó la opción ‘Trabaja con relación a la profesión’ y de la dimensión *carrera* se tildó la opción ‘Farmacia’, por ser la carrera con mayor cantidad de encuestas respondidas y obtener una mejor visualización de los datos.

En la Figura 69 se puede observar el cuadro de resultado obtenido en base a la consulta realizada para la carrera de Farmacia.

				sit_lab	Trabaja en relación con la profesión		
				tipo_entidad	Pública	No Responde	Privada
pais	provincia	ciudad	carreara		cant	cant	cant
Argentina	Misiones	POSADAS	FARMACIA		6	-	7
		GARUPA	FARMACIA		1	-	-
		OBERA	FARMACIA		-	1	2
		LEANDRO N. ALEM	FARMACIA		-	-	1
		ELDORADO	FARMACIA		-	-	1
		GOBERNADOR ROCA	FARMACIA		-	-	1
		DOS DE MAYO	FARMACIA		-	-	1
	Corrientes	ITUZAINGO	FARMACIA		-	-	1
		CORRIENTES	FARMACIA		-	-	1
	Entre Rios	CRESPO	FARMACIA		-	-	1

Figura 69: TA01-Graduados de la carrera de Farmacia que trabajan en relación a su profesión

En la consulta se puede ver una comparativa entre cantidad de graduados que trabajan de su profesión, insertos en ámbitos públicos o privados en la Argentina. Al analizar los resultados se observó una participación activa del sector privado en adquirir profesionales formados en la FCEQyN.

Tema Analítico 1: Situación laboral de los graduados por carrera según su distribución geográfica.

Consulta derivada 2: Cantidad de graduados por carrera con mayor inserción laboral en el mercado local.

Para realizar la consulta se utilizaron las dimensiones *carrera*, *situación laboral* y *residencia*. De la dimensión *situación laboral* se seleccionó las opciones ‘Trabaja con relación a la profesión’ y ‘Trabaja, pero no tiene relación con la carrera’.

Con respecto al mercado local el experto seleccionó la provincia de Misiones para conocer las diez carreras con mayor inserción laboral dentro de la misma, en un período de tiempo abarcado entre 2018 y 2020. En la Figura 70 se puede observar el resultado obtenido.

carreara	cant
FARMACIA	24
PROFESORADO EN MATEMATICA	18
INGENIERIA QUIMICA	15
PROFESORADO EN BIOLOGIA	15
LICENCIATURA EN GENETICA	15
DOCTORADO EN CIENCIAS APLICADAS	11
MAESTRIA EN SALUD PUBLICA Y ENFERMEDADES TRANSMISIBLES	6
BIOQUIMICA	5
INGENIERIA EN ALIMENTOS	5
PROFESORADO UNIVERSITARIO EN COMPUTACION	5

Figura 70: TA01-Carreras con mayor inserción laboral en Misiones

Gráficamente la consulta anterior puede representarse de la siguiente manera (Figura 71):

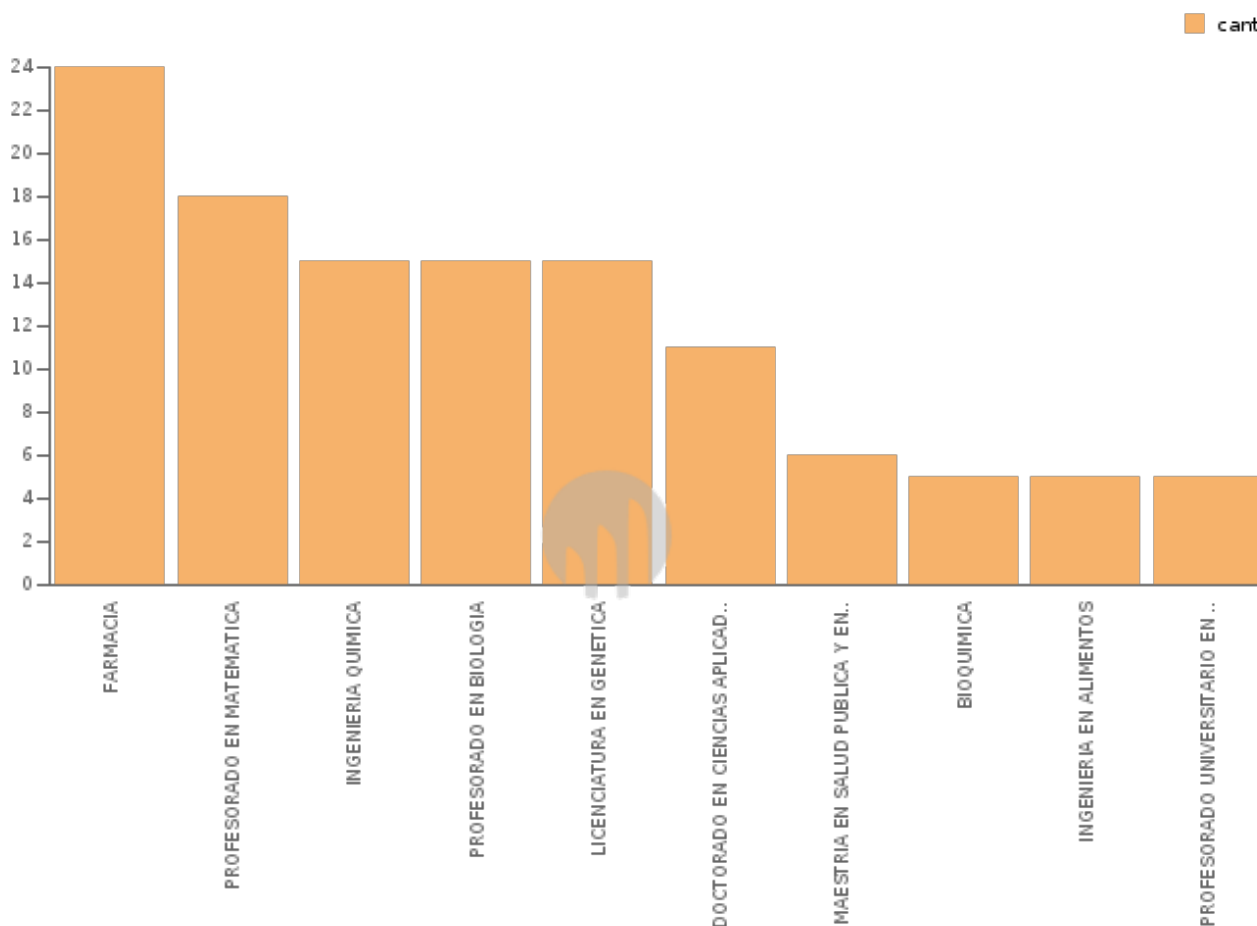


Figura 71: TA01-Gráficamente las diez carreras con mayor inserción laboral en Misiones

Como resultado de la consulta el experto pudo observar que un amplio número de profesiones quedan insertos laboralmente en el mercado local, ejerciendo su profesión.

De las dos primeras carreras con mayor número de graduados insertos en el mercado local, se realizó un desgranamiento por año para conocer la distribución de cada una por cohorte informada. Para ello se agregó la dimensión *tiempo* al análisis, pudiendo observar los resultados gráficamente en la Figura 72.

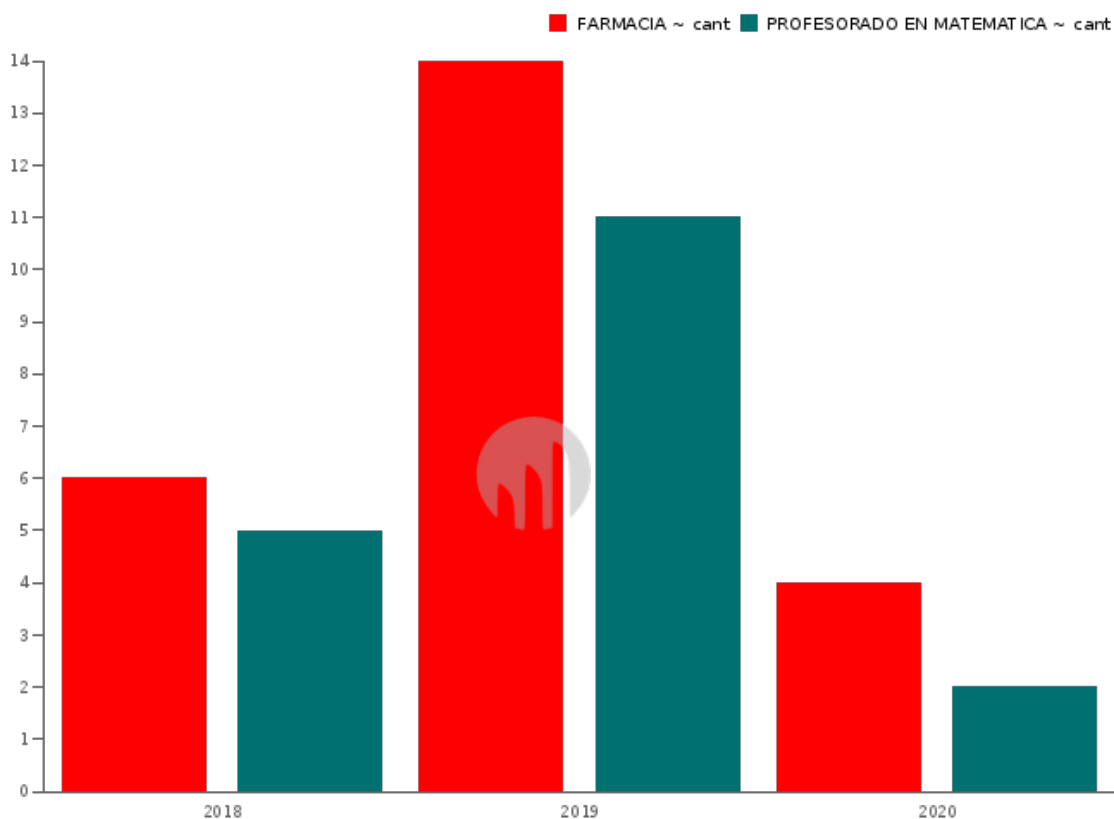


Figura 72: TA01-Cantidad de graduados de Profesorado en Matemática y Farmacia insertos en el mercado local

Como resultado de la consulta se puede realizar una comparación entre ambas carreras con mayor inserción laboral en el mercado local, concluyendo que en el año 2019 se produjo la mayor inserción laboral de los graduados de la FCEQyN.

Tema Analítico 1: Situación laboral de los graduados por carrera según su distribución geográfica.

Consulta derivada 3: Grado de satisfacción de los graduados que se encuentran trabajando por carrera.

Para realizar la consulta se utilizaron las dimensiones *carrera*, *situación laboral* y *satisfacción laboral*. El atributo de la dimensión *situación laboral* se lo agregó en filtro para tener una mejor visualización de los datos.

Para ejemplificar dicha consulta se seleccionaron las carreras de Ingeniería Química e Ingeniería en Alimentos, y se realizó una comparativa entre ambas en un período de tiempo abarcado entre 2018 y 2020. En la Figura 73 se puede observar el resultado obtenido.

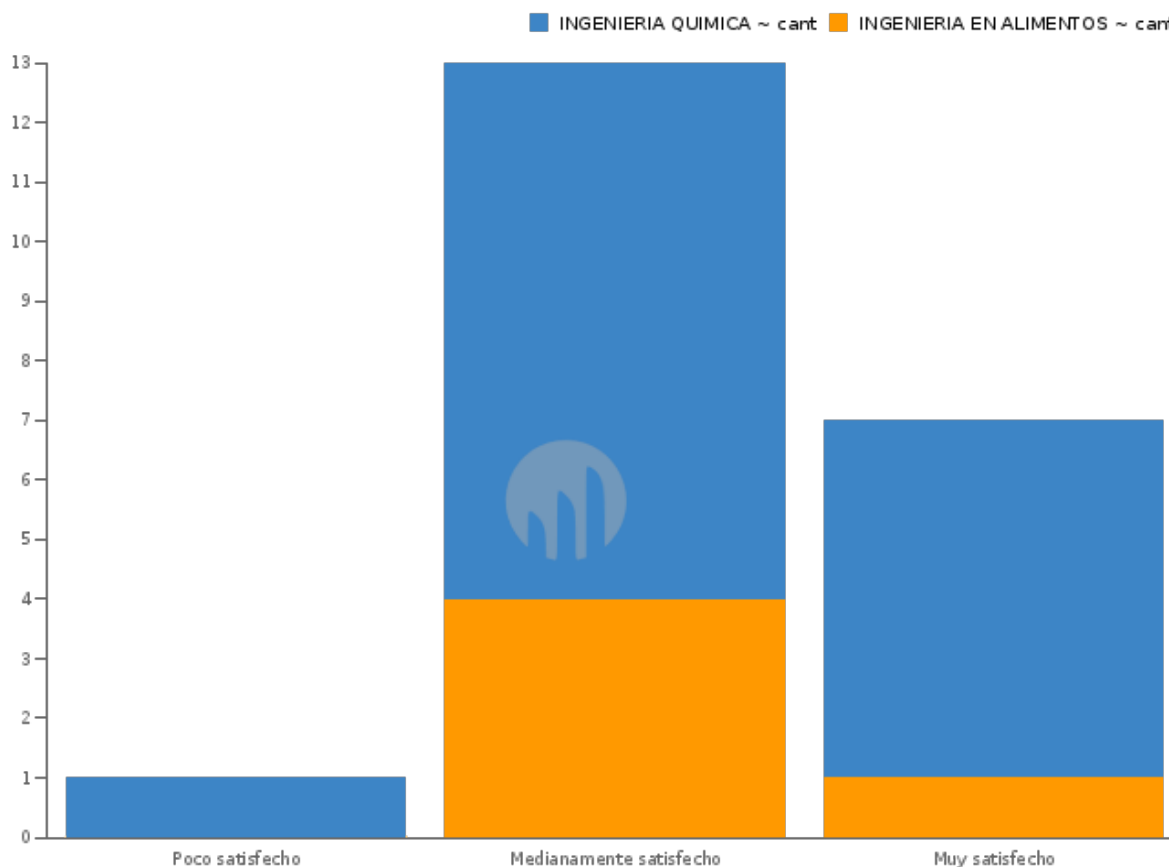


Figura 73: TA01-Grado de satisfacción laboral de las carreras de Ingeniería Química e Ingeniería en Alimentos

Como resultado de la consulta se pudo observar la existencia de un grado de satisfacción laboral no satisfecha en su totalidad por los graduados. Queda de manifiesto la necesidad de trabajar en conjunto con los graduados para conocer los motivos de las opiniones brindadas, y mejorar su inserción en los mercados de trabajo.

Tema Analítico 1: Situación laboral de los graduados por carrera según su distribución geográfica.

Consulta derivada 4: Cantidad de graduados por carrera que poseen una mayor carga horaria laboral.

Para esta consulta se seleccionaron las dimensiones *carrera* y *cantidad de horas semanales* que trabaja un graduado.

Como se observa en la Figura 74, la mayor carga horaria semanal registrada por los graduados es de “Mas de 31 y hasta 45” horas.

cant_hs_sem	cant
Mas de 31 y hasta 45	60
No Responde	56
Mas de 16 y hasta 30	40
Mas de 45	30
Hasta 15	26

Figura 74: TA01-Registro sobre la carga horaria laboral informada por los graduados

En base a la consulta anterior, se selecciono la opción “Mas de 31 y hasta 45” (*Keep only*) y se realizó un desgranamiento por carrera. Se aplicó un ranking descendente para evaluar qué carreras se encontraban involucradas en el análisis.

En la Figura 75 se puede observar el resultado de la consulta.

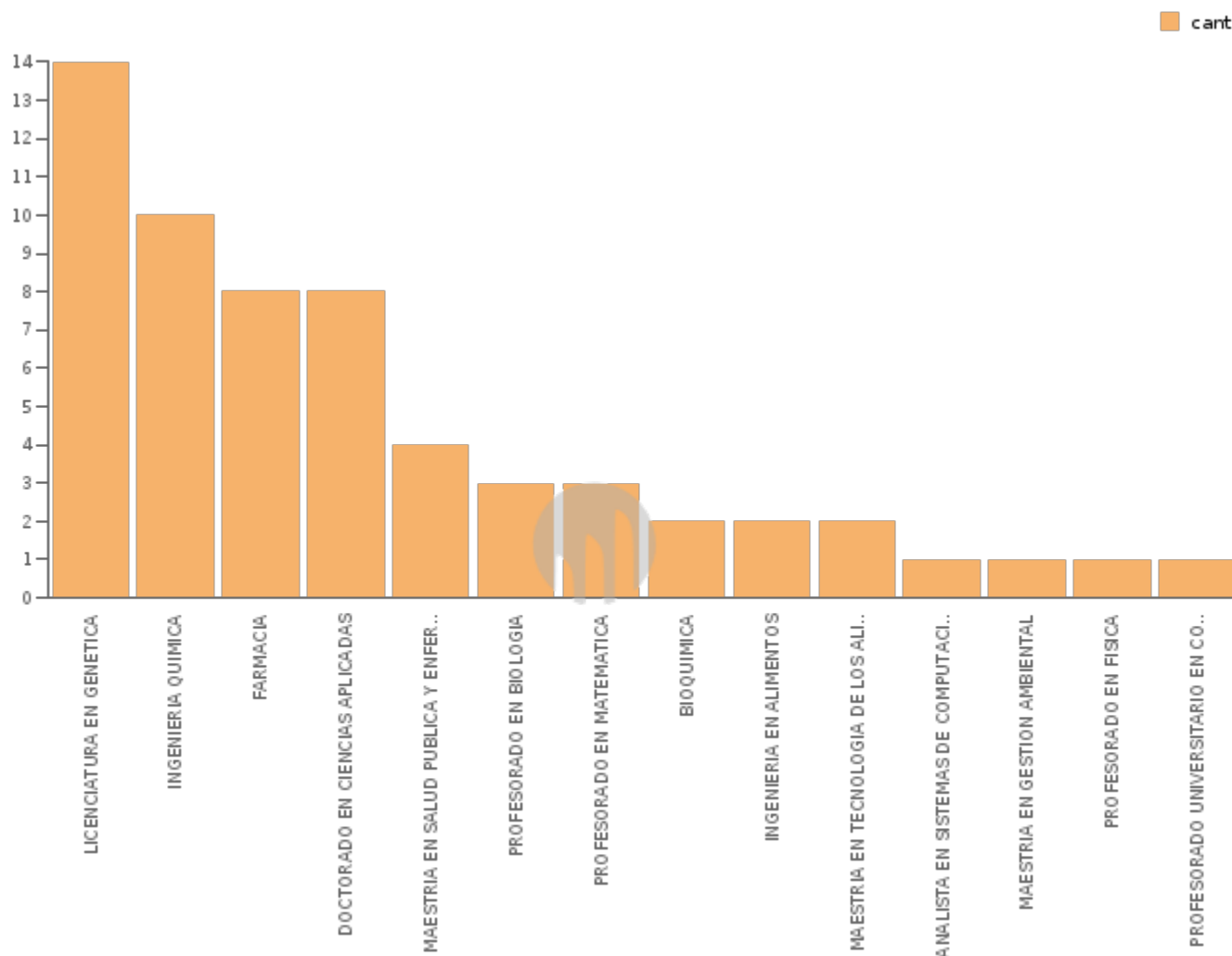


Figura 75: TA01-Carreras que registraron una carga horaria laboral semanal de “Mas de 31 y hasta 45”

Como resultado de la consulta el experto pudo observar las carreras que presentan la mayor carga horaria laboral semanal, manifestada por los graduados. Esta información relevante puede tener numerosas implicancias, las cuales deben ser consideradas y analizadas de acuerdo al contexto en que se utilice dicha información.

Tema Analítico 2: Valoración académica de los graduados por carrera.

Para realizar la consulta el experto utilizó las dimensiones *valoración de la universidad y carrera*.

Como resultado de la consulta se observa que la valoración académica aplicada por los graduados de la FCEQyN refiere al ‘Nivel académico’ que posee la universidad. Esta valoración registra el mayor porcentaje de participación (53%) como puede apreciarse en la Figura 76.

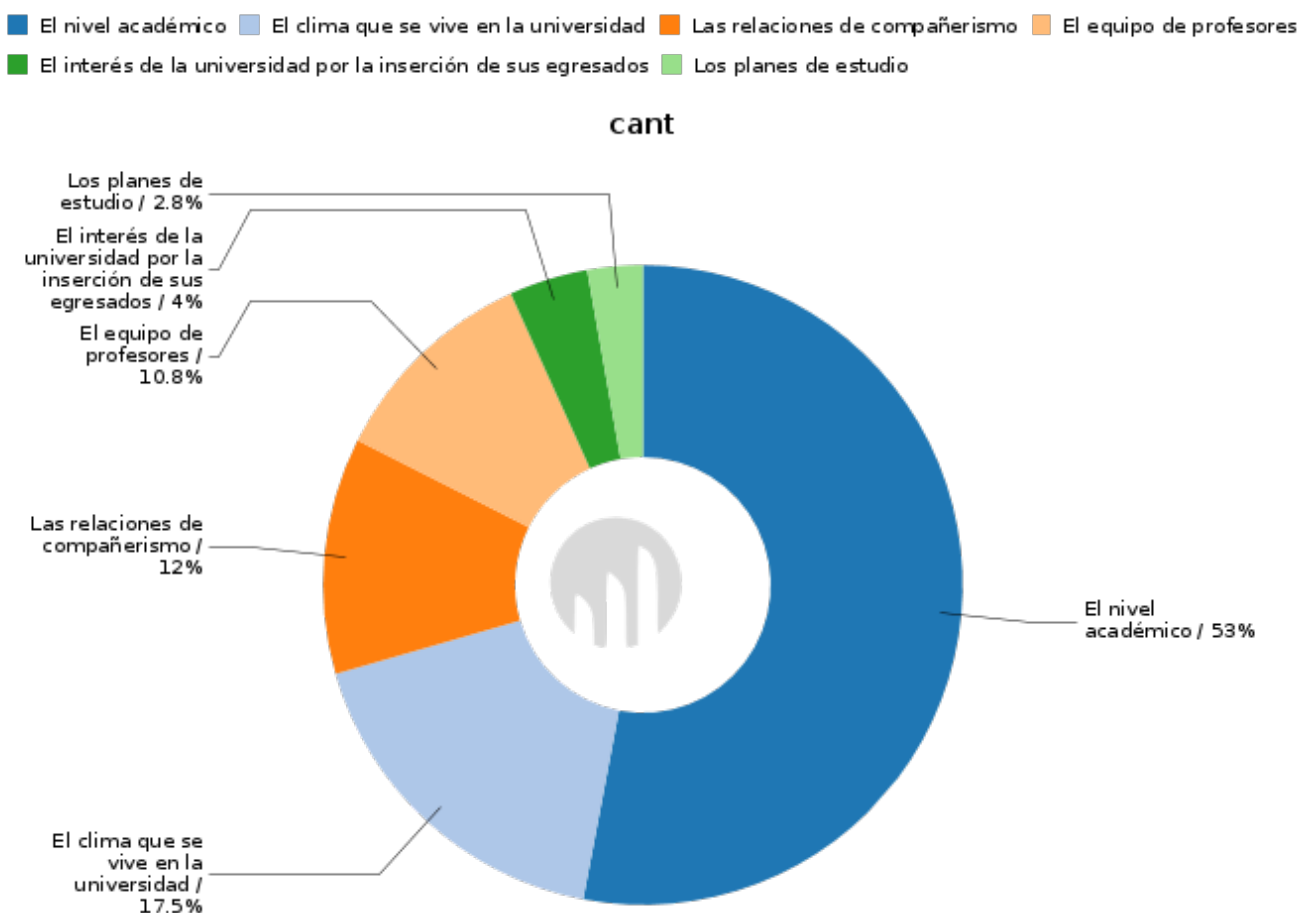


Figura 76: TA02 -Valoración académica de los graduados de la FCEQyN

Luego, se realizó un desgranamiento por año y un ranking descendente, para conocer la distribución de cada valoración registrada por cohorte. Para ello se agregó la dimensión *tiempo* al análisis, pudiendo observar los resultados gráficamente en la Figura 77.

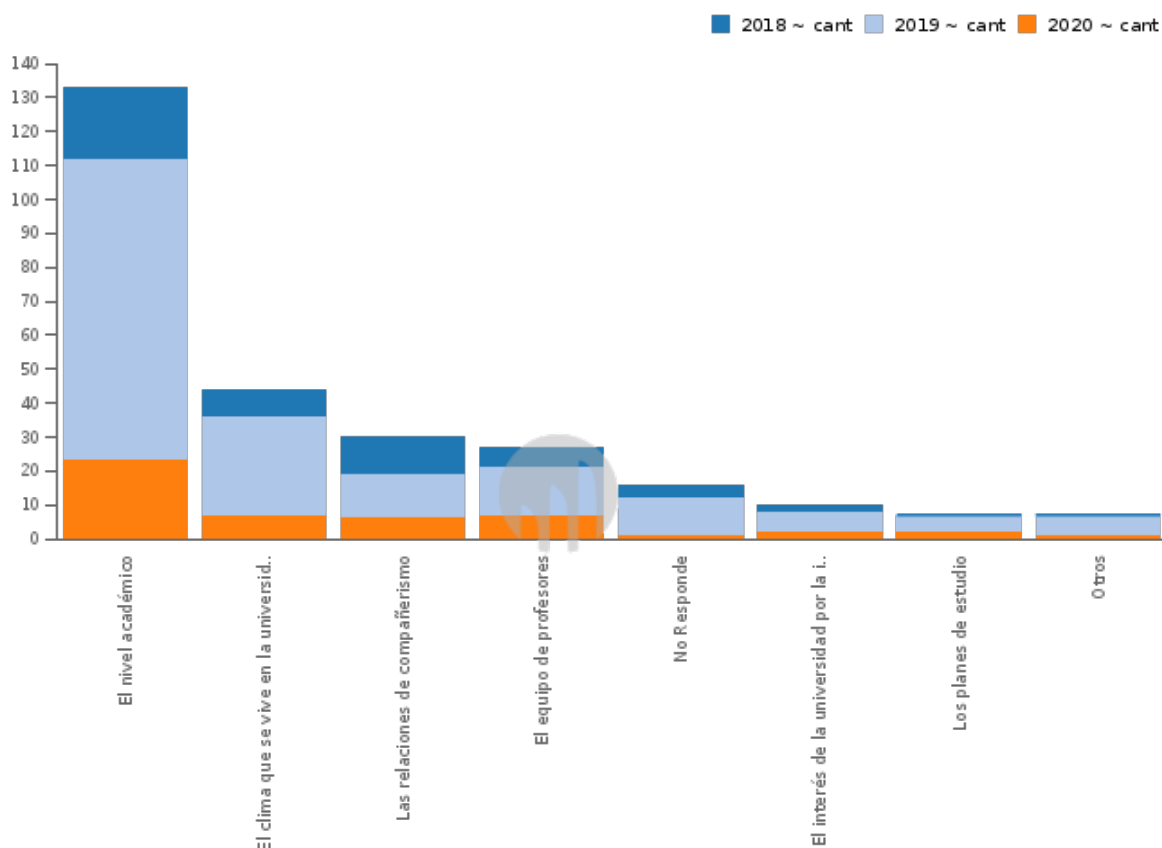


Figura 77: TA02 -Valoración académica de los graduados de la FCEQyN por cohorte

Como resultado de la consulta el experto pudo realizar una comparativa de los distintos criterios de valoración académica, por año. Siendo el ‘Nivel académico’ el criterio con mayor número registrado por año, se concluye que la calidad educativa brindada a los graduados para su formación profesional es lo más valorado por los mismos.

Tema Analítico 2: Valoración académica de los graduados por carrera.

Consulta derivada 1: ¿Qué motivó al graduado a la elección de una carrera?

Para realizar la consulta el experto utilizó las dimensiones *selección de la carrera* y *carrera*.

Como resultado de la consulta se observa que el motivo con mayor participación en la selección de una carrera refirió al ‘Desarrollo profesional’ que posee la FCEQyN. Esta elección registra el mayor porcentaje de participación (24,9%) como puede apreciarse en la Figura 78

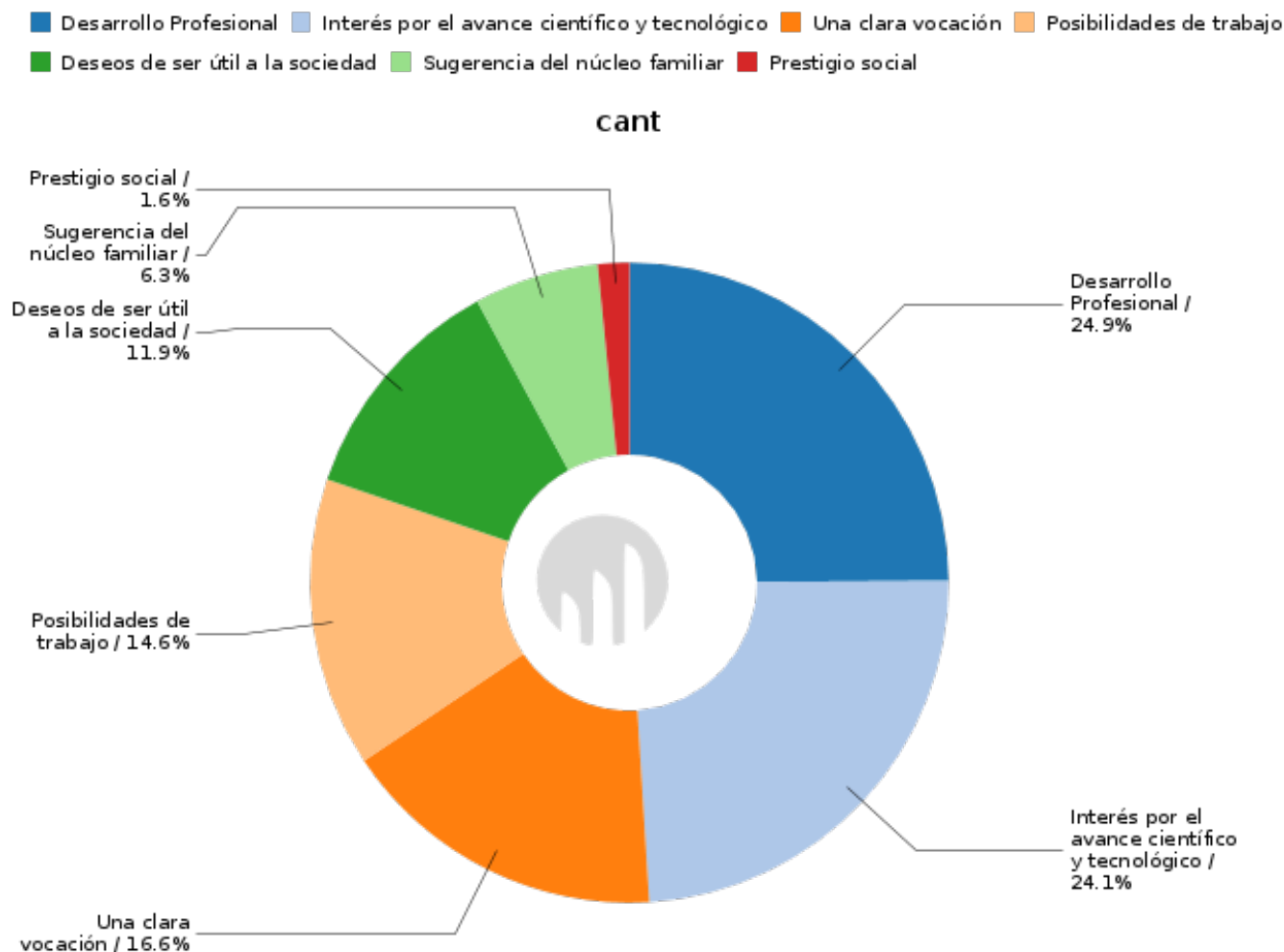


Figura 78: TA02-Valoración de los graduados para la elección de la carrera estudiada

Luego, se realizó un desglose por año y un ranking descendente, para conocer la distribución de cada motivo elegido para la selección de la carrera registrado por cohorte. Para ello se agregó la dimensión *tiempo* al análisis, pudiendo observar los resultados gráficamente en la Figura 79.

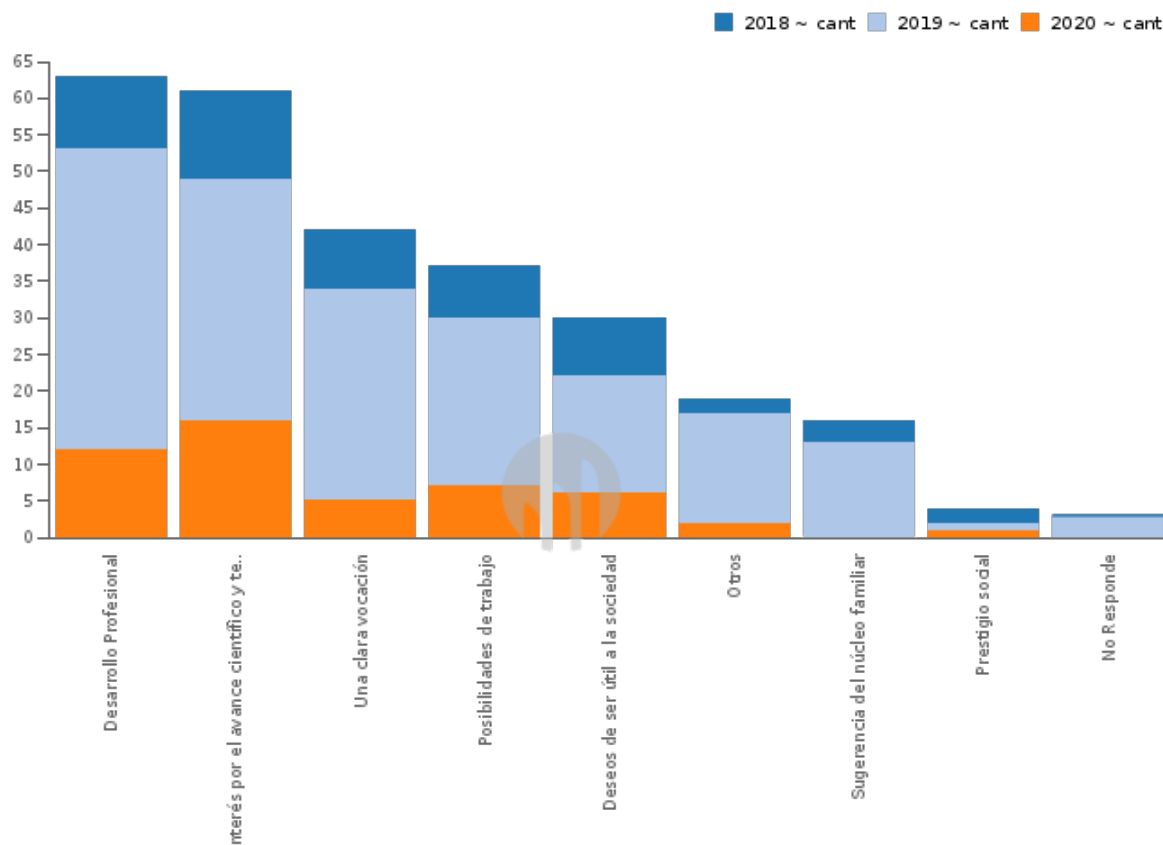


Figura 79: TA02-Valoración de los graduados para la elección de la carrera estudiada por cohorte

Como resultado de la consulta el experto pudo realizar una comparativa de los distintos motivos elegidos para la selección de la carrera, por año. Siendo el ‘Desarrollo profesional’ el criterio con mayor número registrado en general, el interés por el avance científico y tecnológico obtuvo una participación similar, superando dicho registro en el año 2020.

Tema Analítico 2: Valoración académica de los graduados por carrera.

Consulta derivada 2: ¿Cual es la duración real de egreso de una carrera?

Para esta consulta se seleccionaron las dimensiones *carrera* y *tiempo*. Al resultado de la consulta se le realizó un ranking descendente sobre la duración real de egreso en años, entre los graduados que respondieron la encuesta en un período de tiempo abarcado entre 2018 y 2020, por carrera.

La duración real de egreso corresponde al promedio en años, entre la fecha de ingreso del alumno a la carrera y la fecha de egreso a la misma. Se consideró como fecha de ingreso, la fecha de ingreso registrada en el sistema de gestión académica SIU-Guaraní.

El resultado de la consulta puede observarse en la Figura 80, donde para una mejor visualización de los datos se optó por mostrar el nombre corto de cada carrera.

carreara	cant	duracion_real
FA	43	10,163
PB	24	10,083
LSI	1	9
IQ	49	8,776
LG	58	8,5
BI	14	8,286
IA	14	7,5
PM	28	6,893
PF	3	6,667
MGA	2	6,5
MTA	2	6
ASC	8	5,125
DCA	11	5,091
EspBioqCli	1	5
TCyP	2	4,5
MSPyET	7	3,571
PUC	8	3,125
MTI	3	2,667
PUB	1	1

Figura 80: TA02-Ranking descendente por carrera sobre la duración real de egreso en años.

La información obtenida se la analizó y consideró en base a la cantidad de graduados que respondieron la encuesta.

A la consulta anterior se le agregó la duración teórica de la carrera, para obtener una comparativa entre ambos valores y poder deducir el tiempo que le lleva a un alumno recibirse en la carrera elegida. En la Figura 81 se puede observar gráficamente el resultado obtenido.

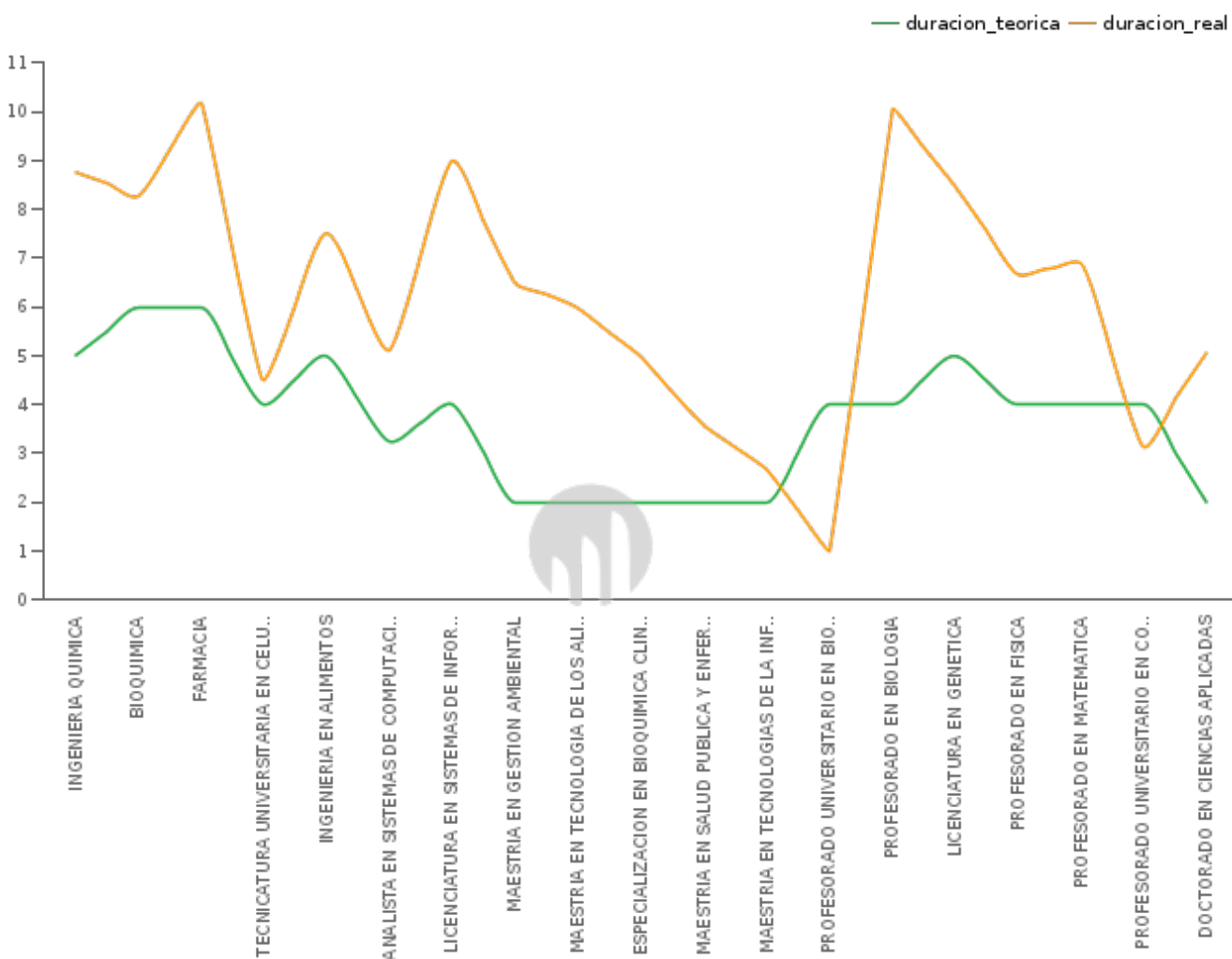


Figura 81: TA02-Duración teórica y duración real de egreso en años, por carrera.

Como resultado de la consulta el experto pudo obtener una comparativa entre la duración teórica y la duración real de egreso, por carrera. Se observa una tendencia constante de la duración real sobre la duración teórica en la mayoría de las carreras de la FCEQyN.

Tema Analítico 2: Valoración académica de los graduados por carrera.

Consulta derivada 3: Graduados de qué carrera terminan sus estudios según la duración teórica del plan de estudios.

Para esta consulta se seleccionó la dimensión *carrera*, junto a la medida desvío duración real. Se recordó al experto que para conocer las carreras cuyos graduados terminaron sus estudios en

base a la duración teórica del plan de estudios, el valor del desvío debe ser igual a cero. Caso contrario, el valor informado corresponde a la cantidad de años de más o de menos que le llevó al alumno graduarse.

El resultado de la consulta puede observarse en la Figura 82, donde para una mejor visualización de los datos se optó por mostrar el nombre corto de cada carrera.

carreara	cant	duracion_teorica	desvio_duracion_real
PB	24	4	6
MGA	2	2	5
LSI	1	5	4
FA	43	5	4
MTA	2	2	4
IQ	49	5	4
LG	58	5	4
DCA	11	2	3
EspBioqCli	1	2	3
PM	28	4	3
PF	3	4	3
IA	14	5	3
BI	14	6	2
ASC	8	3	2
MSPyET	7	2	2
MTI	3	2	1
TCyP	2	3	1
PUC	8	4	-1
PUB	1	4	-3

Figura 82: TA02-Desvío de la duración real en años por carrera

En un análisis preliminar realizado sobre las carreras mostradas con valor negativo, se pudo determinar que más del 40 % de dichas carreras fueron aprobadas por equivalencias.

En la Figura 83 se puede observar gráficamente, el resultado obtenido.

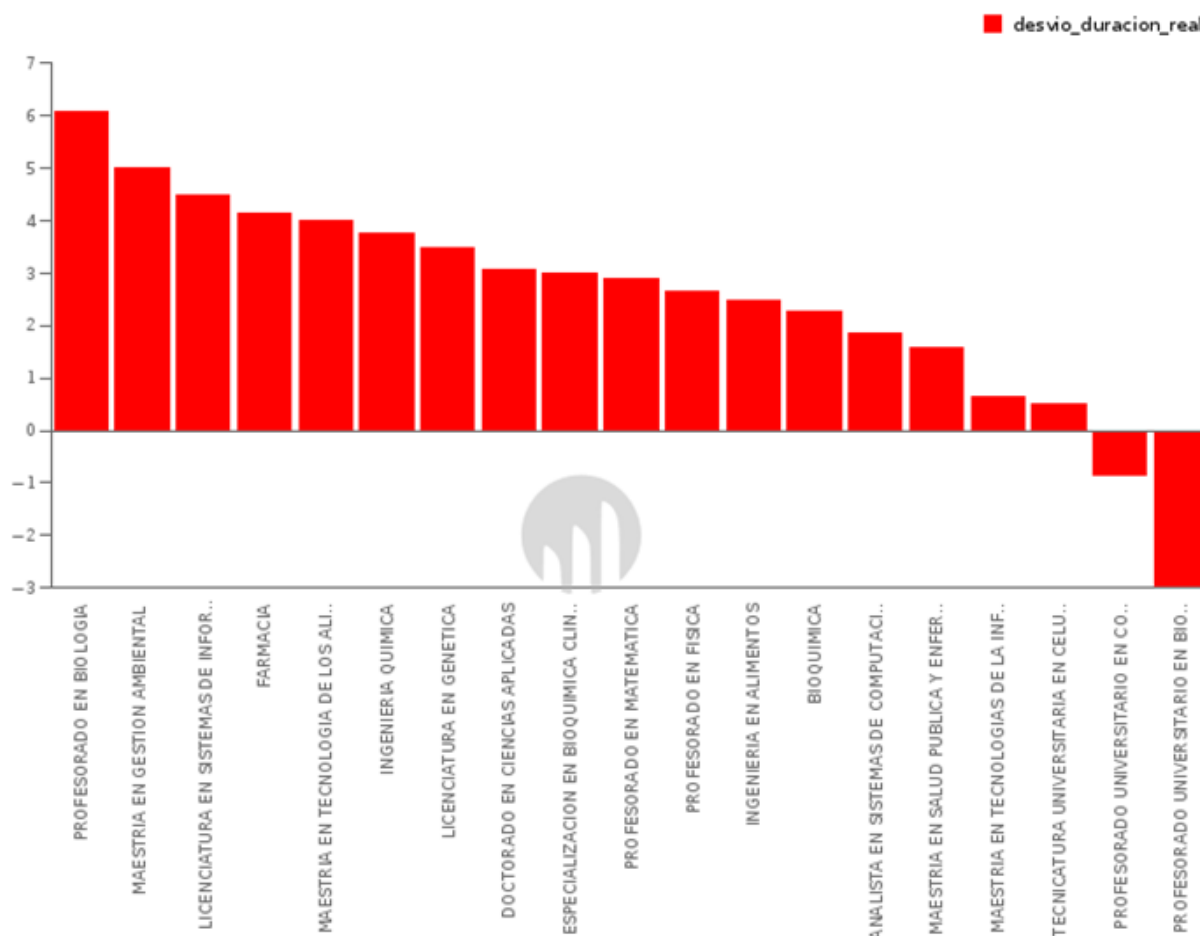


Figura 83: TA02-Representación gráfica del desvío de la duración real del plan de estudio por carrera en años

Tema Analítico 2: Valoración académica de los graduados por carrera.

Consulta derivada 4: ¿Cómo debería ser el plan de estudio de una carrera según su duración y contenido?

Para esta consulta se seleccionó la dimensión *carrera, duración del plan de estudio y contenido del plan de estudio*. Al resultado de la consulta se le realizó un ranking descendente sobre la cantidad de graduados que respondieron la encuesta, en un período de tiempo abarcado entre 2018 y 2020.

La misma puede observarse en la Figura 84.

	plan_corto	No		Si	
	plan_tecnico	Si	No	Si	No
carreara		cant	cant	cant	cant
LICENCIATURA EN GENETICA		33	13	4	-
INGENIERIA QUIMICA		30	5	8	3
FARMACIA		21	9	4	5
PROFESORADO EN MATEMATICA		7	19	-	-
PROFESORADO EN BIOLOGIA		3	16	-	1
BIOQUIMICA		2	5	2	2
INGENIERIA EN ALIMENTOS		10	-	1	-
DOCTORADO EN CIENCIAS APLICADAS		2	2	-	2
ANALISTA EN SISTEMAS DE COMPUTACION		2	2	-	1
PROFESORADO UNIVERSITARIO EN COMPUTACION		1	6	1	-
MAESTRIA EN SALUD PUBLICA Y ENFERMEDADES TRANSMISIBLES		2	4	-	-
MAESTRIA EN TECNOLOGIAS DE LA INFORMACION		1	1	-	-
PROFESORADO EN FISICA		1	2	-	-
TECNICATURA UNIVERSITARIA EN CELULOSA Y PAPEL		-	2	-	-
MAESTRIA EN GESTION AMBIENTAL		-	2	-	-
MAESTRIA EN TECNOLOGIA DE LOS ALIMENTOS		1	1	-	-
ESPECIALIZACION EN BIOQUIMICA CLINICA AREA ENDOCRINOLOGIA		-	1	-	-
PROFESORADO UNIVERSITARIO EN BIOLOGIA		-	1	-	-

Figura 84: TA02-Valoración sobre la duración y contenido de los planes de estudio por carrera

Como resultado de la consulta el experto pudo observar en que carreras la opinión del graduado hace constar que el plan de estudios de la carrera estudiada debería ser revisado en lo que refiere a contenidos, estructura y duración.

Tema Analítico 2: Valoración académica de los graduados por carrera.

Consulta derivada 5: ¿Qué aportes esperan los graduados de una carrera le brinde la universidad?

Para esta consulta se seleccionó la dimensión *carrera y aporte de la universidad*. Se realizó la consulta para el todas las carreras, en un período de tiempo abarcado entre 2018 y 2020, y se obtuvo el porcentaje de participación de cada aporte.

En la Figura 85 se pueden observar el resultado de la consulta.

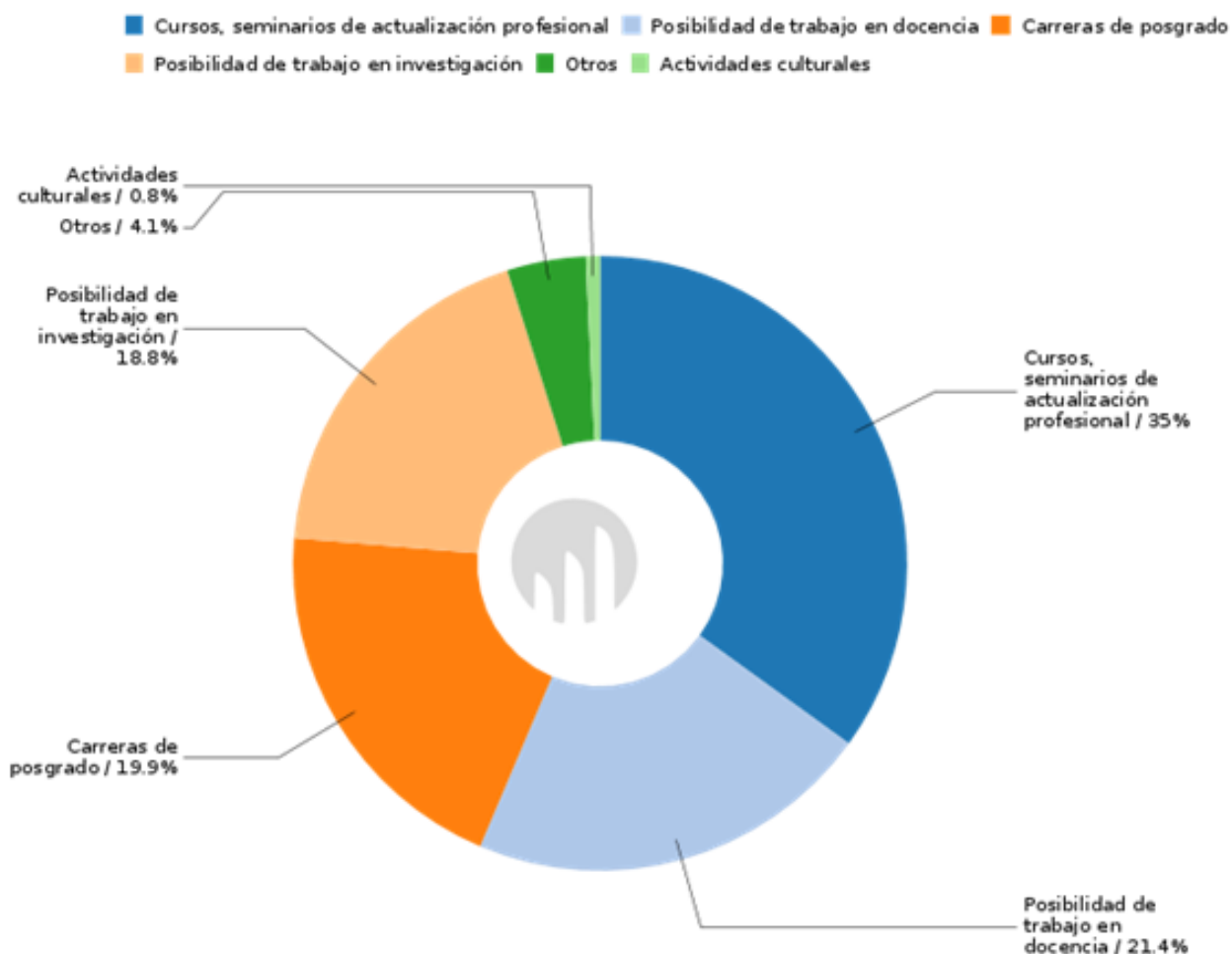


Figura 85: TA02-Aportes valorados por los graduados sean brindados por la Facultad/Universidad

Como se puede observar el aporte que los graduados esperan le brinde la Facultad, es de ‘Cursos, seminarios de actualización profesional’. A partir de esta exploración, se realizó un desglose por carrera, para conocer cuáles participaron del análisis.

De la consulta obtenida se aplicó un ordenamiento descendente sobre la cantidad de graduados, para una mejor visualización de los resultados.

En la Figura 86 se puede observar el gráfico obtenido.

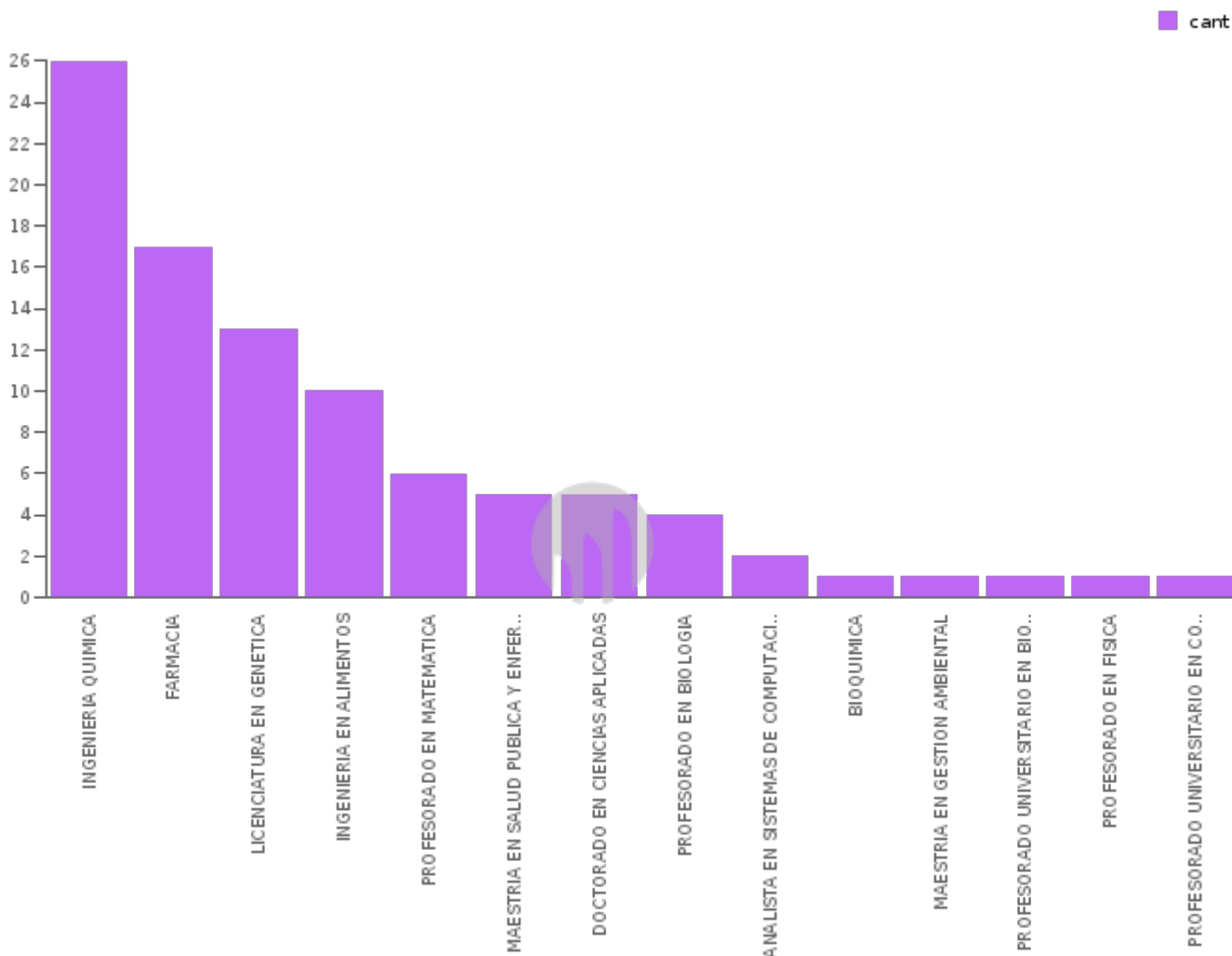


Figura 86: TA02-Cantidad de graduados por carrera que desean recibir ‘Cursos, seminarios de actualización profesional’

Para la validación de los Data Marts desarrollados, se diseñó una encuesta que los expertos completaron para evaluar su nivel de satisfacción.

Para evitar subjetividad al momento de realizar la validación, los expertos respondieron el nivel de satisfacción por cada tema de análisis y consulta derivada definida teniendo en cuenta los requerimientos planteados para cada caso. Las métricas utilizadas para la validación fueron Cumple (C), No cumple (NC) y Cumple parcialmente (CP), y se describen a continuación:

Cumple (C): El Data Mart cumple con el requerimiento planteado en la Etapa 1 (E1.A4- **Temas analíticos**).

No cumple (NC): El Data Mart no cumple con el requerimiento planteado en la Etapa 1 (E1.A4- **Temas analíticos**).

Cumple Parcialmente (CP): El Data Mart cumple parcialmente con el requerimiento planteado en la Etapa 1 (E1.A4- **Temas analíticos**)

En la Tabla 38 se muestran los resultados:

Tabla 38: Validación de los usuarios

Tema analítico	Consulta derivada	Criterio evaluado	Experto / Nivel de satisfacción								
			Coordinador Graduados			Director LSI			Vice Decano		
			C	NC	CP	C	NC	CP	C	NC	CP
Situación laboral	TA1	Requerimiento	X			X			X		
	CD1	Requerimiento	X			X			X		
	CD2	Requerimiento	X			X			X		
	CD3	Requerimiento	X			X					X
	CD4	Requerimiento	X			X					X
Valoración académica	TA2	Requerimiento	X			X			X		
	CD1	Requerimiento	X			X			X		
	CD2	Requerimiento	X			X			X		
	CD3	Requerimiento	X			X					X
	CD4	Requerimiento	X			X			X		
	CD5	Requerimiento	X			X					X

Como resultado de la navegabilidad y usabilidad de la herramienta, se remarca la flexibilidad que brinda la misma por parte de los expertos, permitiendo obtener una visión cada vez más detallada en cada una de las consultas.

A partir de la validación realizada y las respuestas brindadas por los expertos, se pudo observar que en su mayoría, el nivel de satisfacción manifestado *Cumple* con los requerimientos planteados para cada caso. Este nivel de satisfacción es resultado de la participación activa que tuvieron los expertos durante todo el proceso de desarrollo de los Data Marts.

- **Tableros de control**

Una vez que realizada la navegación por los cubos y verificados los mismos en base a las variables e indicadores de interés (cantidad de graduados, promedio edad, duración real de la carrera, desvío de la duración real) se confeccionaron tableros de control donde se pudieron observar indicadores más relevantes.

En la Figura 87, se observa el tablero referido a situación laboral de los graduados, por año y carrera. El mismo permite conocer la cantidad de graduados que respondieron la encuesta en un año particular (gráfico de línea), que al interactuar con el mismo permite obtener un ranking de las cinco carreras que mas graduados respondieron las encuestas (gráfico de barras) y posteriormente, cuál es el porcentaje de participación de los mismos según su situación laboral (gráfico de torta).

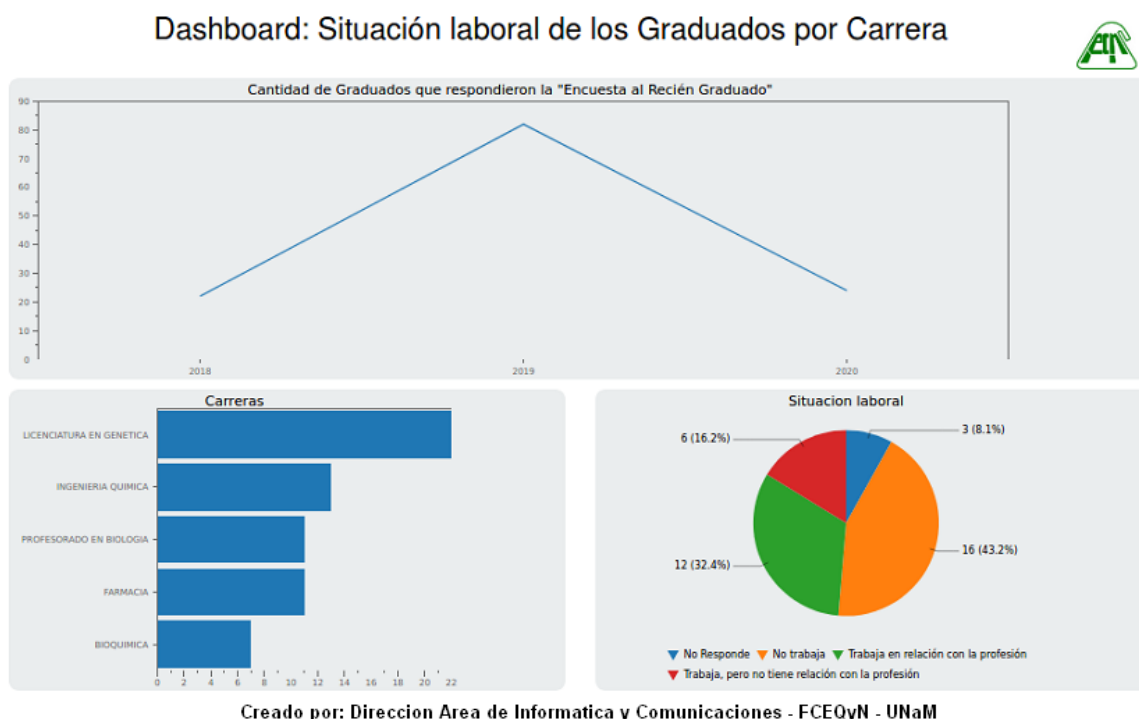


Figura 87: Tablero situación laboral de los graduados por carrera.

Al hacer click en un año se filtra y actualiza automáticamente el gráfico de barras generando un ranking de las cinco carreras que tuvieron mayor cantidad de encuestas contestadas en ese año. Para el año 2019 figuran: Licenciatura en Genética, Ingeniería Química, Profesorado en Biología, Farmacia y Bioquímica.

Luego al seleccionar una de ellas, en el gráfico de torta se muestra como se encuentran distribuidos porcentualmente los graduados según su situación laboral.

En cuanto a la ‘Valoración académica de los graduados’, en la Figura 88 se observa el tablero de control diseñado, por año y por carrera. El mismo permite conocer, para un año en particular, cuál es el criterio mayormente valorado por los graduados en la elección de la carrera estudiada en la FCEQyN -UNaM (gráfico de torta), que al interactuar con el mismo permite obtener las diez carreras que más graduados eligieron este criterio en las encuestas (gráfico de barras).

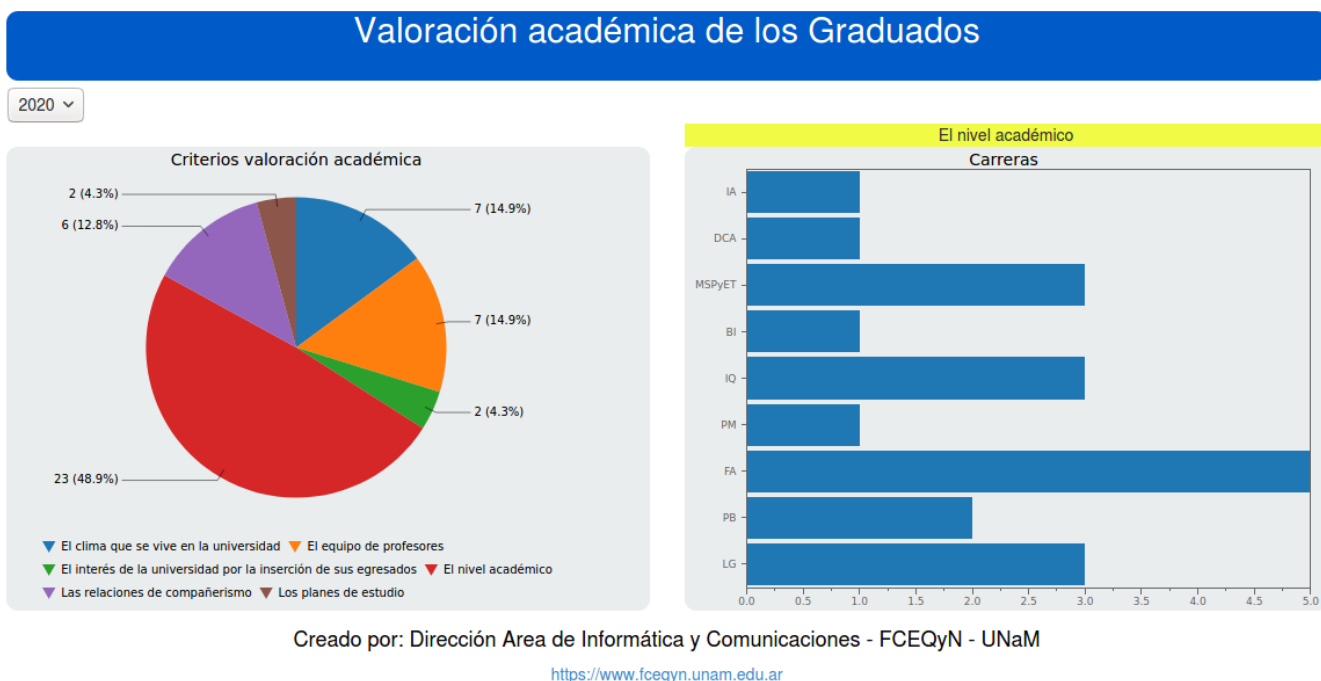


Figura 88: Tablero valoración académica de los graduados por carrera.

Al hacer click en un año se filtra y actualiza automáticamente el gráfico de torta que muestra como se encuentran distribuidos porcentualmente los graduados según el criterio de valoración académica elegido para estudiar una carrera. Luego al seleccionar uno de ellos, se muestran las diez carreras que tuvieron mayor cantidad de encuestas contestadas con el criterio de valoración seleccionado. En el ejemplo de la gráfica se puede observar que para el año 2020 figura “El nivel académico” como el criterio de valoración académica mayoritario, y las carreras: Ingeniería en Alimentos, Doctorado en Ciencias Aplicadas, Maestría en Salud Pública y Enfermedades transmisibles, Bioquímica, Ingeniería Química, Profesorado en Matemática, Farmacia, Profesorado en Biología y Licenciatura en Genética, como las principales en la elección de dicho criterio.

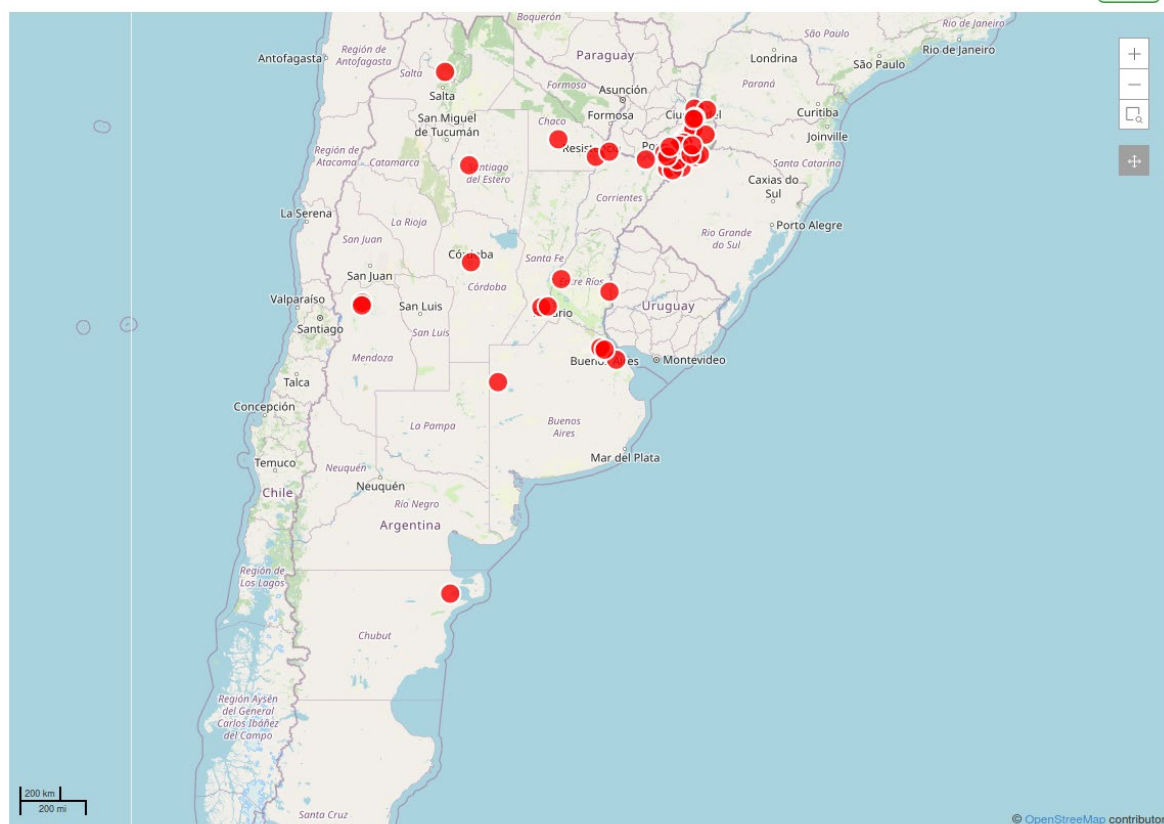
Por otro lado, se diseñó un tablero de control con componentes de mapas (newMapComponent) que facilitó ubicar puntos y zonas en el mapa, referida a la localidad de

residencia de los graduados. La implementación de este componente permitió conocer como se encuentran distribuidos los graduados, en un contexto geográfico.

La implementación de mapas de geolocalización requirió contar con los datos de longitud y de latitud de cada una de las localidades involucradas en el análisis. Estos datos al no encontrarse registrado en ningún sistema fuente, fueron mapeados a las localidades de interés, a través de la herramienta Pentaho Data Integration.

En la Figura 89 se observa la distribución geográfica de los graduados de la FCEQyN en la República Argentina, donde la mayor concentración de graduados se visualizó en la provincia de Misiones.

Distribución geográfica de los Graduados



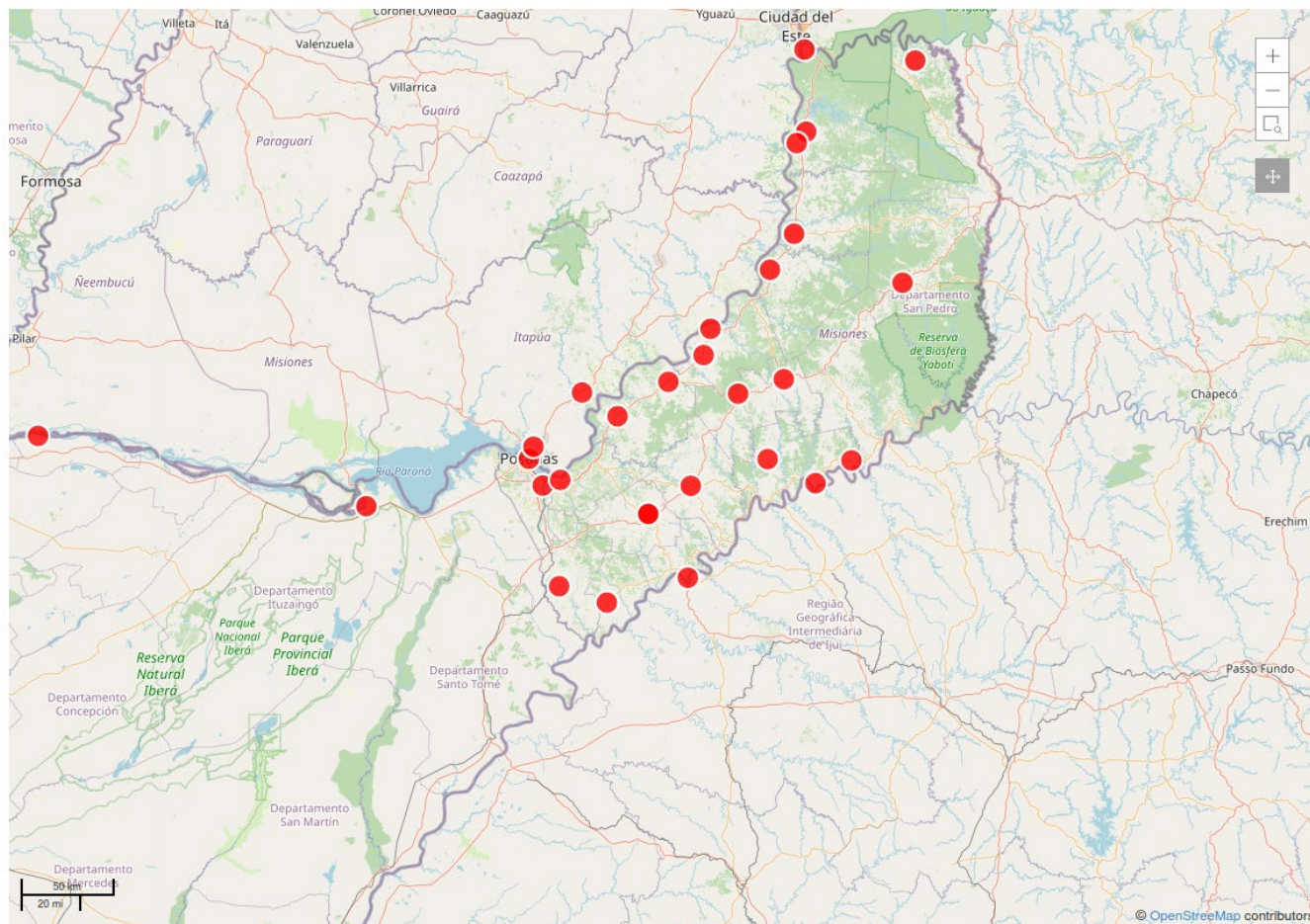
Creado por: Dirección Área de Informática y Comunicaciones - FCEQyN - UNaM

<https://www.fceqyn.unam.edu.ar>

Figura 89: Distribución geográfica de los graduados en la Argentina

Haciendo una aproximación sobre la zona con mayor concentración de graduados, permitió situar en el mapa las localidades involucradas en la visualización general (Figura 90), posibilitando la selección de una localidad en particular.

Distribución geográfica de los Graduados



Creado por: Dirección Área de Informática y Comunicaciones - FCEQyN - UNaM

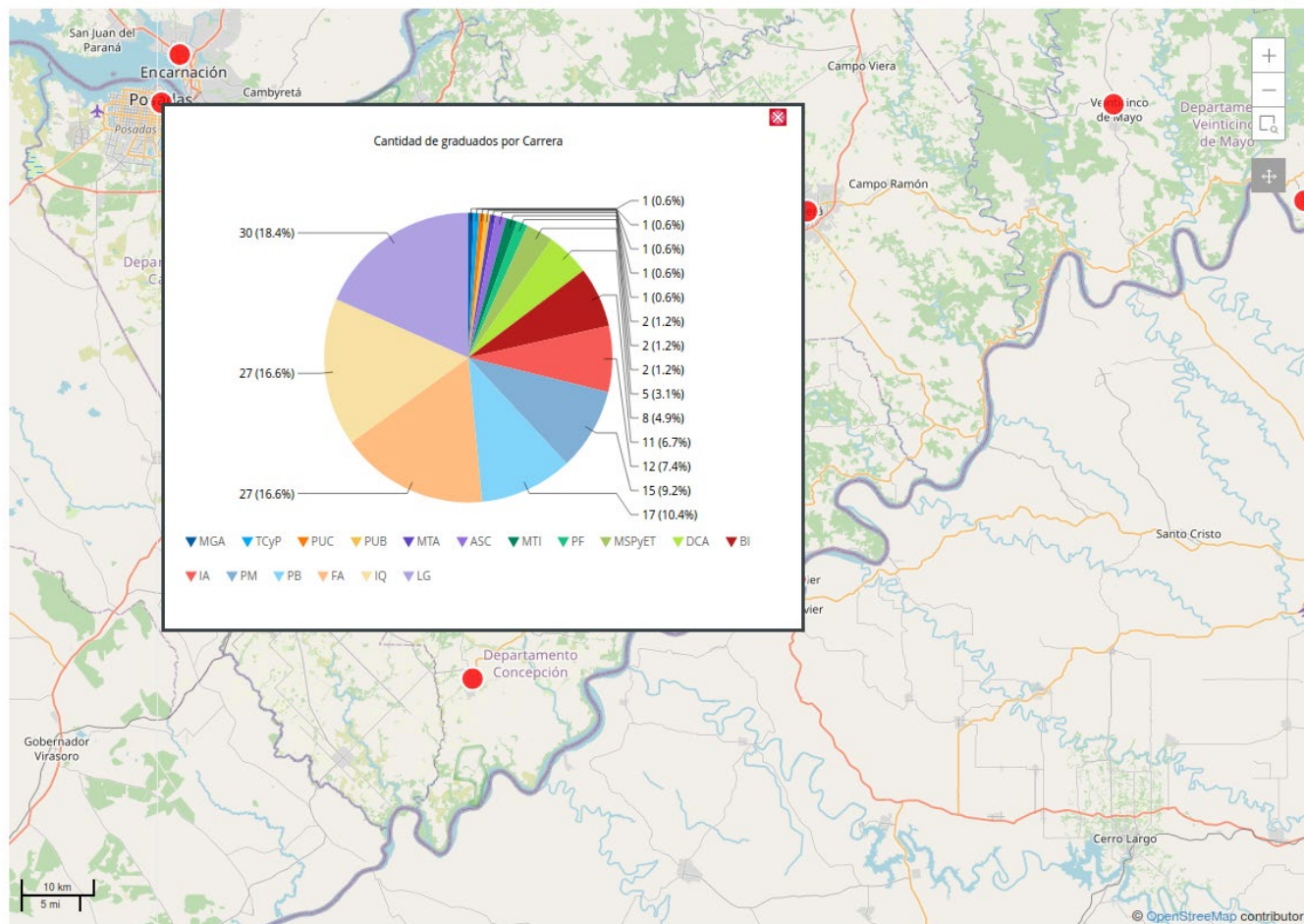
<https://www.fceqyn.unam.edu.ar>

Figura 90: Distribución geográfica de los graduados en Misiones

En la medida que la aproximación se hace más detallada, se pudo visualizar el nombre de cada localidad y seleccionar una de ellas. Al hacer click sobre una localidad se abre una ventana emergente (pop-up window) en la se puede observar un gráfico de torta con el porcentaje de participación de cada carrera en base a la cantidad de graduados que residen en dicha localidad.

En la Figura 91 se puede observar la cantidad de graduados, y el porcentaje de participación por carrera, residentes en Posadas.

Distribución geográfica de los Graduados



Creado por: Dirección Área de Informática y Comunicaciones - FCEQyN - UNaM

<https://www.fceqyn.unam.edu.ar>

Figura 91: Cantidad de graduados por Carrera residentes en Posadas

○ Review sobre la metodología propuesta

Ante la experiencia adquirida a partir del caso de estudio planteado, se realizaron numerosos ajustes en las plantillas diseñadas inicialmente.

Uno de esos ajustes tuvo lugar en las plantillas asociadas a la elicitación de los requisitos (DERU). Esta plantilla contaba inicialmente con un formato propio para ir enumerando las necesidades de los expertos, que luego serían la base para el desarrollo de los Data Marts (Tabla 39).

Tabla 39: Formato inicial de la plantilla DERU

DERU	Ej.: Ventas	
Fecha		
Institución		
Experto consultado		
Cargo		
Tema analítico	Nro. Req.	Requisito de análisis
TA1	R1	
	...	
	Rn	...
TAn	R1	
	...	
	Rn	...

Cuando comenzaron las entrevistas se evidenció dificultad a la hora de plasmar esas necesidades de manera explícitas y enumeradas, debido al volumen de información recibida y a la gran demanda de datos con la que se encontraban los expertos. Muchas de ellas, pudiéndose cubrir mediante simples consultas SQL.

Por ello se ajustó el modelo inicial del DERU, y se creó una nueva plantilla identificada como TA para lograr una mejor organización en la planificación y desarrollo de cada tema analítico.

Por otro lado, también se confeccionaron inicialmente plantillas asociadas a la identificación de perspectivas e indicadores (Tabla 40) y detallar el nivel de granularidad que se pretendía obtener para cada perspectiva. (Tabla 41)

Tabla 40: Plantilla de perspectivas e indicadores

Fecha: Institución: Experto consultado: Cargo:		
Necesidad u Objetivo:		
Nro.Req.	Indicadores	Perspectivas
R1	▪ Indicador 1... ▪ Indicador 2... ▪ Indicador N...	▪ Perspectiva 1... ▪ Perspectiva 2... ▪ Perspectiva N...
...		
Rn		

Observaciones:

Tabla 41: Plantilla sobre el nivel de granularidad.

Perspectiva	Tabla	Campo	Valores	Descripción	Relevante
<i>Perspectiva 1</i>	<i>Tabla X</i>	<i>Campo 1</i>			Si - No
		<i>Campo 2</i>			Si - No
		<i>Campo n</i>			Si - No
<i>Perspectiva 2</i>	<i>Tabla Y</i>	<i>Campo 1</i>			Si - No

Estas plantillas, en base al concepto de mejora continua propuesto por CMMI, fueron rediseñadas y optimizadas para simplificar la gestión de los requisitos. Por un lado, en la plantilla de identificación de las fuentes de datos (FUDA) diseñada se incorporó el detalle de granularidad solicitada por el experto y por otro lado, se diseñó la plantilla de conformación de indicadores (CI) donde queda registrado la manera que se calculan los indicadores, además de su identificación y descripción.

Otro de los ajustes realizado en la metodología propuesta fue la incorporación de una fase de factibilidad y aceptación del modelo, en una etapa temprana (E2.A7 – **Factibilidad y aceptación del modelo**). Esta fase resultó un punto crítico para consolidar los modelos diseñados referidos a los temas de análisis uno y dos, y aportar sugerencias recomendaciones sobre posibles cambios a implementar que permitan dar continuidad al tema de análisis tres.

Por otro lado, para incrementar la participación de los graduados en las encuestas y así disponer de datos actualizados para poblar los Data Marts, se trabajó a nivel institucional en una actualización del circuito administrativo que el alumno debía realizar para tramitar su título universitario. Dicho trámite incorporó como nuevo requisito obligatorio, la presentación del comprobante de completitud de la encuesta por parte del mismo.

El apoyo institucional recibido por parte de las autoridades de la FCEQyN permitió modificar el circuito administrativo y adaptarlo al nuevo contexto de trabajo.

A mediados del año 2018, con la incorporación de dicha obligatoriedad y la creación de la Oficina de Vinculación con los Graduados en el ámbito de la FCEQyN, se produjo una mayor concientización y participación de los graduados en las encuestas a través del tiempo. Previo a estos cambios, la completitud de las mismas no alcanzaba el 10 % de los graduados y eran disparadas sin ningún criterio ni objetivo.

Sin embargo, a pesar que el porcentaje de completitud de las encuestas ha aumentado considerablemente, aún existen graduados que no completan la misma al momento de graduarse sino que lo hacen al momento de retirar el título, generando una diferencia en el porcentaje de completitud considerable.

En la Tabla 42 se describen datos cuantitativos de los datos recabados a través del sistema SIU-Kolla por grupo de graduados, relacionados a la cantidad de graduados por colación, la cantidad de graduados de una colación que completaron la encuesta y el porcentaje de completitud de las mismas.

Tabla 42: Graduados que completaron la encuesta por colación SIU-Kolla

Grupo graduados SIU-Kolla	Cantidad graduados por Colación	Cantidad graduados que completaron la encuesta	% de Completitud de la encuesta
Colación 2018	140	85	60%
Colación 2019	199	160	80%
Colación 2020 (Hasta el 15/12/2020)	122	99	81%

○ **Desempeño de la Metodología propuesta**

Con respecto al proceso de desarrollo y a la gestión del proyecto, al no existir hasta el momento una metodología aplicada al desarrollo de Data Marts en la FCEQyN, no se contaba con ninguna métrica para medir la agilidad de un proceso semejante. Por ello, las primeras estimaciones realizadas sobre la planificación del proyecto, se vieron afectadas en varias oportunidades por los cambios producidos durante la elaboración y el desarrollo de la metodología propuesta.

Durante la primera etapa de desarrollo, el Programa de Graduados se encontraba en pleno proceso de fortalecimiento, designación de referentes institucionales y elaboración de pautas a ser desarrolladas en base a los PDI aprobados, en el ámbito de la FCEQyN y de la UNaM. Esto provocó se extienda la duración de dicha etapa a cinco meses y que se deba establecer prioridades sobre las necesidades relevadas, definiendo los temas analíticos a ser abordados.

Por otro lado, la falta de un mecanismo para el relevamiento de los datos asociados a graduados, requirió de reuniones complementarias para definir y actualizar el circuito administrativo de solicitud de título, finalmente implementado al final del año 2019.

En la segunda etapa de la metodología, el análisis de datos y la documentación que se generó para el primer tema analítico, sirvió como base para el diseño y tratamiento de los

siguientes. Esto permitió reducir el tiempo dedicado a cada tema analítico, de tres meses a un mes de trabajo e identificar en una fase temprana de la metodología, la existencia o no de datos necesarios para el análisis.

La Etapa 4 dedicada a la confección de los procesos ETL es la que mayor tiempo duró en todo el desarrollo de la metodología. Durante el desarrollo de la misma intervino una sola persona, cuya dedicación promedio fue de cuatro horas diarias.

A comienzo de esta etapa se realizó la instalación de dos servidores de Pentaho, dedicados al desarrollo y puesta en producción en la FCEQyN.

Entre las actividades descriptas en la Etapa 4, la relacionada al “Tratamiento de los datos” fue la que mayor esfuerzo demandó, seguida de la actividad de “Selección de los datos” por la particularidad propia del contexto.

A continuación se listan algunas de las tareas realizadas:

- Se integraron y unificaron los datos provenientes de dos fuentes de datos: SIU-Kolla y SIU-Guaraní, esta última implementada de manera descentralizada.
- Se seleccionaron y exportaron los datos necesarios para el análisis: Treinta atributos correspondientes al SIU-Kolla y veintiséis atributos correspondientes al SIU-Guaraní.
- Se creó un DSA para el tratamiento y depuración de los datos.
- Se confeccionaron cincuenta y nueve procesos ETL encargados de capturar, limpiar, transformar, calcular y cargar los datos en los Data Marts diseñados.
- Se diseñaron dos Data Marts.

En la Figura 92 se puede observar el tiempo de duración de las principales etapas del proyecto.

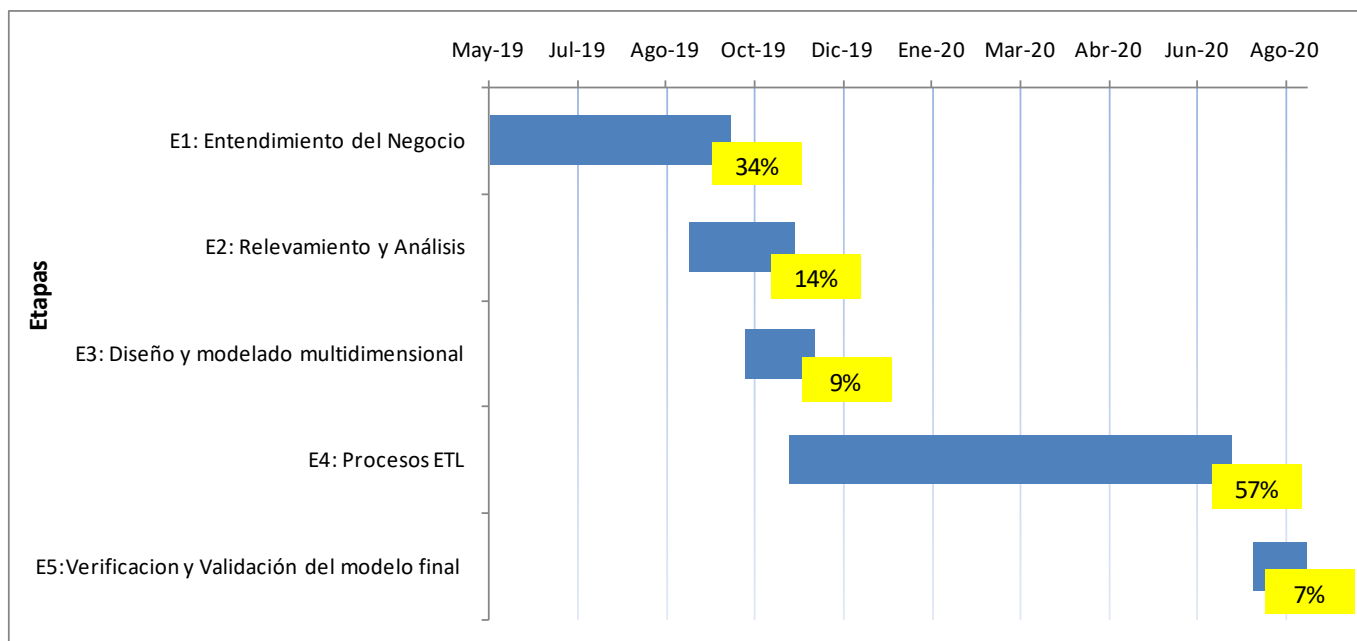


Figura 92: Duración del proyecto [Elaboración propia].

Si bien el proyecto se sigue desarrollando e incorporando nuevas peticiones y actividades, se generaron un total de cuarenta y cinco tareas. En la Figura 93 se pueden observar algunas de ellas.

Sistemas » Business Intelligence

Búsqueda:

» Business Intelligence

VistazoActividadPeticionesNueva peticiónGanttCalendarioNoticiasDocumentosWikiArchivosRepositorioConfiguración

Peticiones

▼ Filtros

☒ Estado

abierta

▼

☒ Tipo

igual

▼

Tarea

▼

✕

Añadir el filtro

▼

► Opciones

☒ Aceptar

☐ Anular

Peticiones

Ver todas las peticiones
Resumen
Calendario
Gantt
Importar

#	▼	Tipo	Estado	Prioridad	Asunto	Asignado a	Actualizado	
☐		345	Tarea	New	Normal	TA03: Necesidades de capacitación del graduado según su distribución geográfica	Daniela Rolon	2020-08-31 15:21
☐		344	Tarea	New	Normal	[BI] Edad: spc_f_edad (date) - TA01	Daniela Rolon	2020-08-25 11:41
☐		343	Tarea	New	Normal	[BI] Tiempo: spc_dim_tiempo (dim_tiempo) - TA01	Daniela Rolon	2020-08-25 11:31
☐		342	Tarea	New	Normal	[BI] Rango Etario: spc_dim_edad (dim_edad) - TA01	Daniela Rolon	2020-08-25 11:29
☐		333	Tarea	New	Normal	[BI] Modelo Físico (TA01)	Daniela Rolon	2020-08-24 15:11
☐		332	Tarea	New	Normal	[BI] Modelo conceptual ampliado (TA01)	Daniela Rolon	2020-08-24 15:13
☐		331	Tarea	New	Normal	[BI] Modelo conceptual preliminar (TA01)	Daniela Rolon	2020-08-24 14:52
☐		321	Tarea	New	Normal	[BI] Dataset F - planes de estudio (TA02)	Daniela Rolon	2020-08-05 15:58
☐		320	Tarea	New	Normal	[BI] TA02: Valoración académica del graduado	Daniela Rolon	2020-08-05 14:57
☐		319	Tarea	New	Normal	[BI] Carga inicial: ETL (TA01)	Daniela Rolon	2020-08-05 15:28
☐		318	Tarea	New	Normal	[BI] ETL Generales dw (TA01)	Daniela Rolon	2020-08-05 15:27
☐		317	Tarea	New	Normal	[BI] ETL Generales dsa (TA01)	Daniela Rolon	2020-08-05 15:07

Figura 93: Lista de tareas en Redmine FCEQyN

A continuación se observa el detalle del Ticket Nro.: 294 que dio origen al primer tema analítico definido, con la descripción del **Cómo**, **Quiero** y **Para** (Figura 94).

Sistemas >> Business Intelligence

Búsqueda: » Business Intelligence

Vistazo Actividad **Peticiones** Nueva petición Gantt Calendario Noticias Documentos Wiki Archivos Repositorio Configuración

Tarea #294

[BI] TA01: Situación laboral del graduado según su distribución geográfica

« Anterior | 31/35 | Siguiente »

Añadido por Daniela Rolon hace 6 meses. Actualizado hace 2 minutos.

Estado: New Fecha de inicio: 2020-03-05

Prioridad: Normal Fecha fin: 2020-07-16

Asignado a: Daniela Rolon % Realizado: 100%

Descripción

TEMA ANALÍTICO TA01: Situación laboral del graduado según su distribución geográfica

DERU Nro.: 1

Cómo: Coordinador de la Oficina de Graduados

Quiero: Conocer la situación laboral de los graduados y su distribución geográfica.

Para: Generar espacios laborales y poder articularlos con distintos sectores socio-productivos e industriales del sector.

Subtareas

Peticiones

Ver todas las peticiones

Resumen

Calendario

Gantt

Importar

Seguidores (1)

Daniela Rolon

Figura 94: Detalle de un requerimiento en Redmine FCEQyN

○ Validación de la Metodología propuesta

Como mecanismo de validación de la propuesta metodológica se utilizó un instrumento de recolección de datos (encuesta), compuesta de ocho preguntas basadas en aspectos generales de la metodología. La misma fue distribuida entre dos referentes técnicos del equipo SIU-Wichi del SIU y un profesional técnico experto en Inteligencia de Negocios de la UNaM.

La encuesta fue diseñada utilizando la escala de Likert [58] como herramienta de medición, para conocer la opinión de los encuestados sobre la propuesta metodológica desarrollada.

Las opciones disponibles utilizadas para la medición fueron: Totalmente de acuerdo, De acuerdo, Ni de acuerdo ni en desacuerdo, En desacuerdo, Totalmente de acuerdo. Para evaluar el criterio sobre complejidad, se seleccionaron las opciones: Muy alta, Alta, Ni alta ni baja, Baja, Muy baja.

Los resultados obtenidos de las encuestas, fueron positivos y enriquecedores para la metodología propuesta.

Entre los aportes realizados por los referentes técnicos del equipo SIU-Wichi del SIU, se puede mencionar la sugerencia de reflejar mediante flechas, en el gráfico general de la metodología el feedback existente entre las distintas etapas; como así también aclarar en la etapa de factibilidad económica, que no necesariamente implica un costo nuevo el desarrollo de un Data Mart en el caso

que ya exista implementado un DW en la universidad. Por último, la incorporación de un esquema temporal (TMP) que contenga los datos de origen sin procesar, para luego hacer las transformaciones e integración en el esquema DSA, para finalmente cargar los datos integrados en el Data Mart.

Por otro lado, observaron como aspecto positivo en la definición de roles la incorporación de los expertos en cada dominio abordado. Esto fue considerado una contribución fundamental para que el producto final resuelva los requerimientos inicialmente planteados.

Todos los aportes realizados sobre la metodología, fueron incorporados a la misma, contribuyendo a mejorar el entendimiento y la aplicabilidad en futuros desarrollos.

La encuesta que se elaboró y se envió a los referentes técnicos se observa en la Figura 95 y Figura 96:

Metodología para el desarrollo de Data Marts en Universidades Nacionales

Facultad de Ciencias Exactas, Químicas y Naturales - UNIVERSIDAD NACIONAL DE MISIONES

Encuesta de validación de la metodología propuesta.

Dirección de correo electrónico *

Nombre y Apellido *

1-Cómo considera la complejidad de la metodología. *

- ☐ Muy alta
- ☐ Alta
- ☐ Ni alta ni baja
- ☐ Baja
- ☐ Muy baja

2- La definición de las etapas y tareas, son comprensibles. *

- ☐ Totalmente de acuerdo
- ☐ De acuerdo
- ☐ Ni de acuerdo ni en desacuerdo
- ☐ En desacuerdo
- ☐ Totalmente en desacuerdo

3 - Considera que la metodología puede ser adaptada a otras Universidades Nacionales. *

- ☐ Totalmente de acuerdo
- ☐ De acuerdo
- ☐ Ni de acuerdo ni en desacuerdo
- ☐ En desacuerdo
- ☐ Totalmente en desacuerdo

Figura 95: Encuesta de validación de la metodología propuesta.

4- La metodología es precisa en la definición de sus etapas y tareas. *

- ☐ Totalmente de acuerdo
- ☐ De acuerdo
- ☐ Ni de acuerdo ni en desacuerdo
- ☐ En desacuerdo
- ☐ Totalmente en desacuerdo

5- La definición de Roles y Responsabilidades, se encuentran definidos con claridad. *

- ☐ Totalmente de acuerdo
- ☐ De acuerdo
- ☐ Ni de acuerdo ni en desacuerdo
- ☐ En desacuerdo
- ☐ Totalmente en desacuerdo

6- Considera que la metodología propuesta puede ser adecuada para el entorno de una Universidad Nacional. *

- ☐ Totalmente de acuerdo
- ☐ De acuerdo
- ☐ Ni de acuerdo ni en desacuerdo
- ☐ En desacuerdo
- ☐ Totalmente en desacuerdo

7- La documentación descriptiva de la metodología es suficiente. *

- ☐ Totalmente de acuerdo
- ☐ De acuerdo
- ☐ Ni de acuerdo ni en desacuerdo
- ☐ En desacuerdo
- ☐ Totalmente en desacuerdo

8- Observaciones

Tu respuesta

Figura 96: Encuesta de validación de la metodología propuesta (cont.)

Los resultados obtenidos de cada pregunta se observan en las figuras que van desde la Figura 97 hasta la Figura 103.

1-Cómo considera la complejidad de la metodología.

3 respuestas

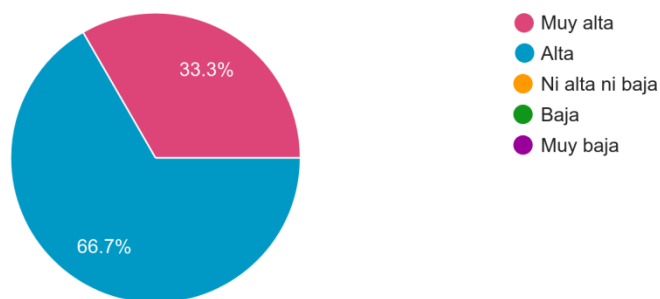


Figura 97: Resultados de la encuesta de validación de la metodología propuesta (Pregunta 1)

2- La definición de las etapas y tareas, son comprensibles.

3 respuestas

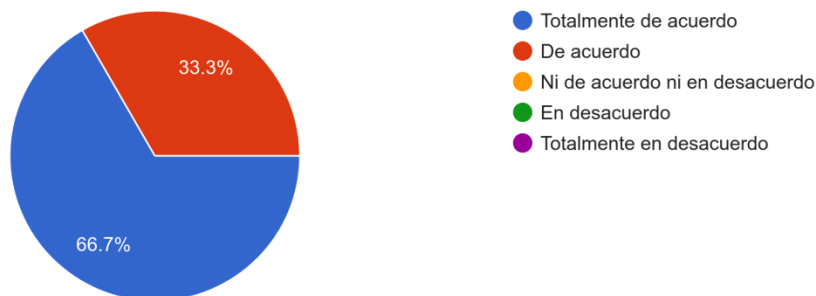


Figura 98: Resultados de la encuesta de validación de la metodología propuesta (Pregunta 2)

3 - Considera que la metodología puede ser adaptada a otras Universidades Nacionales.

3 respuestas

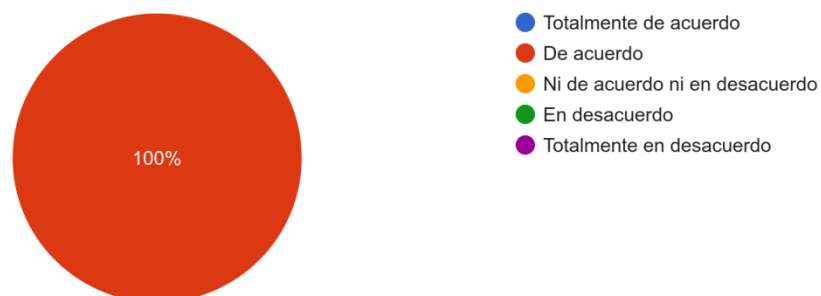


Figura 99: Resultados de la encuesta de validación de la metodología propuesta (Pregunta 3)

4- La metodología es precisa en la definición de sus etapas y tareas.

3 respuestas

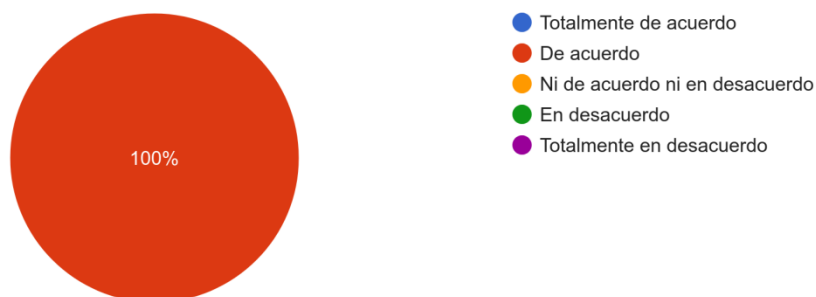


Figura 100: Resultados de la encuesta de validación de la metodología propuesta (Pregunta 4)

5- La definición de Roles y Responsabilidades, se encuentran definidos con claridad.

3 respuestas

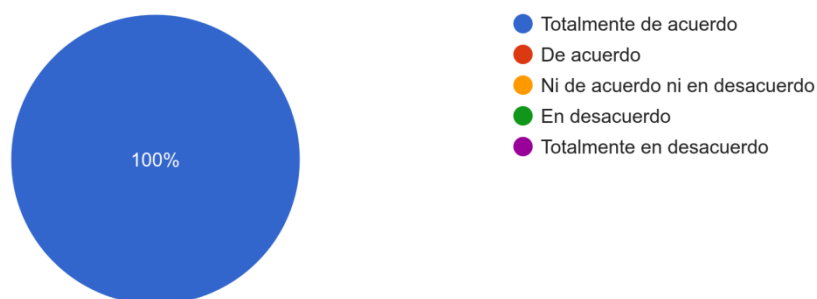


Figura 102: Resultados de la encuesta de validación de la metodología propuesta (Pregunta 5)

6- Considera que la metodología propuesta puede ser adecuada para el entorno de una Universidad Nacional.

3 respuestas

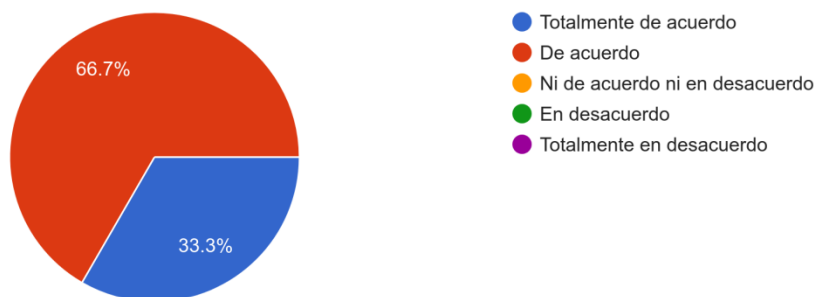


Figura 101: Resultados de la encuesta de validación de la metodología propuesta (Pregunta 6)

7- La documentación descriptiva de la metodología es suficiente.

3 respuestas

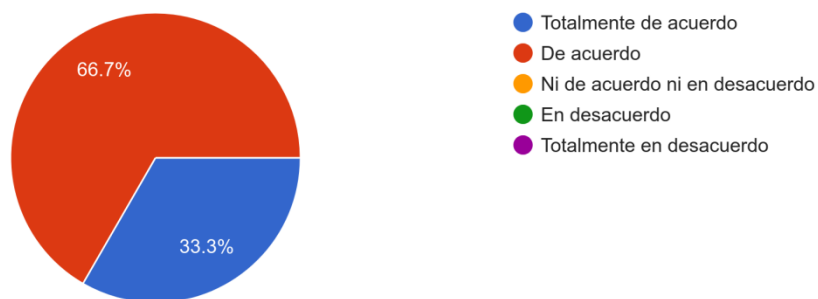


Figura 103: Resultados de la encuesta de validación de la metodología propuesta (Pregunta 7)

Capítulo 7

Discusión

7. Discusión

En esta sección se presentan las interpretaciones de los resultados de la evaluación de la metodología propuesta aplicado al caso de estudio planteado.

Cuando se comenzó con esta investigación el Consorcio SIU no poseía ningún Data Mart desarrollado asociado a graduados. Con los avances en las versiones del sistema SIU-Guarani3 y las demandas recibidas por parte de las distintas universidades nacionales, el SIU incorporó un Data Mart basado en el sistema SIU-Kolla que permite analizar las distintas encuestas realizadas [59].

Este Data Mart, se construye mediante un modelo general y de forma automática de acuerdo a las variables seleccionadas y a la cantidad de respuestas que posee cada encuesta. Siendo su única medida, la cantidad de graduados.

Lo que se buscó en este trabajo es desarrollar una herramienta metodológica para el desarrollo de Data Marts en universidades nacionales del Consorcio SIU, que sirva de guía para el desarrollo de soluciones de Inteligencia de Negocios que se adapten a las necesidades propias de cada institución y que permita obtener indicadores que aporten un mayor nivel de detalle ante un requerimiento planteado por las autoridades. Además, de servir de guía para personalizar un Data Mart existente o adaptarlo a un nuevo contexto.

La incorporación de atributos de calidad a la metodología propuesta permitió obtener un producto final que contempla las expectativas del dominio en base a las necesidades de información de la institución.

La participación activa de los actores involucrados durante todo el proceso permitió contar con el compromiso de los mismos sobre los requisitos relevados. Especialmente en las fases que abarcaron la evaluación de la factibilidad, la verificación y la validación de los modelos diseñados, permitiendo dar un marco de continuidad al desarrollo de cada uno de ellos.

Respecto al cambio realizado en el circuito administrativo para el trámite de un título, el acompañamiento institucional para aplicar dicho cambio fue fundamental. Dicho cambio favoreció la carga de las encuestas por parte de los graduados y disponer de datos actualizados para poblar los Data Marts.

A pesar que el porcentaje de completitud de las encuestas aumentó considerablemente con los cambios realizados en el circuito administrativo, poder alcanzar el 100% de completitud de las mismas en menor tiempo podría requerir de nuevos cambios en los plazos de entrega de los requisitos, dentro del circuito administrativo.

Es importante dar a conocer la importancia que implica contar con una base de conocimiento de los graduados y los beneficios que podrían obtener los mismos al contar con sus datos actualizados, para la implementación de las otras encuestas propuestas por el SIU.

A medida que se realizaron las entrevistas con los expertos, surgieron por cada una de ellas numerosas necesidades de información relacionadas a la temática tratada. Para cubrir esta necesidad se han ido adaptando las plantillas diseñadas inicialmente, de tal manera de poder identificar los interrogantes más relevantes y darles un tratamiento individual a cada uno de ellos.

Cabe mencionar que las plantillas asociadas a la especificación de requisitos de usuarios y definición de temas analíticos implementadas en esta metodología podrían ser incorporadas en la gestión diaria de la DAlyC, dado que es habitual ante un pedido de requerimiento en dicho sector, sean insuficientes los datos de especificación de los mismos (período de tiempo, calidad de un alumno, ¿año calendario o año académico?, etc.), detalles de quién o quienes lo solicitaron, cuando y para qué, imposibilitando hacer un correcto seguimiento del mismo.

La evaluación de la metodología propuesta resultó eficiente y recomendada para cada necesidad objetiva. La incorporación de nuevas necesidades, como fue el análisis de la “trayectoria laboral del graduado” no contemplada en una etapa temprana de desarrollo, y surgida en la última etapa de uno de los proyectos, requirió de un análisis más minucioso con la nueva temática planteada. Esta nueva necesidad surgió durante el proceso de validación del Data Mart asociado a la “Situación laboral” de los graduados, cuyo objetivo era conocer las experiencias, habilidades y prácticas adquiridas por el graduado en base a la carrera estudiada, de manera tal, de poder convocarlos y darles la oportunidad de brindar charlas y capacitaciones a colegas de la región que demanden fortalecer y afianzar los conocimientos adquiridos, en un proceso de formación profesional continua.

En cuanto al tratamiento de los campos descriptos como editables, se analizó otra alternativa de solución: depurar y validar los datos existentes mediante consultas SQL. Siendo ésta, desestimada por la falta de certeza que se detectó a la hora de determinar si el dato informado por el graduado correspondía a la encuesta que se estaba analizando.

Capítulo 8

Conclusiones

8. Conclusiones

En este capítulo se describen las conclusiones alcanzadas en base a los objetivos planteados y las principales contribuciones del trabajo. Finalmente se describen las futuras líneas de investigación.

8.1 Principales contribuciones del trabajo

El desarrollo de sistemas de soporte para decisión basado en herramientas de ingeniería de software, constituyen una deuda pendiente en el ámbito de las universidades nacionales de la República Argentina.

La metodología propuesta de desarrollo de Data Marts descripta en este trabajo, representa un esfuerzo por contribuir al desarrollo de soluciones informáticas de calidad en contextos de universidades nacionales que utilizan herramientas provistas por el SIU.

Esta solución considera las características particulares de las universidades nacionales, utilizando recursos ya disponibles como son: los sistemas informáticos, recursos tecnológicos y estructura organizacional. Esto facilita la adaptación de la misma a otros contextos universitarios, y contribuye a la obtención de los resultados planteados por los interesados.

Para el diseño de la metodología propuesta se opta por HEFESTO como marco de referencia principal, siendo su etapa de análisis de requerimientos un aspecto fundamental tenido en cuenta para la educación de requisitos y la definición de los temas de análisis. El formato propuesto para la definición de los mismos ha contribuido al seguimiento de los requisitos planteados, permitiendo tener de manera breve una descripción de los mismos, quién o quienes lo han solicitado, el objetivo de esa solicitud y el beneficio de lo solicitado.

Por otro lado, se incorpora el concepto de las matrices de bus propuesto por Kimball para identificar dimensiones ya definidas, que puedan ser reutilizadas en otros temas de análisis. Esto ha contribuido a la creación de un repositorio de dimensiones compartidas, que resulta eficiente para cubrir nuevas necesidades y optimizar los tiempos de respuestas.

A la solución propuesta se han incorporado técnicas y herramientas para la gestión de requisitos basadas en las buenas prácticas del modelo CMMI v1.3. La implementación de estas prácticas, ha incrementado la participación de los expertos a lo largo de todo el desarrollo del proyecto.

Esta participación ha facilitado detallar y comprender los requisitos del mismo, como así también validar los resultados obtenidos, contribuyendo al desarrollo de un producto de calidad.

El trabajo en conjunto del analista con los expertos, durante la evaluación del caso de estudio, ha permitido corregir errores e incidentes a tiempo; como así también, detectar en una etapa temprana de la metodología si se cuenta con los datos necesarios para cubrir los requerimientos de los mismos. Esto evita realizar esfuerzos innecesarios por parte de los integrantes del equipo de trabajo.

El uso de instrumentos estructurados para la documentación de los procesos de la metodología propuesta, ha contribuido a definir con mayor claridad cuestiones relacionadas a cada etapa de desarrollo, mejorando la calidad de los procesos y ayudando a mantener la trazabilidad de los requisitos a lo largo de todo el proyecto. Además de constituir un repositorio que pueda ser utilizado para implementar Data Marts diseñados en otras universidades o personalizar alguno existente.

La calidad de los datos en sistemas de apoyo a la toma de decisiones es un componente clave. El uso de campos configurados como editables en herramientas utilizadas para la recolección de datos, impactan negativamente en la calidad de los datos.

Siendo el sistema SIU-Kolla fuente principal de datos, requiere de sistemas fuentes adicionales como el SIU-Guaraní, para el proceso de selección y depuración de los mismos. Dicha depuración requiere procesos ETL complementarios de conexión al sistema SIU-Guaraní para extraer de allí los datos relacionados a los graduados, en base a la encuesta analizada.

Estos procesos requieren mayor atención en etapas de mantenimiento y actualización de los datos o de reingeniería de las encuestas, para mantener la calidad de los datos y contribuir a su posterior reutilización en un sistema de gestión gerencial.

Adaptar un circuito administrativo a las necesidades de una institución, suele ser necesario para contar con mayor cantidad de datos para el análisis. Incorporar como requisito obligatorio el voucher de completitud de la “Encuesta al Recién Graduado” tuvo un impacto efectivo (80%) sobre la cantidad de datos que se generaron y que han servido para los temas de análisis abordados. No obstante se deben establecer mecanismos para asegurar la generación de información constante para realizar seguimientos a lo largo del tiempo, como por ejemplo la implementación de la “Encuesta a los 5 años de Graduado”.

La disponibilidad de herramientas de código abierto contribuye a los objetivos de las universidades nacionales, y considera sus realidades.

Finalmente, es importante destacar que una representación adecuada de los resultados obtenidos, junto a la participación y opinión de los expertos, son factores claves para el éxito de un proyecto de estas características.

8.2 Futuras líneas de investigación

Entre las futuras líneas de investigación se pueden sugerir las siguientes:

- Indagar en el campo de la Ingeniería de Requerimientos Distribuida para ser aplicada en la metodología propuesta, con la finalidad de abordar el desarrollo de Data Marts de manera conjunta con instituciones que se encuentren trabajando en forma descentralizada y/o geográficamente distante.
- Evaluar la metodología propuesta en otros escenarios institucionales, para poder analizar la adaptabilidad de la misma a otros contextos, como ser:
 - Servicio Médico Asistencial
 - Área de Prevención y Atención a las Víctimas de Violencia de Género
 - Comedor, entre otros.
- Adecuar la metodología a procesos de Ciencias de Datos, con la finalidad de extraer patrones y tendencias en forma automática sin intervención de los usuarios sobre el gran volumen de datos que se genera en el ámbito educativo de las universidades nacionales.

Bibliografía

Bibliografía

- [1] Sistema de Informacion Universitaria. (2020) Sitio Oficial - Consorcio SIU. [Online]. <http://www.siu.edu.ar>
- [2] Comision Nacional de Evaluacion y Acreditacion Universitaria. (2020) CONEAU. [Online]. <http://www.coneau.gob.ar/archivos/resoluciones/Res382-11.pdf>
- [3] Facultad de Ciencias Exactas, Quimicas y Naturales. (2019, Diciembre) Plan de Desarrollo Institucional - Digesto - FCEQyN. [Online]. http://digesto.fceqyn.unam.edu.ar/digestopub/documentos/resoluciones/2019/RES_00082019_compressed.pdf
- [4] Facultad de Ciencias Exactas Quimicas y Naturales. (2020) FCEQyN, Digesto. [Online]. <https://www.fceqyn.unam.edu.ar/digesto-convenios/>
- [5] Universidad Nacional de Misiones (UNaM). (2020) Digesto UNaM. [Online]. <https://digesto.unam.edu.ar/index.php/normativa-relevante/resoluciones>
- [6] J. C. Trujillo, J. R. Mazon, and J. Pardillo, *Diseño y Explotación de Almacenes de Datos: conceptos básicos y modelado multidimensional*. Alicante: Club Universitario, 2009, ISBN: 978-84-8454-902-4.
- [7] H. P. Luhn, "A business intelligence system," *IBM Journal of Research and Development*, vol. 2, no. 4, pp. 314-319, October 1958.
- [8] H. Dresner, *Cultura Orientada al Desempeño: El Eslabon Perdido en Inteligencia de Negocios*. Mexico, Mexico: Grupo Patria, 2010, ISBN: 9786074382587.
- [9] R. D. Bernabeu, *Data Warehousing: Investigación y Sistematización de Conceptos - HEFESTO: Metodología propia para la Construcción de un Data Warehouse*. Córdoba, Argentina: Free Software Foundation, 2007.
- [10] S. Ramos, *Microsoft Business Intelligence: vea el cubo medio lleno*. Alicante, España: SolidQ TM Press, 2011, ISBN:978-84-936417-8-8.
- [11] W. H. Inmon, *Building the Data Warehouse*, Cuarta Edicion ed.: Wiley Publishing, Inc., 2005, ISBN:0-7645-9944-5.
- [12] R. Kimball, *The Data Warehouse Toolkit: practical techniques for building dimensional data warehouse.*: John Wiley & Sons, Inc., 1996.
- [13] S. Chaudhuri and U. Dayal, *An Overview of Data Warehousing and OLAP Technology.*: ACM Sigmod record, 1997, vol. 26.

- [14] David F. Nettleton, *Analisis de datos comerciales*, ISBN: 84-7978-593-4, Ed. Madrid, España: Diaz de Santos S.A., 2003.
- [15] C. J. Date, *Introduccion a los Sistemas de Bases de Datos*, Septima Edicion ed., ISBN:9789684444195, Ed. California: Pearson Educación, 2001.
- [16] R. Kimball, M. Ross, W. Thornthwaite, J. Mundy, and B. Becker, *The Data Warehouse Lifecycle Toolkit*, Second Edition ed. Indianapolis, Indiana: Wiley Publishing, Inc, 2008, ISBN: 978-0-470-14977-5.
- [17] D. L. La Red Martinez. (2020, Mayo) Facultad de Ciencias Exactas y Naturales y Agrimensura (UNNE). [Online]. <http://exa.unne.edu.ar/depar/areas/informatica/dad/DAD0.htm>
- [18] D. L. La Red Martinez. (2020, Mayo) Facultad de Ciencias Exactas y Naturales y Agrimensura (UNNE). [Online]. <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/SOF.htm>
- [19] F. P. Brooks, *The Mythical Man–Month: Essays on Software*, Engineering Anniversary Edition ed.: Addison–Wesley, 1995, ISBN 0-201-83595-9.
- [20] B. W. Boehm. (1981) Software Engineering Economics. [Online]. extras.springer.com/2002/978-3-642-59413-7/4/rom/pdf/Boehm_new.pdf
- [21] J. C. Leite, "A Survey on Requirements Analysis," Irvine, California, Technical report 1987.
- [22] Real Academia Española. (2019) Real Academia Española. [Online]. <http://www.rae.es/>
- [23] Institute of Electrical and Electronics Engineers, IEEE Software Engineering Standards Collection, 1997, <https://www.ieee.org/>.
- [24] R. S. Pressman, *Ingenieria del software. Un enfoque práctico.*, Séptima Edición ed.: McGraw-Hill Companies, Inc., 2010, ISBN: 978-607-15-0314-5.
- [25] J. C. Leite, "Software Evolution, The Requirements Engineering View," keynote address de XXVI JAIIO - SoST'97 Simposio en Tecnología de Software, Buenos Aires, Argentina, 1997.
- [26] J. Rumbaugh, "Getting Started: Use Cases to Capture Requirements," <https://www.semanticscholar.org/>, Journal of Object Oriented Programming, Vol.23, pp. 8-12 1994.
- [27] J. C. Leite, Engenharia de Requisitos, 1994, Repositorio Digital Universidade de São Paulo. Teaching material of the Requirements Engineering course.

<https://edisciplinas.usp.br/acessar/>.

- [28] J. Carrol. (1995) Scenario Based Design: Envisioning Work and Technology in System Development. [Online]. <https://dl.acm.org/doi/book/10.5555/209227>
- [29] J. C. Leite, "Eliciting requirements using a natural language based approach: the case of the meeting scheduler problem," Pontificia Universidad Católica de Río de Janeiro, Río de Janeiro - Brasil, Monografias em Ciência da Computação. Nro.13 1993. [Online]. <https://scholar.google.com.ar/>
- [30] P. Sawyer, I. Sommerville, and S. Viller, "Requirements Process Improvement through The Phased Introduction of Good Practice. Software Process – Improvement and Practice," LANCASTER UNIVERSITY, <https://www.lancaster.ac.uk/>, Tecnico Vol 3, Issue 1, pp. 19-34, 1997.
- [31] I. Sommerville and P. Sawyer, *Requirements Engineering: A Good Practice Guide*. New York, NY United States: John Wiley & Sons, Inc. 605 Third Ave., 1997, ISBN:978-0-471-97444-4.
- [32] B. W. Boehm, P. Bose, E. Horowitz, and M. J. Lee, "Software Requirements as Negotiated Win Conditions," *ResearchGate*, pp. 74-84, Mayo 1994.
- [33] K. Pohl, "Requirements Engineering: An Overview," Informatik V, RWTH Aachen, Ahornstr, Aachen - Germany, 1997. [Online]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.67.1394&rep=rep1&type=pdf>
- [34] I. C. Society, P. Bourque, and R. E. Fairley, *Guide to the Software Engineering Body of Knowledge (SWEBOK(R))*, Tercera Edición ed. Los Alamitos, CA, EEUU: IEEE Computer Society Press, 2014.
- [35] Equipo del Producto CMMI, "CMMI para Desarrollo, Versión 1.3. CMMI-DEV, V1.3," Carnegie Mellon University, TECHNICAL REPORT 2010.
- [36] J. E. Torres Romo, "Propuesta metodológica de una solución de inteligencia de negocios aplicada al sistema informático integrado de talento humano y sistema de registro de contratos y actas de finiquito. Caso: Ministerio del Trabajo-MDT," Universidad de las Américas, Quito - Peru, masterThesis 2016. [Online]. [Link: http://dspace.udla.edu.ec/jspui/bitstream/33000/3385/1/UDLA-EC-TMGSTI-2015-16%28S%29.pdf](http://dspace.udla.edu.ec/jspui/bitstream/33000/3385/1/UDLA-EC-TMGSTI-2015-16%28S%29.pdf)
- [37] C. Rojas Muñoz and V. Saquicela Galarza, "Hacia un sistema de ayuda a la decisión para Universidades: Caso de uso de la Universidad de Cuenca," Facultad de Ingeniería,

- Universidad de Cuenca, Cuenca - Ecuador, Tesis de Maestria 2017. [Online]. <https://dspace.ucuenca.edu.ec/bitstream/123456789/28278/3/Trabajo%20de%20tiulaci%C3%B3n.pdf>
- [38] Wilson Bissi, "SCRUM - Metodologia de desenvolvimento agil," *Campo Digital*, vol. 2, no. 1, 2007, Centro Universitário Integrado. Campo Mourão PR/BRASIL.
- [39] E. I. Brizuela and Y. Castro Blanco, "Metodologias para desarrollar Almacen de Datos," *Revista de Arquitectura e Ingeniería. Universidad Autónoma del Estado de México*, vol. 7, no. 3, p. 12, Diciembre 2013. [Online]. <http://www.redalyc.org/pdf/1939/193930080003.pdf>
- [40] S. Luján-Mora and J. Trujillo, "A Data Warehouse Engineering Process," in *Yakhno T. (eds) Advances in Information Systems. ADVIS 2004. Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, 2004, vol. 3261, pp. 14-23. [Online]. https://link.springer.com/chapter/10.1007%2F978-3-540-30198-1_3
- [41] W. Castillo-Rojas, Fernando Medina Q., and Francisco Farina M., "Propuesta de una Metodologia para un Proceso Data WareHousing. Caso de uso: Generacion de Indicadores de productividad Academica de una Universidad," IEEE, 6th International Conference on Software Process Improvement (CIMPS), Zacatecas - Mexico, INSPEC Accession Number: 17417257, Octubre 2017. [Online]. <https://ieeexplore.ieee.org/document/8169943/>
- [42] F. Cuadrado and A. Oliveros, "Elicitacion de requerimientos de stakeholders distribuidos," XXII Congreso Argentino de Ciencias de la Computacion, La Plata - Buenos Aires, Red de Universidades con Carreras en Informática (RedUNCI) ISBN: 978-950-34-1539-9, 2017.
- [43] H. Cobo, S. Baisi, A. M. Canero Pero, and M. Menendez, "Implementacion de soluciones OLAP en Universidades - SIU-Wichi," in *46JAIIO -AGRANDA - ISSN:2451-7569*, 2008, p. 9.
- [44] Sistema de Informacion Universitaria. (2020) Boletin Electronico - Comunidad SIU. [Online]. <https://www.siu.edu.ar/infosiu/>
- [45] Sistema de Informacion Universitaria. (2020) Foro - Comunidad SIU. [Online]. <https://portal.comunidad.siu.edu.ar/>
- [46] Sistema de Informacion Universitaria. (2020) SIU-Wichi - Mapa de implementacion. [Online]. <https://www.siu.edu.ar/mapa-de-implementaciones-siu-wichi/>
- [47] Universidad Nacional de Misiones. (2019, Diciembre) Plan de Desarrollo Institucional - Digesto - UNaM. [Online]. <https://digesto.unam.edu.ar/index.php/2019/item/495-resolucion-cs-n-018-19>

- [48] A. Oliveros and R. L. Antonelli, "Tecnicas de elicitation de requerimientos," in *XXI Congreso Argentino de Ciencias de la Computacion (CACIC 2015)*, Buenos Aires, Octubre 2015.
- [49] Facultad de Ciencias Exactas, Químicas y Naturales. (2020, Mayo) Programa de Seguimiento y Asistencia al Graduado - Digesto - FCEQyN. [Online]. http://digestoanterior.exactas-unam.dyndns.org/admin/documentos/2012/disposicion/DIS_1828_2012_05122012.htm
- [50] Facultad de Ciencias Exactas, Químicas y Naturales. (2020, Mayo) Oficina de Vinculación con los Graduados - Digesto - FCEQyN. [Online]. <http://digesto.fceqyn.unam.edu.ar/digestopub/documentos/disposiciones/2019/03862019.PDF>
- [51] Facultad de Ciencias Exactas, Químicas y Naturales. (2019, Oct) Estructura Organica Funcional Administrativa-Tecnica - Digesto - FCEQyN. [Online]. http://digestoanterior.exactas-unam.dyndns.org/admin/documentos/2009/disposicion/DIS_1620_2009_16112009.htm
- [52] Boletín Oficial de la República Argentina. Decreto 792/2020. (2020) [Online]. <https://www.boletinoficial.gob.ar/detalleAviso/primera/235931/20201012>
- [53] Portal Oficial Estado Argentino. (2020, Agosto) Ley 25326 habeas data. [Online]. <https://www.argentina.gob.ar/normativa/nacional/ley-25326-64790>
- [54] Sistema de Información Universitaria. (2020) SIU-Wichi - Guía Técnica 6.7.0. [Online]. <https://documentacion.siu.edu.ar/wiki/SIU-Wichi/Version6.7.0/instalacion>
- [55] Sistema de Información Universitaria. (2017) SIU-Wichi - Personalizaciones. [Online]. <http://documentacion.siu.edu.ar/wiki/SIU-Wichi/Version6.3.1/personalizacion>
- [56] Vantara Hitachi. (2019, Diciembre) Pentaho Platform. [Online]. <https://help.pentaho.com/Documentation/8.3/Products>
- [57] Universidad Nacional de Misiones. (2020) Resolución 533/20. [Online]. <https://www.fceqyn.unam.edu.ar/suspension-de-actividades-unam-hasta-el-31-de-diciembre/>
- [58] R. Likert, A technique for the measurement of attitudes, 1932, Archives of psychology, 140, 5-55.
- [59] Sistema de Información Universitaria. (2020) SIU-Wichi - Version 6.4.0. Documentación Wichi-Kolla. [Online]. http://documentacion.siu.edu.ar/wiki/SIU-Wichi/Version6.4.0/Documentacion_wichikolla

- [60] redmine.org. (2019) Overview Redmine. [Online]. <http://www.redmine.org>
- [61] Sistema de Informacion Universitaria. (2019) SIU-Kolla - Modelo de Datos 3.7.0. [Online]. [/SIU-Kolla/Versión 3.7/Documentación](#)
- [62] Sistema de Informacion Universitaria. (2020) SIU-Guarani - Portal de Ayuda. [Online]. <http://portalguarani.siu.edu.ar/trac/Portal-G3/wiki/G2/Versiones>

Anexos

Anexo 1

El Anexo 1 presenta el Modelo de la Encuesta al Recién Graduado²² diseñado por el SIU.

ENCUESTA DE RECIÉN GRADUADO

Bloque de Preguntas: DATOS PERSONALES

- Nombre y Apellido (Campo editable)
- Documento, Tipo (Campo autocompletable)
- Número de Documento (Campo editable numérico)
- Fecha de Nacimiento (Campo editable)
- Sexo (Campo autocompletable)
- Nacionalidad (Campo autocompletable)
- Correo electrónico (Campo editable)
- Estado civil (Campo autocompletable)
- Cantidad de hijos (Campo autocompletable)
- Grupo de convivencia (Campo autocompletable)
- Título (Campo editable)
- Año de Ingreso (Campo editable)

Bloque de Preguntas: RESIDENCIA

- Ciudad (Campo editable)
- Provincia (Campo editable)
- País (Campo editable)

Bloque de Preguntas: PROCEDENCIA

- Ciudad (Campo editable)
- Provincia (Campo editable)
- País (Campo editable)

Bloque de Preguntas: HÁBITOS

- Hábito de lectura (Campo autocompletable)
- Tipo de lectura (Campo autocompletable)

²² Encuesta Recién graduado (http://documentacion.siu.edu.ar/wiki/SIU-Kolla/Version_4.0.0/Encuestas_precargadas/Encuestas_de_seguimiento_de_graduados/Encuesta_al_recien_graduado)

- Otros hábitos (Campo autocompletable)

Bloque de Preguntas: NIVEL DE ESTUDIO ALCANZADO POR

- Padre (Campo autocompletable)
- Madre (Campo autocompletable)
- Pareja (Campo autocompletable)

Bloque de Preguntas: DATOS LABORALES

- Condición actual (Campo autocompletable)
- ¿Busca trabajo? (Campo autocompletable)
- Expectativas laborales para los próximos 6 meses (Campo autocompletable)
- ¿Qué características de los graduados universitarios de su área cree usted que son valoradas por los empleadores? (Campo autocompletable)
- En caso de haber seleccionado Otras capacidades, mencione cual (Campo editable)
- Mencione qué organismos, empresas o instituciones visualiza como fuente de trabajo de su área para egresados de la Universidad (Campo editable)
- ¿Le interesaría recibir información de la bolsa de trabajo de la Universidad? (Campo autocompletable).

Bloque de Preguntas: TRABAJO ACTUAL

- ¿Recibe remuneración? (Campo autocompletable)
- Si recibe remuneración ¿realiza aportes jubilatorios? (Campo autocompletable)
- ¿Qué tipo de relación laboral tiene? (Campo autocompletable)
- ¿Que tipo de contratación posee? (Campo autocompletable)
- Cantidad de horas semanales de trabajo (Campo autocompletable)
- ¿Está satisfecho con su trabajo? (Campo autocompletable)
- La entidad donde trabaja es.. (Campo autocompletable)
- Con relación a su trabajo actual y por haberse recibido... (Campo autocompletable)

Bloque de Preguntas: SINO TRABAJA

- Razón por la que no trabaja (Campo autocompletable)

Bloque de Preguntas: EN CASO DE ESTAR BUSCANDO TRABAJO

- Sector preferido (Campo autocompletable)
- Actividad preferida (Campo autocompletable)
- Exigencias con respecto al lugar de trabajo (Campo autocompletable)

Bloque de Preguntas: RELACIÓN CON LA UNIVERSIDAD

- Motivos por los cuales eligió su carrera universitaria (Campo autocompletable)
- ¿Por qué eligió estudiar en esta Universidad? (Campo autocompletable)
- Los contenidos estudiados durante su carrera fueron interesantes (Campo autocompletable)
- Los contenidos estudiados durante su carrera cumplieron sus expectativas (Campo autocompletable)
- Nivel de exigencia de la carrera (Campo autocompletable)
- ¿El plan de estudios debería ser más corto? (Campo autocompletable)
- ¿El plan de estudios debería ser más largo? (Campo autocompletable)
- ¿El plan de estudios debería ser un perfil más técnico? (Campo autocompletable)
- ¿El plan de estudios debería ser más generalista? (Campo autocompletable)

Bloque de Preguntas: SU RELACIÓN EN LA UNIVERSIDAD CON RELACIÓN A:

- Alumnos (Campo autocompletable)
- Aulas (Campo autocompletable)
- Biblioteca (Campo autocompletable)
- Laboratorio (Campo autocompletable)
- Administración (Campo autocompletable)
- ¿Volverías a estudiar en la Universidad Pública? (Campo autocompletable)
- ¿Cuáles son los aspectos de la Universidad que usted más aprecia con relación a la carrera que estudió? (Campo autocompletable)
- ¿Cuál es su percepción del reconocimiento de la sociedad a esta Universidad? (Campo autocompletable)
- ¿Qué desearía que le brinde esta Universidad? (Campo autocompletable)

Bloque de Preguntas: OTROS ESTUDIOS

- Aparte de la carrera universitaria que finalizó ¿realizó otros estudios en educación superior? (Campo autocompletable)
- Tipo de institución (Campo autocompletable)
- Nombre de la carrera (Campo editable)
- ¿Finalizó la carrera? (Campo autocompletable)
- Conocimientos de Inglés (Campo autocompletable)
- Conocimientos de Francés (Campo autocompletable)
- Conocimientos de Alemán (Campo autocompletable)

- Conocimientos de Italiano (Campo autocompletable)
- Conocimientos de Portugués (Campo autocompletable)
- Conocimientos de Herramientas informáticas (Campo autocompletable)
- ¿Piensa seguir estudiando? (Campo autocompletable)
- Si piensa seguir estudiando ¿qué le gustaría hacer? (Campo autocompletable)
- ¿Dónde cree que lo hará? (Campo autocompletable)
- Si va a estudiar en otra universidad o en el extranjero ¿por qué? (puede marcar más de una opción) (Campo autocompletable)
- ¿Estaría interesado en cursos de formación continua? (Campo autocompletable)

Anexo 2

El Anexo 2 presenta el presupuesto del servidor adquirido por la FCEQyN (Figura 104).



Junin 2550 Tel./Fax (376)4429000
http://www.electro.com.ar
ELECTRO MISIONES S.A. C.U.I.T.: 30-61540204-0
CBU: CITI BANK 0168888100008278200166

Proforma Número: 0001-00375545

Hoja: 1
Fecha: 14/04/2020
Hora: 13:09:46

Comprobante NO Válido como Factura

Señor/es : FAC.CIENC.EXACTAS,QUIM.Y NAT-UNAM
Domicilio: F. DE AZARA N° 1552 2° PISO Piso 0
Localidad: (3300) POSADAS, MISIONES (ARGENTINA) - TELEFONO : -0376-4435088
e-mail: -

IVA :Exento
CUIT :30-59821809-5

Código	Descripción	Marca	UM	Cant	Entr.	P.Unit	P.Total	D.
1-C	SUPERMICRO CHASIS 1U CSE-813MF2QC-R804CB		UNI.	1.00	Inm.	90327.000	90327.000	
2-C	MB SUPERMICRO 2011 X10DRL-I		UNI.	1.00	Inm.	49498.000	49498.000	
3-C	MEM DDR4 ECC RDIMM 32GB 2400MHZ SAMSUNG		UNI.	2.00	Inm.	15598.000	31196.000	
4-C	CPU INTEL 2011 XEON E5-2683 V4		UNI.	1.00	Inm.	110321.000	110321.000	
5-C	HDE 3.5 SATA3 7.2K 1TB 128MB 512N ST1000NM000A		UNI.	4.00	Inm.	11913.010	47652.040	
6-C	SUPERMICRO HEATSINK 1U 2011 SNK-P0047PD		UNI.	1.00	Inm.	3051.000	3051.000	

C.S:9661875487519898799979-929:9a9e9f5d-n5a:e8df809r9p90

TOTAL \$ 332045.04

El descuento ya se encuentra incluido en el Precio Unitario.
Estos Precios ya incluyen el I.V.A.
Los Precios arriba cotizados son indicativos, en caso de Licitación / Concurso / Compulsas; estos serán revalorizados.

Forma de Pago : Según Pliego
Las condiciones comerciales aquí expuestas serán aplicadas siempre y cuando el Cliente abone la operación dentro del plazo establecido y de la manera prevista en esta Proforma. Para el caso de operaciones en Cta. Cte., las condiciones válidas serán las vigentes en la misma.

Mantenimiento de Oferta : Según Pliego Realizo : DEANDREA, FERNANDO Interno : 273
 Plazo de entrega : Según Pliego E-mail : FDeAndrea@electro.com.ar

Observaciones : ENTREGA: 7/10 DIAS

COMUNICADO: Electro Misiones S.A. informa que únicamente tomará como válido el e-mail que sea recepcionado y/o enviado de los dominios @electromisiones.com.ar y @electro.com.ar. No será válido ningún e-mail enviado y/o recepcionado en otro dominio, por ejemplo de @hotmail.com, @yahoo.com.ar u otros.

Cualquier inquietud o comentario enviar un E-Mail a InfoGanar@ElectroMisiones.com.ar www.electromisiones.com.ar

Aceptamos la Cotización Numero 0001-00375545 del 14/04/2020, sirviendo el presente de ORDEN DE COMPRA por Artículos, Precios y Condiciones descriptas en la misma, tomando como fecha de origen la aclarada junto a la Firma.

_____ Fecha	_____ Firma	_____ Aclaración	_____ Documento
----------------	----------------	---------------------	--------------------

Figura 104: Presupuesto del servidor adquirido por la FCEQyN.

Anexo 3

El Anexo 3 presenta la definición de las Tablas de Dimensiones correspondientes a los TA 01 y TA02.

Definición de TABLAS de DIMENSIONES			
Tema analítico	TA01-Situación laboral del graduado por carrera según su distribución geográfica.		
Perspectiva	Carrera		
Dimensión	d_carrera		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
carrera_id	Integer	sga_carreras	Carrera
nom_carrera	Character varying	sga_carreras	Nombre
unidad_academica	Character varying	sga_carreras	Nombre

Tema analítico	TA01-Situación laboral del graduado por carrera según su distribución geográfica.		
Perspectiva	Residencia		
Dimensión	d_residencia		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
ciudades_id	Integer		
nom_ciudad_res	Character varying	mug_localidades	Nombre
nom_provincia_res	Character varying	mug_provincias	Nombre
nom_pais_res	Character varying	mug_paises	Nombre
nom_dpto_res	Character varying	mug_dptos_partidos	Nombre

Tema analítico	TA01-Situación laboral del graduado por carrera según su distribución geográfica.		
Perspectiva	Procedencia		
Dimensión	d_procedencia		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
ciudadproc_id	Integer		
nom_ciudad_proc	Character varying	mug_localidades	Nombre
nom_provincia_proc	Character varying	mug_provincias	Nombre
nom_pais_proc	Character varying	mug_paises	Nombre
nom_dpto_proc	Character varying	mug_dptos_partidos	Nombre

Tema analítico	TA01-Situación laboral del graduado por carrera según su distribución geográfica.		
Perspectiva	Satisfacción laboral		
Dimensión	d_satisf_trabajo		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
satisf_trabajo_id	Integer		
satisfaccion_trabajo	Character varying	recién_graduados	satisfecho_trabajo

Tema analítico	TA01-Situación laboral del graduado por carrera según su distribución geográfica.		
Perspectiva	Situación laboral		
Dimensión	d_situacion_laboral		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
sit_lab_id	Integer		
sit_laboral	Character varying	recién_graduado	cond_actual

Tema analítico	TA01-Situación laboral del graduado por carrera según su distribución geográfica.		
Perspectiva	Rango horas semanales		
Dimensión	d_cant_hs_sem		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
cant_hs_sem_id	Integer		
cant_hs_semanal	Character varying	recién_graduado	cant_hs_semanal

Tema analítico	TA01-Situación laboral del graduado por carrera según su distribución geográfica.		
Perspectiva	Tipo entidad laboral		
Dimensión	d_entidad_trabaja		
Campo Dimensión	Tipo Dimensión	Tabla origen	Atributo origen
entidad_id	Integer		
entidad_trabaja	Character varying	recién_graduados	entidad_trabaja

Tema analítico	TA01-Situación laboral del graduado por carrera según su distribución geográfica.		
Perspectiva	Tiempo		
Dimensión	d_tiempo		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
tiempo_id	Date		
mes	Character varying	recién_graduados	fecha_fin_enc
anio	Integer	recién_graduados	fecha_fin_enc

Tema analítico	TA01-Situación laboral del graduado por carrera según su distribución geográfica.		
Perspectiva	Sexo		
Dimensión	d_sexo		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
sexo_id	Integer		
sexo	Character varying	recién_graduados	sexo

Tema analítico	TA01-Situación laboral del graduado por carrera según su distribución geográfica.		
Perspectiva	Estado civil		
Dimensión	d_estado_civil		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
estadocivil_id	Integer		
estado_civil	Character varying	recién_graduados	estado_civil

Tema analítico	TA01-Situación laboral del graduado por carrera según su distribución geográfica.		
Perspectiva	Grupo convivencia		
Dimensión	d_grupo_conv		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
grupoconv_id	Integer		
grupo_conv	Character varying	recién_graduados	grupo_conv

Tema analítico	TA01-Situación laboral del graduado por carrera según su distribución geográfica.		
Perspectiva	Rango etario		
Dimensión	d_redad		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
edad_id	Integer	sga_personas	fecha_nacimiento
rango_edad	Character varying		

Tema analítico	TA02- Valoración académica de los graduados por carrera.		
Perspectiva	Carrera		
Dimensión	d_carrera		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
carrera_id	Integer	sga_carreras	Carrera
nom_carrera	Character varying	sga_carreras	Nombre

Tema analítico	TA02- Valoración académica de los graduados por carrera.		
Perspectiva	Nivel exigencia		
Dimensión	d_nivel_exigencia		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
nivel_exig_id	Integer		
nivel_exig	Character varying	recién_graduado	nivel_exigencia_carrera

Tema analítico	TA02- Valoración académica de los graduados por carrera.		
Perspectiva	Selección carrera		
Dimensión	d_selec_carrera		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
selec_carrera_id	Integer		
selec_carrera	Character varying	recién_graduado	motivo_elec_carrera

Tema analítico	TA02- Valoración académica de los graduados por carrera.		
Perspectiva	Expectativa contenido		
Dimensión	d_expec_cont		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
expec_cont_id	Integer		
expec_cont	Character varying	recién_graduados	cont_carrera_expec

Tema analítico	TA02- Valoración académica de los graduados por carrera.		
Perspectiva	Interes contenido		
Dimensión	d_interes_cont		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
interes_cont_id	Integer		
interes_cont	Character varying	recién_graduado	cont_carrera_interes

Tema analítico	TA02- Valoración académica de los graduados por carrera.		
Perspectiva	Plan corto		
Dimensión	d_plan_corto		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
plan_corto_id	Integer		
plan_corto	Character varying	recién_graduado	plan_estudio_corto

Tema analítico	TA02- Valoración académica de los graduados por carrera.		
Perspectiva	Selección carrera		
Dimensión	d_selec_carrera		
Campo Dimensión	Tipo Dimensión	Tabla origen	Atributo origen
selec_carrera_id	Integer		
selec_carrera	Character varying	recién_graduados	motivo_elec_carrera

Tema analítico	TA02- Valoración académica de los graduados por carrera.		
Perspectiva	Tiempo		
Dimensión	d_tiempo		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
tiempo_id	Date		
mes	Character varying	recién_graduados	fecha_fin_enc
anio	Integer	recién_graduados	fecha_fin_enc

Tema analítico	TA02- Valoración académica de los graduados por carrera.		
Perspectiva	Sexo		
Dimensión	d_sexo		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
sexo_id	Integer		
sexo	Character varying	recién_graduados	sexo

Tema analítico	TA02- Valoración académica de los graduados por carrera.		
Perspectiva	Valoracion universidad		
Dimensión	d_valor_univ		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
valor_univ_id	Integer		
valor_univ	Character varying	recién_graduados	aspectos_univer_aprecia

Tema analítico	TA02- Valoración académica de los graduados por carrera.		
Perspectiva	Plan técnico		
Dimensión	d_plan_tecnico		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
plan_tecnico_id	Integer		
plan_tecnico	Character varying	recién_graduados	plan_estudio_tecnico

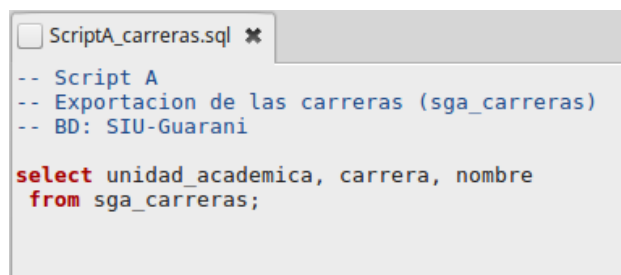
Tema analítico	TA02- Valoración académica de los graduados por carrera.		
Perspectiva	Plan generalista		
Dimensión	d_plan_gen		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
plan_gen_id	Integer		
plan_gen	Character varying	recién_graduados	plan_estudio_gral

Tema analítico	TA02- Valoración académica de los graduados por carrera.		
Perspectiva	Universidad sociedad		
Dimensión	d_valor_univ_socie		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
valor_univ_socie_id	Integer		
valor_univ_socie	Character varying	recién_graduados	reconom_univer_socie

Tema analítico	TA02- Valoración académica de los graduados por carrera.		
Perspectiva	Aporte universidad		
Dimensión	d_aporte_univ		
Campo Dimensión	Tipo	Tabla origen	Atributo origen
aporte_univ_id	Integer		
aporte_univ	Character varying	recién_graduados	brindar_univer

Anexo 4

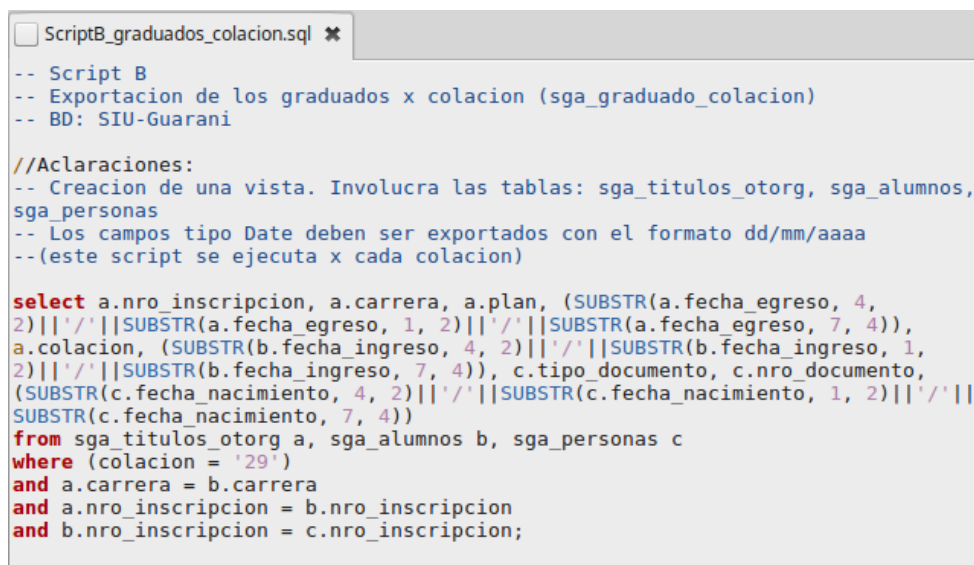
El Anexo 4 presenta las consultas SQL generadas para la extracción de los datasets (A, B, C, D y F) de los sistemas fuentes SIU-Guaraní y SIU-Kolla. Estas consultas se observan en las figuras que van desde la Figura 105 hasta la Figura 109.



```
ScriptA_carreras.sql ✖
-- Script A
-- Exportacion de las carreras (sga_carreras)
-- BD: SIU-Guarani

select unidad_academica, carrera, nombre
from sga_carreras;
```

Figura 105: Consulta SQL para el *dataset A*.



```
ScriptB_graduados_colacion.sql ✖
-- Script B
-- Exportacion de los graduados x colacion (sga_graduado_colacion)
-- BD: SIU-Guarani

//Aclaraciones:
-- Creacion de una vista. Involucra las tablas: sga_titulos_otorg, sga_alumnos,
sga_personas
-- Los campos tipo Date deben ser exportados con el formato dd/mm/aaaa
--(este script se ejecuta x cada colacion)

select a.nro_inscripcion, a.carrera, a.plan, (SUBSTR(a.fecha_egreso, 4,
2)||'/'||SUBSTR(a.fecha_egreso, 1, 2)||'/'||SUBSTR(a.fecha_egreso, 7, 4)),
a.colacion, (SUBSTR(b.fecha_ingreso, 4, 2)||'/'||SUBSTR(b.fecha_ingreso, 1,
2)||'/'||SUBSTR(b.fecha_ingreso, 7, 4)), c.tipo_documento, c.nro_documento,
(SUBSTR(c.fecha_nacimiento, 4, 2)||'/'||SUBSTR(c.fecha_nacimiento, 1, 2)||'/'||
SUBSTR(c.fecha_nacimiento, 7, 4))
from sga_titulos_otorg a, sga_alumnos b, sga_personas c
where (colacion = '29')
and a.carrera = b.carrera
and a.nro_inscripcion = b.nro_inscripcion
and b.nro_inscripcion = c.nro_inscripcion;
```

Figura 106: Consulta SQL para el *dataset B*.

```
*ScriptC-localidades.sql ✖
-- Script C
-- Exportacion de las localidades (mug_paises, mug_localidades,
-- mug_dptos_partidos, mug_provincias)
-- BD: SIU-Kolla

SELECT
    mug_localidades.localidad,
    mug_localidades.nombre,
    mug_localidades.dpto_partido,
    mug_dptos_partidos.nombre,
    mug_dptos_partidos.provincia,
    mug_provincias.nombre,
    mug_provincias.pais,
    mug_paises.nombre
FROM
    kolla.mug_paises,
    kolla.mug_localidades,
    kolla.mug_dptos_partidos,
    kolla.mug_provincias
WHERE
    mug_localidades.dpto_partido = mug_dptos_partidos.dpto_partido AND
    mug_dptos_partidos.provincia = mug_provincias.provincia AND
    mug_provincias.pais = mug_paises.pais and
    (mug_paises.nombre = 'Argentina' or mug_paises.nombre = 'Brasil');
```

Figura 107: Consulta SQL para el *dataset* C.

```
ScriptD_colaciones.sql ✖
-- Script D
-- Exportacion de las colaciones (sga_colaciones)
-- BD: SIU-Guarani

select nro_colacion, descripcion, anio colacion
from sga_colaciones;
```

Figura 108: Consulta SQL para el *dataset* D.

```
ScriptF_planes_estudios.sql ✖
-- Script F
-- Exportacion planes de estudio por carrera
-- BD: SIU-Guarani

select unidad_academica, carrera, plan, vigencia_minima
from sga_planes
where estado = 'V'
order by carrera, plan
```

Figura 109: Consulta SQL para el *dataset* F.

Glosario de acrónimos

Acrónimo	Significado
A	Actividad
ACM	Association for Computing Machinery (Asociación para Maquinaria de Computación)
BI	Business Intelligence (Inteligencia de Negocios)
CI	Conformación de Indicadores
CMMI	Capability Maturity Model Integration (Integración del Modelo de Madurez de Capacidad)
CONEAU	Comisión Nacional de Evaluación y Acreditación Universitaria
DAIyC	Dirección Área de Informática y Comunicaciones
DD	Diccionario de Datos
DERU	Documento de Especificación de Requisitos de Usuario
DM	Data Mart (Almacén de Datos)
DSA	Data Staging Area (Área de Almacenamiento Temporal)
DSS	Decision Support Systems (Sistemas de Soporte a la Decisión)
DTD	Definición de Tablas de Dimensiones
DTH	Definición de Tablas de Hechos
DW	Data Warehouse (Almacén de Datos)
DWEP	Data Warehouse Engineering Process (Proceso de Ingeniería de Almacenamiento de Datos)
E	Etapas
EIS	Executive Information Systems (Sistemas de Información Ejecutiva)
ETL	Extract, Transform and Load (Extracción, Transformación y Carga)
FCEQyN	Facultad de Ciencias Exactas, Químicas y Naturales
FUDA	Documento de Identificación de las Fuentes de Datos

HOLAP	Hybrid Online Analytical Process (Procesamiento Analítico en línea Híbrido)
IEEE	Institute of Electrical and Electronics Engineers (Instituto de Ingenieros Eléctricos y Electrónicos)
MOLAP	Multidimensional On Line Analytic Processing (Procesamiento Analítico Multidimensional en línea)
OLAP	On-Line Analytical Processing (Procesamiento Analítico en Línea)
OLTP	OnLine Transaction Processing (Procesamiento de Transacciones en Línea)
PDI	Programa de Desarrollo Institucional
ROLAP	Relational On Line Analytic Processing (Procesamiento Analítico en línea Relacional)
SIU	Sistema de Información Universitaria
SQL	Structured Query Language (Lenguaje de Consulta Estructurado)
TA	Tema Analítico
UNaM	Universidad Nacional de Misiones