

Universidad Nacional de Misiones

Facultad de Ciencias Exactas, Químicas y Naturales

Tesis de Maestría en Tecnologías de la Información

**Replicación de datos para dar soporte al Programa
de Expansión Territorial de la Universidad Nacional
de Misiones**

**Caso de validación: Legajo Digital de la Facultad de
Ciencias Económicas (LeDig)**

Autor: Lic. Jesús Germán Andrés Pautsch

Director: Dr. Silvio Gonnet

Co-Director: Dr. Eduardo Zamudio

Año 2019

Dedicatorias

*A mi familia,
por su ayuda y sacrificio, en el comienzo de mis estudios.*

*A mi amada esposa y a mis tres pequeñas: Abril, Sofia y Delfina ,
por regalarme sus sonrisas y su ternura a pesar de mis ausencias.*

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

*Para todos los que en busca del equilibrio y entienden que ningún éxito profesional,
justifica un fracaso familiar.*

Resumen

La Universidad Nacional de Misiones (UNaM) mantiene el “Programa de Expansión Territorial”. Este programa tiene impacto en los procesos de gestión, y en consecuencia en los sistemas de información. Esta situación obliga a redefinir algunos de estos procesos y a establecer estrategias TICs de acompañamiento. En este trabajo de tesis se desarrolla e implementa una arquitectura basada en replicación de datos, para mejorar el modo en que se almacena, busca, actualiza la información en las distintas dependencias de la UNaM. Se propone una alternativa a la actual arquitectura cliente servidor, para permitir asegurar la continuidad de las operaciones. Como caso de validación se aborda la gestión de los legajos digitales de los alumnos (sistema LeDig). Inicialmente se resuelve la digitalización de los legajos físicos para luego replicarlos y así asegurar y agilizar el acceso en las distintas sedes. Este proceso ha tenido un impacto directo para la recuperación gradual del espacio físico, crítico en la actualidad. La utilización del LeDig ha permitido aumentar la productividad de los sectores que gestionan y administran los legajos en las distintas sedes de la UNaM, así como también disminuir costos relacionados con la compra de activos como: muebles archivadores, librería y papelería.

Palabras claves: legajos, alumnos, digitalización, replicación de datos, sistemas distribuidos.

Abstract

The National University of Misiones (UNaM) maintains the “Territorial Expansion Program” that has an impact on the management processes, and consequently on the information systems. This situation forces us to redefine some of these processes and to establish ICT support strategies. In this thesis work an architecture based on data replication is developed and implemented so as to improve the way in which the information is stored, searched and updated in the different areas of the UNaM. In order to ensure the continuity of operations, an alternative to the current client server architecture is proposed. As a case of validation, the management of students' digital files (LeDig system) is addressed. Initially, the digitalization of the physical files is resolved replicating and ensuring the access at different locations. This process has had a direct impact on the gradual recovery of physical space, which is a critical issue today. The use of the LeDig has allowed the increment of the productivity of the sectors that manage and administer the files in the different headquarters of the UNaM, as well as to reduce the costs related to the purchase of assets such as: filing cabinets, stationery and paper work.

Keywords: files, students, digitization, data replication, distributed systems.

Agradecimientos

A mis tutores, por brindarme su tiempo y conocimientos para el desarrollo del presente trabajo.

A las autoridades de la Facultad de Ciencias Económicas y a mis compañeras de la Dirección de Estadísticas por permitirme realizar este trabajo y brindarme su apoyo en todo momento.

A mis compañeros de maestría con los cuales fuimos superando los distintos desafíos.

Al equipo técnico de la Facultad de Ciencias Económicas y de la Facultad de Ciencias Exactas Químicas y Naturales quienes me han asistido en las implementaciones y las pruebas realizadas.

A todos los docentes de la maestría, tanto de la UNNE como a los de la UNaM, sobre todo a los que con la creación de la Maestría en Tecnología de la Información nos han dado la posibilidad de perfeccionarnos sin estar lejos de nuestras familias.

A la UNaM que me ha formado como alumno y dado la oportunidad de desarrollarme como profesional. Allí tengo amigos, colegas y gente que, sin duda, ha marcado mi vida.

Índice

1.	Introducción.....	17
2.	Marco teórico	25
2.1.	Arquitecturas centralizadas.....	26
2.2.	Arquitecturas cliente servidor.....	27
2.3.	Arquitectura de sistemas paralelos	28
2.4.	Arquitectura Distribuida	30
2.4.1.	Almacenamiento distribuido de datos	34
2.4.2.	Reglas de Date.....	34
2.4.3.	Replicación de datos.....	35
2.4.4.	Tipos de replicación	37
2.4.5.	Técnicas distribuidas	38
2.4.6.	Fragmentación.....	39
2.5.	Digitalización de documentos.....	42
2.5.1.	Porque digitalizar documentos	42
2.5.2.	Antecedentes sobre digitalización	44
2.5.4.	Proceso de digitalización.....	49
2.5.5.	Firma digital	52
2.5.6.	Normas ISO para la gestión de documentos electrónicos	56
2.6.	Antecedentes sobre replicación de datos	59
3.	Gestión de Legajos. Situación Actual en la Facultad de Ciencias Económicas. 67	
3.1.	Características de la Institución.....	67
3.1.1.	Objetivos de la Institución.....	68
3.1.2.	Programa de Expansión Territorial	69

3.2.	Arquitectura Actual de la Gestión de Legajos.....	71
3.3.	Circuito Administrativo Relevado.....	72
3.4.	Resumen de la problemática.....	74
4.	Modelado de la solución	79
4.1.	Recursos disponibles	79
4.2.	Metodología.....	80
4.3.	Circuito administrativo propuesto	82
4.4.	Análisis sobre las herramientas de replicación.....	86
4.4.1.	Elección de la herramienta de replicación.....	88
4.5.	Almacenamiento de documentos en <i>PostgreSQL</i>	91
4.6.	Arquitectura de datos para el legajo digital	95
4.7.	Consideraciones sobre seguridad.....	96
5.	Desarrollo e implementación.....	101
5.1.	Requerimientos de software y hardware.....	101
5.2.	Modelo de Datos LeDig	102
5.3.	Desarrollo de software.....	107
5.4.	Sistema de Archivos	108
5.5.	Implementación de la Replicación.....	112
5.5.1.	Selección de las tablas a replicar.....	114
5.5.2.	Creación del <i>clúster</i> con <i>slonik</i>	116
5.5.3.	Activación del <i>clúster</i> con <i>slon</i>	119
5.6.	Sistema LeDig	121
5.6.1.	Login LeDig	121
5.6.2.	Búsqueda del legajo del alumno.....	123
5.6.3.	Interfaz con el SIU Pre Inscripción	123

5.6.4.	Digitalización de la documentación	125
5.6.5.	Gestión de legajos digitales	126
5.7.	Consideraciones sobre la replicación.....	129
5.8.	Seguridad y Auditoría.....	132
5.8.1.	Comunicaciones	132
5.8.2.	Perfiles Funcionales	133
5.8.6.	Firma digital	139
6.	Resultados	143
6.1.	Digitalización de la información a replicar.	143
6.2.	Liberación de espacio físico	145
6.3.	Proyecto, desarrollo e implementación.....	147
6.4.	Replicación de datos	150
6.5.	Alta disponibilidad de la información	153
6.6.	Resguardo y recuperación.....	157
6.7.	Agilidad en el acceso, búsqueda y actualización de la información.....	158
6.8.	Interfaz SIU Guaraní y SIU Pre Inscripción.....	160
7.	Discusión	165
8.	Conclusiones	175
8.1.	Objetivos alcanzados	175
8.2.	Contribuciones adicionales.....	176
8.3.	Limitaciones detectadas.....	177
8.4.	Futuras líneas de investigación	178
	Bibliografía.....	181
	Anexo I. Creación del proyecto LeDig	191
	Creación del proyecto LeDig	191

Definición de las fuentes de datos	193
Diseño de las operaciones del sistema	195
Interfaz con el sistema SIU Guaraní.....	199

Índice de Figuras

Figura 1. Arquitecturas centralizadas.....	26
Figura 2. Arquitecturas Cliente-Servidor	28
Figura 3. Arquitectura de Sistemas Paralelos.....	29
Figura 4. Arquitectura Distribuida	31
Figura 5. Fragmentación horizontal	40
Figura 6. Fragmentación vertical.....	41
Figura 7. Fragmentación mixta	41
Figura 8. Normas ISO para la Gestión de Documentos Electrónicos	57
Figura 9. Presencia de la Universidad Nacional de Misiones en la provincia de Misiones	70
Figura 10. Arquitectura actual sobre la gestión de legajos.....	72
Figura 11. Circuito administrativo actual sobre la gestión de legajos.....	74
Figura 12. Criticidad en el espacio físico	76
Figura 13. Propuesta de circuito administrativo.....	84
Figura 14. Carga de requisitos cumplidos en SIU Guaraní.....	85
Figura 15. Arquitectura de datos	96
Figura 16. Esquema de seguridad del sistema LeDig	97
Figura 17. Modelo de datos del sistema LeDig.....	103
Figura 18. Estructura del sistema de archivos LeDig.....	108
Figura 19. Sincronización del sistema de archivos	110
Figura 20. Sistema de archivos. Activación manual	111
Figura 21. Método desacarga_doc.....	112
Figura 22. <i>Application Stack Builder</i> . Lista de complementos compatibles.....	113
Figura 23. Verificación de la creación del <i>clúster</i> de replicación	118
Figura 24. Login del sistema LeDig.....	122
Figura 25. Pantalla principal del sistema LeDig	122
Figura 26. Pantalla de búsqueda de legajos digitales	123
Figura 27. Recuperación de registros de pre inscripción	124
Figura 28. Interfaz entre los sistemas LeDig y Pre Inscripción	124
Figura 29. Gestión de metadatos	125

Figura 30. Presentación de la documentación digitalizada	127
Figura 31. Consulta del legajo digital	128
Figura 32. Gestión de las presentaciones realizadas	129
Figura 33. Replicación de datos en el nodo Sede.....	130
Figura 34. Acceso al catalogo <i>Slony-I</i>	131
Figura 35. Configuración de los perfiles funcionales - LeDig.....	133
Figura 36. Configuración de restricciones funcionales - LeDig	134
Figura 37. Asociación entre los perfiles funcionales y las restricciones.....	135
Figura 38. Tablas de auditoría.....	136
Figura 39. <i>Triggers</i> de auditoría.....	137
Figura 40. Panel de control - <i>Proxmox</i>	137
Figura 41. Arquitectura de la plataforma SIU Araí [66].	140
Figura 42. Legajos digitalizados por cohorte	143
Figura 43. Legajos digitalizados. Evolución por mes	144
Figura 44. Análisis espacio requerido	145
Figura 45. Espacio liberado en la Dirección de Estadísticas.....	146
Figura 46. Product Backlog proyecto LeDig	149
Figura 47. Detalle de un requerimiento en Redmine	149
Figura 48. Duración del proyecto LeDig.....	150
Figura 49. Tráfico demandado por la sincronización.....	151
Figura 50. Replicación de tuplas con el sistema en producción.....	152
Figura 51. Caída vínculo. Consola <i>slon</i> nodo Maestro	152
Figura 52. Caída vínculo. Consola <i>slon</i> nodo Sede.....	153
Figura 53. Tráfico en puerto 80 (Apache).....	154
Figura 54. Tráfico diario en puerto 5432 (<i>PostgreSQL</i>)	154
Figura 55. Tráfico semanal en puerto 5432 (<i>PostgreSQL</i>)	155
Figura 56. Transmisión de archivos (<i>ssh</i>) desde el sitio Maestro al nodo Sede.....	157
Figura 57. Transmisión de archivos con compresión (<i>ssh -C</i>).	157
Figura 58. Comparativa sobre la cantidad de legajos consultados por semana	159
Figura 59. Digitalizaciones con tamaño que superan la media	166

Índice de Tablas

Tabla 1. Comparativa de herramientas de replicación	88
Tabla 2. Actividad sobre las herramientas de replicación.....	91
Tabla 3. Descripción de atributos - Tabla ledig	104
Tabla 4. Descripción de atributos - Tabla leg_presentaciones	104
Tabla 5. Descripción de atributos - Tabla leg_depositarios	105
Tabla 6. Descripción de atributos - Tabla leg_tipo_doc	105
Tabla 7. . Descripción de atributos - Tabla sga_carreras	106
Tabla 8. Descripción de atributos - Tabla leg_config	106
Tabla 9. Descripción de atributos - Tabla int_ledig	107
Tabla 10. Análisis de frecuencia de modificación de datos	115
Tabla 11. Estimación sobre la periodicidad de replicación.....	115
Tabla 12. Barra de herramientas - Interfaz Pre Inscripción	124
Tabla 13. Barra de herramientas - Gestión digital de legajos	127
Tabla 14. Tablas del catálogo <i>Slony-I</i>	131
Tabla 15. Atributos herramienta <i>RRDtool</i>	150
Tabla 16. Tiempos de espera por archivo.....	155
Tabla 17. Cantidad Legajos Importados desde el SIU Guaraní	160
Tabla 18. Cantidad Legajos Importados desde pre inscripción.....	160

Capítulo 1

Introducción

1. Introducción

La Universidad Nacional de Misiones en su búsqueda permanente de incrementar las oportunidades de acceso a la educación superior ha procurado extender la cobertura territorial en la provincia de Misiones. Con esta finalidad existe el “Programa de Expansión Territorial de Educación Superior” [1].

En este contexto, resulta indispensable abordar el rol de los sistemas de información y su arquitectura, ya que la actual (cliente servidor) muchas veces deriva en la interrupción de las operaciones diarias ante la falta de conectividad con la sede central en Rectorado.

Este presente trabajo de tesis propone desarrollar e implementar una arquitectura para mejorar el modo en que se gestiona la información en las distintas dependencias de la Universidad Nacional de Misiones para, de este modo, asegurar la continuidad de las operaciones.

Como caso de estudio y validación se toma la gestión de legajos digitales de la Facultad de Ciencias Económicas de la UNaM.

1.1. Descripción del problema y justificación

En el año 2017 la Universidad Nacional de Misiones atravesó el proceso de acreditación institucional. Como resultado de ello, la CONEAU¹ remarcó como una fortaleza el “Programa de Expansión Territorial de Educación Superior” y aprobado en acuerdo plenario N°802/11 del CIN². El mismo tiene la finalidad de “incrementar las oportunidades de acceso, permanencia y egreso a la educación superior, extendiendo la cobertura territorial para la atención de las necesidades y demandas de formación superior” [1].

¹ CONEAU: Comisión Nacional de Evaluación y Acreditación Universitaria

² CIN: Consejo Interuniversitario Nacional

No obstante y teniendo en cuenta la importancia de lo antes expuesto, no existe una propuesta basada en TIC ³ que acompañe y de soporte a este plan de expansión. Cuando surge la necesidad de consultar y/o actualizar la información en los diferentes sistemas, muchas veces el personal de la universidad se encuentra con varias limitaciones fijadas por la actual arquitectura cliente servidor, como ser: la falta de conectividad con la unidad central (rectorado) y por ende imposibilidad de acceder a los sistemas y las bases de datos (BD), la interrupción del suministro de energía eléctrica en la sede central y otros inconvenientes propios de la arquitectura actual de los sistemas informáticos.

1.2. Objetivos

Este trabajo de tesis busca desarrollar e implementar una arquitectura para mejorar el modo en que se almacena, busca y actualiza la información en las distintas dependencias de la Universidad Nacional de Misiones. Se propone una alternativa a la actual arquitectura cliente servidor, ya que al estar en la unidad central de rectorado a kilómetros de distancia de los usuarios finales y debiendo siempre operar en línea, no permite asegurar la continuidad de las operaciones.

Como caso de validación se aborda la gestión de los legajos digitales de los alumnos de la Universidad Nacional de Misiones.

Principalmente se propone:

- Desarrollar una arquitectura y herramienta software, basada en la réplica de la información.
- Digitalizar la información, pudiendo así, optimizar el espacio físico, el cual es crítico en la actualidad.
- Brindar alta disponibilidad de la información a las distintas áreas y sedes (interior de la provincia) de la Universidad Nacional de Misiones, a fin de asegurar la continuidad de los trámites y operación de los sistemas informáticos.

³ TIC: Tecnologías de la Información y la Comunicación

- Agilizar el acceso (conurrencia), la búsqueda y la actualización de la información (legajos).
- Aumentar la productividad de los sectores que gestionan y administran esta información en forma física y no digital. Por ejemplo: actualmente el Departamento de Estadísticas (Inscripciones) solo se dedica a la registración de los ingresantes y expendio de certificaciones, lo cual insume la mayor parte del tiempo, relegando las estadísticas a segundo plano.
- Disminuir costos, compras de activos (muebles archivadores, librería y papelería) para la administración, almacenamiento y transporte de los archivos físicos.

1.3. Elección del caso

En la sesión del día 15 de noviembre del año 2017, el Honorable Consejo Directivo de la Facultad de Ciencias Económicas aprobó por unanimidad y resolvió mediante instrumento 190/2017 [2], aprobar la ejecución del proyecto de “Digitalización de Legajos Estudiantiles (LeDig)” para los ingresantes a partir de la cohorte 2018.

Las expectativas son poder ejecutar el proyecto y medir su avance a fin de poder replicarlo en las demás facultades e instituciones que componen la Universidad Nacional de Misiones. Se debe mencionar que los legajos a digitalizar corresponden a alumnos inscriptos de todas las carreras, sin embargo la intención detrás del proyecto es ofrecer la digitalización de cualquier documentación cuya disposición y almacenamiento sea necesaria por su importancia en la gestión institucional.

Debido a que muchas veces estos trámites conllevan la necesidad de administrar documentos (por ejemplo: legajos y constancias), también se aborda la viabilidad técnica de la digitalización de documentos, analizando diversos aspectos, como disponibilidad, acceso concurrente, disponibilidad, resguardo y recuperación ante incidentes, integridad documental, marco legal y firma digital, entre otros.

Este proyecto permite contar con un caso que sirve como experiencia y

validación sobre un entorno en el cual no existe ninguna herramienta desarrollada para la gestión, pero sin embargo se dispone del acceso a los datos requeridos y al personal No Docente que tiene el conocimiento en la gestión de los legajos.

1.4. Contexto del presente trabajo

El presente trabajo ha sido desarrollado para la obtención del título de la “Maestría en Tecnologías de la Información”. En parte fue financiado por una beca del Programa Estratégico de Formación de Recursos Humanos en Investigación y Desarrollo (PERHID) otorgada por el Consejo Interuniversitario Nacional [3].

El tema escogido se encuentra incluido en las líneas de investigación del Instituto de Investigación, Desarrollo e Innovación en Informática (IIDII) de la Facultad de Ciencias Exactas Químicas y Naturales de la Universidad Nacional de Misiones.

El trabajo se desarrolló en la Facultad de Ciencias Económicas de la Universidad Nacional de Misiones. Se cuenta con el apoyo de la institución para llevar a cabo el proyecto [2] y también está señalado como prioritario por la gestión actual.

El área de la Dirección de Tecnología para la Gestión de la Facultad de Ciencias Económicas cuenta con el equipamiento y el personal calificado para llevar adelante el proyecto. Casi todos los integrantes del equipo (80%) trabajan en el mismo lugar, tienen buena relación y experiencia, constituyendo un equipo multidisciplinario.

1.5. Estructura del documento

Este documento se encuentra dividido en ocho capítulos. Primeramente se brindando en el capítulo 1 una descripción del problema, la justificación del trabajo, sus objetivos y la elección del caso. En el capítulo 2 se presenta el marco teórico que se ha utilizado como base para el desarrollo del trabajo, abarcando temas como la

arquitectura de los sistemas de información, la digitalización de documentos y algunos antecedentes sobre replicación de datos.

En el capítulo 3, se describe lo relevado respecto a la gestión actual de los legajos en la Facultad de Ciencias Económicas de la Universidad Nacional de Misiones. A continuación, en el capítulo 4 se ofrece el modelado de la solución, para luego en el capítulo 5 detallar las cuestiones referentes al desarrollo y la implementación de la misma.

En los capítulos 6 y 7 se describen los resultados y la discusión respectivamente, sobre las experimentaciones realizadas para, finalmente, en el capítulo 8 resumir las principales conclusiones en contraste con los objetivos que han sido establecidos, limitaciones del enfoque, sugerencia y las futuras líneas de investigación que pueden llevarse a cabo.

Capítulo 2

Marco teórico

2. Marco teórico

En este capítulo se exponen los marcos teóricos de cómo se encuentra en la actualidad el estado de las cuestiones que luego se abordarán en el trabajo.

Primeramente se describe lo concerniente a la arquitectura y las tecnologías de BD para luego ofrecer, en forma complementaria como lo requiere el presente trabajo, un marco de referencia sobre la gestión documental.

La arquitectura de las BD están directamente ligadas, por lo general, a las características de las aplicaciones software a implementar como así también a la arquitectura de las computadoras y las redes de comunicación de datos [4].

Estas redes facilitan que muchas tareas se ejecuten en un sistema servidor y que otras lo hagan en los sistemas clientes, lo cual ha derivado en las arquitecturas denominadas cliente servidor.

Por otro lado la programación y la posibilidad del procesamiento paralelo, en la actualidad permiten acelerar la ejecución de las tareas, reduciendo así los tiempos de respuesta y proporcionando a las BD la capacidad de ejecutar más transacciones por segundo. El auge de esta tecnología condujo al desarrollo de los sistemas de BD paralelos [5].

La globalización de las organizaciones, la existencia de sedes en los distintos países y los departamentos que necesitan datos que no residen donde han sido generados, llevaron a la necesidad de distribuir la información. Una arquitectura de base de datos distribuida (BDD) asegura la continuidad de los negocios con el solo hecho de guardar copias de las BD en diferentes sitios, permitiendo que se pueda continuar con las transacciones con la BD aunque algún nodo sufra un desperfecto o desastre natural [5].

Una arquitectura de BDD implica que la aplicación debe operar en forma transparente sobre los datos que están dispersos en diferentes BD heterogéneas y administradas por distintos sistemas de gestión de BD (DBMS), que se encuentran en

servidores diferentes, manejadas por varios sistemas operativos diferentes y conectados a una variedad de redes de comunicación distintas. Aquí el término “transparente” significa que la aplicación opera desde un punto de vista lógico como si todos los datos fueran manejados por un único DBMS y ejecutados en un solo sitio [6].

A continuación se realizará un análisis detallado de las principales arquitecturas, profundizando en las arquitecturas distribuidas de BD.

2.1. Arquitecturas centralizadas

Los sistemas de BD centralizados son aquellos que se ejecutan en un único sistema informático (o unidad central de procesamiento - UCP) y sin interacción con otro ordenador. Ejemplo de estos entornos son los sistemas de BD mono usuario ejecutándose en computadoras personales. También los sistemas de BD de alto rendimiento que se ejecutan en *mainframes*⁴, los cuales son consultados por terminales sin ninguna capacidad de procesamiento ni almacenamiento (Figura 1). Esta arquitectura se diferencia de los sistemas cliente-servidor en que tienen la funcionalidad dividida entre el servidor y múltiples sistemas clientes [5].



Figura 1. Arquitecturas centralizadas

⁴ *Mainframe*: computadora de alto rendimiento

Las arquitecturas de los DBMS evolucionaron de forma similar a las arquitecturas de los sistemas de computación. Las arquitecturas basadas en mainframes se encargaban del procesamiento principal para todas las funciones del sistema, incluyendo las aplicaciones, programas de interfaz de usuario y el DBMS. Esto debido a que los usuarios accedían a estos sistemas por medio de terminales carentes de potencia de procesamiento y con sólo capacidades de visualización, por lo que todo el procesamiento se realizaba remotamente en el *mainframe*, y se visualizaba en las terminales conectadas con el computador central a través de diferentes tipos de redes de comunicaciones [4].

2.2. Arquitecturas cliente servidor

Con la disminución de los precios del hardware, los usuarios fueron reemplazando sus terminales por estaciones de trabajo. Al inicio, los DBMS utilizaban esos computadores de manera similar a las terminales de visualización, por lo que el DBMS continuaba siendo centralizado. Luego los DBMS comenzaron a aprovechar la capacidad de procesamiento disponible del lado del usuario, lo que llevó a las arquitecturas DBMS cliente/servidor [4].

Como resultado, en forma gradual, los sistemas se han ido distanciando de la arquitectura centralizada. Las terminales, llamadas bobas, conectadas a los sistemas informáticos con arquitectura centralizada, han sido reemplazadas completamente por computadoras personales (PC) [5]. Cuestiones como la interfaz de usuario, que las gestionaba el sistema central, comenzaron a ser gestionadas por las PC, en consecuencia, los sistemas centralizados en realidad hoy operan como sistemas servidores que satisfacen las peticiones generadas por las PC conectadas (clientes) derivando en la actualidad en la separación del servidor de BD de la aplicación cliente [7]. En la Figura 2 se ilustra la arquitectura general de un sistema cliente-servidor.

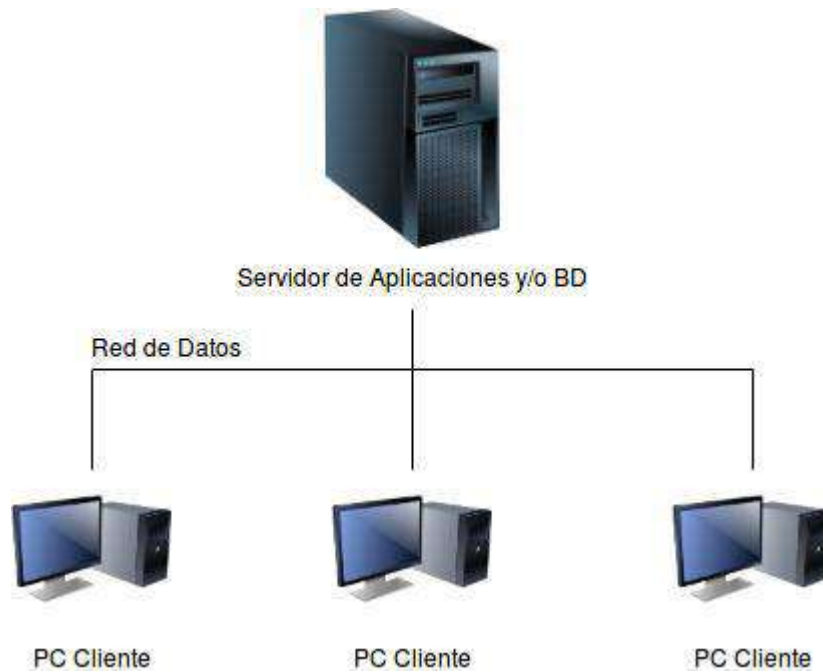


Figura 2. Arquitecturas Cliente-Servidor

Las arquitecturas cliente/servidor se desarrollaron para entornos en los que una gran cantidad de PCs, servidores de archivos, impresoras, servidores WEB y otros equipos están conectados a través de una red. La idea es definir servidores especializados con funcionalidades específicas, como por ejemplo las BD [7].

2.3. Arquitectura de sistemas paralelos

Este tipo de arquitecturas involucra muchas unidades de almacenamiento de datos y procesadores que operan en paralelo para incrementar el rendimiento [7]. Están destinadas a mejorar la velocidad de procesamiento y de E/S mediante la utilización de recursos, UCP y discos, en forma paralela. *Hardware* con varios procesadores, procesadores con varios núcleos, lenguajes de programación paralela, hace que cada vez sea más importante el estudio de los sistemas de BD paralelos. Tecnologías ya instaladas en las organizaciones como las *datawarehouse*, *dataminig* y *big data* han impulsado a los sistemas paralelos de BD para permitir a las aplicaciones manejar estas BD extremadamente grandes (en el orden de los terabytes)

ya que los sistemas de BD centralizados o cliente-servidor por lo general no poseen la potencia para soportar tales aplicaciones [5].

Para establecer una métrica sobre el rendimiento de los DBMS paralelos existen dos medidas principales: la productividad, que refiere al número de tareas que pueden completarse en un intervalo de tiempo determinado, y el tiempo de respuesta, que establece la cantidad de tiempo que necesita para completar una única operación desde el momento en que se envía. Un sistema que procesa un gran número de transacciones puede mejorar su productividad realizando estas transacciones en paralelo (Figura 3). Un sistema que procesa transacciones de una mayor duración, también puede mejorar los tiempos de respuesta realizando en paralelo las distintas sub tareas de cada transacción. En este aspecto la ganancia de velocidad se refiere a la ejecución en menos tiempo de una tarea mediante el incremento del grado de paralelismo y la ampliabilidad que refiere al manejo de transacciones más grandes, cuya duración se desea acortar mediante el incremento del grado de paralelismo [5].

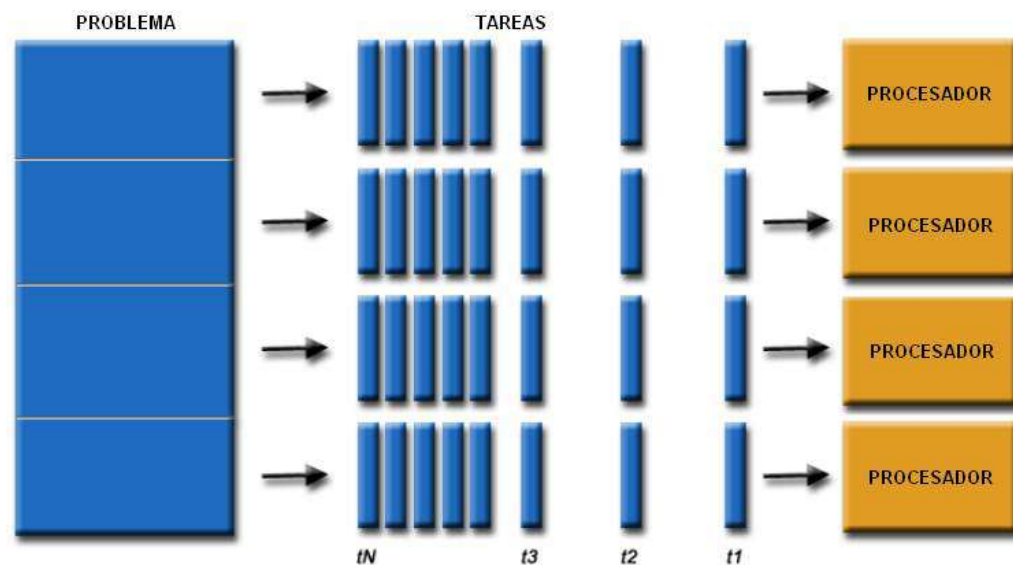


Figura 3. Arquitectura de Sistemas Paralelos

La división del problema puede realizarse por dominio o función. Las unidades de procesamiento pueden compartir memoria o bien comunicarse a través de una interfaz de paso de mensajes.

2.4. Arquitectura Distribuida

Las arquitecturas distribuidas han surgido debido a la unión de otras dos tecnologías: base de datos y comunicación de datos. Esto último refiere a los avances de las tecnologías alámbricas e inalámbricas, comunicaciones por satélite y celulares, la estandarización de protocolos como Ethernet, TCP/IP, ATM, así como la explosión de Internet [4].

Mientras que en la década del ochenta las arquitecturas de los DBMS se estancaban en la centralización, produciendo BD monolíticas, la tendencia al final de esa década era la contraria, provocando más descentralización para garantizar la autonomía de procesamiento de datos [4].

Un sistema de BDD es un conjunto de BD lógicamente relacionadas y distribuidas sobre una red de ordenadores. Los sistemas de gestión de BDD suministran los mecanismos de acceso que hace la distribución transparente al usuario [7]. En un sistema de BDD la información se almacena en diferentes servidores, generalmente usados por la misma aplicación.

Algún vínculo de comunicación, por lo general redes de alta velocidad, se conectan a estas computadoras del sistema distribuido (SD). Estas máquinas son totalmente independientes, es decir, no comparten memoria ni discos. Un SD puede ser muy heterogéneo pudiendo abarcar desde las estaciones de trabajo hasta grandes servidores. Existen diferentes nombres para referirse a estas computadoras tales como sitios o nodos. En la Figura 4 se muestra la estructura general de un SD [5].



Figura 4. Arquitectura Distribuida

Las principales diferencias con las BD paralelas son que las BDD se encuentran en lugares distintos geográficamente hablando, se administran de forma totalmente independiente y poseen una interconexión. Otra cuestión es que en un sistema de BDD existen dos tipos de transacciones, las locales y las globales. Una transacción local es que accede sólo a los datos del sitio en el cual se inició la transacción, a diferencia de una transacción global que accede a datos situados en un sitio diferente o accede a datos de varios sitios distintos de aquel en el que se inició la transacción [5].

Las razones por las cuales adoptar este tipo de arquitectura se fundamenta en la necesidad de [5]:

- Compartir datos. Es una de las principales ventajas que ofrece esta arquitectura ya que provee un entorno donde los usuarios pueden acceder a los datos que residen en otras ubicaciones. Por ejemplo, una universidad podría

permitir que un usuario de una sede acceda a los datos de otra sede o de la unidad central de Rectorado.

- Autonomía. A diferencia de un sistema centralizado, en el que el administrador de BD de la ubicación central controla toda la BD, en esta arquitectura existe un administrador de bases de datos global, el cual es responsable de todo el sistema. Una de sus tareas es delegar al administrador de BD local de cada sitio. Dependiendo del diseño, cada administrador puede tener diferentes grados de autonomía. Esta posibilidad es una de las grandes ventajas que otorgan las BDD.
- Disponibilidad. En el caso que algún sitio falle, los demás pueden seguir trabajando si los datos están replicados en varios sitios. De esta forma, una transacción que necesite de un dato, puede encontrar al mismo en otro sitio. Esta es otra de las principales ventajas de esta arquitectura ya que, el fallo de un nodo, no compromete la continuidad de las operaciones. Existen aplicaciones que puede detectar el fallo de un nodo, y establecer acciones para informar y recuperarse del fallo. El entorno debe prescindir de la utilizando los servicios ofrecidos por el sitio que falló, y luego que el servicio del nodo que falló sea restablecido, deben existir mecanismos que permitan su integración y normal operatoria. Esta complejidad otorga la capacidad de continuar trabajando, características que son importantes en aplicaciones de tiempo real y organizaciones con una marcada expansión territorial cuya imposibilidad de realizar sus transacciones materializaría la pérdida de mucho dinero.

Como contrapartida, pueden presentarse los siguientes inconvenientes [7]:

- Solución más compleja: Partiendo desde el diseño de la BD, la coordinación que debe existir entre los sitios, la implementación de políticas de seguridad y la dependencias entre los distintos elementos de la arquitectura distribuida deriva indefectiblemente en soluciones con una elevada complejidad
- Análisis de costos: Deben contemplarse los costos referentes al mantenimiento de este tipo de sistemas, el contrato de personal especializado

y la conectividad necesaria para su correcto desempeño debido, por ejemplo, a la sobrecarga de procesamiento de mensajes. También cabe destacar que actualmente con los entornos basados en *Cloud Computing*⁵ los costos pueden reducirse considerablemente.

- Falta de estándares: La falta de estándares y metodologías pueden dificultar este tipo de implementaciones.
- Las empresas en su gran mayoría poseen implementaciones que involucran a BD centralizadas.

Con los avances en la computación y el procesamiento distribuido de los sistemas operativos (SO), la comunidad de investigación de las BD ha trabajado para resolver los problemas derivados de la distribución de los datos, las consultas distribuidas y el procesamiento de las transacciones, la administración de los metadatos de las BDD y otros temas, lo que provocó el desarrollo de muchos prototipos. Sin embargo, los sistemas de gestión de BDD totalmente operativos que incluyen la funcionalidad y técnicas propuestas por estas investigaciones, nunca surgieron como un producto comercialmente viable. Los principales desarrolladores no centraron sus esfuerzos en obtener estos productos, sino en la generación de sistemas basados en la arquitectura cliente-servidor, o a través de tecnologías para el acceso a fuentes de datos distribuidas de manera heterogénea [4].

Sin embargo las organizaciones han estado muy interesadas en la descentralización del procesamiento y la posterior integración de las fuentes de información. Con los avances en los sistemas de comunicación ahora existe una aceptación general del esquema cliente-servidor en el desarrollo de una aplicación, el cual asume muchos de los temas DDB y dentro de estos últimos la replicación de datos [4].

⁵ *Cloud Computing* – Computación en la Nube

2.4.1. Almacenamiento distribuido de datos

Existen dos formas de almacenar una relación R en un sistema de BDD:

- Réplica. Con este método, se conserva copias exactas de R y se almacena la misma en diferentes sitios.
- Fragmentación. Con esta alternativa se realiza una división de R en varios fragmentos para que sean almacenadas en diferentes sitios.

Ambas técnicas pueden combinarse: Durante la implementación, R puede fragmentarse y replicarse en varios sitios.

2.4.2. Reglas de Date

J.C. Date [6] describe doce reglas que deben cumplir los sistemas distribuidos:

1. Autonomía local: los sitios deben ser autónomos en el mayor grado posible.
2. No dependencia de un sitio central: todos los sitios deben ser tratados con la misma importancia.
3. Operación continua: un SD debe aumentar la confiabilidad y la fiabilidad. No deberían realizarse cortes de servicios no planificados.
4. Independencia de localización: para el usuario final la localización física de los datos debe ser transparente.
5. Independencia de fragmentación: los usuarios finales no necesitan saber cómo los fragmentos físicos están divididos, ni conocer cada colección lógica de datos.
6. Independencia de replicación: los usuarios finales no necesitan saber si los datos están replicados o no.
7. Procesamiento de consultas distribuidas: el SD debe disponer de mecanismos para optimizar las consultas con la finalidad de reducir la carga de tráfico generada.

8. Gestión de transacciones distribuidas: el SD debe disponer de protocolos para el control de concurrencia y la recuperación de transacciones distribuidas.
9. Independencia del hardware: el sistema debe poder ejecutar el mismo DBMS en sitios con diferentes plataformas hardware.
10. Independencia del sistema operativo: también se debe lograr ejecutar el mismo DBMS en sitios con diferentes sistemas operativos.
11. Independencia de la red: el SD debe operar en entornos con diferentes redes de comunicaciones.
12. Independencia del DBMS: cada sitio debe poder funcionar con un DBMS diferente, incluso con modelos de datos diferentes.

2.4.3. Replicación de datos

La replicación de la BD es el proceso de mantener múltiples copias de elementos de datos en diferentes ubicaciones llamadas réplicas. La motivación para hacer esto es doble: en primer lugar, se puede aumentar la disponibilidad ya que el sistema de BD puede tolerar mejor las fallas, y por otra parte, se puede aumentar el rendimiento de los sistemas, reduciendo los tiempos de respuesta al distribuir la carga de transacción entre las réplicas [8].

Con los servicios y aplicaciones WEB, la computación grid, la exponencial generación de datos y en definitiva, la necesidad de acceso a la información, cada vez se requiere una mayor demanda de datos, por lo que las BD presentan importantes desafíos de disponibilidad y escalabilidad. La replicación es la solución más común, pero implica concesiones difíciles, tales como escalabilidad y consistencia. A menudo, se debe hacer una elección entre estos atributos ya que si se requiere consistencia total, el precio es una pérdida de escalabilidad [9]

Investigaciones recientes en el área de aplicaciones WEB transaccionales han dado lugar a muchas propuestas, pero todas adolecen de una variedad de problemas. Los sistemas que ponen los datos al borde de la red pueden reducir los tiempos de respuesta pero renuncian a la consistencia. Algunos enfoques [10], trabajan en la capa

de middleware y necesitan una pre declaración del patrón de acceso de todas las transacciones para permitir una programación eficiente, y otros [11] ofrece escalabilidad y consistencia; sin embargo, el diseñador de la base de datos se ve obligado a dividir los datos de forma estática. Esto prohíbe ciertas consultas y restringe la evolución libre de la aplicación. En este sistema, una escalabilidad razonable es posible solo si la carga se ajusta exactamente a la partición de datos elegida.

Tomando el ejemplo de la sección anterior, cuando se guarda la copia de la relación R , la misma se copia a dos o más sitios. En situaciones extremas y de alta disponibilidad, se tiene una réplica completa de R en cada sitio del sistema. Respecto a estas alternativas debemos mencionar algunas ventajas y desventajas [5]:

- Disponibilidad. Relacionado al hecho de que aunque se produzca un fallo, la relación R puede encontrarse en otro nodo diferente. De esta forma los sistemas pueden continuar funcionando pese al fallo del sitio.
- Paralelismo incrementado. Las operaciones de solo lectura sobre R pueden realizarse en paralelo, y cuantas más réplicas de R haya, mayor será la probabilidad que los datos se encuentren en el sitio donde se ejecuta la transacción. La réplica de los datos minimiza, o fácticamente, “no necesita” el movimiento de los datos entre los sitios.
- Sobrecarga (actualización). Para que no se produzcan errores de cómputo en las operaciones, el sistema debe asegurar la consistencia de todas las réplicas de R , lo cual provoca que cualquier operación de actualización sobre esta relación, se propague a los demás sitios que contienen réplicas de R . Como resultado de esto, se genera una sobrecarga de los sistemas y las comunicaciones. Por ejemplo, en un banco donde se replica la información de las cuentas, debe asegurarse que el saldo de cada una de ellas concuerde en todos los sitios.

En general, si bien la réplica mejora el rendimiento de las operaciones de lectura y aumenta la disponibilidad de los datos, para transacciones que involucran actualizaciones generan una sobrecarga para el sistema ya que el control de este tipo

de operaciones resulta más complejo. Esta complejidad puede simplificarse, seleccionando una copia de R como principal, y así como en el ejemplo del banco, las cuentas pueden asociarse con el sitio en que fueron abiertas [5].

2.4.4. Tipos de replicación

Existen dos tipos de sistemas de replicación: ansioso y perezoso [9]. Los sistemas ansiosos (o síncronos) mantienen las réplicas sincronizadas dentro de los límites de la transacción. Se ajustan a la serialización de una copia y las planificaciones resultantes son equivalentes a una programación en serie en una única BD. Aunque los sistemas ansiosos ofrecen las mismas garantías de corrección que las instalaciones de BD únicas, el concepto rara vez se utiliza en productos comerciales de replicación. Esto se debe al hecho de que los protocolos ansiosos convencionales tienen importantes inconvenientes en cuanto a rendimiento y escalabilidad [12] [13]. En primer lugar, la sobrecarga de comunicación entre las réplicas es muy alta, lo que lleva a tiempos de respuesta largos y, en segundo lugar, la probabilidad de los interbloqueos son proporcionales a la tercera potencia del número de réplicas. Es por eso que estas implementaciones no pueden escalar más allá de algunas réplicas.

Para evitar estos problemas, los diseñadores de BD comenzaron a desarrollar sistemas perezosos (o asíncronos) en los que las réplicas se pueden actualizar fuera de los límites de la transacción. Este tipo de replicación permite mejor escalabilidad, pero conduce a nuevos problemas:

Las transacciones pueden leer datos obsoletos y los conflictos entre las transacciones de actualización son posiblemente detectados de manera tardía, introduciendo la necesidad de resolución de conflictos. Lamentablemente, estos problemas no se pueden ocultar fácilmente al usuario y, a menudo, la aplicación del cliente debe resolverlos.

Las propiedades insatisfactorias de los protocolos perezosos han llevado a una continuación en la investigación sobre protocolos ansiosos. Un resultado importante

fue la idea de que la distribución de las declaraciones completas de actualización de SQL, como se hace a menudo en la actualización ansiosa en todas partes, no es óptima. El rendimiento puede mejorarse significativamente ejecutando sentencias SQL solo una vez y luego propagando los cambios resultantes de la BD (*writesets*) a otras réplicas [14].

2.4.5. Técnicas distribuidas

Las técnicas distribuidas aplican la actualización en la copia local del sitio donde se origina la transacción de actualización, y luego las actualizaciones se propagan a los otros sitios de réplica. Estas se denominan técnicas distribuidas, ya que diferentes transacciones pueden actualizar copias diferentes del mismo elemento de datos ubicado en diferentes sitios. Son apropiados para aplicaciones colaborativas con centros de decisión u operación distributivos. Pueden distribuir la carga de manera más equitativa y proporcionar mayor disponibilidad del sistema si se combinan con técnicas de propagación diferida [7].

Una complicación que surge en estos sistemas es que diferentes réplicas de un elemento de datos pueden actualizarse en diferentes sitios maestros al mismo tiempo. Si las técnicas distribuidas están acopladas por métodos de propagación ansiosos, entonces los métodos de control de concurrencia distribuidos pueden abordar adecuadamente el problema de las actualizaciones concurrentes. Sin embargo, si se utilizan métodos de propagación perezosa, las transacciones pueden ejecutarse en diferente orden en diferentes sitios, lo que provoca un historial global que no es serializable. Además, varias réplicas pueden no sincronizarse correctamente. Para gestionar estos problemas, se aplica un método de conciliación que consiste en deshacer y rehacer transacciones de tal manera que la ejecución de la transacción sea la misma en cada sitio, pero esto depende generalmente de la aplicación.

Existen diferentes maneras de distribuir los datos. Principalmente podemos describir dos clases:

- Distribución cliente servidor.
- Distribución punto a punto (o distribución completa).

La distribución cliente servidor concentra las tareas de gestión de datos en los servidores, mientras que los clientes se centran en proporcionar el entorno de la aplicación, incluida la interfaz de usuario. Los deberes de comunicación se comparten entre las máquinas cliente y los servidores. Los DBMS representan un compromiso práctico para distribuir la funcionalidad. Hay una variedad de formas de estructurarlas, cada una de las cuales proporciona un nivel diferente de distribución, pero lo que se desea resaltar es que los sitios en una red se distinguen como "clientes" y "servidores" y su funcionalidad es diferente. En cambio, en los sistemas punto a punto, no hay distinción entre máquinas cliente y servidores. Cada máquina tiene una funcionalidad de DBMS y puede comunicarse con otras máquinas para ejecutar consultas y transacciones.

2.4.6. Fragmentación

Una de las cuestiones más importantes es como se van a distribuir los datos a través de los sitios. Para ello es necesario tener en cuenta:

- Como se va a fragmentar la información.
- Definir si la información va a estar duplicada en varios sitios.
- En que sitio almacenar cada fragmento.

Existen dos estrategias fundamentales de fragmentación: horizontal y vertical. Pero también existe la posibilidad de anidar fragmentos de forma híbrida [7].

La fragmentación horizontal divide una relación a lo largo de sus tuplas. De este modo, cada fragmento tiene un subconjunto de las tuplas de la relación (Figura 5). Existen dos versiones de partición horizontal:

- Primaria: en la que la fragmentación horizontal primaria de una relación se realiza utilizando predicados que se definen en esa relación.

- Derivada: es la partición de una relación en que los resultados de los predicados se definen en otra relación.



Figura 5. Fragmentación horizontal

Una fragmentación vertical de una relación R produce fragmentos $R_1, R_2, \dots, R_r$, cada uno de los cuales contiene un subconjunto de los atributos de R , así como la clave primaria de R (Figura 6). El objetivo de la fragmentación vertical es dividir la relación en un conjunto de relaciones más pequeñas para que muchas de las aplicaciones de usuario se ejecuten como un solo fragmento.

En este contexto, una fragmentación óptima es aquella que produce un esquema de fragmentación que minimiza el tiempo de ejecución de las aplicaciones de usuario.

La fragmentación vertical se ha investigado ampliamente en el contexto de los DBMS centralizados y distribuidos. Su motivación dentro del contexto centralizado es como una herramienta de diseño, que permite que las consultas de los usuarios utilicen relaciones más pequeñas, lo que provoca un número menor de accesos a las páginas de memoria [15]. Esto ofrece la posibilidad de que las sub relaciones más activas se identifiquen y ubiquen en un subsistema de memoria más rápido en aquellos casos en que se admiten las jerarquías de memoria [16].



Figura 6. Fragmentación vertical

En algunos casos la fragmentación horizontal o vertical de un esquema de BD puede no ser suficiente para resolver los requisitos de las aplicaciones de usuario. Entonces existe la posibilidad que la fragmentación vertical pueda ir seguida de una horizontal, o viceversa, lo cual produce una partición estructurada en árbol (Figura 7). Dado que los dos tipos de estrategias de partición se aplican una tras otra, esta alternativa se llama fragmentación híbrida, mixta o anidada.

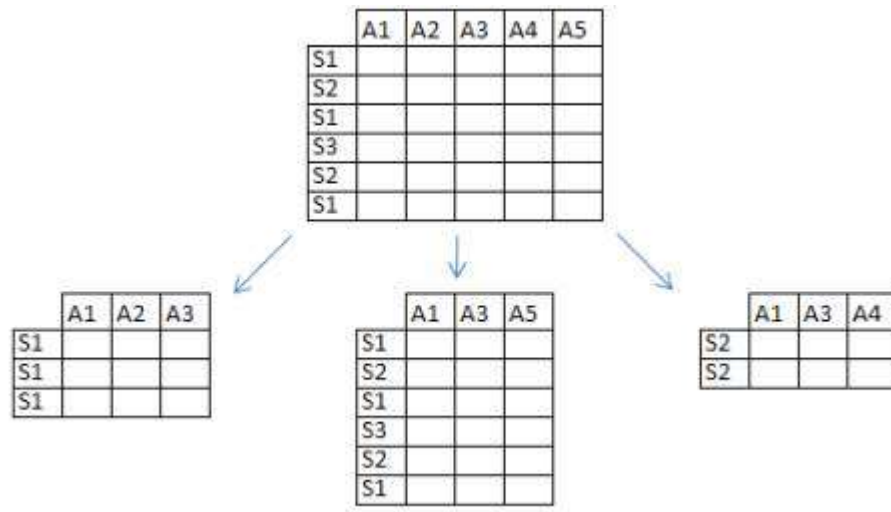


Figura 7. Fragmentación mixta

La cantidad en los niveles de anidación puede ser grande, pero ciertamente es finita. En el caso de la fragmentación horizontal, uno tiene que detenerse cuando cada fragmento consta de una sola tupla, mientras que el punto de terminación de la fragmentación vertical es un atributo. Sin embargo, estos límites son bastante académicos, ya que los niveles de anidamiento en la mayoría de las aplicaciones

prácticas no superan los dos niveles. Realizar demasiadas fragmentaciones puede generar un alto costo a la hora de volver a hacer las uniones (*join*).

2.5. Digitalización de documentos

Si bien la replicación de los sistemas y sus datos otorgan alta disponibilidad dando el soporte necesario para asegurar la continuidad de las operaciones, se debe contemplar si el circuito administrativo consulta o genera documentación en soporte papel, como pueden ser los legajos de los alumnos, actas, notas, expedientes, etc.

Para intentar dar una solución integral al Plan de Expansión Territorial es que se realizó un relevamiento sobre las experiencias, leyes, reglamentación y modo en que esta documentación puede digitalizarse para, de esta manera, poder ser accedida desde diferentes sitios.

2.5.1. Porque digitalizar documentos

En la actualidad, en los diversos organismos públicos del estado, los trámites generan una gran cantidad de documentos físicos de diversos tipos. Estos documentos son necesarios para el registro y normal funcionamiento, y para muchos organismos poseen un valor crítico, ya que su pérdida puede perjudicar al Estado o a las personas [17].

Existen varios inconvenientes vinculados a la naturaleza física de los documentos, como por ejemplo:

- **Volumen**: Al tratarse de documentos físicos, en organismos con un alto número de documentos en tránsito, es necesario un gran volumen para almacenarlos, volumen que va en constante aumento y que requieren de una infraestructura cada vez mayor. Dado el costo que tiene ampliar la infraestructura de un establecimiento o la imposibilidad de hacerlo, los

documentos terminan transfiriéndose físicamente a otro lugar, con potenciales riesgos asociados, como ser extravío, daño, pérdida, etcétera [18].

- Acondicionamiento: Los documentos requieren de un ambiente con clima controlado para evitar el deterioro físico de los mismos. Actualmente en la Administración Pública Nacional (APN), esta condición no siempre se cumple, ya que los documentos siempre están almacenados en un lugar transitorio y esto hace que sufran deterioros con el paso del tiempo [19].
- Acceso: Los documentos necesitan ser físicamente leídos por el personal. Al estar en un formato físico, no es posible que más de una persona tenga acceso a un documento al mismo tiempo, además, si el personal que desea acceder al documento no se encuentra geográficamente cerca del mismo, es necesario que se desplace hacia donde se encuentra el documento o viceversa si existe la posibilidad que el documento pueda ser transferido. También hay que mencionar que al ser manipulados por el personal, los documentos tienden a estropearse, perder integridad o extraviarse [18].
- Estadísticas: Algunos organismos necesitan hacer cálculos estadísticos periódicamente sobre algunos datos contenidos en grupos de documentos que al tratarse de documentos físicos, es necesario emplear personal específico para llevar a cabo esta tarea. El costo en los recursos humanos implicados aumenta a medida que aumenta el número de documentos a analizar, también aumenta la probabilidad de obtener errores en los cálculos estadísticos. Como existe una tendencia en constante aumento de la población de documentos en todos los organismos públicos, en algunos sectores ya no es viable ejecutar análisis estadísticos, debido a que implican muchos recursos y los resultados tienen un margen de error cada vez mayor [18].

Los puntos expresados anteriormente ponen en contexto la situación que se vive en la actualidad con la gestión de documentos en formato físico. Esta situación es grave y tiende a empeorar, originando la necesidad de plantear un cambio integral en materia de gestión documental.

2.5.2. Antecedentes sobre digitalización

A continuación se mencionan algunas experiencias dentro del territorio nacional y a nivel Latinoamérica:

- Centros de digitalización y acceso a la información pública: la experiencia de la Provincia de Santa Fe: Los "Centros de digitalización y acceso a la información pública", es una experiencia realizada en la provincia de Santa Fe respecto a bases de datos documentales y digitalización de los documentos con la finalidad de brindar transparencia en la gestión pública. El objetivo principal fue la creación de "Centros de digitalización" para así exhibir los actos y trámites de gobierno. En este trabajo, se elaboró un prototipo de centro de digitalización en la Dirección Provincial de Anticorrupción y Transparencia del Sector Público para posteriormente replicar las experiencias en las demás jurisdicciones. La finalidad de estos centros es materializar la digitalización y transformar los documentos físicos (papel) a un formato digital para luego ser cargado en la Biblioteca Digital de Normativas (BDN). La BDN no solo obra como repositorio de la documentación digitalizada, sino que es un sistema integral que permite la publicación, recuperación y preservación de los documentos digitales. La mencionada implementación fue desarrollada con el software libre Dspace [20].
- Digitalización y Sistematización del Proceso de Ejecuciones Fiscales en la Provincia de Río Negro: En la provincia de Río Negro existe una experiencia exitosa del trabajo conjunto entre el Poder Judicial y la Agencia de Recaudación Tributaria. Para la ejecución del proyecto se involucró a personal de tres de las ocho gerencias, la Gerencia de Asuntos Legales, la Gerencia de Tecnologías de la Información y la Gerencia de Recaudación y Cobranzas. En dichos organismos se implementó una solución transparente y eficiente para la gestión judicial de las deudas del tipo impositivas en el marco del plan estratégico y como parte del eje de planificación "Acciones

Tendientes al Cobro”. Básicamente, se realizó una reingeniería del proceso de selección de deudores morosos, cuyas deudas reunían determinadas condiciones para su posterior judicialización, el análisis del personal involucrado en el proceso y su capacitación, el desarrollo de un módulo del sistema principal de la Agencia, las implementaciones que resultaron necesarias para poder contar con una comunicación directa y segura entre los organismos, la incorporación de la firma digital para otorgarle legalidad a las actuaciones, así como la posterior explotación de la información mediante la utilización de una herramienta de *business intelligence* valiéndose de un repositorio de datos para la toma de decisiones. La ejecución de este proyecto ha permitido que la provincia de Río Negro se convierta en la primera jurisdicción nacional en digitalizar y “despapelizar” la totalidad de un proceso judicial como son las ejecuciones fiscales, representando así un salto cualitativo en la temática tributaria. (Digitalización y Sistematización del Proceso de Ejecuciones Fiscales en la Provincia de Río Negro) [21].

- Experiencia en Sudamérica: En Colombia, a causa del aumento exponencial de documentos en el distrito capital, se ha planteado un cambio que trajo consigo una óptima gestión de documentos a las organizaciones actuales, ya sean de carácter institucional, privadas o mixtas. El trabajo se sitúa en la época que nos caracteriza vivir, marcada por la globalización, Internet, la necesidad de adoptar estándares y normativas de carácter internacional en cuanto a la gestión documental en formato digital. El estudio también plantea abrir una línea de investigación para la definición de normativas relacionadas a la gestión electrónica de documentos. Dicha normativa estaría enfocada al distrito capital de Colombia, con la visión de expandirse a todo el territorio nacional en las próximas décadas. [18]

A nivel global también existen experiencias que a continuación se detallan.

- Evolución de las bases de datos jurídicas en España: En España, en el ámbito jurídico, se ha desarrollado una importante evolución en cuanto a la difusión de documentos digitales representando un camino claro hacia la digitalización documental. Este avance resultó en la creación de bases de datos documentales de uso jurídico, donde se publican no solamente documentos de este tipo, sino también publicaciones relacionadas. Además, dada la importancia de la información contenida en estas bases de datos, se ha desarrollado un marco jurídico para proteger la integridad documental contenida en el soporte digital [22].
- Copy detection mechanisms for digital documents: Este es un trabajo realizado por el Departamento de Ciencias de la Computación de la Universidad de Stanford, y plantea mecanismos para prevenir la copia de archivos digitales en bases de datos documentales. El mismo tiene como justificación proteger normas de *copyright* sobre archivos alojados en bases de datos documentales, como así también detectar posibles copias, ya sean parciales o totales, sobre documentos.

Como resultado de este trabajo los colegas han desarrollado un prototipo para dar solución a la problemática planteada, estableciendo métricas para evaluar dicho prototipo [23].

2.5.3. Digitalización en Argentina

En Argentina, la entidad que lleva adelante lo referente a la digitalización de documentación es la Secretaría de Modernización Administrativa del Ministerio de Modernización, la cual está facultada por Decreto de la Presidencia de la Nación, N° 13 del 5 de enero de 2016 a: “entender en las propuestas e iniciativas de transformación, innovación, mejora continua e integración de los procesos transversales y sistemas centrales de soporte de gestión del Sector Público Nacional, a partir del desarrollo y coordinación de políticas, marcos normativos, capacidades, instrumentos de apoyo y plataformas tecnológicas” [24].

El pasado primero de marzo de 2016 el Decreto Nacional N° 434, aprueba el PLAN DE MODERNIZACIÓN DEL ESTADO, el cual incluye el PLAN DE TECNOLOGÍA Y GOBIERNO DIGITAL que propone: "implementar una plataforma horizontal informática de generación de documentos y expedientes electrónicos, registros y otros contenedores que sea utilizada por toda la administración a los fines de facilitar la gestión documental, el acceso y la perdurabilidad de la información, la reducción de los plazos en las tramitaciones y el seguimiento público de cada expediente" [25].

Este decreto contempla como segundo eje del PLAN DE MODERNIZACIÓN DEL ESTADO, que tiene como objetivo: "la Gestión Integral de los Recursos Humanos, identificando como instrumento al Registro Único de Personal y Prestadores Informatizado, previendo la implementación del Legajo Único Personal Informatizado para toda persona que mantenga vinculación en la prestación de servicios personales, así como la digitalización y validación mediante la firma electrónica de aquella documentación requerida por exigencias legales o de auditoría" [25].

En ese sentido la Resolución 22-E/2016 [26] de la Secretaría de Modernización Administrativa, aprueba la utilización del Legajo Único Electrónico (LUE) con la finalidad de: "minimizar los documentos en papel, sin menoscabo alguno a la seguridad jurídica". Este módulo es concebido como el único medio de creación, registro, guarda y archivo de la totalidad de la documentación y datos correspondientes a las personas que prestan servicios en virtud de cualquier modalidad en el ámbito establecido en el artículo 8 de la Ley N° 24.156 [27]. En el inciso a) de dicho artículo dispone que esta ley "será de aplicación en todo el Sector Público Nacional, el que a tal efecto está integrado por: a) Administración Nacional, conformada por la Administración Central y los Organismos Descentralizados" incluyendo a las Universidades Nacionales y por estatuto [28], la Universidad Nacional de Misiones.

Luego de los instrumentos citados el 6 de abril de 2016, por Decreto N° 561 [29], la Presidencia de la Nación, aprueba la implementación del sistema de Gestión

Documental Electrónica (GDE) como: "sistema integrado de caratulación, numeración, seguimiento y registración de movimientos de todas las actuaciones y expedientes del Sector Público Nacional, actuando como plataforma para la implementación de gestión de expedientes electrónicos". Este instrumento "ordena la utilización del sistema GDE para la totalidad de las actuaciones administrativas, de acuerdo al cronograma que fije el MINISTERIO DE MODERNIZACIÓN" a todas las entidades mencionadas en el artículo 8° de la Ley N°24.156 ya citado.

Posteriormente, el 28 de octubre de 2016 y por Decreto Nacional N° 1131 [30], se reguló la digitalización y archivo de documentos electrónicos, facultando en ese mismo acto a la Secretaría de Modernización Administrativa del Ministerio de Modernización a redactar las normas operativas y aclaratorias referentes a la conservación y plazos de guarda de la documentación digital. También allí se especifica el proceso de digitalización, archivo y conservación de documentos en soporte papel en el ámbito del sector público nacional.

Finalmente el 30 de diciembre de 2016 por Resolución 44-E/2016 [31] la Secretaría de Modernización Administrativa resuelve, implementar los mecanismos necesarios para "mejorar la gestión pública en términos de calidad y eficiencia, avanzando en el uso de herramientas tecnológicas que permitan una mejor utilización de la información de la Administración". Dicha resolución establece:

- Tabla de Plazos Mínimos de Conservación y Guarda de Actuaciones Administrativas.
- Procedimiento de Guarda Temporal, Archivo y Recuperación de Expedientes Electrónicos.
- Proceso de digitalización.

Este último proceso se explica en forma detallada en la siguiente sección.

2.5.4. Proceso de digitalización

El Anexo III de la Resolución 44-E/2016 de la Subsecretaría de Gestión Administrativa - Ministerio de Modernización [32] establece el: “Proceso de Digitalización de Documentos y Expedientes en Soporte Papel para su Archivo Electrónico e Incorporación al Repositorio Único de Documentos Oficiales – RUDO del sistema de Gestión Documental Electrónica - GDE” para que la documentación en papel sea correctamente digitalizada.

El mencionado procedimiento consta de 10 pasos y se transcribe a continuación:

- 1.- Revisar los documentos y expedientes: Para elegir y definir el medio apropiado para su escaneo, revise el documento/expediente respecto al estado físico, tamaño, plegado, encuadernación y calidad.
- 2.- Desarmar el expediente: Despojar la documentación que conforma el expediente de todo componente metálico, gomoso, y/o cualquier otro material que altere la documentación original o el escaneo.
- 3.- Seleccionar dispositivo (escáner): Seleccione el dispositivo que mejor se ajuste a las características del expediente, definido en el punto 1.
 - a) Si se trata de documentos/expedientes en buen estado de conservación, con características de tamaño de hoja (máximo A3/A4 según el dispositivo) y de impresión (simple/doble faz) similares para todas sus hojas, deberá escanearse el lote utilizando el mecanismo de alimentación automático.
 - b) Si el documento/expediente, tamaño máximo A3/A4 (según admita el dispositivo) por su estado de conservación debe ser tratado con cuidado, o el cuerpo documental presenta hojas de características disímiles, deberá usarse la pantalla fija del escáner.

c) Si el tamaño de alguna parte del documento/expediente es mayor al admitido en los dispositivos de escritorio, deberán utilizarse dispositivos de escaneo de gran formato.

d) Si el documento/expediente por el estado de conservación no admite su manipulación, deberá utilizarse un escáner de tipo “satelital”.

4.- Configurar el dispositivo (escáner) y verificar superposición de márgenes: Configure el dispositivo (escáner) de digitalización teniendo en cuenta los documentos/expedientes a ser escaneados (conforme la tarea definida en los puntos 1 y 2); según los valores indicados a continuación:

a) Verifique la adecuada superposición de márgenes. En caso de escaneo en secciones, se sugiere UN (1) centímetro de cada borde.

b) Valores estándar de configuración de los dispositivos de escaneo:

➤ Resolución: 150 ppp (píxeles por pulgada).

➤ Densidad: 300 dpi (puntos por pulgada)

c) Alimentación: manual, simple/doble faz, o automático, simple/doble faz (según lo definido en el punto 3).

d) Tamaño de la hoja: respetar el tamaño real de los documentos que componen el expediente.

e) Color: escala de grises 8 bits (256 niveles), salvo que en el expediente figuren textos en colores claros en cuyo caso deberá utilizarse la escala de color 32 bits.

5.- Escanear el documento/expediente y revisar el contenido: Escanee el documento/expediente con el fin de obtener una imagen íntegra y revise que todo el expediente fuera escaneado, sin omisión alguna. En caso de omisiones, deberán incorporarse las mismas.

6.- Comprobar el orden: Compruebe que se respetó el orden de todas las piezas documentales aun cuando se trate de notas adjuntadas o pegadas y que todas las hojas

estén orientadas en el mismo sentido. En caso de detectarse alguno de estos errores, re-escanear las partes que contengan el defecto, reemplazando la versión incorrecta por la correcta.

7.- Verificar la calidad de las imágenes: Verifique la calidad de las imágenes obtenidas de modo que todo el contenido del expediente se pueda apreciar correctamente.

a) Criterios generales de calidad a tener en cuenta:

- Legibilidad del expediente en general.
- Claridad en la captura de detalles (texto de tamaño tipográfico pequeño, diferenciación de comas, puntos decimales en las cifras numéricas, líneas segmentadas en gráficos y la existencia de textos con colores claros).
- Ausencia de manchas o motas presentes en el expediente original, producidas por el dispositivo de escaneo (por suciedad en la pantalla, por ejemplo).
- Completitud de la imagen, verificando especialmente la zona de bordes.
- Densidad de las zonas negras sólidas similar al documento original.

b) En caso de que el resultado obtenido no se ajuste a los criterios enunciados y se dificulte su visualización, optimice las imágenes seleccionando:

- Ajuste de contraste y/o brillo.
- Aumento de resolución.
- Cambio de color del escaneo.

8.- Almacenar temporariamente el documento/expediente en el dispositivo (computadora):

Almacene el expediente o cuerpo de documentos en un archivo digital con formato que crea conveniente (se aconseja pdf con texto, tif, pdf, jpg, jpeg, bmp, gif), en un directorio que pueda ser accedido mediante el Módulo Generador Electrónico de Documentos Oficiales (GEDO).

9.- Incorporar el documento/expediente digital al Repositorio Único de Documentos Oficiales (RUDO):

- a) Seleccionar el tipo de documento, cargar referencia y adjuntar documento: ingrese al módulo GEDO del GDE.
- b) Cargar datos o metadatos: En caso que corresponda, según el tipo de documento, cargue los datos o metadatos predefinidos para esta instancia.

Realizadas las acciones descriptas anteriormente, el documento/expediente obtenido quedará almacenado en el RUDO.

10.- Registrar datos del sistema en el documento/expediente en soporte papel que se remita al ARCHIVO GENERAL DE LA NACIÓN: En aquellos casos en que se remite el documento/expediente al ARCHIVO GENERAL DE LA NACIÓN, se deberá:

- a) Elaborar el Acta de Transferencia por el GEDO.
- b) Agregar el Acta de Transferencia al expediente en soporte papel.
- c) Digitalizar el documento/expediente con el Acta de Transferencia incluida.
- d) A fin de facilitar la búsqueda del documento/expediente digitalizado y almacenado en el RUDO, y permitir una sencilla auditoría de los mismos, registre en el documento/expediente en papel el número y fecha consignado por el sistema GDE al documento digital, y el nombre, apellido y firma ológrafa del operador que ha realizado la tarea o por cualquier otro medio de identificación que estuviere integrado con el módulo GEDO.

2.5.5. Firma digital

En la República Argentina la Ley N° 25.506 de Firma Digital sancionada el 14 de noviembre de 2001 [33], reconoce: "la eficacia jurídica del documento electrónico, la firma electrónica y la firma digital". En el artículo 48 de esta ley se

establece que: "el Estado Nacional, dentro de las jurisdicciones y entidades comprendidas en el artículo 8° de la Ley N°24.156, promoverá el uso masivo de la firma digital de tal forma que posibilite el trámite de los expedientes por vías simultáneas, búsquedas automáticas de la información y seguimiento y control por parte del interesado, propendiendo a la progresiva despapelización".

La ley establece un plazo máximo de 5 (cinco) años para su implementación en la totalidad de las leyes, decretos, decisiones administrativas, resoluciones y sentencias emanados de las jurisdicciones y entidades comprendidas en el artículo 8° de la Ley 24.156. En su artículo 49 establece también que: "El Poder Ejecutivo deberá reglamentar esta ley en un plazo no mayor a los 180 (ciento ochenta) días de su publicación en el Boletín Oficial de la Nación".

El anexo de la Ley N° 25.506 detalla varios puntos que a continuación se transcriben:

Procedimiento de verificación: proceso utilizado para determinar la validez de una firma digital. Este proceso debe considerar al menos:

- a) Que dicha firma digital ha sido creada durante el período de validez del certificado digital del firmante;
- b) Que dicha firma digital ha sido creada utilizando los datos de creación de firma digital correspondientes a los datos de verificación de firma digital indicados en el certificado del firmante;
- c) La verificación de la autenticidad y la validez de los certificados involucrados.

Datos de creación de firma digital: datos únicos, tales como códigos o claves criptográficas privadas, que el firmante utiliza para crear su firma digital.

Datos de verificación de firma digital: datos únicos, tales como códigos o claves criptográficas públicas, que se utilizan para verificar la firma digital, la integridad del documento digital y la identidad del firmante.

Dispositivo de creación de firma digital: dispositivo de hardware o software técnicamente confiable que permite firmar digitalmente.

Dispositivo de verificación de firma digital: dispositivo de hardware o software técnicamente confiable que permite verificar la integridad del documento digital y la identidad del firmante.

Políticas de certificación: reglas en las que se establecen los criterios de emisión y utilización de los certificados digitales.

Técnicamente confiable: cualidad del conjunto de equipos de computación, software, protocolos de comunicación y de seguridad y procedimientos administrativos relacionados que cumplan los siguientes requisitos:

1. Resguardar contra la posibilidad de intrusión y/o uso no autorizado;
2. Asegurar la disponibilidad, confiabilidad, confidencialidad y correcto funcionamiento;
3. Ser apto para el desempeño de sus funciones específicas;
4. Cumplir las normas de seguridad apropiadas, acorde a estándares internacionales en la materia;
5. Cumplir con los estándares técnicos y de auditoría que establezcan la Autoridad de Aplicación.

Clave criptográfica privada: En un criptosistema asimétrico es aquella que se utiliza para firmar digitalmente.

Clave criptográfica pública: En un criptosistema asimétrico es aquella que se utiliza para verificar una firma digital.

Integridad: Condición que permite verificar que una información no ha sido alterada por medios desconocidos o no autorizados.

Criptosistema asimétrico: Algoritmo que utiliza un par de claves, una clave privada para firmar digitalmente y su correspondiente clave pública para verificar dicha firma digital.

Recientemente la Ley 27446 - SIMPLIFICACIÓN Y DESBUROCRATIZACIÓN DE LA ADMINISTRACIÓN PÚBLICA NACIONAL dispuesta por el Senado y la Cámara de Diputados de la Nación Argentina el 18 de junio del 2018 [34] deroga varios artículos de la Ley 25.506 (artículos 4°, 18, 28, 35 y 36) y sustituye otros (artículo 10). Paso seguido, en su artículo 7° establece que: "los documentos oficiales electrónicos firmados digitalmente, expedientes electrónicos, comunicaciones oficiales, notificaciones electrónicas y domicilio especial constituido electrónico de la plataforma de trámites a distancia y de los sistemas de gestión documental electrónica que utilizan el sector público nacional, las provincias, el gobierno de la Ciudad Autónoma de Buenos Aires, municipios, poderes judiciales, entes públicos no estatales, sociedades del Estado, entes tripartitos, entes binacionales, Banco Central de la República Argentina, en procedimientos administrativos y procesos judiciales, tienen para el sector público nacional idéntica eficacia y valor probatorio que sus equivalentes en soporte papel o cualquier otro soporte que se utilice a la fecha de entrada en vigencia de la presente medida, debido a su interoperabilidad que produce su reconocimiento automático en los sistemas de gestión documental electrónica, por lo que no se requerirá su legalización" para todas las jurisdicciones y entidades contempladas en el artículo 8° de la ley 24.156.

La Universidad Nacional de Misiones se adhirió a esta iniciativa en el año 2012 y estableció para la misma su "Manual de procedimientos y recomendaciones para la Firma Digital" [35] en el cual se especifica:

- Ámbito de aplicación de la autoridad de certificación.
- Autoridad de Registro para la adquisición de dispositivo criptográfico.
- Documentación necesaria para el proceso de solicitud del certificado de la firma.
- Revocación digital del certificado.
- Procedimiento para la solicitud de certificados.
- Procedimiento de revocación de los certificados.

2.5.6. Normas ISO para la gestión de documentos electrónicos

Las normas ISO⁶ pueden ayudar a las organizaciones a implementar las mejores prácticas en un sistema de gestión de documentos electrónicos. Durante la última década se han publicado diferentes estándares abarcando variadas perspectivas y requisitos que las organizaciones deben observar para la correcta gestión de los registros [36]. A continuación se presentan las principales normas.

Respecto a la gestión de documentos se tiene:

- ISO 15489 Gestión de documentos.
- ISO 16175 Principios y requisitos funcionales para documentos en entorno de oficina electrónica.
- ISO 30300 Sistemas de gestión para documentos. Fundamentos y vocabulario.
- ISO 30301 Sistemas de gestión de documentos. Requisitos.
- ISO 30302 Sistemas de gestión de documentos. Guía de implementación.

Las normas específicas para la digitalización:

- ISO 13028 Directrices para la implementación de la digitalización de documentos.
- ISO 15801 Información almacenada electrónicamente. Recomendaciones sobre veracidad y fiabilidad.

Si bien en este trabajo se estudiaron las normas antes mencionadas, existen otras las cuales son importantes tenerlas en cuenta como ser las referidas a:

Seguridad:

- ISO 27001 Sistemas de gestión de la Seguridad de la Información.
- ISO 27002 Código de buenas prácticas para la gestión de la seguridad de la información.

Depósito y almacenamiento:

⁶ ISO: International Organization for Standardization - Organización Internacional para la Estandarización

- ISO 23081 Metadatos para la gestión de documentos.
- ISO 19005 Formato de fichero de documento electrónico para la conservación a largo plazo.
- ISO 15836 Conjunto de elementos de metadatos. Dublín Core.
- ISO 25964 Tesoros e interoperabilidad con otros vocabularios.

Transferencia y migración:

- ISO 13008 Proceso de migración y conservación de documentos electrónicos.
- ISO 17933 GEDI. Intercambio de documentos electrónicos genéricos.
- ISO 14721 Sistemas de información de archivos abiertos.

Proceso

- ISO 26122 Análisis de los procesos de trabajo para la gestión de documentos.

En la Figura 8 se ilustran las normas ISO relacionadas a la gestión de documentos electrónicos.

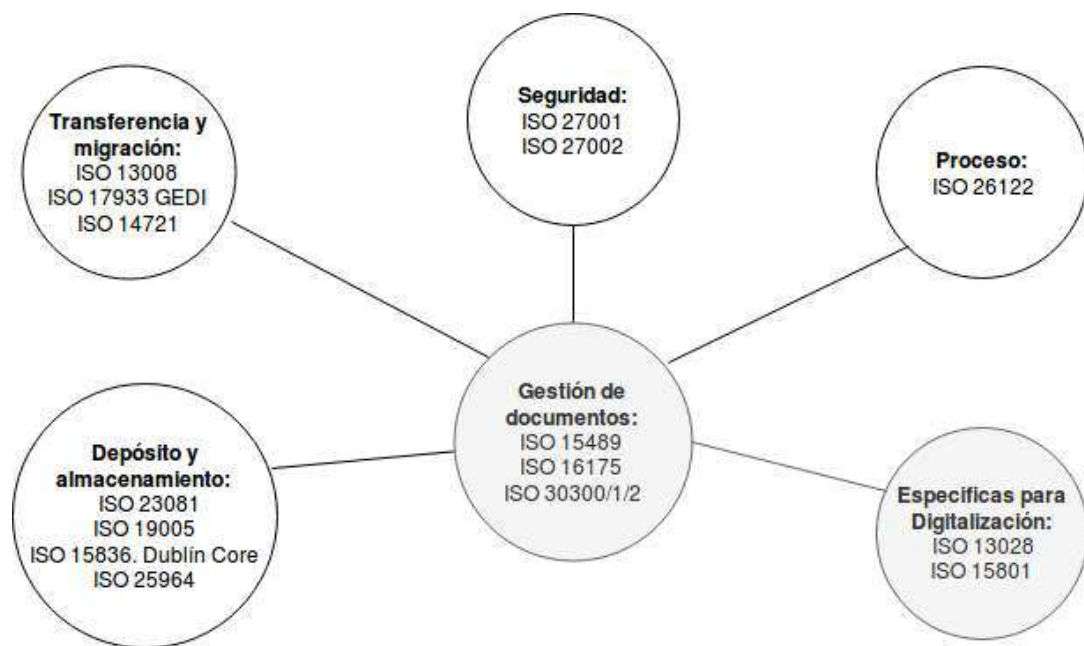


Figura 8. Normas ISO para la Gestión de Documentos Electrónicos

Específicamente la norma ISO 15489 [37], presentada en Ottawa (Canadá) en el año 2000 fue uno de los pilares respecto a la importancia y reconocimiento de la “Gestión Documental”. La misma detalla cómo deben gestionarse los documentos en un sistema de archivos, teniendo como premisa la mejora en la eficacia y eficiencia de los organismos.

Esta norma ha recibido reconocimiento internacional, en la misma se menciona que la gestión de documentos en una institución debería incluir:

- a) la confección de políticas y normas.
- b) la asignación de responsabilidades y competencias.
- c) el establecimiento y difusión de procedimientos.
- d) servicios relacionados con su gestión y uso.
- e) el diseño, la implementación de sistemas especializados de gestión de documentos.
- f) la integración de la gestión documental en los procesos de la organización.

Las organizaciones deberán crear y mantener documentos genuinos, fiables y proteger la integridad de los mismos. Para lograr esto, se deben establecer y llevar a cabo un programa de gestión documental en donde:

- a) se determinen los documentos que deberían ser creados en cada circuito administrativo, y la información que es necesaria incluir.
- b) se decida el modo y la estructura en que la documentación debe ser creada e incorporada al sistema, así como las tecnologías que deberán utilizarse.
- c) se definan los metadatos que describen a cada documento y a los procesos relacionados con el mismo; y cómo se vincularán y gestionarán los metadatos. La norma ISO 23081 [38], es una guía para entender e implementar estos metadatos.
- d) se establezcan los requisitos para recuperar, utilizar, preservar y transferir documentos entre los usuarios del sistema y los procesos administrativos.

- e) se resuelva cómo estructurar y organizar los documentos para de esta forma satisfacer los requisitos necesarios para su correcta utilización.
- f) se evalúen los riesgos que se derivarían de la ausencia de documentos que testimonien las actividades realizadas.
- g) se conserven los documentos y se permita el acceso a lo largo del tiempo, con el objeto de satisfacer las necesidades y expectativas de la sociedad.
- h) se cumplan los requisitos legales, reglamentos, normas y la política de la organización.
- i) se garantice la conservación de los documentos durante el período de tiempo establecido.
- j) se evalúen las posibilidades para mejorar la eficacia y eficiencia de los procesos de la gestión de los documentos.
- k) se implemente un programa de capacitación para el personal involucrado en la gestión documental.

2.6. Antecedentes sobre replicación de datos

En la Facultad de Informática de la Universidad Nacional de la Plata se encuentra el Instituto de Investigación en Informática, que en colaboración con la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC) llevaron a cabo el proyecto de investigación en Sistemas de Software Distribuidos y Bases de Datos Distribuidas [39].

Entre algunos trabajos realizados se encuentran:

- Un proyecto FONTAR de la Agencia Nacional de Promoción Científica y Tecnológica. Este es un trabajo realizado para la Facultad de Veterinaria de la Universidad Nacional de La Plata que se basa en el análisis y diseño del un banco de datos genético, centralizado y distribuido, de ganado vacuno y equino de la provincia de Buenos Aires, para la trazabilidad de los ejemplares.

El sistema desarrollado permite la interacción con los potenciales usuarios (sitio WEB) que solicitan a la unidad de procesamiento de los análisis a realizar, envían las muestras a procesar, consultan sobre el estado de avance de la solicitud y reciben los resultados. Luego otro sistema (local) interactúa con un secuenciador que analiza las muestras de ADN y envía los resultados.

- La ampliación y escalabilidad del sistema del Museo de Ciencias Naturales de la Universidad Nacional de la Plata que integra la Red Nacional de Colecciones Biológicas. Como resultado de este proyecto se diseñó un sitio WEB que permite el tratamiento de bases de datos heterogéneas y físicamente distribuidas. La diversidad biológica de los elementos a almacenar, la localización física de los mismos y la distribución de los usuarios, conforman un escenario que ha requerido la implementación de un sistema distribuido de características complejas. Actualmente se está trabajando sobre 35 colecciones que involucran a 2.400.000 registros.
- El análisis, diseño e implementación de un sistema de información para la gestión estratégica de recursos humanos del Ministerio de Seguridad de la Provincia de Buenos Aires. Esta línea de investigación, también forma parte del Instituto de Investigación en Informática de la Universidad Nacional de la Plata, y es un sistema de información que contempla los cuerpos policiales de dicha provincia contemplando:
 - Policías comunales
 - Policías de distrito
 - Policías de custodia de objetivos fijos
 - Policías de traslado de detenidos
 - Policías de investigaciones sobre tráfico de drogas ilícitas
 - Central de atención telefónica de emergencias
 - Cuerpo de baqueanos.

Con este sistema es posible dar tratamiento a los datos de todos los regímenes estatutarios sirviendo, de este modo, a la gestión de unidades organizativas centralizadas y descentralizadas del Sistema de Seguridad Pública.

Otro estudio hallado fue la “Replicación de bases de datos en modalidad Maestro Esclavo, caso de estudio: *Firebird SQL Server*”. Es un estudio específico sobre *FBJReplicator* que utiliza el gestor de base de datos de *Firebird* para replicación en diferentes sistemas operativos. Dispone de librerías que permiten el desarrollo de *drivers* para estos entornos de desarrollo, posibilitando desarrollar aplicaciones clientes servidor. La versión 2.5 incorpora el algoritmo 2PC haciendo posible la ejecución de comandos SQL en forma remota y luego verificar del impacto de la transacción distribuida. Esto permite la implementación sistemas de base de datos distribuidas basado en disparadores para la replicación de tablas en varios servidores *Firebird/Interbase* en la modalidad maestro esclavo [40].

A nivel Sud América también se han relevado varias experiencias, entre ellas la utilización de *Symmetric* para la replicación de datos en la plataforma educativa ZERA.

La plataforma educativa ZERA perteneciente al Centro de Tecnologías para la Formación (FORTES), en la Universidad de las Ciencias Informáticas de Cuba (UCI) fue desarrollada teniendo presente las necesidades educativas existentes en varios países de Latinoamérica. Incluye la gestión de los programas de estudios, periodos lectivos y todo el sistema de evaluación con el que va a contar la entidad educativa. Brinda la posibilidad al usuario de continuar sus actividades en la plataforma fuera de la institución, situación que se dificulta en establecimientos donde no se puede asegurar la conectividad a los servidores. Para resolver esto, la información es almacenada en bases de datos locales conectadas a través de la red a una base de datos central, formando así un sistema de base de datos distribuidas. Teniendo en cuenta los volúmenes de datos que son generados por las distintas instituciones, surge la necesidad de utilizar herramientas como *SymmetricDS*, para la optimizar la replicación de datos. La correcta identificación de las tablas a replicar posibilita la adecuada configuración, asegurando la consistencia de los datos y disminuyendo los errores en la replicación [41].

Otra experiencia es el “Proceso de réplica de datos con *Microsoft SQL Server* para el replicador de datos Reko”. El replicador de datos Reko es un *software* que

surge debido a la necesidad de la distribución de los datos en el Sistema de Gestión Penitenciario Venezolano (SIGEP) durante el año 2007. Reko cubre las principales necesidades relacionadas con la distribución de datos entre los gestores de bases de datos como *PostgreSQL*, *MySQL* y *Oracle* en la protección, recuperación, sincronización y transferencia de datos entre los distintos sitios de las diferentes localidades. Como ejemplos de su utilización se pueden mencionar:

- Sistema DATAFAR, de las Fuerzas Armadas Revolucionarias de la República de Cuba
- Sistema de Gestión Hospitalaria de PDVSA
- Sistema Nacional de Gestión de Ingreso a la Educación Superior Universidad de las Ciencias Informáticas de Cuba (UCI).

Actualmente Reko no permite la replicación de datos utilizando el gestor de base de datos *Microsoft SQL Server*, por lo cual, se reduce su utilización en los sistemas que requieren el empleo de este gestor como es el caso del Sistema Nacional de Gestión de Ingreso a la Educación Superior. Para poder efectuar la réplica en estos gestores se recurren a métodos tradicionales, como la copia de *backups* y *scripts* hacia las distintas bases de datos, lo que causa una gestión de la información lenta y los datos corren el riesgo de poseer errores de coherencia y consistencia. En [42] se desarrolló el dialecto para el replicador de datos Reko y se agregaron funcionalidades para sincronizar, enviar y recibir datos en ambientes con conexión y sin conexión utilizando el gestor de bases de datos *Microsoft SQL Server* [42].

Otro de los trabajos relevados, presenta un modelo de servicios WEB, para soportar replicación de instantáneas en múltiples motores de bases de datos. Utiliza un catálogo maestro de replicación (CMR) para almacenar los datos a replicar. Se utiliza la herramienta *Mercury* que ayuda en el monitoreo de las bases de datos distribuidas. Empresas como *Oracle Corp.*, *Informix* y el *ASK Group*, incluyeron agentes *Mercury* en sus productos. Sin embargo, esta herramienta parte de que los servicios de replicación ya estén definidos en el motor de base de datos. En [43] se logró modelar y desarrollar una herramienta basada en cuatro componentes

principales: un conjunto de servicios WEB, dos agentes, un catálogo maestro de replicación y una consola de administración que trabaja de manera independiente a las aplicaciones que deseen implementar la replicación de datos. Para su implementación se debe seguir un conjunto de lineamientos como por ejemplo, seleccionar las tablas de fácil replicación con tipos de datos del estándar *SQL99* para garantizar un correcto mapeo de los datos [43].

Capítulo 3

Gestión de Legajos.

Situación Actual en la Facultad de Ciencias Económicas

3. Gestión de Legajos. Situación Actual en la Facultad de Ciencias Económicas

En este capítulo, inicialmente se describen cuáles son las características de la institución en donde se pretende implementar la solución propuesta, y la importancia del Programa de Expansión territorial dentro del Plan de Desarrollo Institucional.

Luego, se analiza la arquitectura actual de los sistemas informáticos, referente a la gestión de los legajos y cuáles son las tareas involucradas en el circuito administrativo.

Por último, se realiza un resumen sobre los principales inconvenientes detectados.

3.1. Características de la Institución

La UNaM, es una institución universitaria de derecho público, autónoma en lo académico e institucional y autárquica en lo económico y financiero. Adopta una organización regionalizada en la jurisdicción provincial y de concepción operativa descentralizada [44].

Impulsa la integración e interrelación con instituciones afines, gubernamentales y no gubernamentales de la provincia y la región, nacionales e internacionales que compartan o coincidan con sus fines y objetivos.

Constituye una comunidad en la búsqueda de la excelencia y calidad institucional sobre la base de la formación y capacitación permanente de sus integrantes.

3.1.1. Objetivos de la Institución

Son fines permanentes de la Universidad Nacional de Misiones:

- La preservación, promoción y difusión de la cultura universal con énfasis en lo nacional y regional.
- El resguardo, acrecentamiento y difusión del conocimiento universal y del generado en su propio ámbito.
- La organización, instrumentación y evaluación de la enseñanza-aprendizaje en los niveles de su competencia y su articulación con los otros sectores del sistema educativo.
- La aplicación del conocimiento a la solución de problemas del desarrollo humano en la provincia, la región y el país.
- El compromiso con la conservación y preservación del medio ambiente y los recursos naturales.
- El de constituirse en un ámbito de formación ciudadana y ejercicio democrático.

Para el cumplimiento de sus fines la Universidad Nacional de Misiones desarrolla como funciones:

- Formar y capacitar a sus integrantes en todos los ámbitos y niveles de actuación, para lograr su adecuación a los avances del conocimiento, de acuerdo con las demandas individuales y los requerimientos nacionales y regionales.
- Facilitar el desarrollo equilibrado de las especialidades existentes y la promoción de aquellas que respondan a los intereses sociales de la provincia, la región y el país.
- Fomentar e impulsar el compromiso solidario con la sociedad mediante un accionar eficaz y eficiente en la prestación de servicios dentro del ámbito de su competencia.

- Proveer y sostener la acción social en la comunidad universitaria a efectos de garantizar los principios de equidad e igualdad de oportunidades.

Si se analizan en detalle estas funciones, vale destacar que muchas de ellas, como así también sus fines, están relacionadas y serian facilitados por la implementación de una arquitectura informática como la que se plantea en este trabajo.

3.1.2. Programa de Expansión Territorial

En el mes de octubre del año 2018, con la participación de representantes de las distintas unidades académicas, se aprobó por resolución número 081/2018, el Plan de Desarrollo Institucional 2018-2026 (PDI) de la Universidad Nacional de Misiones [45] con el objetivo de establecer las políticas institucionales para el planeamiento y la gestión en las Unidades Académicas, Escuelas e Institutos que de ella dependen.

En el mismo se detalla la particularidad de una administración descentralizada y desplegada en toda la provincia: quedando la Universidad Nacional de Misiones compuesta por seis facultades, una escuela y el rectorado. En la ciudad de Posadas se encuentran las facultades de Ciencias Económicas; Ciencias Exactas, Químicas y Naturales con la Escuela de Enfermería, la Facultad de Humanidades y Ciencias Sociales y el Rectorado. En la ciudad de Oberá, las facultades de Arte y Diseño, e Ingeniería; y en Eldorado la sede de la Facultad de Ciencias Forestales junto a la Escuela Agrotécnica Eldorado, esta última de nivel medio. En dicho PDI y a partir de su Programa de Expansión Territorial se amplía la presencia de la Universidad Nacional de Misiones con otras ofertas académicas en las localidades de San Pedro, San Vicente, Santa Ana, Puerto Iguazú, Puerto Rico, San Antonio, Leandro N. Alem y San Vicente (Figura 9).

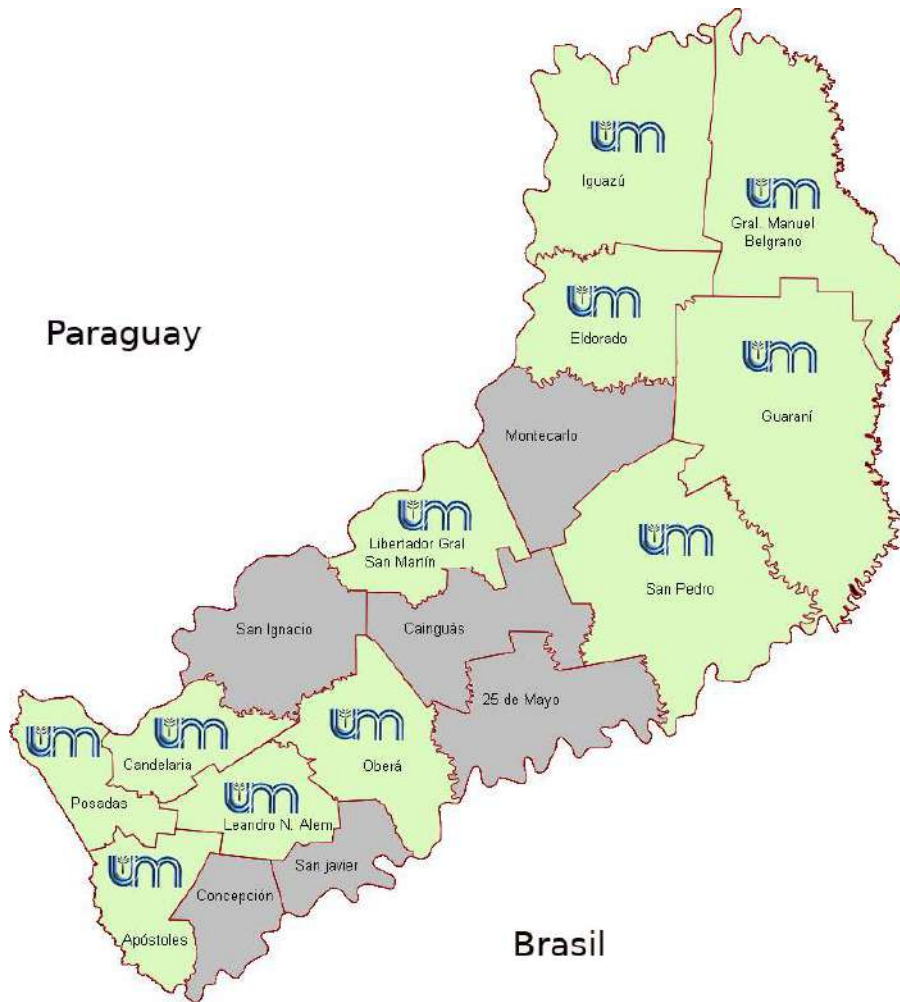


Figura 9. Presencia de la Universidad Nacional de Misiones en la provincia de Misiones

EL Programa de Expansión Territorial [1] fue creado por resolución N°1366 de la SPU⁷ y aprobado en acuerdo plenario N°802/11 CIN⁸. El mismo tiene la finalidad de “incrementar las oportunidades de acceso, permanencia y egreso a la educación superior, extendiendo la cobertura territorial para la atención de las necesidades y demandas de formación superior”.

El programa tiene como objetivos ofrecer carreras en áreas prioritarias para el desarrollo regional del noreste del país y:

⁷ SPU: Secretaría de Políticas Universitarias

⁸ CIN: Consejo Interuniversitario Nacional

- Contribuir al acompañamiento, fortalecimiento y ampliación de las propuestas de formación académica y profesional, que ofrece la Universidad, a través de las sedes de Iguazú y San Vicente.
- Alcanzar mayores niveles de articulación y coordinación, criterios de trabajo y líneas de acciones comunes, entre las sedes y carreras.
- Incrementar y favorecer las posibilidades de impacto del Sistema de Educación Superior universitario, en el desarrollo regional y local.
- Coordinar y acompañar, conjuntamente con las autoridades académicas de las distintas facultades, las acciones desarrolladas en el marco de este Programa, informando periódicamente al Consejo Superior y al Rector de sus avances, proyecciones y necesidades.

3.2. Arquitectura Actual de la Gestión de Legajos

Para la gestión de los legajos los principales sistemas involucrados son:

SIU Guaraní: permite administrar la gestión académica de los alumnos. También contempla servicios para los docentes, personal No Docente y autoridades del Sistema Universitario Nacional [46].

En la Facultad de Ciencias Económicas este sistema posee las siguientes características:

Versión de sistema: 2.9.3

Lenguaje de programación: Power Builder

Motor de BD: Informix

SIU Pre Inscripción: este sistema permite que los aspirantes a cursar una carrera en la universidad puedan pre inscribirse a la misma vía WEB. Posee un módulo que permite consultar el grado de avance sobre el estado de las pre inscripciones y algunas consultas estadísticas. El objetivo más importante de este sistema es permitir al SIU Guaraní recuperar los datos de cada aspirante mediante un código de barras

impreso en la ficha que emite como resultado del registro exitoso de la pre inscripción [47].

En la Facultad de Ciencias Económicas este sistema posee las siguientes características y la arquitectura del circuito se exhibe en la Figura 10:

Versión de sistema: 2.9.2

Lenguaje de programación: PHP

Motor de BD: *PostgreSQL*

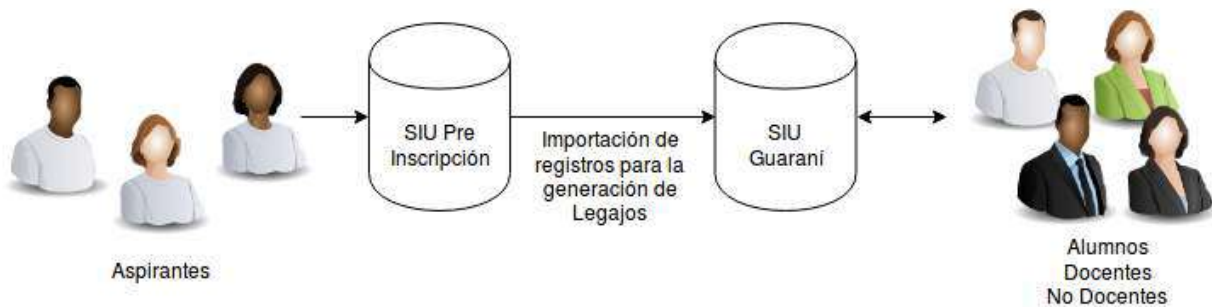


Figura 10. Arquitectura actual sobre la gestión de legajos

3.3. Circuito Administrativo Relevado

Actualmente, el trámite de inscripción de un alumno se realiza íntegramente en la Dirección de Estadísticas de la facultad. Los aspirantes que desean cursar una carrera en la facultad, antes deben pre inscribirse vía WEB al SIU Pre Inscripción (<http://preinscripcion.fce.unam.edu.ar>). Allí el aspirante debe cargar los datos solicitados en el sistema para luego dirigirse personalmente a la Dirección de Estadísticas con los requisitos solicitados para finalizar su pre inscripción.

Una vez que en la Dirección de Estadísticas se verifica y completa el registro de pre inscripción y, si el aspirante posee toda la documentación obligatoria para la inscripción a la carrera, se procede con la grabación del registro en el SIU Guarani, pasando el aspirante del estado “Pre Inscripto” al estado “Inscripto”.

Si el aspirante no cumple o no puede completar en el acto de pre inscripción con alguna documentación o requisito obligatorio, la grabación en el SIU Guaraní no se realiza y el aspirante continúa en su estado de pre inscripto o como se lo denomina en la facultad, queda como un “Alumno Condicional”.

La documentación con la cual el aspirante debe cumplir para pasar a ser un “Inscripto”, y con la cual el legajo quedará inicialmente conformado es:

- Fotocopia del DNI.
- Partida de nacimiento (original y copia).
- Título del nivel medio o polimodal legalizado, original y copia, o constancia de título en trámite, copia original, con un máximo de dos materias pendientes de aprobación, situación que deberá superarse indefectiblemente antes del 31 de agosto del año de ingreso.
- Una foto 4x4.
- Extranjeros: pasaporte con visa (original y copia).
- Una Carpeta colgante.
- Formulario de pre inscripción impreso.

Cabe aclarar que la Universidad Nacional de Misiones no retiene ninguna documentación original, salvo la constancia de título en trámite, la cual luego de ser exhibida a los agentes de la Universidad Nacional de Misiones es devuelta al propietario, quedando almacenada en el legajo solo la copia de cada ejemplar.

Finalmente y luego de ser grabado, el legajo del alumno se almacena físicamente en estanterías con carpetas colgantes (Figura 11).

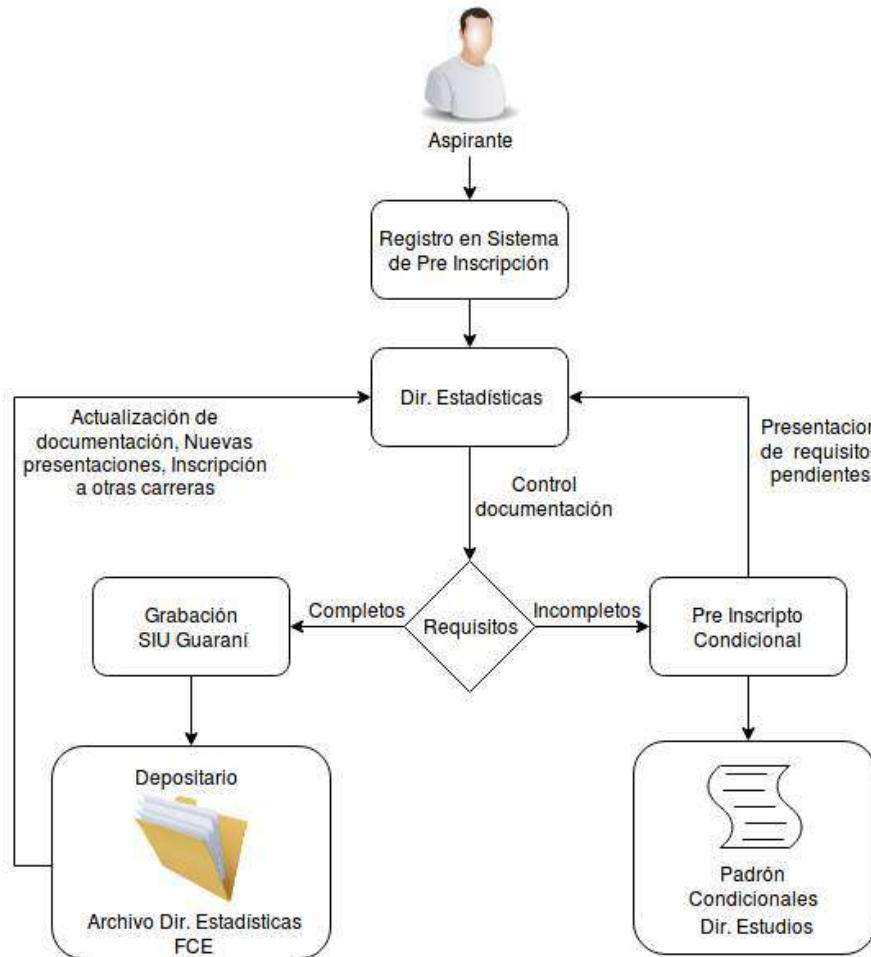


Figura 11. Circuito administrativo actual sobre la gestión de legajos

3.4. Resumen de la problemática

Básicamente, cuando algún agente No Docente o autoridad necesita consultar y/o actualizar el legajo de un alumno necesariamente debe dirigirse físicamente a la Dirección de Estadísticas y solicitar al personal que se encuentre en ese momento. Luego, esperar que se busque el legajo físicamente en las estanterías de carpetas colgantes, para finalmente gestionar sobre el mismo. Todo esto, con previa autorización de la directora del área la cual no queda registrada, salvo que el legajo indefectiblemente deba salir de la Dirección de Estadísticas. Esta situación es crítica, sobre todo cuando la necesidad surge de alguna extensión áulica ubicada en el

interior de la provincia. Debido a este modo de operar, también han ocurrido casos de pérdidas de legajos.

No se posee acceso concurrente a la información de los legajos, esto obligatoriamente genera que todos los trámites, por ejemplo expedición de título, sean secuenciales. Esto por supuesto genera muchas más demoras, comparando con la hipotética posibilidad de que cada sector autorizado pueda acceder a la información de los legajos.

También existen cuestiones técnicas no resueltas. Una de ellas es la imposibilidad de gestionar o modificar los registros en el sistema de pre inscripción. Esto provoca que cuando un aspirante carga en forma incorrecta algún dato, lo cual suele pasar regularmente, el mismo no pueda ser corregido desde el sistema. Para solucionar estos casos el personal de la Dirección de Estadísticas debe iniciar un incidente para que el personal de la Dirección de Tecnología para la Gestión que gestiona la BD realice la corrección o en su defecto, debido a que muchas veces se desconoce la causa del inconveniente, se proceda a la eliminación del registro para que el aspirante vuelva a cargar toda la información nuevamente al sistema.

No existe un procedimiento respecto a la gestión y cumplimiento de requisitos que los alumnos deben cumplir. El control de consistencia entre lo que indica el SIU Guaraní y las carpetas físicas de los legajos de los alumnos se realizan en forma manual, legajo por legajo. Si bien el sistema ofrece pistas de auditoría sobre la información manipulada, no puede asegurarse que la misma sea fidedigna respecto a los legajos físicos.

Otro de los principales inconvenientes es la falta de espacio físico para situar los archivadores de los legajos. Debido a que en la oficina de la Dirección de Estadísticas ya no hay más lugar, muchas cajas con legajos, cohortes nuevas, fueron depositadas arriba de los mismos archivadores en cajas sueltas (Figura 12) y otras, cohortes más antiguas, fueron almacenadas en otro depósito situado cerca del aula magna y hasta en los depósitos del ex comedor de la facultad. El resultado de esto, al no ser un lugar preparado para albergar documentación, es un deterioro acelerado de los legajos allí situados (hojas amarillentas y con manchas de hongos).



Figura 12. Criticidad en el espacio físico

Debido a todo lo antes expuesto, es necesario realizar una re ingeniería de los sistemas y del circuito administrativo, cuestión que los propios agentes de la facultad y autoridades han reconocido y solicitado.

Capítulo 4

Modelado de la solución

4. Modelado de la solución

Como ya se mencionó, no existe una solución basada en TIC para dar soporte a este plan de expansión. En este capítulo se plantea un caso para la gestión de los legajos de los alumnos en las distintas dependencias de la universidad utilizando técnicas de replicación de datos.

En este capítulo se propone un modelo para la implementación del Legajo Digital (LeDig) de los alumnos y su respectivo circuito administrativo como método para la gestión integral de los legajos. También se muestra la arquitectura para facilitar tanto la administración como la consulta, de manera concurrente, de los legajos. Por último se propone una solución a la escalabilidad respecto al resguardo, distribución y seguridad de la información.

4.1. Recursos disponibles

Debido a que el proyecto fue instruido por el Honorable Consejo Directivo de la Facultad de Ciencias Económicas y que los sistemas involucrados residen y operan en la Dirección de Tecnología para la Gestión, se posee libre acceso a estos y sus respectivas BD.

Para lograr la alta disponibilidad de los diferentes sistemas de la Universidad Nacional de Misiones y ponerlos al servicio de las expansiones territoriales, es necesario proporcionar la replicación de los sistemas y sobre todo de los datos. Una forma de proporcionar esta redundancia es mediante el uso de métodos de replicación de BD, como se detalló en el capítulo 2, sección 2.4.3. Replicación de datos.

El consorcio SIU [48] es proveedor de los sistemas informáticos, la instalación de los mismos es relativamente sencilla y se encuentra correctamente documentada [49]. En su totalidad están desarrollados con el *framework* basado en PHP llamado SIU Toba [50]. Este entorno de trabajo está desarrollado y es

distribuido a las universidades por el mismo consorcio SIU, utilizando como motor de BD *PostgreSQL* [51].

Debido a lo antes mencionado, teniendo en cuenta cuestiones de compatibilidad y pericia adquirida por el personal de la Dirección de Tecnología para la Gestión sobre los mismos, la solución propuesta será desarrollada teniendo en cuenta estos aspectos.

4.2. Metodología

La investigación llevada a cabo, se encuentra dentro del tipo “Investigación Aplicada” y su diseño se encuentra definido en las siguientes fases:

- Relevamiento de la situación actual y definición de los requisitos que se deben atender.
- Análisis sobre la digitalización de la documentación y fundamentalmente determinar cuál es el mejor modo de almacenar los mismo en una base de datos.
- Modelo de Datos LeDig: se procederá a modelar la arquitectura de los datos y el desarrollo e implantación del sistema, contemplando la importación de los datos con el SIU Guaraní y el almacenamiento de los legajos digitales en el Sistema de Archivos.
- Desarrollo del software: Luego de analizar las distintas alternativas y viabilidad, se desarrollará un prototipo, con la finalidad de facilitar el relevamiento de requisitos, comenzando con la experimentación, digitalización de los legajos y abordaje de la reducción del archivo físico, crítico para la unidad académica. El prototipo deberá contemplar: interfaz con el Sistema de Pre Inscripción, re ingeniería del circuito administrativo en la Dirección de Estadísticas, resguardo y digitalización de los legajos.
- Análisis y selección de las tablas a replicar, creación de los nodos e implementación del *clúster* de replicación.

- Esquema de Seguridad, Auditoría y Recuperación: Debido a que el resguardo, seguridad e integridad de la información son aspectos críticos, deberán estudiarse y definirse los mecanismos necesarios, *backups*, pista de auditoría, recuperación ante incidentes, para su disponibilidad.
- Validación de los resultados: Verificación de las herramientas software desarrolladas y del funcionamiento del sistema con los usuarios finales. En forma conjunta con las autoridades de la unidad académica, se definirán las pautas y los períodos necesarios para realizar la validación de todo lo expuesto. Finalmente el trabajo concluirá con un análisis de los resultados esperados contrastados con la situación actual.

Cada una de estas actividades se desarrollará implementando elementos de metodologías ágiles y utilizando algunos elementos de Kanban [52]. Ambos implementados en partes con la herramienta de gestión de proyectos Redmine [53], que habitualmente es utilizada por el equipo de la Dirección de Tecnología para la Gestión.

Algunos de los roles definidos en este sentido fueron:

- **Usuarios:** Personal del Área de la Dirección de Estadísticas – Facultad de Ciencias Económicas.
- **Stakeholders:** Autoridades de la Facultad de Ciencias Económicas (Decana - Vicedecano).
- **Product Owner:** Secretaria Académica.
- **SCRUM Team:** miembros del equipo de la Dirección de Tecnología para la Gestión de la Facultad de Ciencias Económicas.

Los sectores involucrados reconocen la problemática y tienen voluntad para trabajar en equipo y promover un cambio y la Dirección de Tecnología para la Gestión cuenta con el equipamiento y el personal calificado para llevar adelante el proyecto.

Los integrantes del equipo trabajan en el mismo lugar, tienen buena relación, tienen experiencia en la ejecución de proyectos y constituyen un equipo multidisciplinario.

No solo que se cuenta con el apoyo de la institución para llevar a cabo el proyecto, también está señalado como prioritario por la gestión actual, solicitado el 12 de septiembre del año 2017 por la Secretaría Académica, e instrumentado por el Honorable Consejo Directivo de la Facultad de Ciencias Económicas a través de la resolución número 190/2017 [2].

4.3. Circuito administrativo propuesto

Si bien se constató la inexistencia de un circuito administrativo formal para dicha gestión, el relevamiento realizado en el capítulo anterior ayudó a identificar claramente los aspectos a tener en cuenta para el desarrollo de la solución y los aspectos críticos a mejorar.

A continuación se describe el circuito administrativo para realizar la gestión de los legajos y así dar una solución a los inconvenientes detectados y descriptos anteriormente.

1. Registro en Sistema de Pre Inscripción: en esta etapa el aspirante que desea ser alumno de la facultad debe completar la ficha de pre-inscripción, la misma se encuentra disponible en el siguiente enlace: preinscripcion.fce.unam.edu.ar. Luego de validar todos los datos, el sistema permitirá al aspirante imprimir una ficha, equivalente a la anterior ficha impresa SUR-1⁹, que luego deberá llevar y presentar en la Dirección de Estadísticas, junto con los demás requisitos necesarios.
2. Dirección de Estadísticas: una vez que el aspirante se presenta con la documentación requerida, la Dirección de Estadísticas realizará el control de cada requisito verificando su completitud y validez. Si los requisitos están completos se procede con la etapa de Digitalización (LeDig), de lo contrario el aspirante pasa a ser un inscripto “Condicional” hasta que logre cumplimentar con todos los requisitos obligatorios. Con los pre inscriptos que no completaron la documentación se conforma un “Padrón de Condicionales”,

⁹ SUR-1: Ficha de pre inscripción emitida por el sistema

que luego será remitido a la Dirección de Estudios para su posterior tratamiento.

3. Digitalización LeDig: luego, si el aspirante cumple con todos los requisitos para ser alumno de la entidad, el personal de la Dirección de Estadísticas procederá a digitalizar toda la documentación presentada, foliará y luego devolverá la carpeta (legajo físico) al alumno. El circuito también contempla la posibilidad de que los alumnos realicen actualizaciones y nuevas presentaciones. En este caso los alumnos que presente o actualicen su documentación deberán traer su legajo (carpeta) a la Dirección de Estadísticas. El personal de dicha área deberá: foliar la nueva documentación en orden correlativo a la última presentación realizada, digitalizar y anexarla al legajo físico. La finalidad de este procedimiento es asegurar que la versión del legajo digital sea idéntica a la del legajo físico.
4. Firma Digital: Debido a que mayoritariamente el legajo está compuesto por fotocopias de la documentación original, esta etapa se marca en la Figura 13 como opcional. Pero si luego de la digitalización de los documentos, existe documentación que requiera ser rubricada, el agente de la facultad realiza la firma digital de la documentación del legajo que requiera esta acción.
5. Grabación SIU Guaraní: esta etapa no sufre ninguna modificación respecto al circuito anterior, simplemente se recupera la información del sistema de pre inscripción a través del código de barras impreso en la ficha, se graba y genera el legajo del alumno en el SIU Guaraní como habitualmente se lo venía realizando.
6. Depositorio: necesariamente esta etapa ya fue descripta en la fase de digitalización, y es el mayor cambio de paradigma, al menos para el personal de la Dirección de Estadísticas, ya que se plantea que, luego de que la presentación sea digitalizada, firmada y garantizada la correspondencia entre el legajo digital y el legajo físico, en el mismo acto de presentación de la documentación, el legajo físico será devuelto al alumno quedando el mismo como depositario.

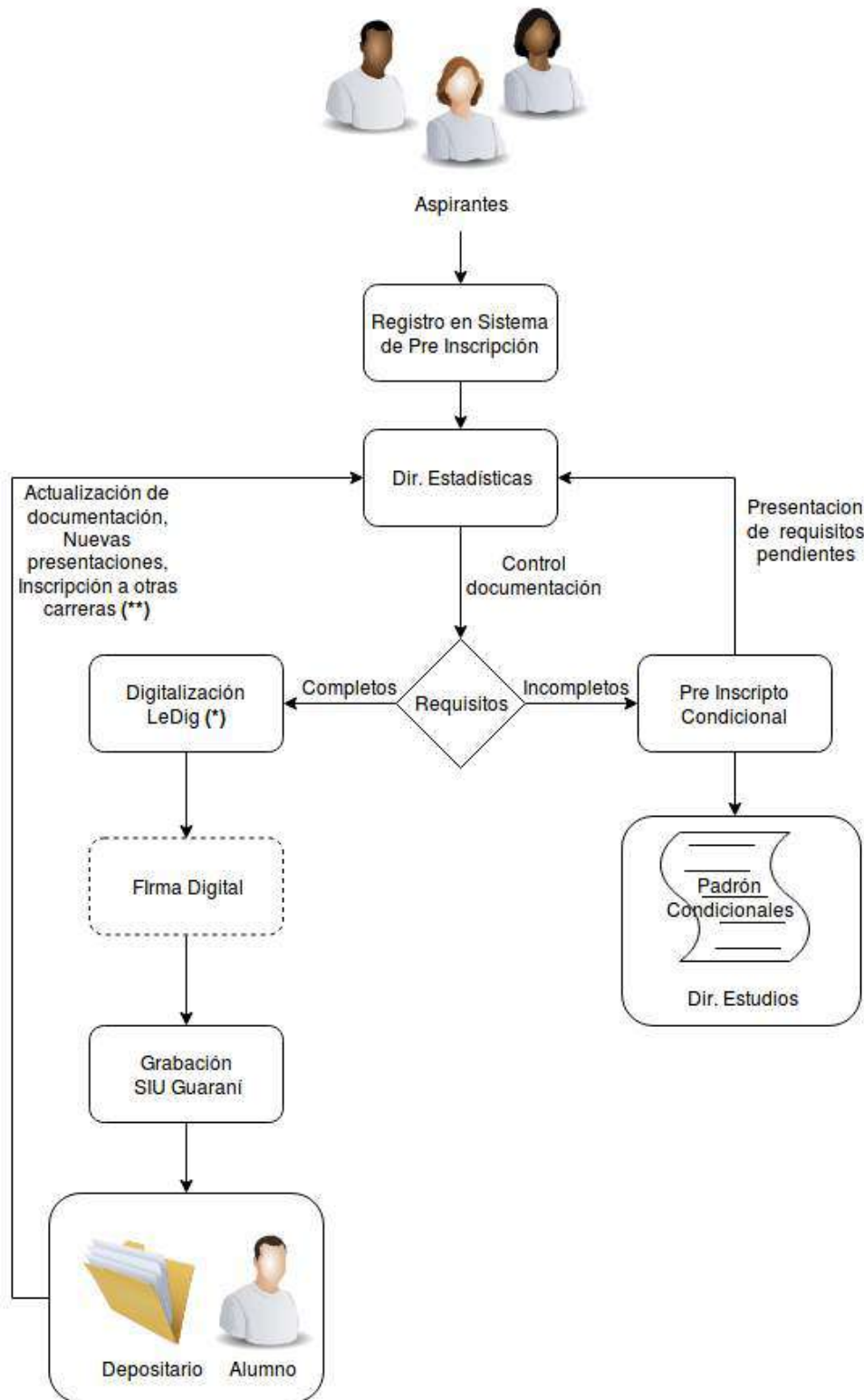
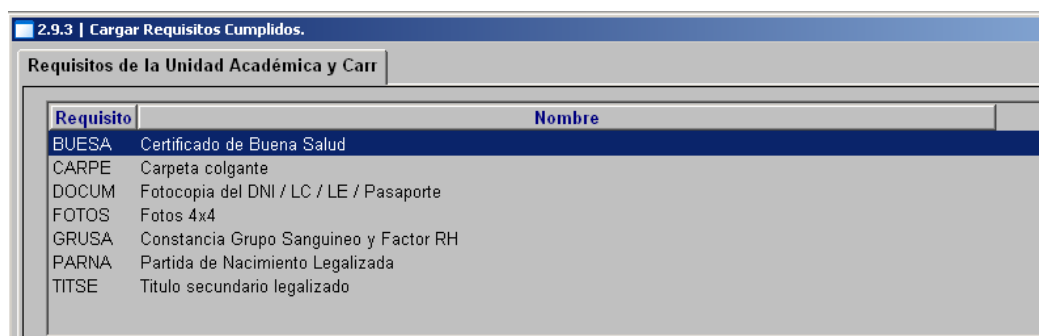


Figura 13. Propuesta de circuito administrativo

Es importante que la información de cada sistema quede en un estado consistente para así poder confiar en los datos que proveen y con la finalidad de evitar los mencionados e ineficientes controles contra el archivo físico.

A continuación se resumen los pasos para el control y consistencia del legajo digital:

1. Cuando un aspirante o alumno presente alguna documentación, la misma se debe registrar en el SIU Guaraní con el tilde de requisitos correspondientes. El sistema tiene prevista esta funcionalidad (Figura 14).



The screenshot shows a web application window titled "2.9.3 | Cargar Requisitos Cumplidos." Below the title bar is a tab labeled "Requisitos de la Unidad Académica y Carr". The main content area contains a table with two columns: "Requisito" and "Nombre". The table lists several requirements, with the first one, "BUESA", highlighted in blue.

Requisito	Nombre
BUESA	Certificado de Buena Salud
CARPE	Carpeta colgante
DOCUM	Fotocopia del DNI / LC / LE / Pasaporte
FOTOS	Fotos 4x4
GRUSA	Constancia Grupo Sanguíneo y Factor RH
PARNA	Partida de Nacimiento Legalizada
TITSE	Título secundario legalizado

Figura 14. Carga de requisitos cumplidos en SIU Guaraní

2. Luego se debe foliar la nueva documentación en orden correlativo a la última presentación realizada.
3. Paso seguido, se realiza la digitalización de la documentación presentada en el sistema LeDig.
4. Si la documentación lo requiere, firmar digitalmente la presentación.

Respetar estos simples pasos facilitará el control y posterior auditoría, asegurando que el legajo digital sea una copia fiel del legajo físico. Por otro lado, cada acción realizada en el sistema LeDig (alta, baja o modificación de un registro) debe generar una pista de auditoría en la BD en la cual, aparte de los datos manipulados, se registra la fecha, horario, evento y usuario que realizó la acción en el sistema.

Esto permitirá:

- Saber quién ingresó la información al sistema y por consiguiente, quien realizó la verificación o validación de la copia fiel.
- Reconstrucción del legajo digital a una fecha determinada.
- Determinar responsabilidades sobre las acciones ejecutadas en el sistema.

4.4. Análisis sobre las herramientas de replicación

Existen diversas herramientas, tanto de código abierto como propietarias, que permiten la replicación de contenido en una BD. A continuación se detalla un relevamiento y se exhibe una comparativa acerca de las diferentes herramientas *open source* de replicación existentes en el mercado para el mencionado motor de BD.

- Postgres-R: Este complemento proporciona una replicación eficiente de la base de datos para *clústers* conservando la consistencia mediante una extensión de la base de datos *PostgreSQL*. Es transparente, estable y seguro, con lo que es posible escalar el proceso de replicación. *Postgres-R* permite agregar o eliminar nodos cuando el sistema está en estado de ejecución. Los nodos bloqueados se detectan y eliminan automáticamente sin afectar las operaciones de la base de datos. *Postgres-R* usa bloqueo de dos fases para control de concurrencia con bloqueo en el nivel de relación [54].
- Slony-I: Es un sistema de replicación maestro-esclavo capaz de replicar grandes bases de datos a un número limitado de sistemas esclavos (umbral). Si se supera el umbral, el costo de comunicación puede aumentar, haciendo ineficiente la replicación. Su objetivo es mantener centros de datos y sitios de respaldo donde los datos estén seguros y disponibles todo el tiempo, sobre todo si los nodos son propensos a fallas. *Slony-I* es un *plug-in* que puede añadirse en el mismo proceso de instalación del motor, lo cual asegura la compatibilidad con el mismo [55].

- ESCADA Replication Server: Este sistema ofrece replicación de bases de datos portátiles, confiables y flexibles. El servidor de replicación ESCADA utiliza una replicación maestro esclavo síncrona y una replicación entre clústeres sin pérdida de datos a través de la WAN¹⁰. La herramienta también proporciona replicación consistente en varios maestros. ESCADA está diseñado e implementado en tecnología basada en componentes, lo que permite combinar con varios componentes interoperables para proporcionar una gama de soluciones de replicación [56].
- Pgpool-II: Es un middleware que se conecta como cliente a las bases de datos cliente y maestro para realizar la replicación. Permite manejar múltiples maestros y múltiples esclavos, lo que ayuda a proporcionar un mecanismo eficiente de tolerancia a fallas. Existe un límite en la cantidad de conexiones simultáneas con *PostgreSQL*, y las conexiones excedentes son rechazadas. Elevar la cantidad máxima de conexiones aumenta el consumo de recursos y afecta el rendimiento del sistema. *Pgpool-II* puede configurarse para que estas conexiones adicionales se pongan en cola para que las mismas no sean rechazadas o generen un error [57].
- DB Replicator: Es una herramienta de código abierto desarrollada en Java para la replicación de base de datos multi maestro. Proporciona una API¹¹ para realizar una actualización asíncrona en cualquier parte de la base de datos a replicar. Admite la sincronización bidireccional de datos entre diferentes nodos. Permite la sincronización entre las bases de datos heterogéneas, con tablas de estructura diferentes, gracias a una función de programación integrada. Esto ayuda a programar la operación para su ejecución. Independientemente de las fallas, la operación se ejecuta en un momento particular. Además, posee una estrategia de detección y resolución

¹⁰ WAN: *Wide Area Network* - Red de Área Amplia

¹¹ API: *Application programming Interfaz* – Interfaz de Programación de Aplicaciones

automática de conflictos con una participación mínima del administrador de bases de datos [58].

A continuación, en la Tabla 1, se presenta un cuadro comparativo con las diferentes herramientas antes mencionadas: [59]

	<i>Postgres-R</i>	<i>Slony-I</i>	ESCADA	<i>DB Replicator</i>	<i>Pgpool-II</i>
Tipo de replicación	Asíncrona	Asíncrona	Asíncrona	Asíncrona	Síncrona
Actualización de cualquier dato	Si	No	Si	Si	Si
División de la carga de trabajo	Si	No	Si	No	Si
Arquitectura maestro/esclavo	Si	Si	Si	Si	Si
Soporte WAN	Si	Si	Si	Si	Si
Manejo de conflictos	Si	No	No	Si	Si
Detección automática de fallo en nodo	Si	No	Si	No	Si

Tabla 1. Comparativa de herramientas de replicación

4.4.1. Elección de la herramienta de replicación

En esta sección se realiza el análisis de las herramientas de replicación, algunas expuestas previamente, teniendo en cuenta para la elección factores como la documentación disponible, últimas versiones de cada herramienta, así como también la curva de aprendizaje necesaria en base a la experiencia del equipo de trabajo. Por

último se presenta una conclusión y la elección de la herramienta más adecuada para implementar en el proyecto.

Acceso a la documentación

Para este ítem se ha tenido en cuenta la documentación ofrecida por el equipo de desarrollo de cada herramienta, la documentación y los tutoriales en línea y wikis entre otros. También se analizaron aspectos relacionados a la extensión de la documentación respecto a detalles sobre las implementaciones, claridad de las mismas y ejemplos de configuración. De las herramientas analizadas, salvo el caso de ESCADA, todas las demás poseen documentación oficial disponible en línea, aunque algunas se destacan por ser más extensas y detalladas. La documentación oficial de *Pgpool-II* es una de las más abundantes [60], con guía de inicio, ejemplos de configuración, y hasta un tutorial para reportar errores.

Slony-I ofrece una documentación extensa, con guías paso a paso de cómo replicar una BD por primera vez con la herramienta, lo cual puede reducir la curva de aprendizaje cuando no se dispone de conocimientos previos por parte del equipo de trabajo.

En el caso de *DB Replicator*, en su sitio oficial posee una guía de instalación, manual de usuario y manual del desarrollador. El manual del usuario es claro y posee ejemplos prácticos, por lo que se puede entender fácilmente, pero no es tan detallado como se señaló en el caso de las dos herramientas anteriores.

Por último, la herramienta *Postgres-R* también posee documentación oficial [61], aunque se verificó que es más reducida que todas las expuestas anteriormente.

Últimas versiones

Analizar el presente de cada herramienta, sus actualizaciones y versiones son factores claves para decidir sobre cuál es la que mejor alternativa para adoptar en un proyecto. En el análisis se ha tenido en cuenta la última versión lanzada, las

disponibles, repositorios, versionados y la frecuencia en las actualizaciones de los mismos.

La herramienta con el desarrollo más activo es *Pgpool-II* y posee el mayor número de versiones, estando el proyecto activo desde el año 2006, según el resumen del repositorio oficial [62]. En el caso de *Slony-I*, en el resumen de su repositorio se constató que la versión 2.2 fue liberada recientemente. Cuenta con versiones estables, ya integradas y probadas en el motor de *BD PostgreSQL*, y su desarrollo se encuentra activo.

La herramienta *Postgres-R* tenía un desarrollo muy activo, entre tres y cinco versiones por mes entre los años 2008 y 2010, año del que data la última versión disponible para descargar. No se tiene acceso al repositorio del proyecto, ya que no está publicado. La herramienta *ESCADA*, posee dos versiones, ambas del año 2017. Dicho repositorio tampoco está publicado.

Por último, la herramienta *DB Replicator* posee solamente una versión lanzada (v.0.21.4), la cual fue publicada en el año 2011. El repositorio publicado en la página del proyecto ya no está en línea.

Experiencia con las herramientas

En relación a este aspecto, el equipo viene trabajando con *Slony-I* desde el año 2015 y ha implementado soluciones con dicha herramienta. Recién este el año 2019, el equipo de trabajo ha incorporado *Pgpool* como herramienta alternativa para enseñarla en la cátedra de las carreras de informática que se dictan en la Universidad Nacional de Misiones.

Con respecto a las demás herramientas, el equipo todavía no ha tenido la oportunidad de trabajar con ellas ni tampoco se las propuso como tal para enseñarlas en las cátedras debido a su des actualización y desarrollo inactivos.

Conclusión

Como conclusión de lo expuesto y a pesar de que existen diferentes opciones a la hora de elegir herramientas de replicación para las *BD PostgreSQL*, muchas están

en desuso, o con un desarrollo inactivo. Para tener un panorama más claro, en la siguiente tabla (Tabla 2) se muestra un resumen de las diferentes herramientas analizadas.

Herramienta	Documentación	Desarrollo	Experiencia
<i>Postgres-R</i>	Reducida	No Activo	No
<i>Slony-I</i>	Abundante y con tutoriales	Activo	Si
ESCADA	No posee	No Activo	No
<i>DB Replicator</i>	Reducida	No Activo	No
<i>Pgpool-II</i>	Abundante y con tutoriales	Activo	Poco

Tabla 2.Actividad sobre las herramientas de replicación

De la anterior tabla comparativa, queda expuesto que tanto *Pgpool-II* como *Slony-I* son las herramientas con desarrollo activo y bien documentadas, características que son deseables para adoptarlas en cualquier proyecto. *Slony-I* puede instalarse como complemento en el mismo proceso de instalación del motor de BD *PostgreSQL*. Por otra parte el equipo de trabajo posee conocimiento y experiencia de haber trabajado con la misma.

4.5. Almacenamiento de documentos en *PostgreSQL*

Para dar una solución al almacenamiento de los legajos físicos, se plantea la posibilidad de digitalizar los documentos que forman parte del legajo, resguardar los archivos generados en la base de datos y devolver el legajo físico al alumno. De esta manera se evitaría que la facultad quede como depositaria del legajo físico, liberando así espacio físico en la Dirección de Estadísticas, el cual es crítico como se señaló en el capítulo anterior.

Con respecto a cuál es la mejor manera de almacenar los archivos en la base de datos en *PostgreSQL*, existen varias alternativas. En esta sección se describen las ventajas y desventajas de almacenar la ruta al archivo binario, junto con otros metadatos, en lugar de almacenar el archivo binario directamente en la base de datos.

El estándar SQL define un tipo de dato binario al *blob* o *binary large object*. En *PostgreSQL* el tipo de datos que permite almacenar archivos binarios es el *bytea* [63].

Respecto a este tipo de datos, el tamaño demandado para almacenar el atributo será el mismo que el objeto que se almacena. Para la inserción, actualización, eliminación y selección del objeto se utilizan sentencias del lenguaje SQL estándar [64].

Entre sus ventajas podemos mencionar:

- El almacenamiento y acceso utiliza la misma interfaz que cualquier otro tipo de datos.
- No es necesario rastrear el OID¹² de un objeto creado.
- Seguridad y control de acceso simplificado.
- El control de versiones es más fácil.
- Garantiza las propiedades ACID¹³ para transacciones.
- Las copias de seguridad no necesitan rastrear archivos externos.

Por otra parte, posee las siguientes limitaciones:

- Bytea utiliza *TOAST*¹⁴.
- Posee limitaciones de tamaño por entrada (1GB).
- Se deben codificar los datos binarios antes de enviarlos a la base de datos y luego hacer lo contrario después de recuperarlos.

¹²OID: *Object ID* – Identificador único del objeto

¹³ACID: *Atomicity, Consistency, Isolation and Durability* - Atomicidad, Consistencia, Aislamiento y Durabilidad

¹⁴TOAST: Mecanismo para evitar que los datos almacenados excedan el tamaño de un bloque de datos (normalmente 8KB)

- Los requisitos de memoria en el servidor pueden ser elevados incluso para un pequeño conjunto de tuplas.
- Afectación del rendimiento al escribir los archivos en la base de datos.
- Requisitos de memoria más altos para la base de datos.
- Las copias de seguridad pueden tomar mucho más tiempo.
- Los cambios en objetos tipo BLOBS no se replican automáticamente.
- Complicaciones en el acceso a los archivos por parte de las aplicaciones. Normalmente, un archivo temporal se copia al cliente para acceder y modificar y luego se debe volver a almacenar el mismo a la base de datos.

Una alternativa es optar por el almacenamiento de metadatos y la ruta o enlace simbólico al archivo binario [64].

Las ventajas de hacer esto en la base de datos son:

- El rendimiento al acceder a un archivo binario omite la capa de acceso a la base de datos.
- Número de archivos limitados solo por sistema de archivos.
- Tamaño limitado solo por el sistema de archivos.

Entre las desventajas se pueden mencionar:

- La necesidad de desarrollar una interfaz para realizar un seguimiento de los archivos.
- Los archivos y la base de datos pueden desincronizarse, existiendo archivos binarios que no registren enlace en la base de datos y viceversa.
- Los ajustes de seguridad entre el sistema de archivos y la base de datos son independientes entre sí.

Analizando lo relevado solo se sugiere almacenar los archivos en la base de datos cuando el sistema deba responder a los requerimientos de ACID, ya que al almacenar los archivos fuera de la base de datos, se dificulta garantizar la atomicidad, la consistencia, el aislamiento y la durabilidad de los mismos.

Por otra parte, se debe tener en cuenta que almacenar los archivos en la base de datos puede comprometer el funcionamiento de las aplicaciones donde el rendimiento y los tiempos de respuesta son críticos ya que la capa de base de datos agrega una gran cantidad de sobrecarga y complejidad que pueden no ser necesarias.

Centralizar la información en la base de datos genera también un único sitio de resguardo y por ende un punto crítico en caso de fallas, situación que se subsanaría almacenando los documentos en un sistema de archivos.

Como conclusión, si bien la opción más simple de implementar el almacenamiento de los documentos digitalizados, es hacerlo directamente en la base de datos utilizando el tipo de datos *bytea*, garantizando así las propiedades de ACID y replicando el atributo en cuestión; se optó por la alternativa de un sistema de archivos con un enlace simbólico y metadatos, principalmente debido a que:

- Luego de grabarse la información de un legajo, la misma solo es consultada, no generando operaciones que requieran un tratamiento transaccional riguroso.
- A medida que se agreguen las sucesivas cohortes, el volumen de datos generados por los documentos digitalizados y almacenados en la BD podría comprometer el funcionamiento del sistema.
- En periodos de inscripción, el sistema está sometido a una carga de trabajo donde el rendimiento y los tiempos de respuesta de la aplicación son importantes.
- Contar con los documentos digitalizados en las sedes es importante, pero no crítica como para requerir la replicación de los mismos.
- Los tiempos requeridos para la restauración del sistema en caso de algún siniestro serían elevados, debido al volumen de los archivos digitales almacenados en la base de datos.

4.6. Arquitectura de datos para el legajo digital

Como se mencionó en la sección 2.5, los documentos digitales deben ir acompañados de metadatos que describan al archivo digital y ayuden a que la búsqueda del mismo sea más eficiente. Para soportar el LeDig estos metadatos serán almacenados en una BD llamada LeDig y los documentos digitales en un sistema de archivos por separado. Si bien en las BD de *PostgreSQL* se pueden almacenar documentos digitales, la fundamentación de separar los metadatos y el sistema de archivos digitales es evitar un crecimiento que luego dificulte o genere demoras en la operatividad del sistema.

Para hacer que el modelo sea escalable y tenga alta disponibilidad es necesario plantear una distribución de los datos. La misma se implementa con el complemento de *PostgreSQL* llamado *Sloni-I*, como ya se explicó anteriormente. La idea es simple, se configura un nodo maestro en la BD del sistema LeDig, el cual será gestionado por la Dirección de Estadísticas, y varios nodos esclavos que pueden ser las sedes u otros sectores dentro de la facultad que requieran acceso para consultar la información y/o la documentación digitalizada de cada legajo, cabe destacar que el acceso es en modo solo lectura. Esta replicación de los datos debe ser transparente para el usuario y no debe requerir ningún accionar por parte del mismo.

En el modelo de la Figura 15 se muestra la interconexión entre las BD y cómo se realizará la replicación de los mismo. Esta arquitectura a su vez, otorga la flexibilidad para que el sistema LeDig recopile los metadatos necesarios desde los datos del SIU Guaraní o desde la BD del SIU Pre Inscripción. Esto evita que el personal No Docente de la Dirección de Estadísticas deba realizar la carga de los metadatos.

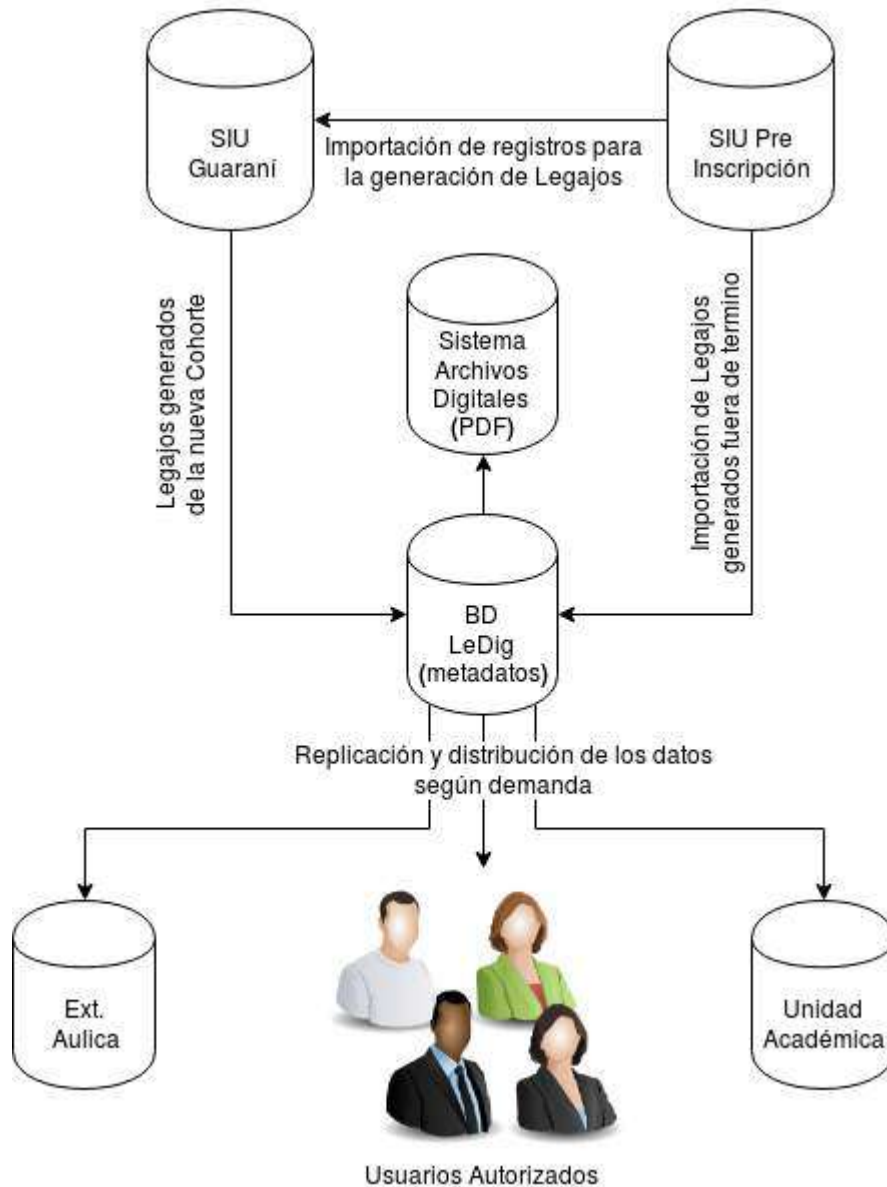


Figura 15. Arquitectura de datos

4.7. Consideraciones sobre seguridad

La seguridad informática consiste en garantizar que los recursos y la información sean utilizados de la manera en que se decidió y que el acceso a la información, así como su actualización, sólo sea posible a las personas que se encuentren acreditadas y dentro de los límites de su autorización.

Respecto al LeDig solamente los agentes de la Dirección de Estadísticas están autorizados a actualizar la información de los legajos, teniendo los demás usuarios y dependencias acceso de solo lectura (Figura 16).

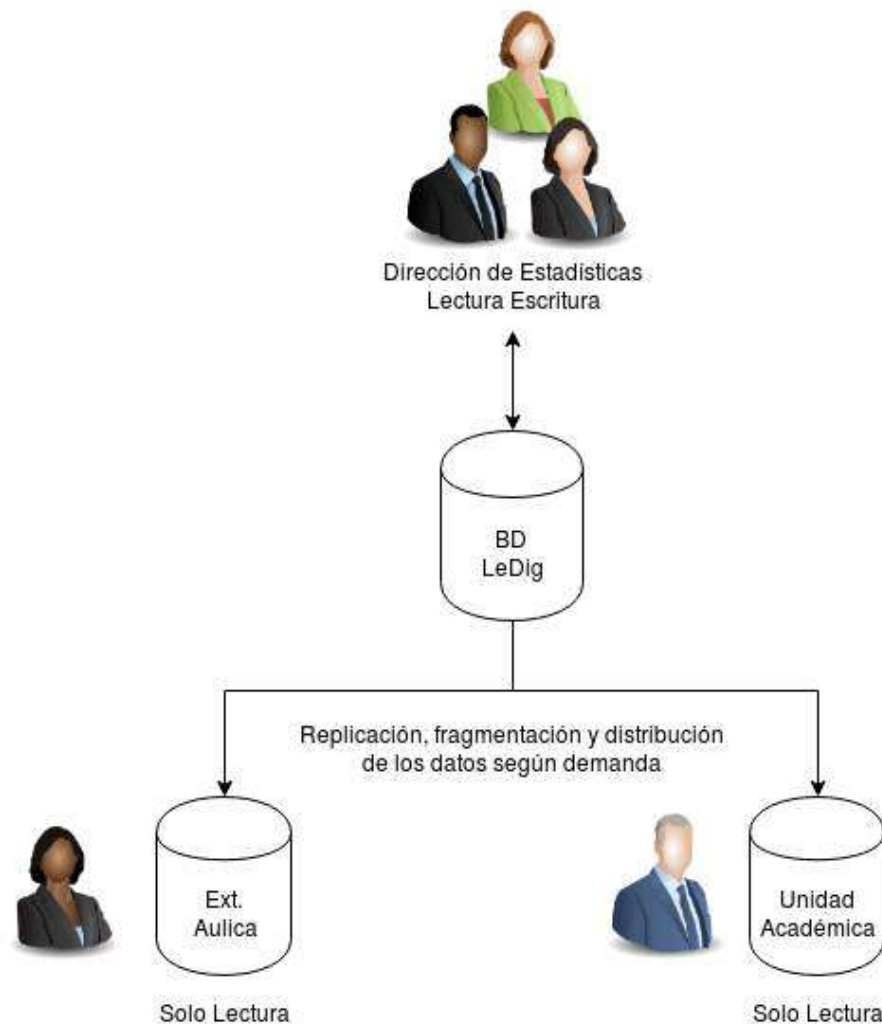


Figura 16. Esquema de seguridad del sistema LeDig

Este esquema será implementado en dos niveles:

- Nivel Aplicación: la aplicación debe ser diseñada y desarrollada con los niveles de seguridad adecuados y estableciendo los perfiles funcionales indicados por las autoridades.

- Nivel BD: se basa en configurar roles y otorgar los privilegios y permisos adecuados a usuarios, o grupos de usuarios, en los que se incluye la capacidad de tener acceso a tablas, registros o campos específicos con un determinado modo (leer, insertar o actualizar).

También se deben tener en cuenta los aspectos relacionados a la seguridad física, es decir, los servidores o sitios de cómputo donde se encuentra el DBMS deben estar protegidos frente al acceso de personal ajeno a la Dirección de Tecnología para la Gestión, y por último considerar cuestiones pertinentes al resguardo seguro y recuperación de la información ante siniestros.

De este modo se logra cumplimentar los objetivos planteado por las autoridades de la unidad académica, contar con la documentación digitalizada y permitir a los usuarios autorizados realizar consultas sobre la información de los legajos, todo a través de la misma aplicación LeDig, pero con las restricciones funcionales según los privilegios que sean configurados.

Capítulo 5

Desarrollo e implementación

5. Desarrollo e implementación

En este capítulo se describe la implementación del Sistema de Legajos Digitales para los alumnos a través de un sistema de BD replicada, el desarrollo se encuentra acotado a una unidad académica, la Facultad de Ciencias Económicas y abarca las distintas Secretarías, Direcciones y Departamentos que necesiten gestionar información sobre los alumnos. El objetivo de esta selección es diseñar un esquema que conste de un nodo maestro instalado en la Dirección de Estadísticas y réplicas del mismo en las distintas sedes que necesiten acceder a la información de los legajos.

5.1. Requerimientos de software y hardware

Para el desarrollo e implementación se utilizaron contenedores virtuales en la plataforma *Proxmox* versión 5.1 con las siguientes características y herramientas:

- *Debian GNU/Linux 9 (stretch)*
- *Framework Toba 2.7.13*
- *PHP 7.3.3*
- *SQL Power Architect Community Edition*
- *PostgreSQL 9.6* y su complemento para replicación *Slony-I 2.2.6*
- *Pentaho Data Integration 8*
- Librerías para *Informix* (ifxtools.jar, ifxjdbc.jar, ifxlang.jar, ifxlsupp.jar)
- *SimpleScan 3.20.0*
- *Sinadura* (Opcional)
- *OpenVPN 2.0*
- *Redmine* versión 2.5.2
- *RRDTool 1.6*

Respecto a los recursos de hardware, se ha trabajado con un equipo con las siguientes características:

- Procesador: 4 núcleos y 4 Ghz. Caché: 8 MB
- Conjunto de instrucciones: 64-bit
- Memoria RAM: 32GB
- Disco duro: Dos serial ATA3 de 4TB cada uno

De estos recursos, los destinados a las maquinas virtuales en la plataforma *Proxmox* han sido:

- Procesador: 2 núcleos.
- Memoria RAM: 2GB
- Disco duro: 82GB

5.2. **Modelo de Datos LeDig**

Para realizar el modelado de los datos se tomaron como referencia algunas tablas de los sistemas ya existentes, como el de pre inscripción, con el propósito de mantener la mayor compatibilidad a la hora de la generación de los metadatos necesarios para el sistema LeDig.

Para la confección del modelo de datos se utilizó la herramienta *open source* “SQL Power Architect Community Edition” y las principales tablas del modelo de datos resultante se ilustra en la Figura 17.

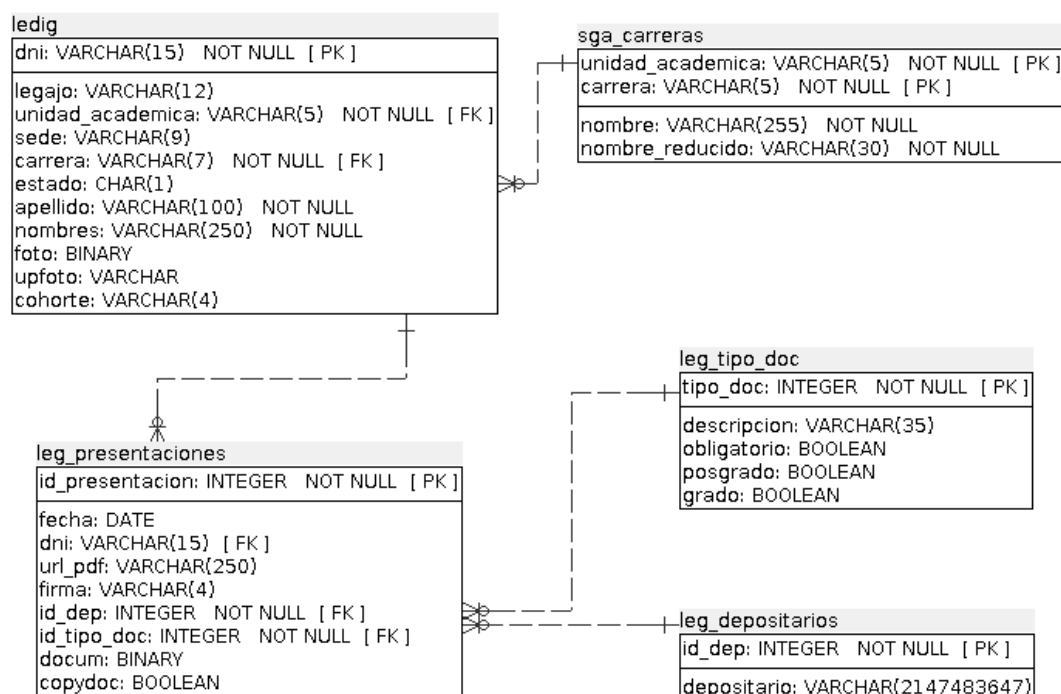


Figura 17. Modelo de datos del sistema LeDig

La tabla ledig contiene los metadatos necesarios para vincular al sistema LeDig con los SIU Guaraní, SIU Pre Inscripción y el sistema de archivos donde se almacenan los documentos digitalizados. Los atributos y su descripción se detallan a continuación en la Tabla 3:

Atributos tabla: ledig	Descripción
dni	Número de Documento Nacional de Identidad del Alumno. Utilizado para identificar unívocamente a una persona en el sistema
legajo	Número de Legajo del Alumno. Puede ser vacío cuando la carpeta que se digitaliza es de un pre inscripto
unidad_academica	Código identificador de la unidad académica en donde se dicta la carrera
sede	Código identificador de la sede en donde se dicta la carrera
carrera	Código identificador de la carrera en donde se inscribió el alumno
estado	Estado en el que se encuentra el registro en el sistema de pre inscripción. Puede tener los siguientes valores: A: Usuario activo

	P: Usuario pendiente de activación C: Registro completo I: Registro grabado en el SIU Guaraní O: Pre inscripto en forma condicional
apellido	Apellido del alumno o pre inscripto
nombres	Nombres del alumno o pre inscripto
foto	Foto tipo dni que se mostrará en la pantalla del sistema
cohorta	Año en que el alumno se inscribió por primera vez a una carrera

Tabla 3. Descripción de atributos - Tabla ledig

En la tabla leg_presentaciones se guardan otros metadatos más específicos, referentes al sistema de archivos. Cada vez que un alumno o aspirante agregue documentación a su legajo, la misma se digitalizará y se agregará una tupla a esta tabla. Los atributos y su descripción se detallan a continuación en la Tabla 4:

Atributos tabla: leg_presentaciones	Descripción
id_presentacion	Numero que identifica unívocamente a la presentación de alguna documentación realizada por el alumno
fecha	Fecha de la presentación
dni	Número de dni del alumno a la cual pertenece la documentación presentada
url_pdf	Contiene la ruta donde se ubica el archivo con la documentación digitalizada en el Sistema de Archivo
firma	[opcional] Indica si el archivo con la documentación digitalizada esta firmado digitalmente por algún agente de la facultad
id_dep	Ver detalle Tabla leg_depositarios
tipo_doc	Ver detalle Tabla leg_tipo_doc
copydoc	Indica si el archivo PDF fue copiado al sistema de archivos del nodo Sede

Tabla 4. Descripción de atributos - Tabla leg_presentaciones

En la tabla leg_depositario se podrán definir quién posee la documentación física o, en su defecto, los lugares en donde actualmente la documentación está

resguardada. El término “depositario” fue extraído del Sistema de Gestión Documental (ComDoc) y refiere a quién posee el documento físicamente. Los atributos y su descripción se detallan a continuación en la Tabla 5:

Atributos tabla: leg_depositarios	Descripción
id_dep	Código que refiere al depositario de la documentación
depositario	Descripción que indica donde o quien queda con la documentación física una vez digitalizada. Los valores que pueden tomar el atributo son: “Alumno”, “Dir. Estadísticas”, "Ex Comedor".

Tabla 5. Descripción de atributos - Tabla leg_depositarios

A la hora que se agregue al legajo una documentación se le podrá asignar una etiqueta a la misma según el tipo de documentación digitalizada. Estas etiquetas se definen en la tabla leg_tipo_doc y los atributos y su descripción se detallan a continuación en la Tabla 6:

Atributos tabla: leg_tipo_doc	Descripción
tipo_doc	Código que refiere al tipo de documentación presentada
descripcion	Descripción que se le puede dar a cada presentación de documentación realizada por un alumno. Los valores que puede tomar el atributo son: Legajo Carpeta Básico Título Secundario Documentación de Identidad Ficha Inscripción SUR1 Constancia Título Trámite Reconocimiento Académico Licencia Estudiantil

Tabla 6. Descripción de atributos - Tabla leg_tipo_doc

La unidad académica que desee utilizar el sistema podrá definir sus carreras en la tabla sga_carreras. Los atributos y su descripción se detallan a continuación en la Tabla 7:

Atributos tabla: sga_carreras	Descripción
unidad_academica	Código identificador de la unidad académica en donde se dicta la carrera
carrera	Código que identifica de manera unívoca a cada carrea
nombre	Nombre de la carrera
nombre_reducido	Siglas representativas del nombre de la carrera

Tabla 7. . Descripción de atributos - Tabla sga_carreras

Existen dos tablas adicionales, una para la gestión de la configuración y para facilitar la interfaz con los otros sistemas involucrados (SIU Pre Inscripción y SIU Guaraní).

La tabla leg_config permite establecer parámetros como cohorte por defecto, ruta del sistema de archivos maestro y esclavo o cualquier otro parámetro que se necesite utilizar luego en el sistema LeDig. Su estructura es la siguiente (Tabla 8):

Atributos tabla: leg_config	Descripción
id_config	Número que identifica unívocamente a cada parámetro
param	Nombre de parámetro
valor	Valor del parámetro

Tabla 8. Descripción de atributos - Tabla leg_config

La tabla `int_ledig` (Tabla 9) sirve como DSA¹⁵ de los registros de legajos generados por la nueva cohorte de ingresantes en el SIU Guaraní que luego serán importados al sistema LeDig a la tabla de metadatos (`ledig`) para posteriormente asociarlos en el proceso de digitalización.

Atributos tabla: int_ledig	Descripción
dni	Número de Documento Nacional de Identidad del Alumno. Utilizado para identificar unívocamente a una persona en el SIU Guaraní
apellido	Apellido del alumno
nombre	Nombres del alumno
legajo	Número de legajo que el alumno posee en el SIU Guaraní
carrera	Código que identifica a la carrera que el alumno está inscripto en el SIU Guaraní

Tabla 9. Descripción de atributos - Tabla `int_ledig`

5.3. Desarrollo de software

Para el desarrollo del sistema LeDig se utilizó el *framework* SIU Toba, provisto por el consorcio SIU a las Universidades Nacionales. Los detalles sobre la creación del proyecto LeDig y el desarrollo de la aplicación se explican en profundidad en el Anexo I, y su código fuente puede descargarse desde el siguiente enlace: <https://github.com/gpautsch/ledig>.

Para facilitar el acceso a la información y evitar que el personal No Docente tenga que realizar la carga manual de los metadatos de los legajos ya existentes de cohortes anteriores a la implementación de LeDig, se confeccionó un flujo de datos y se realizó una importación inicial de los mismos desde el SIU Guaraní utilizando la herramienta *Pentaho Data Integration* en su versión 8, que permite integrar datos entre distintos motores de base de datos.

¹⁵ *Data Staging Area*: Área de almacenamiento de datos

5.4. Sistema de Archivos

El sistema de archivo es empleado como repositorio de los legajos digitalizados, los cuales son almacenados en formato PDF. La asociación de estos, con el legajo del alumno, se realiza a través de los metadatos definidos en la tabla *ledig*.

La estructura del sistema de archivos es muy simple, teniendo como primera jerarquía la cohorte y luego, anidado a cada una de ellas, un directorio correspondiente al DNI de cada alumno. Dentro de esta última carpeta se sitúan los archivos digitalizados (Figura 18).

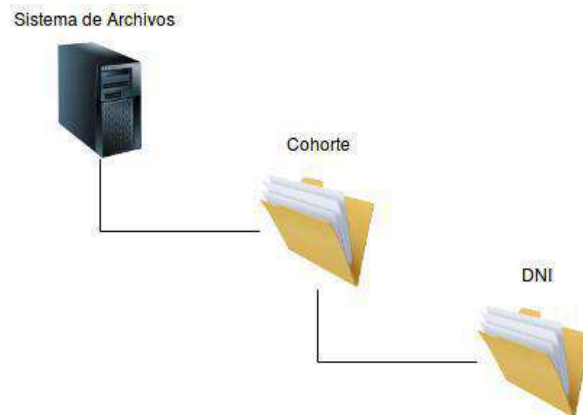


Figura 18. Estructura del sistema de archivos LeDig

Esta estructura en el sistema de archivos no solo ordena el almacenamiento sino que también facilitará cualquier cambio de servidor o partición en el almacenamiento de los mismos.

Si bien la replicación entre los nodos Maestro y Sede, otorga la disponibilidad de los datos y la actualización automática de los mismos, como ya se explicó en la sección 4.5, los atributos de tipo *BLOB*, y por ende el sistema de archivos, no pueden replicarse en *PostgreSQL*.

Luego de varias pruebas realizadas, las cuales se abordan en detalle en los capítulos resultados y discusión, la actualización del sistema de archivos se realiza copiando los PDF utilizando el protocolo *SSH*¹⁶ desde el programador de tareas del nodo Maestro. Para este caso se muestran las configuraciones realizadas en un servidor Debian GNU/Linux 9 (*stretch*).

Para automatizar esta tarea de manera que sea transparente para los usuarios, primeramente se debe generar una llave pública en el nodo maestro. Para esto ejecutamos en una consola la siguiente instrucción:

```
ssh-keygen -b 4096 -t rsa
```

Esto retorna la llave pública encriptada como se muestra a continuación:

```
Your public key has been saved in llavepublica.pub.  
The key fingerprint is:  
SHA256:gZzzZVKD5QL0lJr5KTxZzaArGR/IvSSJvYf1nhwKODqs gpautsch@gpautsch  
The key's randomart image is:  
+---[RSA 4096]-----+  
|  o  +o.      || o O.  o      ||E.o o      |+-----  
[SHA256]-----+
```

Ahora esta llave debe ser copiada al nodo sede. Para ello se ejecuta la siguiente instrucción:

```
ssh-copy-id {usuario}@{ip_nodo_sede}
```

Lo cual retornará en la consola la siguiente información:

```
Number of key(s) added: 1  
Now try logging into the machine, with: "ssh  
{usuario}@{ip_nodo_sede}"  
and check to make sure that only the key(s) you wanted were added.
```

Luego de esto, es posible ejecutar los *script* de transferencia de archivos en forma segura y automatizada. A continuación se explica el funcionamiento de estos *script* de transferencia de archivos entre el sistema de archivos del nodo maestro y el nodo sede. El código fuente puede descargarse desde el siguiente enlace: <https://github.com/gpautsch/ledig>.

¹⁶ SSH: *Secure Shell*. Protocolo que comunicación segura

Inicialmente, durante las pruebas realizadas en el sistema, luego de registrar una presentación en la tabla `leg_presentaciones`, se realiza el copiado del archivo PDF al sistema de archivos del servidor en el sitio Sede con la finalidad que la actualización sea inmediata y transparente para el usuario.

Si bien este método está disponible y documentado en el código de la aplicación, el mismo fue reemplazado por el copiado a través del planificador de tareas (*crontab*) del sistema operativo en el nodo Maestro. Esta decisión está fundada en que la Dirección de Estadísticas manifestó que no es necesaria la actualización inmediata de los legajos y también en que los usuarios del sistema percibieron una demora significativa en la navegabilidad del sistema, al tener que esperar que los archivos PDF sean copiados al nodo Sede.

Respondiendo a las necesidades, esta sincronización puede realizarse varias veces al día o preferentemente en horario donde no se comprometa la utilización del sistema.

Básicamente, desde el *crontab* del sistema operativo se ejecuta el *script copyFS.php*. El mismo recorre las tuplas de la tabla `leg_presentaciones` que tengan en el atributo *copydoc* el valor *false* y actualiza este atributo a *true*. Como salida se genera el archivo *scriptFS.sh* con los comandos *ssh* necesario para luego copiar los PDF al sistema de archivos del nodo Sede (Figura 19).

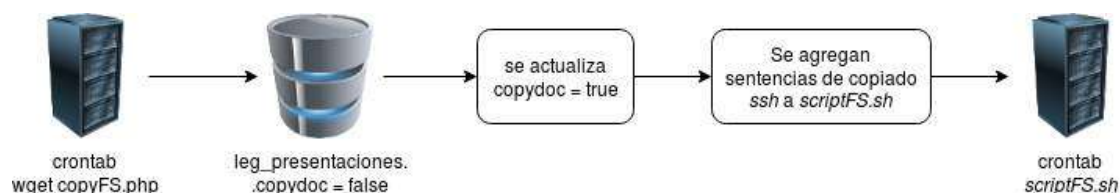


Figura 19. Sincronización del sistema de archivos

Ante la detección de algún inconveniente en la actualización del sistema de archivos, el administrador del sistema puede acceder a una interfaz que le permite

Replicación de datos para dar soporte al Programa de Expansión Territorial de la Universidad Nacional de Misiones

modificar el atributo *copydoc* haciendo clic en el icono de PDF de la presentación para así forzar la posterior copia de los archivos al nodo Sede (Figura 20).

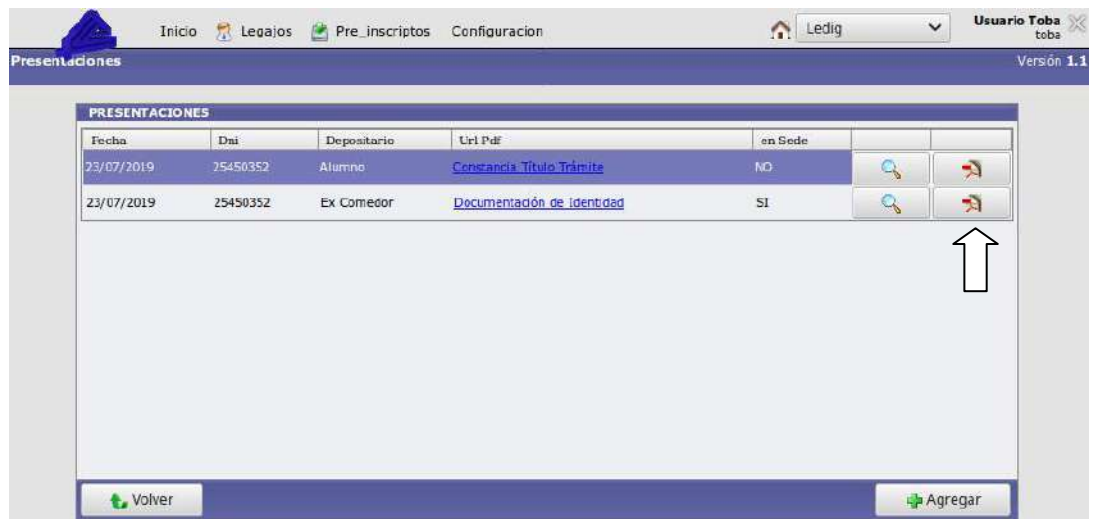


Figura 20. Sistema de archivos. Activación manual

Para asegurar la disponibilidad de los documentos digitalizados en las sedes, se creó en la aplicación de LeDig un método llamado: `descarga_doc`. Este método interactúa con la tabla `leg_config`, más precisamente con dos parámetros definidos allí:

- FS Sede: contiene la ruta al directorio donde opera el sistema de archivos ubicado en la sede donde los datos son replicados
- FS Maestro: es una *URL* que apunta al sistema de archivos ubicado en el nodo maestro

Estando en una Sede, siempre que se realice una solicitud para acceder a un documento digital, este método buscará en primera instancia en el FS Sede. De no encontrar el archivo en esta ubicación, lo cual significa que la digitalización del documento fue generada recientemente y todavía la copia vía *ssh* no está disponible, el método conmutará automáticamente al FS Maestro para poner a disposición el archivo correspondiente (Figura 21).

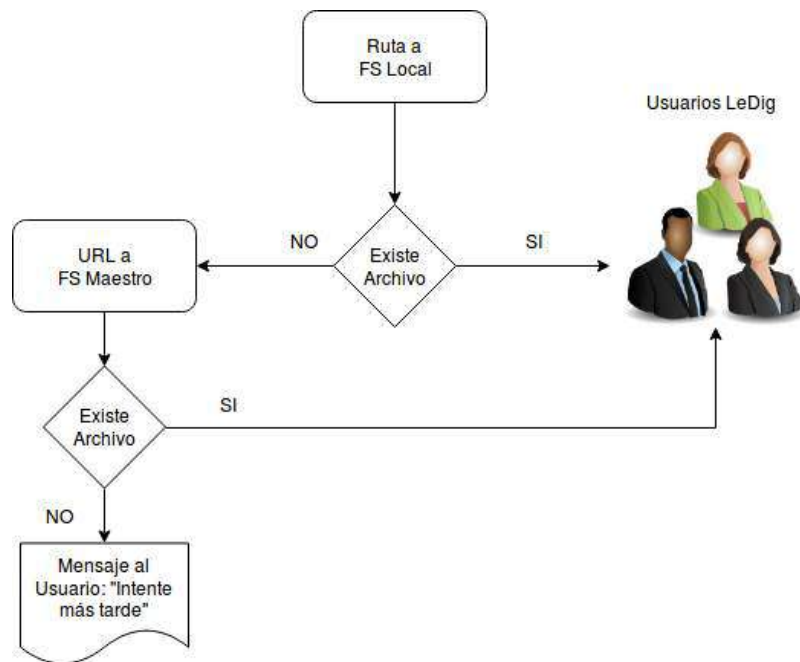


Figura 21. Método descargar_doc

Este método también es el encargado de armar la ruta correspondiente para disponer de los archivos y ocultar la misma para que nunca sea visualizada. La idea de esto es brindar seguridad sobre el sistema de archivo ante posibles accesos no autorizados, ya que la única forma de ejecutar el método `descarga_doc`, es siendo un usuario debidamente autenticado por la aplicación LeDig.

5.5. Implementación de la Replicación

Slony-I es un complemento de *PostgreSQL* que permite la replicación de un nodo maestro a múltiples esclavos. Las características incluyen:

- replicar datos entre diferentes versiones de *PostgreSQL*.
- replicar datos entre diferentes *hardware* y sistemas operativos.
- selección de las tablas a replicar.
- replicar tablas a diferentes nodos esclavos.

- permitir que diferentes servidores de bases de datos sean el origen (maestro) para diferentes tablas.

Para este trabajo, principalmente se utilizaron las siguientes herramientas y versiones:

- *PostgreSQL* 9.6
- *pgAdminIII* v1.22
- *Slony-I* 2.2.6

Luego de instalar estas herramientas, como último paso hay que definir el directorio (*path*) donde se instaló *Slony-I* para que *PostgreSQL* pueda tener acceso a los mismos. Una forma simple de hacerlo es accediendo por *pgAdmin*, desde el menú: “Archivo, Opciones, *BinaryPaths*, *Slony-I Path*”.

Para la instalación en servidores se debe descargar y compilar el código fuente de *Slony-I*, para lo cual se siguió la guía definida en [65].

Una alternativa a esto es utilizar el instalador “*Application Stack Builder*” descargable desde: <http://www.postgresql.org/download>. Una vez instalado el motor de la BD, el mismo instalador ofrece la lista de complementos (*plugins*) compatibles con la versión de *PostgreSQL* instalada (Figura 22). En esta lista puede seleccionarse en la sección *Replication Solution* el complemento *Slony-I* y de esta manera evitar cualquier inconveniente de compatibilidad entre versiones.

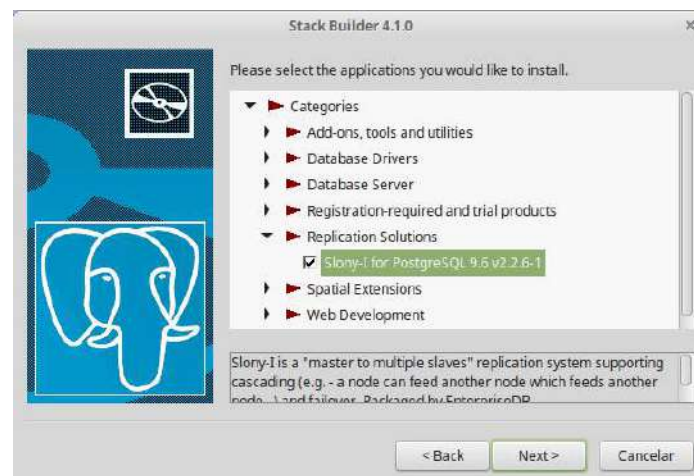


Figura 22. *Application Stack Builder*. Lista de complementos compatibles

Otra cuestión a tener en cuenta es que en todos los nodos (sitios) las tablas a replicar deben tener la misma estructura en la BD. Si bien esto impide cualquier tipo de fragmentación, para el caso de estudio abordado no es una limitante.

Cabe aclarar que si bien en las siguientes secciones se describen las tareas básicas para armar un *clúster* de replicación para otorgar alta disponibilidad a una sede de la facultad, dentro de sus posibilidades *slony* también contempla:

- Herramientas para la configuración del *clúster* de replicación generado.
- Agregar nodos, esquemas, tablas y atributos a la replicación.
- Herramientas de monitoreo y gestión de transacciones.

En la documentación de *Slony-I* [65] se brinda una completa guía de todas las herramientas y una descripción minuciosa sobre todas las posibilidades de configuración que se permite en cada una de ellas.

5.5.1. Selección de las tablas a replicar

Para optimizar la transferencia de datos, es recomendable, evitar tener que replicar todas las tablas de la base de datos, replicando solo las tablas que cambian frecuentemente. Para realizar este análisis se establecieron tres valores según la frecuencia de actualizaciones que sufren los datos en cada tabla por unidad de tiempo. De esta manera se establecen los siguientes valores y descripciones:

- Baja: indicado para tablas secundarias cuyas tuplas, una vez definidas, prácticamente no requieren actualizaciones.
- Media: indica que la tabla puede recibir actualizaciones durante un periodo definido, por ejemplo, por el calendario académico de la facultad.
- Alta: indica que la tabla puede recibir actualizaciones en cualquier momento.

En la Tabla 10, se ofrece un análisis de la frecuencia en que se modifican los datos. Esto tiene como objetivo optimizar la replicación, ya que no es necesario

replicar todo el tiempo tablas que sufren una baja cantidad de operaciones de manipulación de datos.

Tabla	Frecuencia de cambio	Observaciones
ledig	Media	El agregado es frecuente de noviembre a febrero de cada año en periodo de inscripción. Luego es estable durante el transcurso del año.
leg_presentaciones	Alta	Se agrega información a la tabla durante todo el año cada vez que un alumno presente algún requisito o documentación para registrar en su legajo.
leg_depositarios	Baja	Solo en caso que se agregue un nuevo lugar físico para almacenar los legajos. Esta tabla prácticamente no recibe actualizaciones.
leg_tipo_doc	Baja	Salvo que se agregue una descripción en particular, esta tabla prácticamente no recibe actualizaciones.
sga_carreras	Baja	Se actualiza solo cuando se agrega una nueva oferta académica en la facultad.

Tabla 10. Análisis de frecuencia de modificación de datos

Este análisis exhibe que no es necesario replicar todas las tablas todo el tiempo, ya que la mayoría de las transacciones en los sistemas de gestión, por lo general, afectan a un grupo selecto de tablas. Esto no quita que en algún momento las demás tablas deban ser replicadas, sino que la periodicidad con la que es necesario actualizarlas es muy poco frecuente y se debe estimar para cada caso. En la Tabla 11 se muestran las sugerencias para establecer la replicación de tablas en base al análisis previamente realizado.

Tabla	Necesidad de replicación
ledig	Durante el periodo de inscripción a carrera según calendario académico anual. Por lo general de mediados de noviembre a marzo.
leg_presentaciones	Todo el tiempo activa
leg_depositarios	Solo cuando se indique
leg_tipo_doc	Solo cuando se indique
sga_carreras	Solo cuando se indique

Tabla 11. Estimación sobre la periodicidad de replicación

5.5.2. Creación del clúster con *slonik*

En esta sección se detalla la configuración que se realizó entre el nodo Maestro, ubicado en la Dirección de Estadísticas y que posee la información a replicar, y en el nodo Sede, el cual deberá replicar cualquier dato que se modifique en las tablas del nodo Maestro. Para esto, primeramente se creó la misma base de datos del sistema LeDig en el nodo sede, quedando de esta manera los dos nodos en un punto de inicio, uno idéntico a otro.

El *backup* de la base de datos en el nodo Maestro puede hacerse fácilmente con la sentencia:

```
$ pg_dump -U {usuario} -h {ip_nodo_maestro} -d ledig > ledig.sql
```

Luego de crear la base de datos ledig en el nodo Sede, se genera una copia idéntica de la base de datos con la sentencia:

```
$ psql -U {user-name} -h {ip_nodo_sede} -d ledig < ledig.sql
```

Una vez creadas las bases de datos en ambos nodos se pueden comenzar a crear los *script* que luego serán interpretados por *Slony-I* y que definirán el esquema de replicación. Para ello se han definido dos archivos, uno por cada nodo.

Para el nodo Maestro (campus.txt):

```
#DEFINICION CLUSTER UNIDAD ACADEMICA
cluster name = ua;
#DEFINICION DE LOS NODOS MAESTRO(1) Y SEDE(2)
node 1 admin conninfo = 'dbname=ledig host={ip_nodo_maestro} user={usuario}
password={clave}';
node 2 admin conninfo = 'dbname=ledig host={ip_nodo_sede} user={usuario}
password={clave}';
init cluster (id=1, comment = 'nodo maestro');
#DEFINICION DE LAS TABLAS A REPLICAR
create set (id=1, origin=1, comment= 'Aquí van las tablas que deseen replicar');
set add table (set id=1, origin=1, id=1, fully qualified name = 'public.ledig');
set add table (set id=1, origin=1, id=2, fully qualified name =
'public.leg_presentaciones');
store node (id = 2, comment = 'nodo sede, EVENT NODE = 1);
store path (server=1, client=2, conninfo = 'dbname=ledig host={ip_nodo_maestro}
user={usuario} password={clave}');
store path (server=2, client=1, conninfo = 'dbname=ledig host={ip_nodo_sede}
user={usuario} password={clave}');
store listen (origin = 1, provider = 1, receiver = 2);
store listen (origin = 2, provider= 2, receiver = 1);
```

Para el nodo Sede (sede.txt):

```
cluster name = ua;  
node 1 admin conninfo = 'dbname=ledig host={ip_nodo_maestro} user={usuario}  
password={clave}';  
node 2 admin conninfo = 'dbname=ledig host={ip_nodo_sede} user={usuario}  
password={clave}';  
subscribe set (id = 1, provider = 1, receiver = 2, forward = yes);
```

El siguiente paso es ejecutar en una terminal las instrucciones para crear el *clúster* de replicación definido. Esto se realiza con el comando *slonik*, que también permite modificar las configuraciones del *clúster* de replicación. Al momento de la ejecución, *slonik* interpreta los comandos definidos en el *script* y realiza la creación o cambios de configuración en los nodos especificados.

Slonik trabaja con procedimientos almacenados con requisitos especiales en cuanto a qué nodo particular en el *clúster* de replicación lo está invocando. La ausencia de parámetros en dichos procedimientos hace que sea bastante difícil ejecutarlos desde *psql*, ya que carece de la capacidad de mantener múltiples conexiones con transacciones abiertas a múltiples bases de datos.

En el atributo *Exit Status*, *slonik* devuelve “0” si la ejecución terminó normalmente o el código pertinente de existir algún error.

Para el caso de uso abordado la utilización del comando es la siguiente:

```
$ ./slonik campus.txt
```

En respuesta a esta ejecución se retorna por consola:

```
campus.txt:11: waiting for event (1,5000000009) to be confirmed on node 2  
campus.txt:11: waiting for event (1,5000000009) to be confirmed on node 2  
campus.txt:11: waiting for event (1,5000000009) to be confirmed on node 2  
campus.txt:11: waiting for event (1,5000000009) to be confirmed on node 2  
campus.txt:11: waiting for event (1,5000000009) to be confirmed on node 2  
campus.txt:11: waiting for event (1,5000000009) to be confirmed on node 2
```

En ese momento se debe ejecutar en el nodo Sede:

```
$ ./slonik sede.txt
```

La ejecución de lo anterior deberá retornar en la consola algo similar a lo siguiente:

```
waiting for events (2,5000000001) only at (2,0) to be confirmed on node 1
waiting for events (2,5000000001) only at (2,0) to be confirmed on node 1
waiting for events (2,5000000001) only at (2,0) to be confirmed on node 1
waiting for events (2,5000000001) only at (2,0) to be confirmed on node 1
waiting for events (2,5000000001) only at (2,0) to be confirmed on node 1
```

Luego de unos segundos, se confirma la creación y comunicación entre los nodos. A modo de verificación, desde el *pgAdmin* del nodo maestro deberá verse algo parecido a la Figura 23:

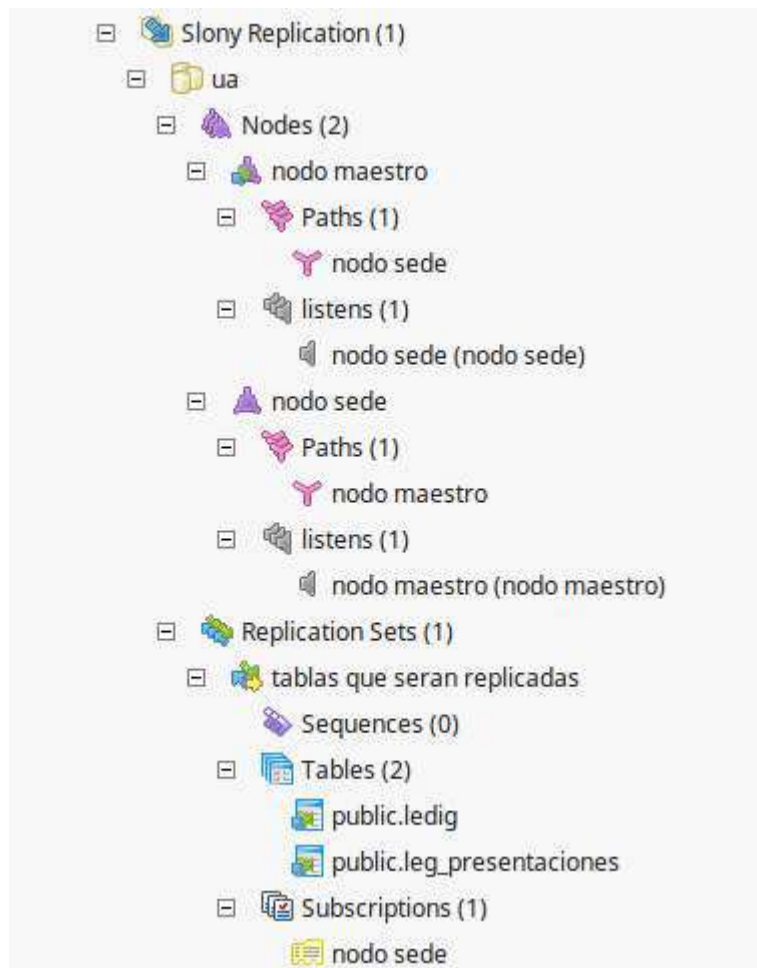


Figura 23. Verificación de la creación del *clúster* de replicación

5.5.3. Activación del *clúster* con *slon*

En este punto, si bien el *clúster* está creado en ambos sitios, la replicación aún no está activa. El comando *slon* es la aplicación que permite activar la replicación y se debe ejecutar una instancia del comando para cada nodo definido en el *clúster*. Existen varios parámetros que se pueden configurar. Entre los principales parámetros de *slon* podemos encontrar:

- -d: Especifica qué niveles de mensajes de depuración debe mostrar la ejecución al registrar su actividad.
- -s: Indica la frecuencia en milisegundos con la que se deben sincronizar los nodos. El valor predeterminado es 10.000 ms.
- -t: Al final de cada período de tiempo *t* definido en milisegundos, se genera una sincronización en el nodo local, incluso si no se han actualizado datos replicables que podrían haber generado una sincronización. . El valor predeterminado es 60.000 ms.
- -g: Controla el tamaño máximo de sincronizaciones. El valor predeterminado es seis (6). Si un nodo en particular está retrasado 200 sincronizaciones, intentará agruparlos según el número configurado. Con esto se espera reducir los costos generados por las transacciones debido a que se agrupan las sentencias *commit*.
- -l: Indica que el nodo debe retrasar a sus proveedores en el intervalo de tiempo especificado. Esto hace que los eventos se ignoren hasta que alcancen el tiempo correspondiente al intervalo.

Los mismos deben ser configurados analizando el contexto, calendario académico y la frecuencia de actualización de las tablas que se desean replicar. En el atributo *Exit Status*, *slon* devuelve “0” si la ejecución finalizó con éxito, de lo contrario devuelve “-1” y a continuación un código de error.

Para una descripción completa de todos los comandos y su parametrización véase [65].

Para iniciar la replicación en este caso de uso es necesario ejecutar desde las consolas los siguientes comandos:

- En el nodo Maestro:

```
$ slon ua "dbname=ledig host={ip_nodo_maestro} user={usuario}  
password={clave}"
```

Lo cual retornará por consola la siguiente información sobre el estado de la replicación.

```
INFO remoteworkerThread_2: SYNC 5000000001 done in 0.015 seconds  
INFO remoteworkerThread_2: SYNC 5000000002 done in 0.015 seconds  
INFO remoteworkerThread_2: SYNC 5000000003 done in 0.003 seconds  
INFO remoteworkerThread_2: SYNC 5000000004 done in 0.002 seconds  
INFO remoteworkerThread_2: SYNC 5000000005 done in 0.002 seconds  
INFO remoteworkerThread_2: SYNC 5000000006 done in 0.002 seconds  
INFO remoteworkerThread_2: SYNC 5000000007 done in 0.002 seconds
```

- En el nodo Sede:

```
$ slon ua "dbname=ledig host={ip_nodo_sede} user={usuario} password={clave}"
```

Luego de pocos segundos aparecerá en la consola la información de las tablas que se están sincronizando, como se muestra a continuación:

```
INFO remoteworkerThread_1: SYNC 5000000012 done in 0.019 seconds  
INFO remoteworkerThread_1: SYNC 5000000014 done in 0.002 seconds  
INFO remoteworkerThread_1: SYNC 5000000018 done in 0.002 seconds  
INFO remoteworkerThread_1: SYNC 5000000025 done in 0.002 seconds  
INFO remoteworkerThread_1: syncing set 1 with 2 table(s) from  
provider 1  
INFO remoteworkerThread_1: SYNC 5000000026 done in 0.005 seconds  
INFO remoteworkerThread_1: syncing set 1 with 2 table(s) from  
provider 1  
INFO remoteworkerThread_1: SYNC 5000000027 done in 0.005 seconds  
INFO remoteworkerThread_1: syncing set 1 with 2 table(s) from  
provider 1  
INFO remoteworkerThread_1: SYNC 5000000028 done in 0.005 seconds  
INFO remoteworkerThread_1: syncing set 1 with 2 table(s) from  
provider 1
```

A partir de este momento los nodos están replicando la información entre ellos y ya se puede notar las réplicas en el nodo Sede de cualquier cambio realizado en el nodo Maestro.

Dentro del directorio *tools* del paquete de instalación de *Slony-I*, se proveen herramientas para el monitoreo del *clúster*, las mismas se describen en detalle en el “Capítulo 5: Consideraciones de Implementación” de la documentación [65].

Entre las más utilizadas en este desarrollo se encuentran:

- *check_slon.sh*: sirve para verificar que *slon* esté en la lista de procesos y, opcionalmente, verifica si hay algún error o mensaje de advertencia en el archivo de registro de *slon* definiendo los siguientes parámetros: el *clustername*, el *dbname* y, opcionalmente la ubicación del archivo de registro. Este script debe ejecutarse en cada nodo del *clúster*
- *check_slony_cluster.sh*: se utiliza para verificar que los nodos receptores activos estén confirmando la sincronización dentro de un tiempo determinado (por defecto, diez segundos). Requiere dos parámetros; el *clustername* y el *dbname*. Este *script* se debe ejecutar sólo en el nodo maestro.

5.6. Sistema LeDig

A continuación se exhibe la validación del circuito administrativo con el sistema LeDig desarrollado.

5.6.1. Login LeDig

La pantalla de autenticación de usuario es necesaria por dos motivos; en primer lugar para garantizar que solo las personas autorizadas puedan acceder a la información de los legajos (Figura 24), y en segundo lugar, para establecer las restricciones funcionales debido a que no todos los usuarios tienen los mismos privilegios.

Replicación de datos para dar soporte al Programa de Expansión Territorial de la Universidad Nacional de Misiones

El Departamento de Estadística de la sede central de la Facultad de Ciencias Económicas, genera los legajos y los actualiza, mientras que las demás dependencias solo poseen privilegios de lectura sobre la información replicada.



Figura 24. Login del sistema LeDig

Luego que el usuario fue validado y las restricciones funcionales ejecutadas la pantalla principal del sistema LeDig se exhibe en la Figura 25.



Figura 25. Pantalla principal del sistema LeDig

5.6.2. Búsqueda del legajo del alumno

Para esta funcionalidad la aplicación provee de un buscador (opción “Legajos” del menú principal) donde se pueden ingresar diferentes criterios, con la finalidad de encontrar en la base de datos el legajo de un alumno (Figura 26).

Dni	Legajo	Unidad Academica	Sede	Carrera	Estado	Apellido	Nombres	Cohorte
25450352	0000000	FCECO	00000	MAEN	A	Pautsch	German	2017

Figura 26. Pantalla de búsqueda de legajos digitales

5.6.3. Interfaz con el SIU Pre Inscripción

De no encontrarse el registro del alumno, el sistema preguntará si se trata de un pre inscripto y ofrecerá al usuario la alternativa de conectarse a la base de datos de sistema de pre inscripción para encontrar la información con los mismos criterios ingresados previamente (Figura 27).



Figura 27. Recuperación de registros de pre inscripción

Elegida esta opción, el sistema ofrece los resultados encontrados en la base de datos de aspirantes pre inscriptos (Figura 28)

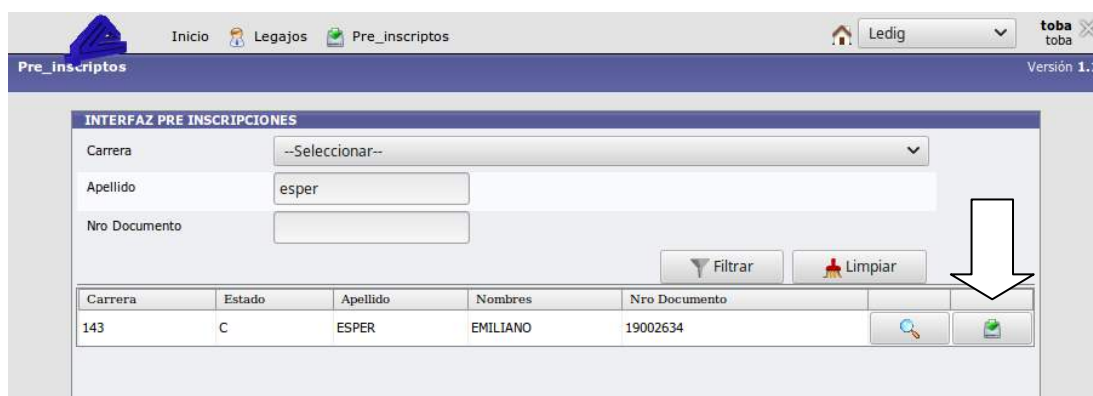


Figura 28. Interfaz entre los sistemas LeDig y Pre Inscripción

Aquí el usuario puede realizar las siguientes acciones (Tabla 12):

Botón	Descripción / Acción
	Accede al registro del pre inscripción. Si se poseen los privilegios necesarios, permite realizar modificaciones
	Importa al LeDig los metadatos necesarios desde la base de datos del sistema de Pre Inscripción.

Tabla 12. Barra de herramientas - Interfaz Pre Inscripción

Si bien el legajo de un alumno debería ser encontrado utilizando alguna de las opciones anteriores, ya que el sistema LeDig se conecta con las bases de datos del SIU Guaraní y SIU Pre Inscripción, también se ofrece la opción de definir y cargar los metadatos en forma manual desde el menú “Legajos / Agregar” (Figura 29).

The screenshot shows a web application titled 'Legajos Digitales' with a version of 1.1. The interface includes a navigation bar with 'Inicio', 'Legajos', and 'Pre_inscriptos' links. The main form contains the following fields: 'Dni (*)' with value '25450352', 'Legajo' with value '000000', 'Nombres (*)' with value 'German', 'Apellido (*)' with value 'Pautsch', 'Carrera (*)' with a dropdown menu showing 'MAESTRÍA EN ADMINISTRACIÓN ESTRATÉGICA DE NEGOCIOS', 'Cohorte (*)' with value '2017', 'Unidad Académica' with value 'FCECO', 'Sede' with value '00000', and 'Estado' with a dropdown menu showing 'A'. There is a 'Foto' section with a placeholder image and a label 'Imagen: 6 KB' with a checkbox 'Cambiar el Archivo'. At the bottom, there are 'Volver' and 'Guardar' buttons.

Figura 29. Gestión de metadatos

5.6.4. Digitalización de la documentación

Para este proceso se ha tomado como referencia el Anexo III de la Resolución 44 - E/2016 de la Subsecretaría de Gestión Administrativa - Ministerio de Modernización [32]. En la misma se especifica el: “Proceso de Digitalización de Documentos y Expedientes en Soporte Papel para su Archivo Electrónico e Incorporación al Repositorio Único de Documentos Oficiales – RUDO del sistema de Gestión Documental Electrónica - GDE”.

Por otro lado, se recuerda la importancia de que la incorporación de un documento digitalizado al legajo, solo debe hacerse si el personal de la Dirección de

Estadísticas ha tenido la documentación física en mano, dando fe así de que la digitalización es copia fiel del documento original.

Para realizar la digitalización se ha propuesto una guía la cual fue presentada, validada y reglamentada por el Honorable Consejo Directivo. Los pasos para la digitalización de los legajos físicos son los siguientes:

1. Desglosar el legajo físico, sacar todos los ganchos y foliar toda la documentación.
2. Abrir aplicación *SimpleScan*, verificar resolución: 150 ppp.
3. Hacer clic en icono Escanear, Todas las Páginas del Alimentador.
4. Verificar el documento digital, eliminar páginas en blanco y rotar las necesarias.
5. Guardar archivo en PC (p.e. dni_del_alumno.pdf).
6. Navegar hacia el sitio: ledig.fce.unam.edu.ar.
7. En la pantalla legajo buscar el registro del alumno, o traerlo de preinscripción si se está en un periodo de inscripción a carreras.
8. Si se realizó la grabación en el SIU Guaraní, anota en ficha SUR física el legajo del alumno.
9. En la grilla donde se muestran los datos hacer clic en icono “Adjuntar”.
10. Se abrirá la pantalla “Presentaciones”. Hacer clic en botón Agregar.
11. Ingresar los datos solicitados y adjuntar el archivo escaneado (p.e. dni_del_alumno.pdf).
12. Hacer clic en guardar y luego en el hipervínculo. A modo de verificación comparar el legajo digitalizado contra el físico.

5.6.5. Gestión de legajos digitales

Una vez obtenido en pantalla los metadatos correspondientes al legajo del alumno por cualquiera de los medios antes mencionados, las acciones que pueden realizarse sobre el mismo son (Tabla 13):

Botón	Descripción / Acción
	Accede a la información principal del legajo. Con los permisos correspondientes, esta opción permite actualizar la información del legajo (metadatos).
	Esta opción permite adjuntar nueva documentación digitalizada al legajo.

Tabla 13. Barra de herramientas - Gestión digital de legajos

La primera opción lleva a la pantalla de edición de metadatos ya vista en la Figura 29. La segunda operación es la que resuelve la necesidad planteada por las autoridades de la unidad académica y tiene como objetivo adjuntar al legajo del alumno la documentación presentada.

Ejecutando esta opción el sistema ofrecerá al usuario una pantalla con las siguientes opciones (Figura 30):

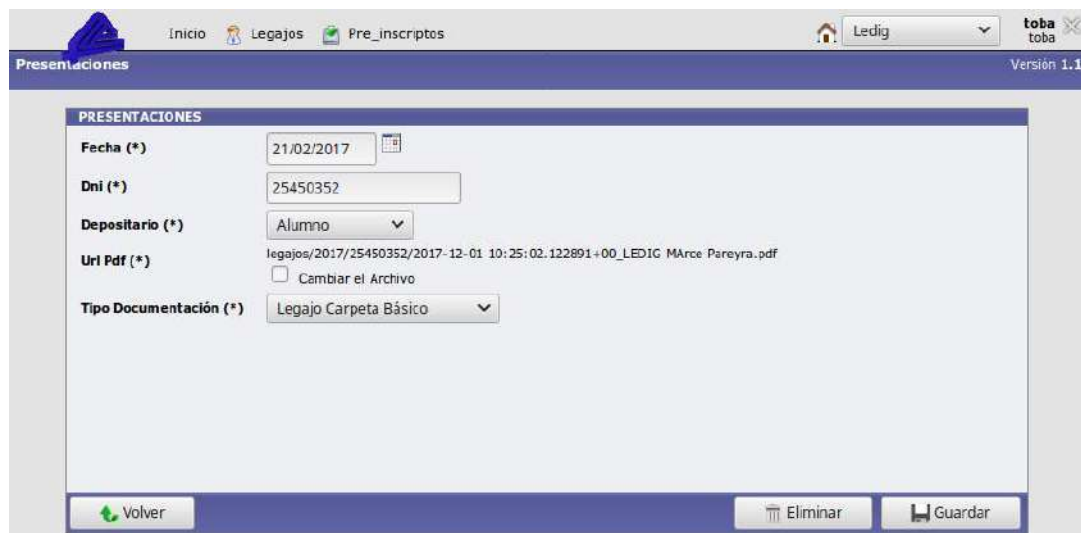


Figura 30. Presentación de la documentación digitalizada

Registrada la presentación, LeDig ofrecerá una pantalla con la confirmación de la presentación y un enlace, al método `descarga_doc`, que obtiene el legajo digitalizado (PDF) accediendo al sistema de archivos (Figura 31).

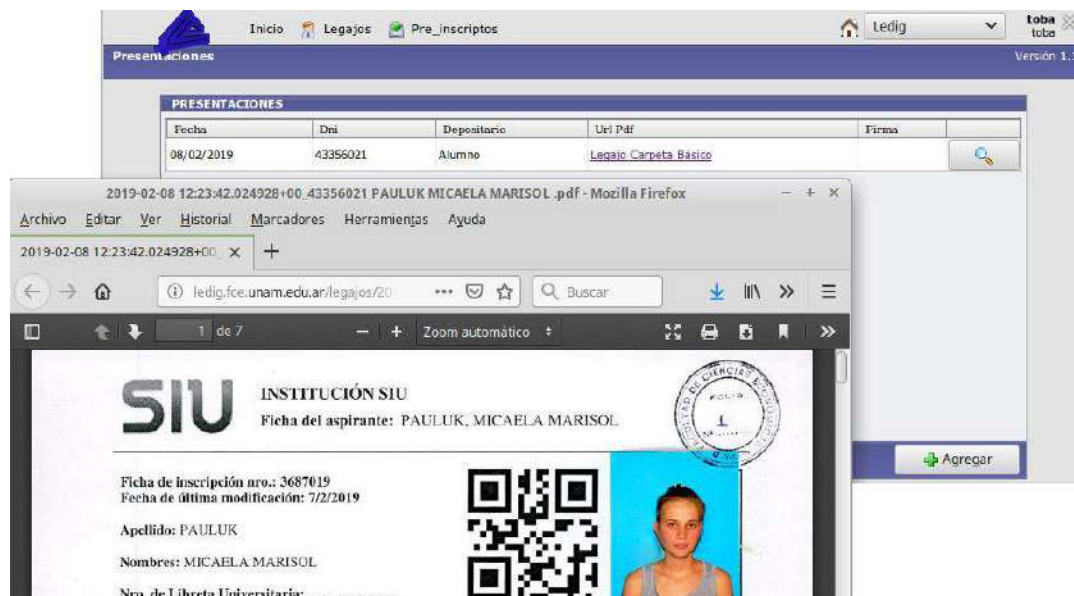


Figura 31. Consulta del legajo digital

El sistema también contempla la posibilidad de que cada alumno pueda realizar la presentación de la documentación en distintas fechas y períodos, por lo cual LeDig ofrece una pantalla (Figura 32) donde se puede visualizar cada presentación realizada por el alumno. En cada registro de presentación el sistema genera el enlace hacia la documentación digital guardada en el sistema de archivos.

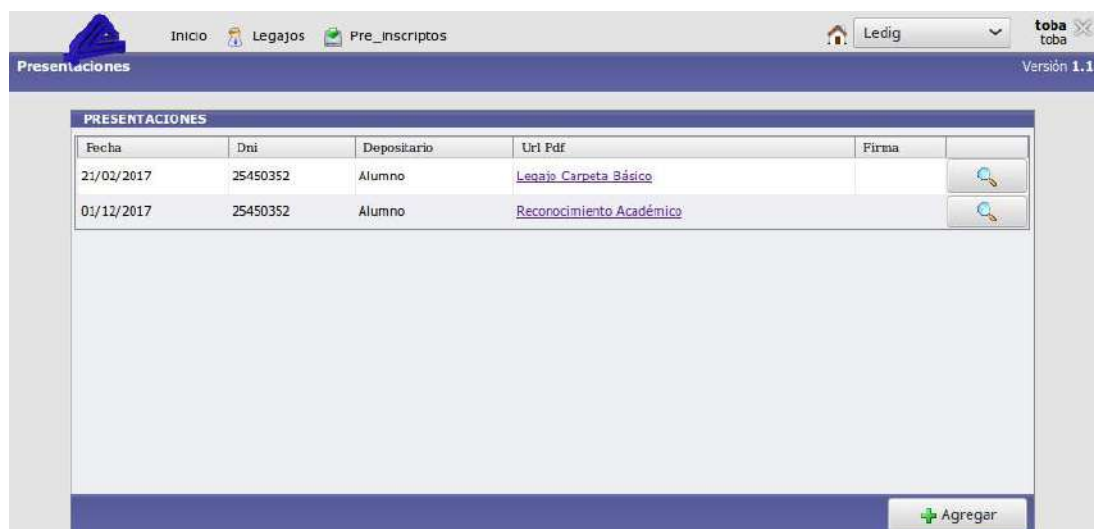


Figura 32. Gestión de las presentaciones realizadas

De este modo el sistema LeDig gestiona los metadatos necesarios para vincular los SIU Guaraní, Pre Inscripción y sistema de archivos digitales para facilitar la gestión de los legajos.

5.7. Consideraciones sobre la replicación

Para verificar el correcto funcionamiento de *slon*, y en consecuencia si la replicación esta activa, dentro de la carpeta *tools* del instalador, *Slony-I* provee varias aplicaciones de monitoreo. Una de ellas es *check_slon* que acompañada de el nombre del *clúster* y la base de datos, retorna el número de proceso que *slon* posee asignado en el sistema operativo.

Por ejemplo:

```
$ check_slon ua ledig
```

Retorna:

```
OK - slon process 417
```

Consultando los procesos activos en el sistema operativo se puede verificar esta información.

Ejecutando:

```
$ ps -aux | grep slon
```

Retorna:

```
postgres 417 jun23 ./slon ua host=10.93.0.49 dbname=ledig
```

Otra forma sencilla de verificar la réplica de los datos, es observar en el sistema LeDig instalado en la sede y ver si los cambios realizados en el nodo Maestro tienen impacto, o a nivel capa de datos, desde la herramienta *pgAdmin* se puede constatar la replicación de datos entre los diferentes sitios conectándose a cada nodo y ejecutando una consulta sobre los datos. En la Figura 33 se muestra como un dato actualizado en el nodo maestro se replica en el nodo sede.

	apellido character varying(100)	nombres character varying(250)	foto bytea	upfoto character
1	PIACENTINI xxxxxxxxxxxxxx	MARCOS TADEU SIMOES	<binary data>	
2	LEONES DE SOUZA	VALDINEI	<binary data>	
3	BACELAR MOTA	EDMARSON	<binary data>	
4	VIANA TEXEIRA	JUEDIR	<binary data>	
5	LEITE CORREIA FILHO			
6	LENGA DE GOLDBERG			
7	AZEVEDO SANTOS			
8	SANTOS CARRERA			
9	CAVALCANTE DOS SANTOS			
10	GUEDES DA CUNHA			
11	GALIAN			
12	VIVEROS			
13	GAUNA			
14	GIRARD			

28117 rows.

Figura 33. Replicación de datos en el nodo Sede

Para un control más detallado se puede consultar las tablas del catálogo *Slony cluster* (Figura 34).

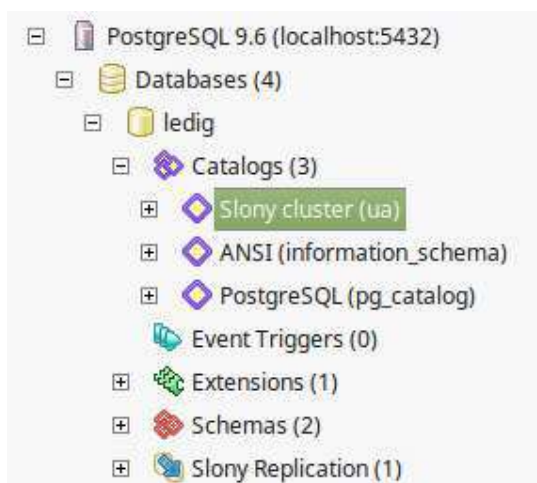


Figura 34. Acceso al catalogo *Slony-I*

Al generarse el *clúster* de replicación también se crean una serie de tablas en este catálogo de la base de datos. A continuación se ofrece una descripción de las más relevantes (Tabla 14) en lo que respecta a poder realizar un seguimiento a los eventos que ocurren durante la sincronización de las réplicas y su confirmación en los diferentes nodos.

Tabla	Descripción
<i>sl_confirm</i> y <i>sl_event</i>	Contiene la confirmación de los eventos de replicación. Después de un período de tiempo, <i>Slony-I</i> elimina los eventos confirmados antiguos tanto de esta tabla como de la tabla <i>sl_event</i> . Esta última contiene información sobre eventos de replicación relacionando el número de evento y el nodo en donde se produjo el mismo.
<i>sl_log_1</i> y <i>sl_log_2</i>	Almacena cada cambio que se propagará a los nodos del suscriptor. <i>Slony-I</i> cambiará frecuentemente entre <i>sl_log_1</i> y <i>sl_log_2</i> (<i>Log Switching</i>). Cuando todas las transacciones en una de estas tablas se hayan replicado y confirmado en todos los nodos <i>Slony-I</i> ira vaciando las tabla.
<i>sl_config_lock</i>	Esta tabla existe únicamente para evitar la superposición de la ejecución de los procedimientos de cambio de configuración y los posibles interbloqueos resultantes.

Tabla 14. Tablas del catálogo *Slony-I*

Slony-I es un motor de replicación basado en disparadores, para cada tupla que se inserte, actualice o elimine en la base de datos, *Slony-I* insertará una tupla en las tablas *sl_log_1* o *sl_log_2*. Esto significa que *Slony-I* tendrá un impacto negativo en el rendimiento del sistema. Medir este impacto es difícil ya que esta pérdida de rendimiento dependerá de la carga de trabajo en la base de datos y las características del *hardware*.

Es probable que las siguientes operaciones de *Slony-I* afecten el rendimiento:

- El impacto de los de datos en una tabla replicada resultará en una sentencia *insert* de tuplas en las tablas *sl_log_1* o *sl_log_2*.
- Cuando *slon* genera una sincronización entre los nodos, deberá agregar un registro a la tabla *sl_event*.
- Cada nodo esclavo deberá consultar las tablas *sl_log_1*, *sl_log_2* y *sl_event* para extraer los datos y replicarlos.

5.8. Seguridad y Auditoría.

Para brindar seguridad al sistema LeDig a continuación se describen las principales consideraciones sobre las cuales se ha trabajado.

5.8.1. Comunicaciones

Respecto a la conectividad de los dos sitios en redes geográficamente distantes, se ha utilizado una VPN¹⁷ implementada con la herramienta de software libre *OpenVPN* en su versión 2.0. Dicha VPN implementa una extensión de red segura de capa dos o tres del modelo OSI¹⁸ utilizando el protocolo SSL¹⁹/TLS²⁰ estándar. Soporta métodos flexibles de autenticación de clientes basados en

¹⁷ VPN: *Virtual Private Network* - Red Privada Virtual

¹⁸ OSI: *Open System Interconnection* - Interconexión de Sistemas Abiertos

¹⁹ SSL: *Secure Sockets Layer* - Capa de Conexiones Seguras

²⁰ TLS: *Transport Layer Security* - Capa de Transporte de Seguridad

certificados. *OpenVPN* ofrecer un modo escalable que permite a varios clientes conectarse a un único proceso de servidor a través de un único puerto TCP²¹.

5.8.2. Perfiles Funcionales

Desde la opción “Administración de Usuarios” de la barra de herramientas, SIU Toba permite la creación de los usuarios que luego accederán a la aplicación así como también la definición de los perfiles funcionales que tendrá el sistema.

Para comenzar a definir los perfiles se debe ejecutar la opción “Perfiles Funcionales” del menú “Perfiles”. Para definir un perfil funcional basta asociar al mismo un identificador (Id Perfil) y un nombre. Luego de esto, desde la solapa de “Opciones Accesibles” se otorgan los accesos necesarios de acuerdo al perfil que se está definiendo. La Figura 35 muestra la definición del perfil funcional para los usuarios de una de las extensiones áulicas, los cuales podrán acceder a los legajos y las presentaciones realizadas por los alumnos, pero no podrán acceder, por ejemplo, a la funcionalidad que conecta con el sistema de pre inscripciones.

Figura 35. Configuración de los perfiles funcionales - LeDig

²¹ TCP: Transmission Control Protocol - Protocolo de control de transmisión

5.8.3. Restricciones Funcionales

Otra posibilidad que ofrece el *framework* SIU Toba es la de definir restricciones funcionales para cada perfil funcional definido anteriormente.

Como el concepto lo indica, esto permite restringir ciertas funcionalidades y atributos, que están contenidos en el controlador de interfaz, al grupo de usuarios que estén contemplados por un perfil funcional determinado (Figura 36).

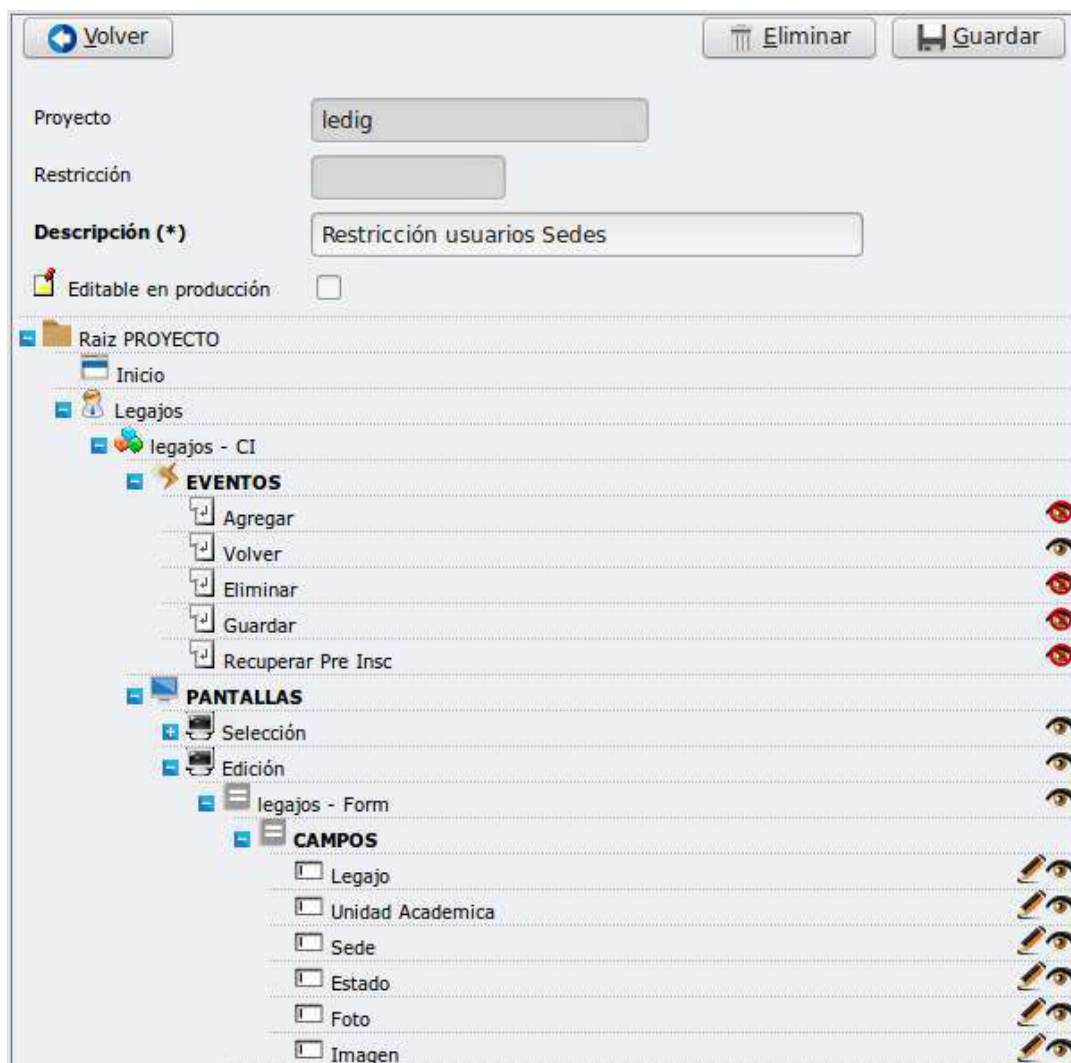


Figura 36. Configuración de restricciones funcionales - LeDig

En la Figura 36 puede apreciarse que si bien los usuarios de las sede pueden acceder a la pantalla de legajos, solo podrán visualizar los datos de los mismos, ya que los eventos de agregar, eliminar o guardar, no van a ser visibles en la pantalla.

También, la configuración permite establecer restricciones de edición y visualización para cada atributo presente en las diferentes pantallas del sistema.

Por último esta restricción se asocia al perfil funcional correspondiente en la solapa “Restricciones Funcionales” de la opción “Perfiles Funcionales” del menú “Perfiles”, como se muestra en la Figura 37.

Figura 37. Asociación entre los perfiles funcionales y las restricciones

5.8.4. Pistas de auditoría

Para la generación de las pistas de auditoría se utilizó la opción que provee el SIU Toba, la cual genera un esquema llamado “auditoria” en la base de datos con las mismas tablas del sistema de gestión desarrollados (Figura 38).

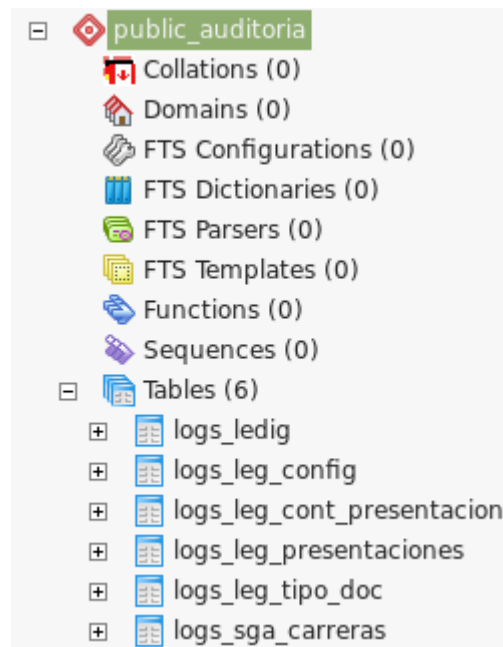


Figura 38. Tablas de auditoría

Esta posibilidad se activa y ejecuta desde la misma pantalla de definición de la fuente de datos y básicamente agrega tres atributos en cada una de las tablas de gestión ya existentes.

Ellos son:

- auditoria_usuario: registra el legajo del agente que realiza la operación.
- auditoria_fecha: registra la fecha en la que se realizó la operación.
- auditoria_operacion: indica si la operación fue una inserción (I), actualización (U) o eliminación (D) de registro.

Las tuplas de estas tablas del esquema de auditoría se completan por medio de disparadores (*triggers*) los cuales el SIU Toba genera automáticamente dentro el mismo esquema de auditoría (Figura 39).

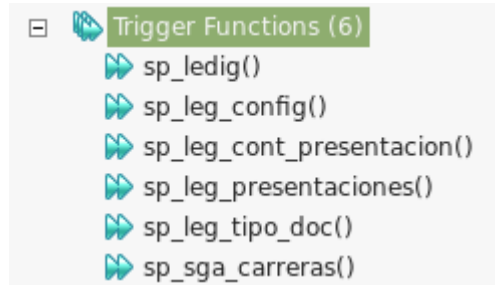


Figura 39. Triggers de auditoría

Contar con estos registros permite saber cuándo y quién realizó alguna acción sobre las tuplas así como también poder consultar el historial de modificaciones, por ejemplo, de un legajo en particular.

5.8.5. Resguardo y Recuperación

LeDig se encuentra implantado en un sistema operativo *Linux* bajo la distribución *Debian GNU/Linux 9* y virtualizado mediante la plataforma *Proxmox*, el cual posee un tablero de control para la administración de contenedores virtuales (Figura 40).

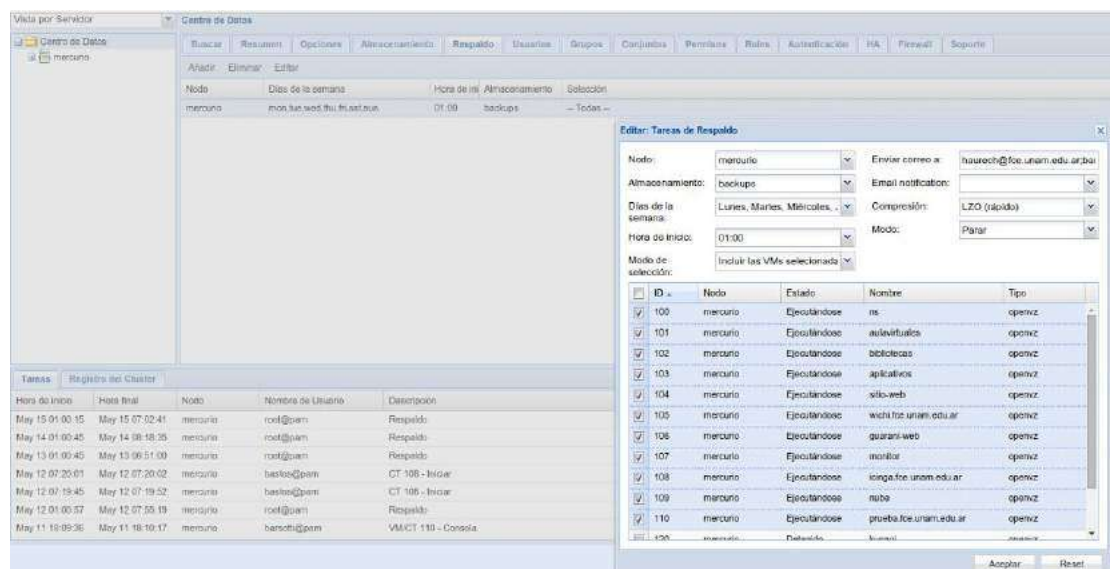


Figura 40. Panel de control - Proxmox

Esta aplicación brinda la posibilidad de ejecutar tareas programadas de resguardo, que generan un archivo comprimido con las configuraciones y datos del sistema operativo a modo de contenedor virtual.

Particularmente en el tratamiento de resguardo del sistema LeDig se genera un archivo de *backup* diario sobre un directorio, y posteriormente un acumulativo que forma parte de otro repositorio histórico, el cual tiene un tiempo de re escritura de siete días.

A fin de mantener la integridad de los datos, el resguardo diario e histórico se almacena en el equipo físico pero en un disco rígido diferente al que utiliza la plataforma *Proxmox* de modo que, bajo una avería del disco sea posible obtener los resguardos situados en el disco complementario.

Existen dos equipos dedicados y con sistemas de resguardos diferentes ubicados dentro y fuera del edificio central (campus), con esto se logra redundancia en cuanto a los datos en caso que algún siniestro afecte al edificio. Estos sistemas de resguardo remoto se diferencian en el método de tratamiento y transporte. El primero de ellos realiza el envío de un bloque total diariamente, en cambio, el segundo genera la actualización de los últimos cambios en forma periódica, además de un resguardo total en forma semanal.

En cuanto a la transferencia entre origen y destino, los datos se encuentran encriptados a fin de que no puedan ser leídos al circular por la red.

El acceso al entorno, nivel usuario y administración del sistema operativo, se encuentran protegidos mediante usuario y contraseña además del control de los pedidos de conexión a nivel de red a fin de evitar DoS²² y escaneos no deseados.

²² DoS: *Denegation of Service* - Denegación de Servicios

5.8.6. Firma digital

El legajo de un alumno está compuesto sólo por fotocopias de la documentación original: título secundario, documentación de identidad, partida de nacimiento y ficha de inscripción. Cuando un documento digitalizado requiera la rúbrica digital, se tomará como base el “Manual de Procedimientos y recomendaciones para la Firma Digital” que propone la Universidad Nacional de Misiones [35]. Cabe aclarar que el mismo está basado en la Ley No 25.506 de “Infraestructura de Firma Digital – REPÚBLICA ARGENTINA ” y el “Manual de Procedimientos Política Única de Certificación de la AC ONTI”.

Si bien la Universidad está trabajando en la implementación de la firma digital, respecto a la implementación actual en los sistemas de información universitaria, el firmador digital de SIU Toba está obsoleto y ya no funciona en las últimas versiones de los navegadores.

El consorcio SIU está trabajando en el sistema SIU Araí [66]. El proyecto busca brindar una plataforma integradora de servicios del SIU. Cada uno de los sistemas SIU son módulos dentro de la plataforma.

Uno de estos módulos contempla la firma digital de documentos y expedientes electrónicos. En la Figura 41 se exhibe la arquitectura general que tendrá la plataforma, así como la vinculación entre las aplicaciones existentes y sistemas externos.

Replicación de datos para dar soporte al Programa de Expansión Territorial de la Universidad Nacional de Misiones

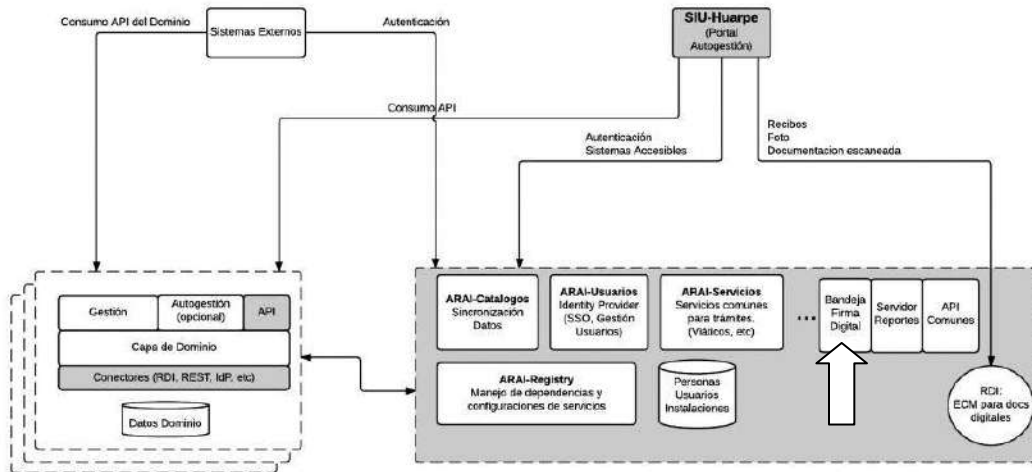


Figura 41. Arquitectura de la plataforma SIU Araí [66].

Es por lo expuesto que a la hora de signar un documento electrónico se sugiere la utilización de un software de tercero como puede ser: *open source Sinadura* [67].

Capítulo 6

Resultados

6. Resultados

En este capítulo se presentan los principales resultados obtenidos teniendo como referencia cada objetivo planteado en el proyecto.

6.1. Digitalización de la información a replicar.

Con la implementación del sistema LeDig, desde el periodo de pre inscripción 2018 se comenzó con la digitalización de toda la documentación de los legajos de los aspirantes de carreras de grado y pre grados. Con el correr de tiempo, también se procedió a digitalizar las cohortes de legajos pertenecientes a años académicos anteriores.

En la Figura 42 se muestra el trabajo llevado a cabo respecto a la digitalización de legajos en el sistema LeDig, comparada con la cantidad de alumnos ingresantes. Estos últimos datos fueron extraídos del SIU Guaraní.

También se exhibe la diferencia entre la cantidad de alumnos ingresantes y la cantidad de legajos digitalizados. La misma se debe principalmente a que, por decisiones administrativas, los legajos de las carreras de posgrado y los de alumnos condicionales todavía no han sido digitalizados.

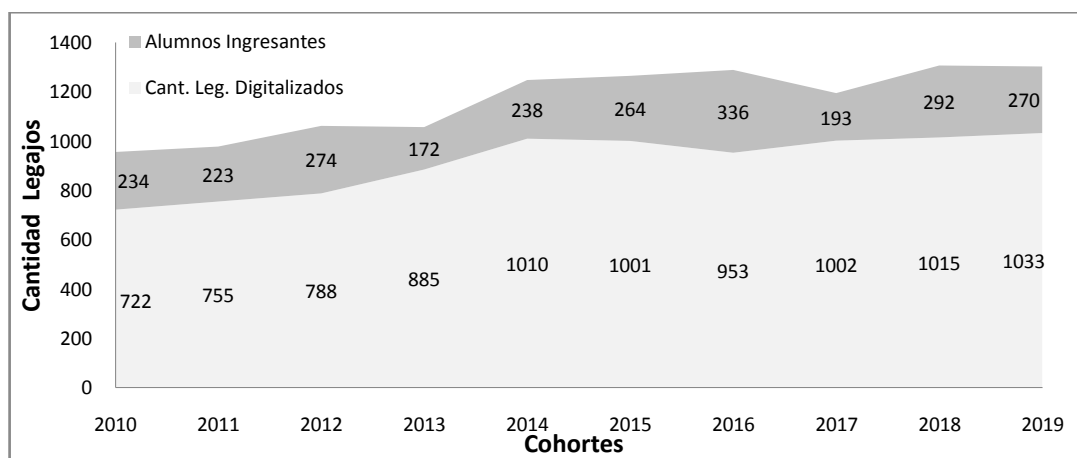


Figura 42. Legajos digitalizados por cohorte

Con el correr de tiempo y la práctica adquirida por los agentes de la Dirección de Estadísticas, fue posible, no solo dar cumplimiento a lo fijado por el Honorable Consejo Directivo, sino que también se procedió a digitalizar las cohortes de legajos pertenecientes a años académicos anteriores, teniendo en la actualidad más de 35 GB de presentaciones digitalizadas, correspondientes a más de trece mil legajos.

En la Figura 43 se muestra la evolución de cómo se fueron digitalizando las cohortes 2018 y 2019. Los datos comienzan a exhibirse desde el mes de noviembre del año anterior debido a que en ese mes comienzan el calendario de pre inscripciones a las carreras. Se puede apreciar que los picos de noviembre del año anterior (nov A-1) y de febrero coinciden con el cronograma de pre inscripción fijado por el calendario académico.

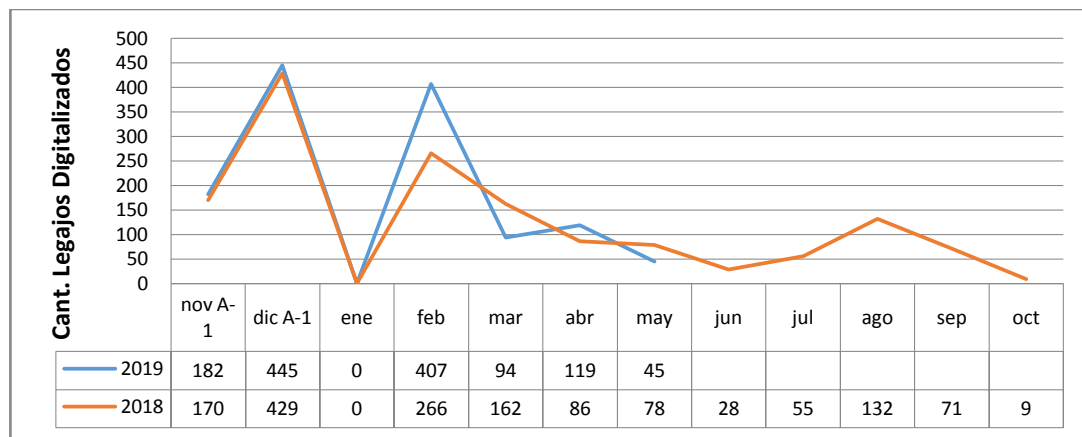


Figura 43. Legajos digitalizados. Evolución por mes

Contemplando todo el procedimiento de digitalización, inclusive el desglose de la documentación física, la digitalización completa en un legajo ronda actualmente entre los 7 y 10 minutos.

En lo que tiene que ver con el espacio en disco requerido, se verificaron los siguientes datos sobre los legajos digitalizados (Figura 44).

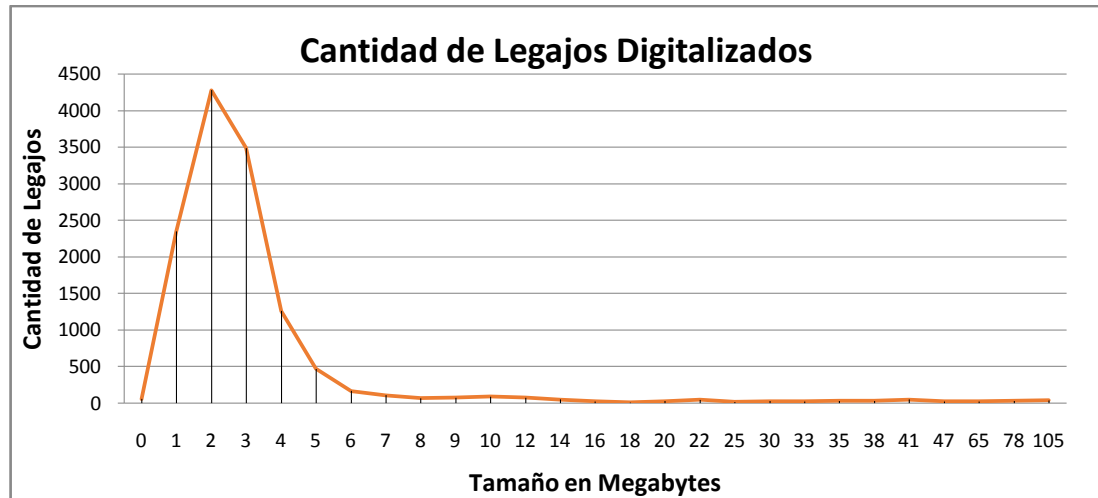


Figura 44. Análisis espacio requerido

De los 12.907 legajos digitalizados al momento de relevar los datos, se pudo verificar que el 88,08% (11.368 legajos) solo requieren, para su almacenamiento, entre 1 y cuatro 4 MB.

Se observan digitalizaciones que superan los 20 MB e incluso algunas mayores a los 100 MB. Esta situación es tratada en detalle en el capítulo de discusión.

6.2. Liberación de espacio físico

Otro de los principales inconvenientes por el cual se inició el presente trabajo, era la disposición del espacio físico, debido a que en las unidades académicas ya no se disponía de más lugares para seguir almacenando los legajos físicos de los alumnos.

En el mes de mayo del año 2018, la Dirección de Estadísticas ya ha comenzado con la devolución de los legajos físicos a los alumnos, una vez que la documentación ha sido digitalizada.

Actualmente se llevan devueltos 2.099 legajos. En proporción a la cantidad de legajos ya digitalizados (12.907), esta cifra representa el 16.26% de legajos devueltos a los alumnos en un periodo de, aproximadamente, diez meses. Se emplea el término de “aproximadamente” ya que en estos meses están incluidos los recesos invernales y de verano en los cuales no hay actividad académica ni administrativa.

De los 81 m² que actualmente tiene la oficina de estadísticas, a medida que se vacíen las estanterías, podrán liberarse 36 m². Este espacio representa el 44,44% de la superficie total de la oficina de la Dirección de Estadísticas (Figura 45).

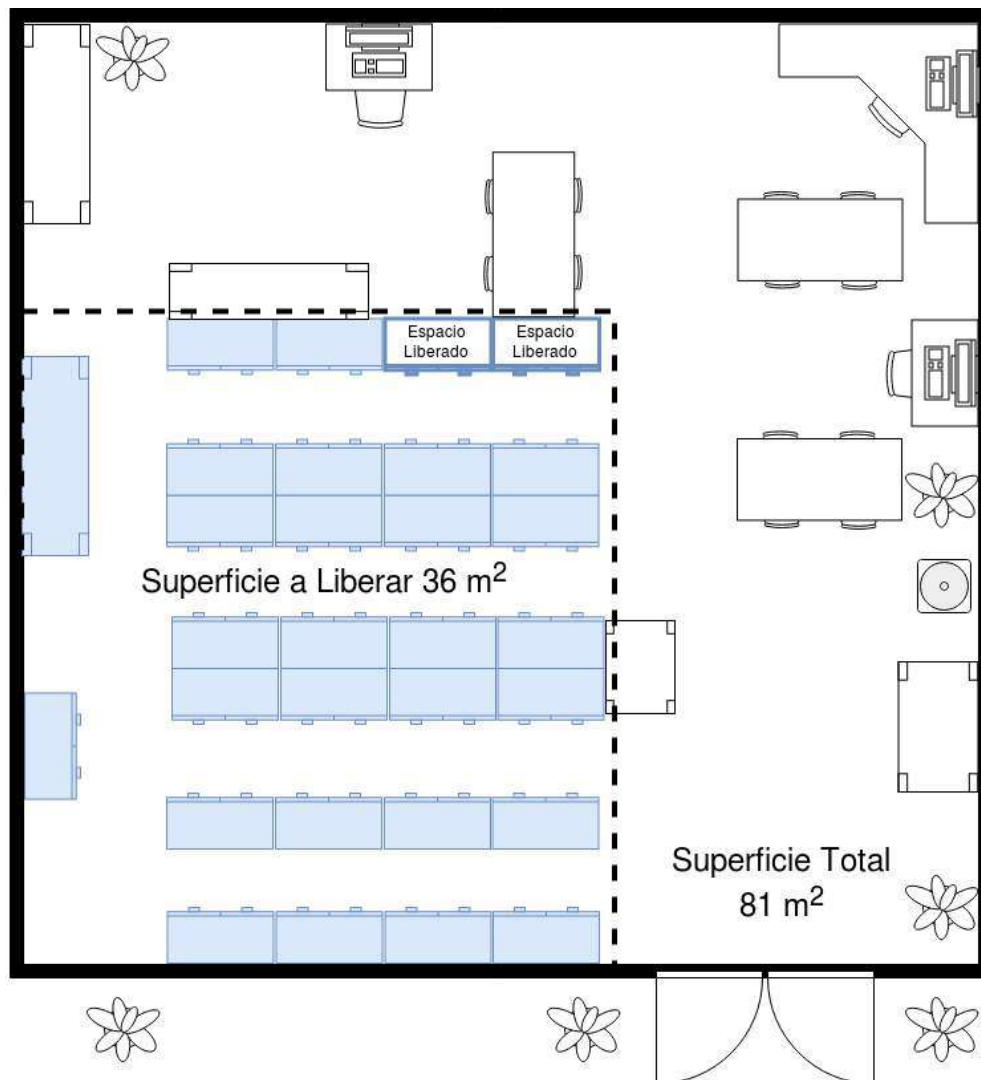


Figura 45. Espacio liberado en la Dirección de Estadísticas

Por medio de la digitalización y distribución de la información de los legajos, las unidades académicas podrán ir recuperando en forma gradual el espacio físico destinado al archivo de la documentación y los legajos.

Respecto a esto, si bien en el año 2018 se ha comenzado con la devolución de los legajos físicos a los alumnos, solo se han devuelto a los alumnos, poco más del quince por ciento (16,26%) de los legajos digitalizados y todavía no se han podido recuperar espacios físicos que sean significativos. Principalmente esto se debe a la falta de información, ya que a los alumnos de los años avanzados no se les comunicó que debían retirar su legajo. Esta situación comenzó a mejorar en agosto del 2019 con la difusión en el sitio WEB institucional, sistema SIU de Auto Gestión de los Alumnos, redes sociales y medios de comunicación oficiales de la facultad. A partir de esa fecha, en 30 días se devolvieron 585 legajos.

Solo en la oficina de la Dirección de Estadísticas, el espacio necesario para almacenar las estanterías con legajos representa casi el 50% de la superficie total de la oficina, el cual puede destinarse a otras necesidades que muchas unidades académicas deben cubrir realizando gastos de alquileres.

Como muestra de esto, vale mencionar que en una superficie equivalente a la actualmente ocupada por los legajos físicos en la Dirección de Estadísticas (36m²), donde antes funcionaba el Honorable Consejo Directivo de la Facultad de Ciencias Económicas, se construyeron tres oficinas donde actualmente funcionan las secretarías de: extensión, administrativa y consejo directivo.

6.3. Proyecto, desarrollo e implementación

En lo que respecta a la gestión del proyecto y el proceso de desarrollo del sistema se puede mencionar que, si bien los primeros borradores del proyecto comenzaron a elaborarse desde el mes de abril del 2017, el mismo fue presentado a la Secretaría Académica el 11 de mayo del mismo año. Desde esta fecha comenzó a trabajarse de manera conjunta con dicha secretaría, la Dirección de Estadísticas y la Dirección de Tecnología para la Gestión.

El 12 de septiembre del 2017, la Secretaría Académica elevó al Consejo Directivo el proyecto para que sea analizado por la comisión de Asuntos Académicos. Esto incluía un primer prototipo del sistema LeDig que solo contemplaba la digitalización de los legajos estudiantiles y que había sido probado digitalizando los legajos de las cohortes 2015, 2016 y 2017.

Luego se realizaron varias demostraciones sobre el prototipo en la comisión de Asuntos Académicos y en la octava sesión del Consejo Directivo, la cual tuvo como resultado que el 15 de noviembre del 2017 se apruebe por unanimidad, y por resolución de CD Nro. 190/2017 [2], se ordenó implementar el proyecto LeDig para los ingresantes a partir de la cohorte 2018 de todas las carreras de la Facultad de Ciencias Económicas.

Si bien en esta primera etapa se realizaron desarrollos de software a modo de prototipo, la mayor parte del tiempo estuvo dedicada a la elaboración y aprobación del proyecto, lo cual desde su inicio hasta su aprobación llevó ocho meses.

A partir de la aprobación en Consejo Directivo se comenzó a trabajar en la mejora constante del sistema LeDig. En diciembre del 2018, se incorpora al sistema la posibilidad de poder contar con réplicas de los legajos en las distintas sedes de la unidad académica.

Debido a que el proyecto solo contaba con un único desarrollador del Departamento de Sistemas de la Dirección de Tecnología para la Gestión, el cual también estaba afectado a resolver otras cuestiones del área, el tiempo de desarrollo se extendió a once meses.

Para registrar las actividades sobre el desarrollo se utilizaron elementos de metodologías ágiles y algunos elementos de Kanban [52]. Ambos implementados en partes con la herramienta de gestión de proyectos Redmine [53], habitualmente utilizada por el equipo de la Dirección de Tecnología para la Gestión.

Si bien en la actualidad se sigue mejorando el sistema, entre las entradas a la plataforma de gestión de proyectos (RedMine) relacionadas con el desarrollo integral

Replicación de datos para dar soporte al Programa de Expansión Territorial de la Universidad Nacional de Misiones

del sistema LeDig, gestión de los legajos, *bugs*, configuración de equipos (escáners), migración de datos y otros, se generaron 78 tickets (Figuras 46 y 47).

ID	Tipo	Estado	Prioridad	Asunto	Asignado a	Actualizado
709	Tareas	Nuevo	Normal	Bug Ledig - Modificación DNI	Germán Pautsch	2019-04-24 08:07
631	Tareas	Resuelto	Baja	ledig	Germán Pautsch	2019-04-24 11:04
584	Tareas	Resuelto	Normal	ledig-cohorte 2007	Germán Pautsch	2019-04-24 11:09
572	Tareas	Cerrada	Normal	Problema de acceso a sistema LeDig	Germán Pautsch	2019-04-24 11:05
562	Tareas	Resuelto	Normal	ledig Cohorte 2016	Germán Pautsch	2019-04-24 11:04
556	Tareas	Cerrada	Normal	ledig	Germán Pautsch	2019-04-24 11:02
536	Tareas	Rechazado	Normal	autorización cambio de cohorte en el ledig	Germán Pautsch	2018-12-05 07:31
525	Tareas	Resuelto	Normal	Cambio de Cohorte en el Ledig	Germán Pautsch	2018-12-05 07:51
438	Tareas	Pendiente de autorización	Normal	ledig	Germán Pautsch	2018-09-27 17:39
437	Tareas	Resuelto	Normal	Modificación de la cohorte en el LeDig	Germán Pautsch	2019-04-24 11:10
265	Tareas	Resuelto	Normal	No puedo recuperar alumnos en el Ledig	Germán Pautsch	2019-04-24 11:11

Figura 46. Product Backlog proyecto LeDig

Tareas #709

Bug Ledig - Modificación DNI

Añadido por Germán Pautsch hace 4 minutos. Actualizado hace 1 minuto.

Estado: Nuevo Fecha de inicio: 2018-04-24
 Prioridad: Normal Fecha fin:
 Asignado a: Germán Pautsch % Realizado: 0%
 Categoría: - Tiempo dedicado: -
 Versión prevista: -

Descripción

Ante una modificación del dni (PK) en tabla ledig desde gestión el campo queda en disabled. Hacer que quede en enabled y que luego se haga un rename de lo carpeta en el file system. Verificar funcionamiento de la URL resultante.

Subtareas Añadir

Peticiones relacionadas Añadir

Figura 47. Detalle de un requerimiento en Redmine

El tiempo de duración de las principales etapas: elaboración y aprobación del proyecto, desarrollo y replicación de datos, se muestra en la Figura 48.

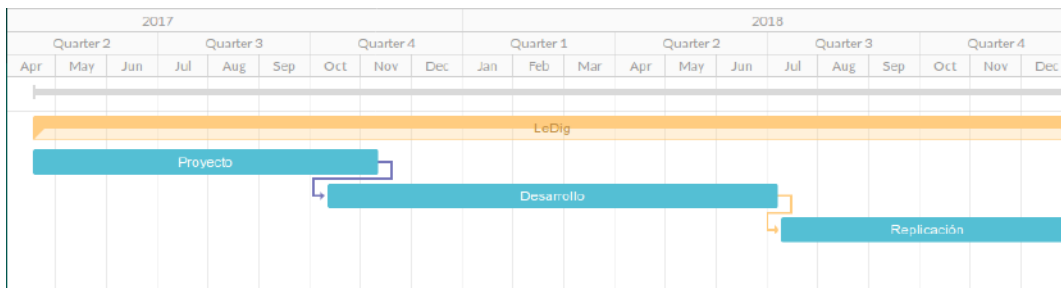


Figura 48. Duración del proyecto LeDig

6.4. Replicación de datos

En lo que respecta a las mediciones sobre el tráfico de red generados por la replicación de los datos, tanto en el puerto de la aplicación (80) como en el de la base de datos (5432), los resultados se han recolectado utilizando la herramienta *RRDtool*²³. La misma provee la posibilidad de: monitorear un puerto, definir una base de datos y representarlos gráficamente a través del tiempo.

Los datos que se representan en la herramienta se detallan en la Tabla 15.

Atributo	Descripción	Unidad
Data In	Volumen de datos que entran por el puerto	bytes
Data Out	Volumen de datos que salen por el puerto	bytes
Total	Resultado de la suma entre Data In y Data Out	bytes
Max In	Máxima velocidad detectada en la entrada del puerto	bits/seg
Max Out	Máxima velocidad detectada en la salida del puerto	bits/seg
Last In	Velocidad de la última medición en la entrada del puerto	bits/seg
Last Out	Velocidad de la última medición en la salida del puerto	bits/seg

Tabla 15. Atributos herramienta *RRDtool*

Las mediciones han sido realizadas en el puerto 5432 (*PostgreSQL*) entre el nodo Maestro y el nodo Sede, con la replicación activa a setenta kilómetros de distancia. Con la finalidad de obtener datos relevantes bajo condiciones pesimistas,

²³ *RRDtool: Round Robin Database Tool*

para realizar dichas mediciones se incorporaron tres tablas más al clúster de replicación, quedando el mismo definido por las siguientes tablas:

- ledig
- leg_presentaciones
- tipo_doc
- sga_carreras
- leg_depositarios

En la Figura 49 se exhiben las mediciones realizadas en el puerto 5432 (*PostgreSQL*) entre el nodo Maestro y el nodo Sede, con la replicación activa pero sin replicar ningún dato. Esto con la finalidad de establecer los bits/segundo que demandan los eventos de sincronización (SYNC) que genera *slon*. Se puede observar que bajo las condiciones descriptas, el tráfico generado es en promedio a 3,18 kbps.



Figura 49. Tráfico demandado por la sincronización

La Figura 50 brinda en contraste con la anterior (Figura 49), datos sobre el monitoreo en la salida del puerto 5432 de nodo Maestro. Se pudo determinar que para la replicación de 20 legajos con sus respectivas presentaciones de documentación, con el sistema ya en producción, en la cual entre las 8 y 12 horas aproximadamente, ha sido requerido en promedio 12,13kbps.

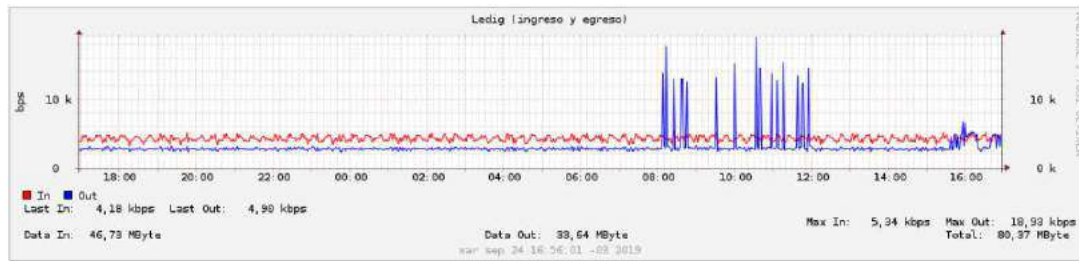


Figura 50. Replicación de tuplas con el sistema en producción

Posteriormente, se verificó que al no detectar un vínculo de red activo entre el nodo Maestro y el nodo Sede, la sincronización en el nodo maestro queda en un estado de espera. La Figura 51 muestra una captura de pantalla al momento de producirse una caída del vínculo entre los nodos.

Primeramente el evento de sincronización “SYNC 5000000017” es el último en generarse en el nodo Maestro y luego de cumplido el tiempo de espera configurado, el servicio de *slon* emite un mensaje de error por *timeout* debido a la falta de comunicación con el nodo Sede.

```
INFO remoteworkerThread_2: SYNC 5000000009 done in 0.005
seconds
INFO remoteworkerThread_2: SYNC 5000000010 done in 0.004
seconds
INFO remoteworkerThread_2: SYNC 5000000011 done in 0.004
seconds
INFO remoteworkerThread_2: SYNC 5000000012 done in 0.005
seconds
INFO remoteworkerThread_2: SYNC 5000000013 done in 0.005
seconds
INFO remoteworkerThread_2: SYNC 5000000014 done in 0.005
seconds
INFO remoteworkerThread_2: SYNC 5000000015 done in 0.003
seconds
INFO remoteworkerThread_2: SYNC 5000000016 done in 0.005
seconds
INFO remoteworkerThread_2: SYNC 5000000017 done in 0.004
seconds
ERROR remoteListenThread_2: timeout (300 s) for event selection
NOTICE: slony-I: cleanup stale sl_nodelock entry for pid=5713
NOTICE: slony-I: cleanup stale sl_nodelock entry for pid=5714
NOTICE: slony-I: cleanup stale sl_nodelock entry for pid=5716
NOTICE: slony-I: cleanup stale sl_nodelock entry for pid=5744
NOTICE: slony-I: Logswitch to sl_log_2 initiated
INFO cleanupThread: 0.017 seconds for cleanupEvent()
ERROR slon_connectdb: PQconnectdb (dbname = sede) failed - could
not connect to server: No route to host
Is the server running on host "sede" and accepting
TCP/IP connections on port 5432?
```

Figura 51. Caída vínculo. Consola *slon* nodo Maestro

De modo similar, en la consola del nodo Sede, también se exhibe información a cerca de la caída del vínculo (Figura 52).

```
INFO remoteworkerThread_1: SYNC 5000000038 done in 0.008 seconds
INFO remoteworkerThread_1: syncing set 1 with 5 table(s) from
provider 1
INFO remoteworkerThread_1: SYNC 5000000039 done in 0.008 seconds
INFO remoteworkerThread_1: syncing set 1 with 5 table(s) from
provider 1
INFO remoteworkerThread_1: SYNC 5000000040 done in 0.008 seconds
INFO remoteworkerThread_1: syncing set 1 with 5 table(s) from
provider 1
INFO remoteworkerThread_1: SYNC 5000000041 done in 0.008 seconds
ERROR remoteListenThread_1: timeout (300 s) for event selection
ERROR slon_connectdb: PQconnectdb(dbname = ledig) failed - could not
connect to server: Network is unreachable
Is the server running on host "maestro" and accepting
TCP/IP connections on port 5432?
```

Figura 52. Caída vínculo. Consola *slon* nodo Sede

La visualización de estos eventos y su confirmación quedan registrados en el catálogo de *Slony-I*, en las tablas *_ua.sl_event* y *_ua.sl_confirm*, respectivamente. Esto permite hacer un seguimiento de cuáles fueron los últimos cambios replicados en el nodo Sede.

En las tablas *_ua.sl_log_1* y *_ua.sl_log_2* quedan registrados datos más específicos sobre los cambios replicados como: nombre de la tabla, atributo y valor.

Con respecto al tiempo necesario para ver reflejado en el nodo esclavo lo actualizado en el nodo maestro, los tiempos de respuesta varían entre 2 y 4 segundos.

6.5. Alta disponibilidad de la información

Para brindar alta disponibilidad de la información a las distintas áreas y sedes de la Universidad Nacional de Misiones, a fin de asegurar la continuidad de los trámites y operaciones y continuando con el desarrollo del tema anterior, cuando el vínculo se establece, *slon* lo detecta, confirma las transacciones pendientes, si las hay, para luego seguir generando los eventos de sincronización.

En la Figura 53 se pueden apreciar las mediciones realizadas en el puerto 80 (Apache) por el cual opera la aplicación LeDig. Se demuestra que durante el periodo

que duró la caída del vínculo, el sistema continuó operativo, específicamente, entre las 7:30 am y 12:30 pm. De este modo se otorga disponibilidad del sistema, aún cuando no hay comunicación entre los nodos, asegurando así la continuidad de los trámites y operaciones en ambos sitios.

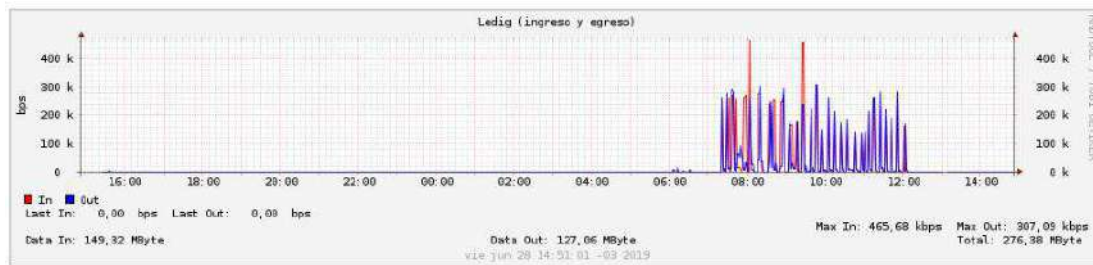


Figura 53. Tráfico en puerto 80 (Apache)

Al restablecerse la comunicación, los datos generados durante la caída del vínculo en el nodo Maestro, son replicados al nodo Sede. En la Figura 54 se muestran el monitoreo realizado sobre el puerto 5432, en la misma franja horaria que la Figura 53.

Se puede observar que aproximadamente a las nueve horas del día siguiente, cuando la conexión de red fue restablecida, se genera un pico en el consumo del ancho de banda, provocado por la replicación de 24 legajos, equivalentes a 13,70 MB, y cuya velocidad alcanza un máximo de 10,97 kbps.

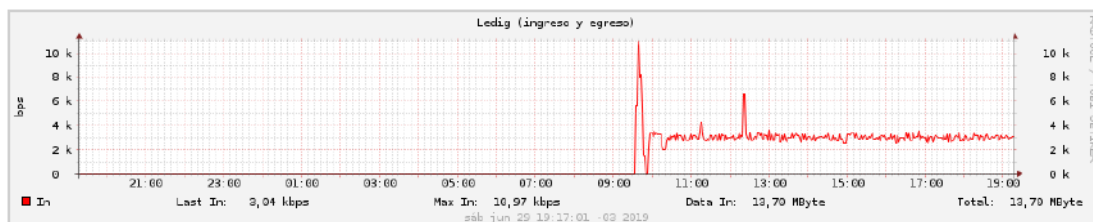


Figura 54. Tráfico diario en puerto 5432 (PostgreSQL)

La Figura 55 muestra varios cortes de red durante la semana 36 del año 2019 debido a reestructuraciones realizadas en el sitio Sede. A la fecha, la demanda detectada sobre el canal de comunicación entre los nodos en el puerto 5432, rara vez supera los 10 kbps.



Figura 55. Tráfico semanal en puerto 5432 (PostgreSQL)

Cuando se produce la caída del vínculo de red entre los nodos, se comprobó que solo basta con monitorear la situación y esperar que el vínculo se restablezca, ya que al activarse la comunicación, la replicación se activa automáticamente sin requerir de ninguna acción por parte del administrador de la base de datos.

Respecto a la gestión del sistema de archivos PDF, se detectó que la copia de los mismos, del nodo Maestro al nodo Sede, al momento en que el alumno realiza la presentación de la documentación, afectaba directamente la navegabilidad, generando tiempos de espera y quejas por parte de los usuarios.

Sobre el monitoreo realizado desde el navegador, sobre la interfaz del sistema LeDig, al momento de cargar un archivo de un legajo digitalizado, se obtuvieron los siguientes datos (Tabla 16):

Tamaño PDF en KB	Tiempo de espera en segundos
2521	00:12
2264	00:10
2572	00:13
2247	00:10

Tabla 16. Tiempos de espera por archivo

Estas demoras mínimamente se duplican si en ese mismo instante se dispara la copia de la presentación digitalizada al sitio Sede.

Debido a esto, se decidió realizar la copia de los archivos a través del protocolo *ssh* desde el planificador de tareas²⁴ para, de este modo, no alterar la calidad percibida en la navegación por parte de los usuarios.

Teniendo en cuenta todo lo anteriormente expuesto y que, desde la Dirección de Estadísticas se manifestó que la disposición inmediata de los legajos en la sede no es crítica para el funcionamiento del sector, se optó por actualizar el sistema de archivos en momentos que el alto tráfico no perjudique la operatividad de los demás sistemas. Esto último resultó en la generación de un *script* llamado *copyFS.php* que actualiza el sistema de archivos de la Sede a las 23 horas de cada día. Para contemplar fallos o cortes en el suministro de la energía eléctrica o en las conexiones de red, posteriormente se agregó al planificador de tareas una segunda ejecución del mismo *script*, que de fallar la primera se dispara a las 5 horas a.m copiando los archivos de los legajos digitalizados que no hayan sido contemplados en la primera instancia.

Los resultados obtenidos con este método se pueden ver en la Figura 56. Para probar el consumo del ancho de banda y el tiempo demandado en ambas ejecuciones se realizó la copia de 50 legajos a las 23 horas y otros 50 legajos a las 5 horas del día siguiente. Se establece realizar la copia de 50 legajos como un umbral que exceda al promedio de legajos digitalizados por día que en promedio es de 30 legajos.

²⁴ En Linux *crontab*



Figura 56. Transmisión de archivos (ssh) desde el sitio Maestro al nodo Sede

Realizando la misma prueba en ambos horarios pero utilizando la compresión de los datos ofrecida por el parámetro “-C” de la instrucción de copiado *scp* vía *ssh* se puede obtener una mejora significativa en los tiempos requeridos en la transmisión de los datos (Figura 57).



Figura 57. Transmisión de archivos con compresión (ssh -C).

6.6. Resguardo y recuperación

Otra premisa importante que se estableció y se pudo satisfacer, fue la mejora en la calidad del resguardo de los legajos. Respecto a esto, tanto la aplicación como el sistema de archivos se encuentran virtualizados en la plataforma *Proxmox*. Se realizan resguardos diarios de todo el sistema en distintas ubicaciones, lo que permite a su vez contar con varias unidades de recuperación de toda la plataforma.

A fines del mes de abril del año 2019 se pudo constatar la eficiencia del sistema de resguardo y recuperación debido a un ataque externo sufrido en la base de

datos por haber dejado el puerto 5432 abierto (Tareas #724: Hackeo en Servidor de Sistemas).

El inconveniente fue reportado por un usuario del sistema el día lunes 29 de abril a las 7:28 am por medio del incidente número #719. Luego de realizar algunas verificaciones en los *logs* del servidor, se constató que el atacante ingresó y eliminó todas las tablas del sistema el día sábado 27 de abril a las 2:07 am.

Debido a que el ataque fue un día no laborable en donde no se generan nuevos datos, con el equipo técnico de la Facultad de Ciencias Económicas se decidió restaurar el *backup* del contenedor virtual del día 26 de abril.

Desde que se abordó el problema hasta que el sistema fue restablecido transcurrieron dos horas. En gran parte, este tiempo fue utilizado en el copiado y luego en el proceso de restauración de la máquina virtual ya que la misma a la fecha mencionada tenía un tamaño de 40 GB correspondiente a 9.164 legajos digitalizados.

6.7. Agilidad en el acceso, búsqueda y actualización de la información

Con la instancia *Hadoop* provista por la herramienta *Pentaho Data Integration*, se generó un proceso para extraer, de los archivos *logs* generados por el servicio de *Apache*, todas las peticiones de tipo *GET* realizadas a archivos con extensión PDF dentro del directorio del sistema de archivos donde opera LeDig.

Con estos datos se pudo relevar la cantidad de consultas semanales durante un año de utilización del sistema. La comparativa entre el promedio de consultas manuales, las realizadas por medio del sistema LeDig (solo en la Dirección de Estadísticas) y el promedio actual de consultas luego de la implementación del sistema, se exhiben en la Figura 58. Cabe aclarar que hasta la semana 23, los datos corresponden al año 2019, y desde allí en adelante el gráfico fue confeccionado con la información del año anterior.

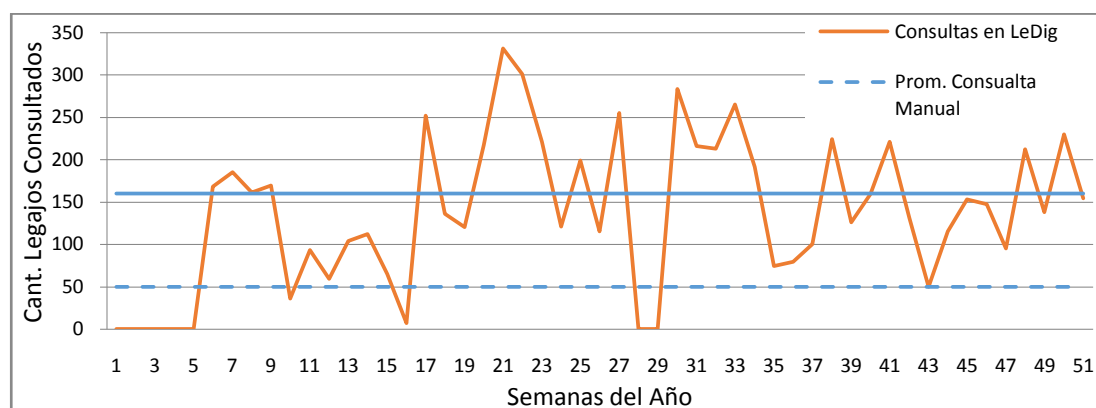


Figura 58. Comparativa sobre la cantidad de legajos consultados por semana

Relacionado al tiempo que demanda la búsqueda y consulta manual de un legajo, si bien no se llevaba un registro, desde la Dirección de Estadísticas se informó que con la implementación del sistema LeDig, el tiempo necesario para la búsqueda de los legajos se ha reducido considerablemente.

Con la ausencia del sistema esta tarea demandaba en promedio diez minutos luego de producirse la solicitud del demandante. Con la implementación del LeDig no solo que el acceso al legajo toma unos pocos segundos, sino que también varios usuarios pueden acceder a los legajos al mismo tiempo, quedando registros de auditoría sobre cualquier acción realizada. Todo esto, sin la necesidad de que las personas se trasladen y fundamentalmente sin necesidad de que el legajo físico sea trasladado, lo cual frecuentemente generaba el deterioro del mismo.

Por otra parte, la agilidad y concurrencia en la consulta, provocó un aumento en promedio de legajos consultados por semana, el cual en la actualidad se elevó en promedio a ciento sesenta legajos. Esto triplica las consultas realizadas en forma manual, lo cual operativamente, con el personal existente en la Dirección de Estadísticas, sería inviable atenderlas.

6.8. Interfaz SIU Guaraní y SIU Pre Inscripción

Para poner a disponibilidad del personal No Docente la mayor cantidad de registros, con la finalidad de no tener que cargarlos en forma manual, se tuvo en cuenta para la importación, todas las carreras, incluso las de pos grado, y tanto los nuevos inscriptos (NI) como los re inscriptos (RI) a cada año académico. Estos últimos son alumnos de la facultad que por algún motivo han abandonado la carrera, pero en ese año académico se han re inscripto a la misma.

La cantidad de registros importados desde el SIU Guaraní se detallan en la Tabla 17.

Año Académico	NI + RI
2010	1404
2011	1504
2012	1711
2013	1517
2014	1746
2015	2026
2016	1926
2017	1643

Tabla 17. Cantidad Legajos Importados desde el SIU Guaraní

A partir del año 2018 en que se implementó el sistema LeDig, los metadatos ya no fueron importados desde el SIU Guaraní, sino que para ello el sistema los importa desde la interfaz que tienen LeDig con el sistema de Pre Inscripción.

La cantidad de registros importados desde el SIU Pre Inscripción se detallan en la Tabla 18.

Año Académico	Pre inscriptos
2018	1015
2019	1044

Tabla 18. Cantidad Legajos Importados desde pre inscripción

La interfaz con el sistema de pre inscripción, ahora también permite, acceder al registro completado por el aspirante posibilitando modificar algunos datos erróneos o completar otros faltantes. Esto evita que desde la Dirección de Estadísticas se levanten incidentes para solucionar estas cuestiones, pudiendo así resolver la situación en el mismo instante en que se presenta el inconveniente y evitando, sobre todo, generar demoras en el proceso de pre inscripción otorgando autonomía al sector. Esta misma interfaz también automatiza la confección del padrón de los alumnos condicionales, que antes se hacía en forma artesanal.

Capítulo 7

Discusión

7. Discusión

En este capítulo se brindan las interpretaciones, significados, impacto y recomendaciones sobre los principales resultados expuestos en el capítulo anterior.

Los resultados obtenidos respecto a la evolución en el tiempo sobre la digitalización de los legajos indica que este procedimiento no provoca retrasos significativos en los procesos administrativos del sector, ya que de lo contrario, la gráfica de la Figura 43 de la sección 6.1, no registraría los picos mencionados, lo cual indicaría que el tiempo demandado por el proceso de digitalización supera a la capacidad de los agentes respecto a la atención de los aspirantes.

Respecto al proceso de digitalización de los legajos, el principal inconveniente que prolonga los tiempos al momento de realizar el escaneo de la documentación, es la tarea de desglosar la documentación física, debido a la utilización de broches y ganchos en las hojas. A fin de agilizar el desglose, es recomendable no utilizar este tipo de sujeción en los legajos físicos. Contemplando esta demora se constató que esta tarea no afecta la atención y operatoria de la Dirección de Estadísticas.

Los datos relevados permitieron proyectar que para una cohorte, de un promedio de mil legajos, se requerirán aproximadamente, 3 GB de espacio en disco. Sin embargo, en el gráfico de la Figura 44 de la sección 6.1, se han detectado legajos digitalizados que superan los 20 MB y otros incluso mayores a 100 MB. En la Figura 59 se exhiben más en detalle la cantidad de legajos con estos rangos de valor.

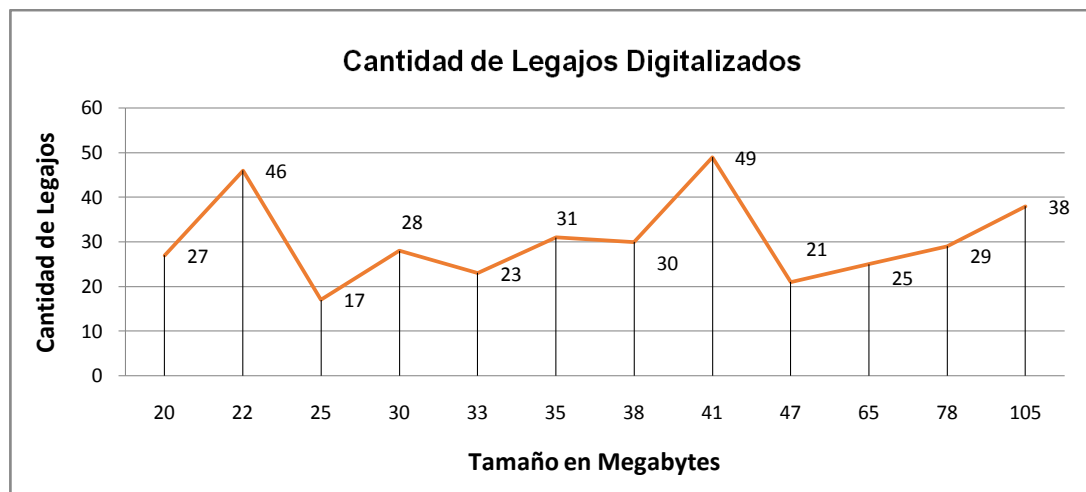


Figura 59. Digitalizaciones con tamaño que superan la media

Accediendo a la documentación digitalizada de estos legajos se pudo corroborar que estos archivos pertenecen a digitalizaciones de expedientes de trámites, como por ejemplo: solicitudes de equivalencias de alumnos que provienen de otras facultades y en cuyos expedientes obran las certificaciones de los programas de las materias que el alumno ha aprobado, pudiendo contener cientos de fojas.

Para evitar digitalizaciones innecesarias como por ejemplo la de los expedientes de solicitud de equivalencias, lo cual demanda un gran esfuerzo del personal, disminuye la productividad y en algunos casos generan archivos con más de 100 MB, se ha sugerido a la secretaría académica, que solicite al Honorable Consejo Directivo, reglamentar qué documentación debe ser digitalizada y cual no. Sobre todo si es necesario digitalizar el proceso, en este caso representado por el expediente de equivalencias, o simplemente basta con digitalizar la disposición que le otorga al alumno el reconocimiento.

Si la disponibilidad de documentos es necesaria en los sistemas a replicar, se pudo demostrar que se puede programar una buena sincronización para poner a disposición de las extensiones territoriales estos archivos. En el caso de validación

desarrollado, la única manera de que un documento digitalizado no pueda estar disponible, es que se haya digitalizado recientemente, que todavía no se haya copiado vía *ssh* al final de la jornada laboral, y que el vínculo con el nodo maestro esté caído. Esta situación puede mejorarse, ya que la sincronización para el copiado de los documentos puede configurarse para adaptarse a las distintas necesidades.

Respecto a la actualización del sistema de archivos en un sitio esclavo, no se deben descartar alternativas como por ejemplo, membrecías de *Dropbox* que por una suma de 12,50 US\$ mensuales ofrece:

- 3 TB (3 072 GB) de espacio
- Recuperación de archivos durante 120 días
- Cifrado AES de 256 bits y SSL/TLS
- Sincronización inteligente
- Herramientas administrativas de *Dropbox Paper*
- Integración con Office 365
- Consola de administración y registro de auditoría
- Permisos detallados de uso compartido
- Grupos administrados por los usuarios y por la empresa
- Borrado remoto de dispositivos
- Autenticación de dos factores (2FA)
- Acceso ilimitado a la API para los socios de la plataforma de seguridad y productividad
- Un millón de llamadas a la API por mes para los socios de transferencia de datos
- Soporte por chat en tiempo real

En lo que tiene que ver con la replicación de las tablas en la base de datos, se observó que cuando la comunicación entre los nodos se restablece, se producen leves picos en el consumo del ancho de banda, lo cual no compromete la operatividad en ninguno de los sitios. Estos picos en el tráfico van a depender del tiempo que dure la

caída del vínculo y de la cantidad de tuplas que se generen en el mismo lapso en el nodo Maestro.

Analizando la frecuencia transaccional que tienen las tablas en la base de datos, se puede realizar una replicación más selectiva, reduciendo de este modo, la complejidad del *clúster* y haciendo la replicación más eficiente.

Es importante remarcar que tanto en el sitio maestro como en los esclavos, se asegura el principio de transparencia de los sistemas distribuidos, ya que los usuarios finales no perciben la complejidad de la replicación y la distribución de los datos, operan el sistema como un único medio.

Luego del ataque recibido en la base de datos el 27 de abril del 2019, quedaron demostradas las ventajas de tener el sistema virtualizado ya que, desde que se abordó el siniestro hasta que el sistema fue restablecido, solo transcurrieron dos horas. Como gran parte de ese tiempo fue consumido en el copiado de la máquina virtual, que a su vez contenía el sistema de archivos con más de doce mil legajos digitalizados, se recomienda en próximas instalaciones, separar el sistema de archivos de la aplicación para, de este modo, facilitar las tareas de mantenimiento y/o restauración del sistema.

Si bien la Dirección de Estadísticas ya ha comenzado con la devolución a los alumnos de las carpetas físicas de los legajos digitalizados, la resistencia al concepto de que la facultad ya no sea la depositaria de los legajos y la falta de conocimiento por parte de alumno de que debe retirar su legajo físico, fueron los principales motivos por el cual este procedimiento se demoró.

Para motivar, agilizar y tener un control preciso sobre la devolución de los legajos, se recomienda para las futuras cohortes, agregar un requisito en el SIU Guaraní que podría ser llamado “Legajo Físico”. A través de la activación de controles sobre los requisitos, se puede generar que el alumno se acerque a la Dirección de Estadísticas para verificar su situación. Es en esta instancia que se le devolvería la carpeta al alumno y se daría como cumplido el requisito.

Si bien la implementación de este requisito podría acelerar la recuperación del espacio físico, debe hacerse en una fecha del calendario académico en la que no perturbe las funcionalidades del sector, ya que esto generará una concurrencia masiva de los alumnos con legajo digitalizado a la Dirección de Estadísticas para retirar su legajo.

Otro de los logros obtenidos con la implementación del sistema LeDig y relacionados con los gastos de insumos, es que el dinero destinado a muebles metálicos para carpetas colgantes y papelería puede destinarse a la cobertura de otras necesidades.

En lo que tiene que ver con el almacenamiento, administración y transporte de los legajos, el sistema LeDig permite la gestión integral de los mismos, ya no siendo necesario el traslado de la documentación física y evitando el posible deterioro o pérdida de la documentación. Los agentes de la facultad pueden tener acceso, incluso en forma concurrente, a la documentación de los legajos sin la necesidad de trasladarse a la oficina de la Dirección de Estadísticas o depender de la presencia o disponibilidad del personal de dicho sector.

En lo que respecta a los gastos de papelería y estantes archivadores para las carpetas colgantes, cubrir estas necesidades le demandaba a la unidad académica un monto superior a \$40.000 pesos por año. Esto teniendo en cuenta solo:

- Mil Carpetas colgantes²⁵:.....\$26.820
- Mueble metálico para carpetas colgantes²⁶:.....\$21.820

Fuente: Mecadolibre.com.ar. Fecha de consulta: 26/08/2019

²⁵ Fuente: https://articulo.mercadolibre.com.ar/MLA-677977522-carpeta-colgante-delta-color-ladrillo-ventana-fija-x-100-un-_JM?quantity=1

²⁶ Fuente: https://articulo.mercadolibre.com.ar/MLA-673267075-armario-metalico-puertas-corredizas-_JM?quantity=1#reco_item_pos=0&reco_backend=machinalis-seller-items&reco_backend_type=low_level&reco_client=vip-seller_items-above&reco_id=4171258f-bfa7-497c-b306-481c2297701f

Relacionado a la agilización del proceso de consultas sobre legajos, si bien no existe un registro de la cantidad de consultas que se producían en la Dirección de Estadísticas, el personal informó que en promedio tenían aproximadamente diez consultas diarias, lo cual hace un acumulado semanal de cincuenta legajos consultados.

Si los picos de consultas exhibidos en la Figura 57 de la sección 6.7 debieran atenderse en forma manual, teniendo en cuenta lo anteriormente mencionado:

- Búsqueda Manual de un Legajo.....10 minutos
- Búsqueda Manual de 160 Legajos.....1.600 minutos (más de 26hs)

Esta proyección permite establecer que un agente que trabaja treinta y cinco horas semanales por contrato, debería estar casi veintisiete horas semanales, solamente atendiendo estas consultas de legajos.

También vale aclarar que para el cálculo del promedio de consulta actual (160 legajos), solo se tuvo en cuenta a los usuarios de la Dirección de Estadísticas.

La conectividad del sistema LeDig con las bases de datos del SIU Guaraní y SIU Pre Inscripción, facilitaron la carga de los metadatos que describen a las presentaciones realizadas (PDF), evitando que los agentes administrativos del sector tenga que realizar la carga manual de los mismos, siendo esto uno de los factores claves para que la digitalización de los ingresantes del año 2019 se realice de forma simultánea al periodo de pre inscripción.

Esto se logró en gran parte con la herramienta *Pentaho Data Integration*, la cual permite una excelente integración de datos entre los distintos motores, pudiendo incluso programar cualquier *script* que la herramienta no posea disponible. Los resultados obtenidos fueron muy satisfactorios, incluso se implementó un flujo de datos que integra los legajos de los alumnos ingresantes situados en la base de datos

Replicación de datos para dar soporte al Programa de Expansión Territorial de la Universidad Nacional de Misiones

del SIU Guaraní, y genera los usuarios correspondientes para el Aula Virtual y el sistema Koha de bibliotecas.

Capítulo 8

Conclusiones

8. Conclusiones

En este capítulo se describen el cumplimiento de los principales objetivos alcanzados así como también otras contribuciones inesperadas. Luego se describen algunas de las limitaciones detectadas en la utilización de este enfoque para, por último, identificar cuales podrían ser las futuras líneas de investigación.

8.1. Objetivos alcanzados

Con el trabajo realizado, se demuestra que los sistemas existentes en la universidad, cuentan con un antecedente que ha permitido superar las limitaciones de la actual arquitectura cliente servidor, facilitando y dando un soporte más adecuado a las expansiones territoriales, asegurando la continuidad de las operaciones diarias incluso ante la ausencia de conectividad.

El desarrollo del enfoque de arquitectura distribuida para la replicación de datos, se presenta como una alternativa válida para dar soporte a procesos específicos de instituciones con objetivos dirigidos a la expansión territorial.

Las mediciones indicaron que cuando la replicación está activa, incluso añadiendo tres tablas más al clúster de replicación que lo habitualmente demandado, no eleva de manera considerable el tráfico entre los nodos, y si se realiza un análisis previo sobre la cantidad de transacciones que posee cada tabla de la base de datos, para luego clasificar qué tablas realmente necesitan formar parte del clúster de replicación, esto permite reducir el tráfico necesario para la actualización de los nodos.

Uno de los principales interrogantes al inicio del proyecto era saber si el procedimiento de digitalización de los legajos era lo suficientemente ágil como para permitir llevarse a cabo de manera simultánea al proceso de pre inscripción a carreras. Con la utilización del sistema LeDig, se pudo validar que el procedimiento de digitalización de los legajos no afecta la atención y operatoria de la Dirección de

Estadísticas, inclusive en épocas críticas del calendario académico, como es el período de inscripción a carreras.

Por otra parte con las configuraciones realizadas en los dispositivos de escaneo, se ha logrado garantizar una calidad, que permite, llegado el caso y dada la necesidad, que la reimpresión del legajo físico con una calidad legible y similar al original.

Asimismo, la evaluación del caso de estudio permite demostrar que la adopción de esta propuesta contribuye al logro de requerimientos no funcionales de los sistemas de información como el acceso concurrente, la alta disponibilidad, seguridad, eficiencia y usabilidad del sistema y los datos gestionados.

El enfoque propuesto permite un aumento de la productividad en los sectores que intervienen en la gestión de los legajos de los estudiantes. En particular, esto se refleja en el incremento de las consultas sobre los legajos de los estudiantes en el sistema LeDig.

Si bien la adquisición de los dispositivos de escaneo representan un costo considerable para las instituciones, el caso de estudio del sistema de legajos de alumnos de la Facultad de Ciencias Económicas, demuestra que la implementación de soluciones tecnológicas junto con una adecuada definición de los procesos, evita la compra de otros activos y permite optimizar el uso de recursos para las organizaciones, por lo general limitados.

8.2. Contribuciones adicionales

La integración de las distintas bases de datos ha originado algunas contribuciones no contempladas en los objetivos del trabajo que son importantes mencionar.

Particularmente la interfaz con el SIU Pre Inscripción permite generar automáticamente el padrón de alumnos condicionales que, anteriormente era

confeccionado en forma artesanal revisando la documentación presentada por el aspirante.

El hecho de permitir al personal de la Dirección de Estadísticas, gestionar algunos datos básicos del registro de pre inscripción del aspirante, posibilita realizar correcciones en la información cargada vía WEB por el aspirante. Esto facilita la resolución de inconvenientes en el mismo momento en que el aspirante se presenta con la documentación para finalizar su pre inscripción.

Al estar integradas todas las bases de datos, una vez que los legajos son grabados en el sistema SIU Guaraní, automáticamente se generan los usuarios para el resto de los sistemas como ser: aula virtual y biblioteca. Esto ha permitido homogeneizar criterios y evitar duplicaciones en los perfiles que los alumnos poseen dentro de cada sistema.

Es importante destacar que luego de la implementación del sistema LeDig en la Facultad de Ciencias Económicas, la Facultad de Ciencias Exactas Químicas y Naturales acompañó esta iniciativa aprobando la implementación del sistema LeDig por unanimidad en la séptima sesión ordinaria de Consejo Directivo por medio de la resolución número 572/2018 [68].

En la actualidad el sistema LeDig se encuentra instalado y operativo en ambas unidades académicas, replicando los datos del Módulo de Apóstoles, donde se dictan las carreras de informática, a 70 kilómetros de distancia de la sede central y con más de tres mil legajos digitalizados.

8.3. Limitaciones detectadas

Entre las limitaciones detectadas durante el desarrollo del trabajo se puede mencionar que necesariamente la estructura de las tablas a replicar, debe ser la misma en ambos nodos. Esto impide realizar replicaciones fragmentadas, ya sea en forma horizontal o vertical. De ser necesaria la fragmentación, podría modelarse tablas

intermedias con la estructura necesaria, y luego incorporar esas tablas intermedias al clúster de replicación en lugar de las tablas base.

Otra de las limitaciones se presenta en los nodos esclavos, ya que la información replicada en ellos, no pueden actualizarse. En este caso podría estudiarse otro tipo de herramientas como *pgpool* que, de acuerdo a la documentación relevada, permite la actualización bidireccional de los datos.

Por último la imposibilidad de replicar archivos o imágenes en la misma base de datos, impide garantizar las propiedades ACID en las transacciones. Si bien la documentación de *slony-I* indica que los atributos de tipo *BLOB* no pueden replicarse, a partir de la versión *PostgreSQL 9.0* se incorporaron modificaciones al tipo de dato *Bytea*. Si bien el formato de entrada es diferente al tipo de datos *BLOB*, las funciones y operadores provistos son en su mayoría los mismos. Sobre este tipo de datos se han realizado pruebas preliminares replicando la imagen de la foto carnet del alumno con resultados positivos.

Otra alternativa es para abordar este problema sería mediante la utilización de estrategias basada en control de versiones sobre el sistema de archivos a replicar.

8.4. Futuras líneas de investigación

Entre las futuras líneas de investigación se pueden sugerir las siguientes:

- Realizar las pruebas y mediciones utilizando para el resguardo de los legajos digitales (PDF), el tipo de dato *Bytea*. Esto con la finalidad de evitar la complejidad de administrar un sistema de archivos separado de la base de datos y para intentar garantizar las propiedades ACID en las transacciones.
- Existe un pedido todavía ya formalizado para implementar el sistema LeDig en el SiPTeD que es un ente autárquico denominado Sistema Provincial de Teleducación y Desarrollo que funciona en la esfera del Ministerio de Cultura y Educación de la provincia de Misiones (<https://sipted.misiones.gob.ar>).

- Implementar *Blockchain*²⁷ para los legajos digitales. De este modo se podría garantizar que todos los sitios y seden poseen una copia valida e inviolable del legajo. Sobre todo, sería interesante incorporar esta tecnología para casos de bases de datos sensibles, como por ejemplo el sistema de contratos de la Facultad de Ciencias Económicas. (<https://bfa.ar/>)
- Anexar al sistema de archivos de LeDig estrategias de control de versiones y de *cloud* que recientemente han sido implementadas en la Facultad de Ciencias Económicas. (<https://nube.fce.unam.edu.ar>)
- Dar soporte al Sistema Institucional de Educación a Distancia (SIED), aprobado el 11 marzo 2019 por resolución número 003/2019 del Consejo Superior y de acuerdo a lo establecido por la resolución del Ministerio de Educación de la Nación (Res 2641-E/2017); generando la posibilidad para que la pre inscripción a carreras sea totalmente a distancia. Incluso que el aspirante sea el que suba a un sitio, la documentación ya digitalizada. Luego el personal de la Dirección de Estadísticas realizaría la validación y la correspondencia entre la documentación que el aspirante ha enviado y los requisitos necesarios para dar cumplimiento al proceso de inscripción.

²⁷ *Blockchain* o Cadena de bloques. Tecnología diseñada para administrar un registro de datos online, caracterizada por ser transparente y prácticamente incorruptible.

Bibliografía

- [1] Consejo Superior. (2016, Dec.) Programa de Expansión Territorial de Educación Superior. [Online]. <https://www.unam.edu.ar/index.php/11-academica/29-programa-de-expansion-territorial-de-educacion-superior>
- [2] Honorable Consejo Directivo de la Facultad de Ciencias Económicas. Universidad Nacional de Misiones. (2017, Dec.) Digesto Digital. [Online]. <http://digesto.fce.unam.edu.ar/documentos/resoluciones/2017/R.190-17.pdf>
- [3] Consejo Interuniversitario Nacional. (2018, July) Becas del Programa Estratégico de Formación de Recursos Humanos en Investigación y Desarrollo (PERHID). Orden 141. [Online]. <https://www.cin.edu.ar/comisiones/posgrado-perhid-becas/#>
- [4] Elmasri R. and Navathe S., *Fundamentos de sistemas de bases de datos*, Quinta ed.: Pearson, 2007.
- [5] Silberschatz A., Korth H., and Sudarshan S., *Fundamentos de bases de datos*, 5th ed.: McGraw-Hill Interamericana, 2007.
- [6] Date J., *Introducción a los sistemas de bases de datos*, Séptima ed.: Pearson Educación, 2001.
- [7] Özsu M. and Valduriez P., *Principles of Distributed Database Systems, Third Edition*, Tercera ed., 978-1-4419-8833-1, Ed.: Springer, 2011.
- [8] Plattner C. and Alonso G., "Scalable replication for transactional web applications," *Proceedings of the 5th ACM/IFIP/USENIX international conference on Middleware*, no. ISBN:3-540-23428-4, pp. 155-174, Oct. 2014.
- [9] Gray J., Helland P., O'Neil P., and Shasha D., "The Dangers of Replication and a Solution," *In Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, pp. 173–182, Jan. 2009.

- [10] Amza C., Cox A., and Zwaenepoel W., "Conflict-Aware Scheduling for Dynamic Content Applications," *In Proceedings of the 4th USENIX Symposium on Internet Technologies and Systems (USITS)*, Mar. 2003.
- [11] Cecchet E., "C-JDBC: a Middleware Framework for Database Clustering," *IEEE Data Engineering Bulletin*, vol. 27, no. 2, June 2004.
- [12] Weikum G. and Vossen G., *Transactional Information Systems.*: Morgan Kaufmann Publishers, 2002.
- [13] Jimenez-Peris R., Patino-Martinez M., Alonso G., and Kemme B., "Are Quorums an Alternative for Data Replication?," *ACM Transactions on Database Systems*, 2003.
- [14] Kemme B. and Alonso G., "Don't be lazy, be consistent: Postgres-R, a new way to implement Database Replication," *In Proceedings of the 26th Interna-*, 2002.
- [15] Navathe S., Ceri S., Wiederhold G., and J. Dou, "Vertical partitioning of algorithms for database design," *ACM Digital Library*, vol. 9, pp. 680–710, 1984.
- [16] Eisner M. and Severance D., "Mathematical Techniques for Efficient Record Segmentation in Large Shared Databases," *ACM Digital Library*, vol. 619-635, Oct. 1976.
- [17] Bernardes I. and Delatorre H., "Gestão Documental Aplicada," *Arquivo Público do Estado de São Paulo*, no. ISBN: 978-85-61599-02-7, 2008.
- [18] Sierra Escobar L., "Gestión Documental enfocada a procesos: una mirada desde la administración pública distrital," *Revista Interamericana de Bibliotecología*, vol. 35, no. ISSN: 0120-0976, p. 243, 2012.
- [19] Lodolini E., "Archivística. Principios y problemas," *ANABAD, Ed. Madrid, España*, no. ISBN: 84-604-6312-5, 1993.

- [20] Coudannes M., Pighin G., and Robledo M., "Centros de digitalización y acceso a la información pública: la experiencia de la Provincia de Santa Fe," *44 JAIIO - SIE 2015*, no. ISSN: 2451-7534, 2015.
- [21] Baroni S. and Sferco L., "Digitalización y Sistematización del Proceso de Ejecuciones Fiscales en la Provincia de Río Negro," *45 JAIIO - SIE 2016*, no. ISSN: 2451-7534, p. 334, 2016.
- [22] Díez A., "Evolución de las bases de datos jurídicas en España," *Anales de Documentación*, vol. 7, no. ISSN 1697-7904, 2004.
- [23] Sergey B., James D., and García-Molina H., "Copy detection mechanisms for digital documents," *ACM SIGMOD Record*, vol. 24, no. ISSN: 0163-5808, p. 398, 1995.
- [24] Ministerio de Justicia y Derechos Humanos. Presidencia de la Nación. (2016, enero) InfoLeg - Información Legislativa. [Online].
<http://servicios.infoleg.gob.ar/infolegInternet/anexos/255000-259999/257556/texact.htm>
- [25] Ministerio de Justicia y Derechos Humanos. Presidencia de la Nación. (2016, marzo) InfoLeg - Información Legislativa. [Online].
<http://servicios.infoleg.gob.ar/infolegInternet/anexos/255000-259999/259082/norma.htm>
- [26] Ministerio de Justicia y Derechos Humanos. Presidencia de la Nación. (2016, marzo) InfoLeg - Información Legislativa. [Online].
<http://servicios.infoleg.gob.ar/infolegInternet/anexos/265000-269999/266504/norma.htm>
- [27] Ministerio de Justicia y Derechos Humanos. Presidencia de la Nación. (1992, octubre) InfoLeg - Información Legislativa. [Online].
<http://servicios.infoleg.gob.ar/infolegInternet/anexos/0-4999/554/texact.htm>

- [28] X Asamblea Universitaria. (2012, mayo) Sitio Oficial - Universidad Nacional de Misiones. [Online].
https://www.unam.edu.ar/images/unam/documentos/estatuto_unam.pdf
- [29] Ministerio de Justicia y Derechos Humanos. Presidencia de la Nación. (2016, abril) InfoLeg - Información Legislativa. [Online].
<http://servicios.infoleg.gob.ar/infolegInternet/anexos/260000-264999/260145/norma.htm>
- [30] Ministerio de Justicia y Derechos Humanos. Presidencia de la Nación. (2016, octubre) InfoLeg - Información Legislativa. [Online].
<http://servicios.infoleg.gob.ar/infolegInternet/anexos/265000-269999/266999/norma.htm>
- [31] Ministerio de Justicia y Derechos Humanos. Presidencia de la Nación. (2016, diciembre) InfoLeg - Información Legislativa. [Online].
<http://servicios.infoleg.gob.ar/infolegInternet/anexos/270000-274999/270156/norma.htm>
- [32] Subsecretaría de Gestión Administrativa - Ministerio de Modernización. (2016, diciembre) InfoLeg - Información Legislativa. [Online].
<http://servicios.infoleg.gob.ar/infolegInternet/anexos/270000-274999/270156/res44-3.pdf>
- [33] Ministerio de Justicia y Derechos Humanos. Presidencia de la Nación. (2001, noviembre) InfoLeg - Información Legislativa. [Online].
<http://servicios.infoleg.gob.ar/infolegInternet/anexos/70000-74999/70749/norma.htm>
- [34] Ministerio de Justicia y Derechos Humanos. Presidencia de la Nación. (2018, junio) InfoLeg - Información Legislativa. [Online].
<http://servicios.infoleg.gob.ar/infolegInternet/anexos/310000-314999/311583/norma.htm>

- [35] Autoridad de Registro. (2012) Manual de procedimientos y recomendaciones para Firma Digital. Universidad Nacional de Misiones. [Online]. https://www.unam.edu.ar/images/unam/documentos/manual_de_procedimientos_para_firma_digital.pdf
- [36] Giménez-Chornet V. (2015, enero) ResearchGate. [Online]. https://www.researchgate.net/publication/307606229_Normas_ISO_para_la_gestion_de_los_documentos_electronicos_buenas_practicas_para_la_gestion_documental_en_las_empresas
- [37] International Organization for Standardization. (2016) ISO - Information and documentation. [Online]. <https://www.iso.org/obp/ui/#iso:std:iso:15489:-1:ed-2:v1:en>
- [38] International Organization for Standardization. (2017) ISO - Information and documentation. [Online]. <https://www.iso.org/obp/ui/#iso:std:iso:23081:-1:ed-2:v1:en>
- [39] Pesado P. et al. (2006, Aug.) "Sistemas de software distribuidos y bases de datos distribuidas". VIII Workshop de Investigadores en Ciencias de la Computación. Repositorio Institucional de la Universidad Nacional de la Plata. ISBN: 950-9474-35-5. [Online]. <http://sedici.unlp.edu.ar/handle/10915/20809>
- [40] Cherencio G., Perello M., and Romero J. (2018, Oct.) Replicación de Bases de Datos en Modalidad “maestro-esclavo”, caso de estudio: Firebird SQL Server. XXIV Congreso Argentino de Ciencias de la Computación. Repositorio Institucional de la Universidad Nacional de la Plata. ISBN: 978-950-658-472-6. 462-470. [Online]. <http://sedici.unlp.edu.ar/handle/10915/73225>
- [41] Guerra Y., Altuna E., Sánchez A., and Pérez Cáceres Y. (2015, Dec.) Experiencias en el uso del symmetric para la replicación de datos en la plataforma educativa ZERA. Revista Cubana de Ciencias Informáticas. ISSN 2227-1899. [Online]. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-

[18992015000400014&lng=es&tlng=es](http://scielo.sld.cu/scielo.php?pid=S2227-18992015000400014&lng=es&tlng=es)

- [42] Peña N., Mata M., Labrada Quiala R., and Jerez G. (2016, July) Proceso de réplica de datos con Microsoft SQL Server para el Replicador de Datos Reko. Revista Cubana de Ciencias Informáticas. ISSN 2227-1899. [Online]. http://scielo.sld.cu/scielo.php?pid=S2227-18992016000300013&script=sci_arttext&tlng=en
- [43] Erazo R., Cobos C., and Mendoza M. (2008, June) Web services model to snapshot replication over multiple database engines. Revista Facultad de Ingeniería. Universidad de Antioquia. [Online]. <http://www.scielo.org.co/pdf/rfiua/n44/n44a14.pdf>
- [44] Universidad Nacional de Misiones. (2017, Feb.) Sitio Oficial - UNaM. [Online]. https://www.unam.edu.ar/images/unam/documentos/estatuto_unam.pdf
- [45] Mgter Bohren A. (2018, Oct.) Sitio Oficial - Universidad Nacional de Misiones. [Online]. <https://www.unam.edu.ar/index.php/blog/1217-aprobaron-el-plan-de-desarrollo-institucional-de-la-unam>
- [46] Consejo Interuniversitario Nacional. (2018, Nov.) Sistema de Información Universitaria - SIU. [Online]. <https://www.siu.edu.ar/siu-guarani/>
- [47] Sistema de Información Universitaria, "Manual de Pre Inscripción a Carreras v2.9.2," Consejo Interuniversitario Nacional, Descripción funcional 2018.
- [48] Sistema de Información Universitaria - Consejo Interuniversitario Nacional. (2018, agosto) SIU. [Online]. <https://www.siu.edu.ar/sistema-de-informacion-universitaria/>
- [49] Sistema de Información Universitaria - Consejo Interuniversitario Nacional. (2018, Mar.) Wiki del SIU - Documentación. [Online]. <http://documentacion.siu.edu.ar/wiki/Inicio>

- [50] Sistema de Información Universitaria - Consejo Interuniversitario Nacional. (2017, octubre) SIU Toba - Ambiente de desarrollo WEB. [Online]. <http://toba.siu.edu.ar/trac/toba>
- [51] The PostgreSQL Global Development. (2018, julio) The World's Most Advanced open source relational database. [Online]. <https://www.postgresql.org/>
- [52] Ladas C., *Scrumban-essays on kanban systems for lean software development*, Lean Series ed.: Modus Cooperandis , 2009.
- [53] redmine.org. (2018) Overview Redmine. [Online]. <https://www.redmine.org/>
- [54] Wanner M. (2017, julio) Proyecto Postgres-R. [Online]. <https://wiki.postgresql.org/wiki/Postgres-R>
- [55] Slony Development Group. (2017, agosto) Slony-I. Enterprise Level Replication System. [Online]. <http://www.slony.info/>
- [56] Oliveira R. (2017, Sep.) ESCADA Replication Server. Step Up to Database Dependability and Performance. [Online]. <http://escada.sourceforge.net/>
- [57] Pgpool Global Development Group. (2017, Oct.) Middleware between PostgreSQL servers and a PostgreSQL database client. [Online]. http://www.pgpool.net/mediawiki/index.php/Main_Page
- [58] Golovkov A. Dnipropetrovs`k National University. (2017, junio) System for asynchronous point-in-time databases replication. [Online]. <http://dbre.sourceforge.net/>
- [59] Salman A., Sailaja P., Venkataswamy G., and Supriya N., "Database Replication: A Survey of Open Source and Commercial Tools," *International Journal of Computer Applications (ISSN 0975-8887)*, vol. 13, no. 6 - Article 1, Januray 2011.

- [60] Pgpool Global Development Group. (2018, febrero) Pgpool docs. [Online]. <http://www.pgpool.net/docs/>
- [61] Postgres Global Development Group. (2018, marzo) Postgres-R. [Online]. <http://postgres-r.org/>
- [62] Ishii T., Peng B., and Hoshiai T. (2017, diciembre) Clustering tool for PostgreSQL. [Online]. <https://git.postgresql.org/gitweb/?p=pgpool2.git;a=summary>
- [63] PostgreSQL Global Development Group. (2019, Feb.) PostgreSQL 9.6.13 Documentation. [Online]. <https://www.postgresql.org/docs/9.6/datatype-binary.html>
- [64] PostgreSQL Web Team. (2019, Feb.) PostgreSQL wiki. [Online]. <https://wiki.postgresql.org/wiki/BinaryFilesInDB>
- [65] Slony Development Group. (2019, Apr.) Slony-I Enterprise Level Replication System. [Online]. <http://www.slony.info/documentation/2.2/index.html>
- [66] Sistema de Información Universitaria - Consejo Interuniversitario Nacional. (2019, Apr.) SIU Araújo. [Online]. <http://documentacion.siu.edu.ar/wiki/SIU-Araujo>
- [67] Sinadura. (2019, May) Sinadura - La firma digital libre. [Online]. <http://www.sinadura.net/es>
- [68] Honorable Consejo Directivo. Facultad de Ciencias Exactas Químicas y Naturales. Universidad Nacional de Misiones. (2018, Nov.) Digesto Digital. [Online]. http://digesto.fceqyn.unam.edu.ar/digestopub/documentos/resoluciones/2018/RES_05722018.PDF

Anexo I

Creación del proyecto LeDig

Anexo I. Creación del proyecto LeDig

Para el desarrollo del sistema LeDig se utilizó el *framework* SIU Toba, provisto por el consorcio SIU a las Universidades Nacionales. Los detalles sobre la creación del proyecto LeDig y el desarrollo de la aplicación se explican en profundidad en este anexo.

Creación del proyecto LeDig

SIU Toba administra los diferentes desarrollos de software en proyecto. Para crear o administrar un proyecto SIU Toba ofrece una serie de comandos que se encuentran en la carpeta “bin” a los cuales podemos acceder ejecutando:

```
./toba
```

Esto retorna la lista de comandos disponibles, como se muestra a continuación:

```
-----
SIU-TOBA ( Ambiente de desarrollo WEB )
-----
Versión: 2.7.13
Directorio de la INSTALACION: /home/gpautsch/toba_2.7.13

Comandos disponibles
-----
instalacion      Administracion de la INSTALACION
base             Administracion de BASES de DATOS
instancia        Administracion de INSTANCIAS
proyecto         Administracion de PROYECTOS
personalizacion  Administracion de PERSONALIZACIONES
test             Ejecucion de baterias de TEST
servicios_web    Administracion de servicios web

-----
TIEMPO: 0,053 segundos
```

La ejecución es muy intuitiva ya que si luego se ejecuta alguna de las opciones ofrecidas, el sistema irá listando los comandos disponibles para cada opción.

Por ejemplo, si ahora se ejecuta: `./toba proyecto` la consola retornará:

```
-----
Administracion de PROYECTOS
-----
INVOCACION: toba proyecto OPCION [-p id_proyecto] [-i id_instancia]
[-p id_proyecto] Asume el valor de la variable de entorno 'TOBA_PROYECTO'
[-i id_instancia] Asume el valor de la variable de entorno 'TOBA_INSTANCIA'

Lista de opciones
-----
crear                Crea un proyecto NUEVO.
                    Opcional: [-d] Directorio donde se creará el
proyecto
-----
TIEMPO: 0,002 segundos
```

Entonces para crear el nuevo proyecto LeDig, se debe ejecutar la siguiente instrucción desde la línea de comandos del sistema operativo:

```
$ /bin/toba proyecto crear -i desarrollo -p ledig
```

Luego de unos segundos toba retornará el mensaje:

```
Creando el proyecto 'ledig' en la instancia 'desarrollo'...OK
```

Paso seguido toba preguntará si se desea agregar el alias del proyecto en el servidor de apache. Luego de seleccionar la opción “s” y hacer una recarga de los servicios de apache, el proyecto ya podrá ser accedido desde el navegador desde la url: http://{ip_server_desarrollo}/toba_editor/2.7 y seleccionando el proyecto creado (Figura 1).

Replicación de datos para dar soporte al Programa de Expansión Territorial de la Universidad Nacional de Misiones



Figura 1. Página principal del *framework* SIU Toba

Definición de las fuentes de datos

El paso siguiente a la creación del proyecto es definir las fuentes de datos según el esquema definido en el Capítulo 4, en la sección 4.6 (Figura 15). Esto se logra desde la solapa “Datos” del *framework* SIU Toba.

Por defecto toba ya ofrece una fuente de datos, a la cual hay que configurar con los datos que comúnmente están asociados a una conexión de *PostgreSQL* (Figura 2).

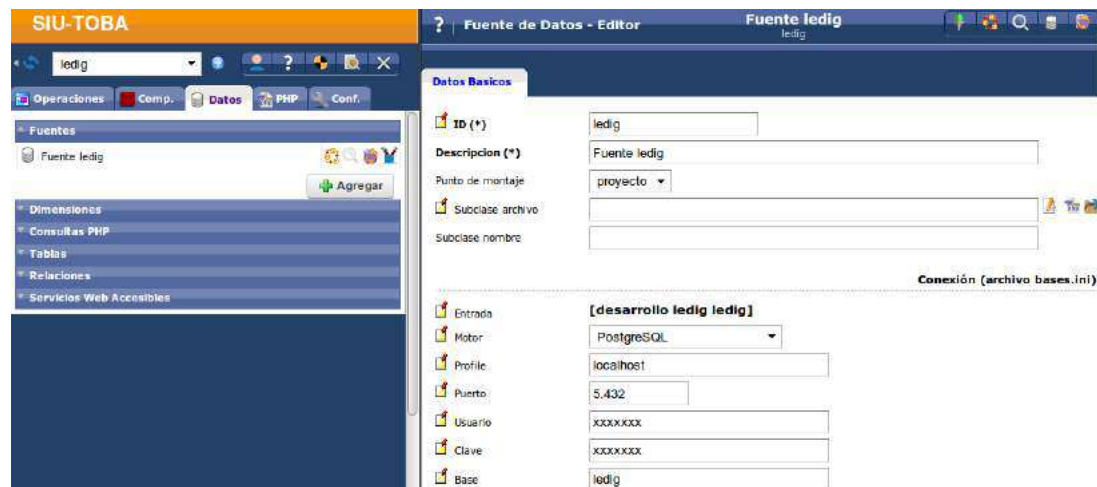


Figura 2. Configuración de las fuentes de datos

A modo de verificación, desde el icono de la lupa, se pueden acceder a la base de datos y visualizar las tablas del esquema (Figura 3).



Figura 3. Verificación de la conexión

Luego, análogamente, desde el botón “Agregar” creamos la conexión a la base de datos del sistema SIU Preinscripción quedando las fuentes de datos de la siguiente manera (Figura 4).



Figura 4. Fuentes de datos del sistema LeDig

Nótese la ausencia de la fuente de datos correspondiente al sistema SIU Guaraní. Esto no es necesario hacerlo desde el desarrollo del sistema LeDig ya que no se necesita acceder en forma permanente a la base de datos del sistema SIU Guaraní si los metadatos ya se encuentran en la base de datos de LeDig. Este

particular se aborda en detalle más adelante en este anexo, en la sección “Interfaz con el sistema SIU Guaraní”.

Diseño de las operaciones del sistema

Para el diseño de las operaciones del sistema se utilizó el asistente que provee el SIU Toba. Esta opción reduce drásticamente el tiempo necesario para desarrollar una aplicación.

Si bien las funcionalidades resultantes son muy simples, permitiendo el alta, baja y modificación de registros, posteriormente puede agregarse la complejidad necesaria a estas operaciones, introduciendo el código fuente que sea necesario.

Una vez creado el proyecto y definidas las fuentes de datos, se puede comenzar con el procedimiento de diseño de las operaciones que va a tener el sistema, utilizando como insumo para la confección de la interfaz con el usuario, las tablas de las fuentes de datos definidas.

Para acceder a este asistente basta con presionar el botón “Asistente” desde la pantalla principal del SIU Toba. Una vez allí, el mismo permite en una serie de pasos, diseñar las operaciones necesarias, de manera muy simple, como se describe a continuación.

Como primer paso se debe dar un nombre a la operación que se está por generar y en la misma pantalla seleccionar entre las tres posibilidades de la Tabla 1:

Tipo Operación	Descripción
ABM Simple	Genera un ABM a través de un formulario
Grilla Simple	Genera un ABM a través de una grilla
Importar Operación	Permite importar una operación previamente generada

Tabla 1. Asistente SIU Toba - Tipos de operación.

Para este caso se ha seleccionado el tipo de operación ABM Simple ya que posibilita configurar y personalizar de manera más detallada cada operación del sistema que se está generando.

En el siguiente paso, SIU Toba nos solicita: nombre del archivo donde se generará el código PHP pertinente, la fuente de datos y la tabla de la base de datos en base a la cual se generará la operación. Completada la información solicitada en este formulario SIU Toba muestra una pantalla como la que se exhibe en la Figura 5.

Información básica
Cuadro
Formulario
Filtro

Punto de Montaje (*)
proyecto

Carpeta archivos (*)
legajos

Anexo nombre clases (*)
_legajos

Fuente de Datos (*)
Fuente ledig

Tabla (*)
ledig

Puede filtrar los datos
☒

Definición de campos

Campo	Etiqueta	Clave	Obligatorio	Tipo	Cuadro	Form.	Filtro
dni	Dni	☒	☒	Cadena	☒	☒	☒
legajo	Legajo	☐	☒	Cadena	☒	☒	☒
unidad_academica	Unidad Academica	☐	☐	Cadena	☒	☒	☐
sede	Sede	☐	☐	Cadena	☒	☒	☐
carrera	Carrera	☐	☒	Cadena	☒	☒	☐
estado	Estado	☐	☐	Cadena	☒	☒	☐
apellido	Apellido	☐	☒	Cadena	☒	☒	☒
nombres	Nombres	☐	☒	Cadena	☒	☒	☐

Recargar

Figura 5. Asistente SIU Toba - Información básica

En esta pestaña de “Información básica” aparecerán automáticamente todos los atributos que posee la tabla. Aquí se realizan las configuraciones referentes a cada uno de ellos, permitiendo establecer su visibilidad, obligatoriedad y si el mismo forma parte de la posibilidad de filtrado y búsqueda de la información.

En la pestaña de “Cuadro” ofrecerá por defecto, en base a las configuraciones realizadas en la pestaña anterior, la consulta SQL con la cual se genera el set de datos para completar la grilla de la pantalla inicial de la operación (Figura 6).

Información básica **Cuadro** Formulario Filtro

Puede eliminar filas ☒

Mensaje vacío

Origen de datos (*)

☒ Componente de persistencia (datos_tabla)

☐ Clase PHP de consultas

SQL (*)

```
SELECT
  t_l.dni,
  t_l.legajo,
  t_l.unidad_academica,
  t_l.sede,
  t_l.carrera,
  t_l.estado,
  t_l.apellido,
  t_l.nombres
FROM
  ledig as t_l
ORDER BY legajo
```

Regenerar SQL

Generar

Figura 6. Asistente SIU Toba - Cuadro

Luego, en la solapa “Formulario” se puede determinar si la selección de datos (grilla cuadro) y el formulario de edición de los mismos, estarán en la misma pantalla o no (Figura 7).

Información básica Cuadro **Formulario** Filtro

Separar selección y edición

Destina distintas pantallas para la selección del elemento a editar (cuadro) y la edición en sí misma (formulario).

Generar

Figura 7. Asistente SIU Toba - Formulario

En la última pestaña “Filtro”, se determinan los criterios de búsqueda que se emplearán (Figura 8) sobre los atributos que se hayan seleccionado para tal fin.

Información básica Cuadro Formulario **Filtro**

Ocultar cuadro hasta que se presione BUSCAR

Requiere ingresar al menos un criterio

Opciones de búsqueda

Columna	Operador de comparación
dni	=
legajo	=
apellido	ILIKE

Generar

Figura 8. Asistente SIU Toba - Filtro

Terminado este paso se han realizado todas las configuraciones necesarias para la generación de la interfaz y funcionalidad de la operación. Esta, junto con otras diseñadas de la misma manera, integrará el sistema legajos digitales (LeDig), que será replicada a las distintas dependencias, otorgando así la alta disponibilidad de la información y continuidad de las operaciones.

Luego de generada la operación, la misma queda disponible en la pantalla inicial de SIU Toba, permitiendo acceder a las propiedades de cada uno de los componentes y permitiendo su personalización (Figura 9).



Figura 9. Asistente SIU Toba - Resultado operación “Legajos”

Interfaz con el sistema SIU Guaraní

Para facilitar el acceso a la información y evitar que el personal no docente tenga que realizar la carga manual de los metadatos de los legajos ya existentes de cohortes anteriores a la implementación de LeDig, se realizó una importación inicial de los mismos desde el sistema SIU Guaraní utilizando la herramienta *Pentaho Data Integration* (PDI) en su versión 4.2, que permite integrar datos entre distintos motores de base de datos.

Primeramente se define una “transformación” para que luego sea integrada a un “flujo de trabajo”. En la Figura 10 se muestra la transformación que exporta los datos necesarios de la base de datos *Informix* (SIU Guaraní) y luego los importa como metadatos a la base de datos *PostgreSQL* del sistema LeDig.

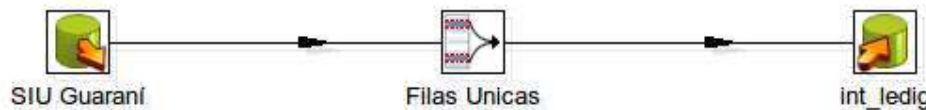


Figura 10. Transformación en PDI para la migración entre Informix y PostgreSQL

En la primera instancia (SIU Guaraní) se realiza la importación de los datos de los legajos ya grabados. La configuración de este objeto puede verse en la Figura 11.

Tabla - Entrada

Nombre paso: SIU Guaraní

Conexión: Informix [▼] [Editar...] [Nuevo...]

SQL: [Obtener consulta SQL...]

```
SELECT nro_documento, legajo, CA.unidad_academica, CA.sede, CA.carrera, apellido, nombres, periodo_inscripcion
FROM sga_carrera_aspira CA
INNER JOIN sga_personas P ON CA.nro_inscripcion = P.nro_inscripcion
INNER JOIN sga_alumnos A ON A.nro_inscripcion = P.nro_inscripcion
WHERE periodo_inscripcion = 2010
```

Line 5 Column 37

Activar conversión perezosa. ☐

¿Reemplazar variables en script? ☐

Insertar datos del paso: [▼]

¿Ejecutar para cada fila? ☐

Limitar tamaño: 0

[OK] [Previsualizar] [Cancelar]

Figura 11. Conexión a *informix* – SIU Guaraní

En la cláusula *where* del *script sql*, el atributo *periodo_inscripcio* recibe un parámetro *\${cohorta}*, el cual proviene de la ejecución realizada por línea de comando del sistema operativo o en su defecto del planificador de tareas.

En el caso que un alumno esté inscripto a varias, esta consulta a la base de datos de *informix* retornará varios registros por alumno, por lo que para generar un solo legajo por registro, hay que limitar estos últimos a un registro por alumno. Esto se hace en la herramienta PDI con la instancia “Filas Unicas” como muestra la Figura 12.

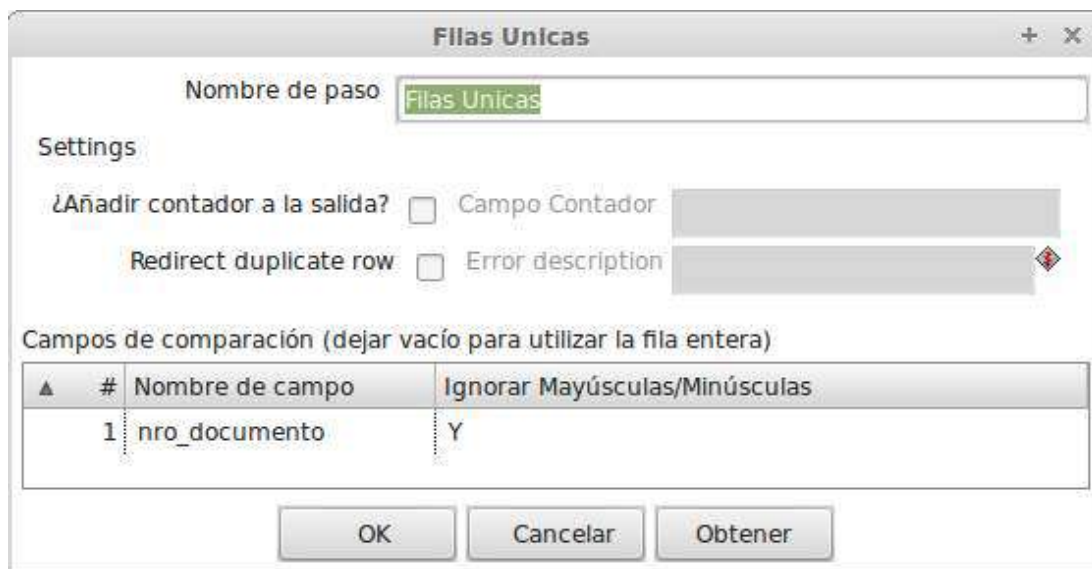


Figura 12. PDI - Operador "Filas Unicas"

Luego de la ejecución de esta instancia, el resultado que se obtiene es un listado con todos los alumnos de una cohorte específica, el cual contienen un registro por alumno ya que solo se deja un registro por “nro_documento” cuyo atributo contiene el valor del número de documento del alumno.

Lo siguiente es importar ese listado a la tabla de interfaz (*int_ledig*) del motor *postgreSQL* para que los mismos queden disponibles y puedan ser procesados como metadatos por el sistema LeDig. En la Figura 13 se muestra la configuración del objeto “*int_ledig*” donde se realiza la importación de los datos mencionados y la

correspondencia entre los atributos obtenidos entre la consulta *sql* realizada en el motor *informix* y los atributos de la tabla *int_ledig* del motor *portgreSQL*.

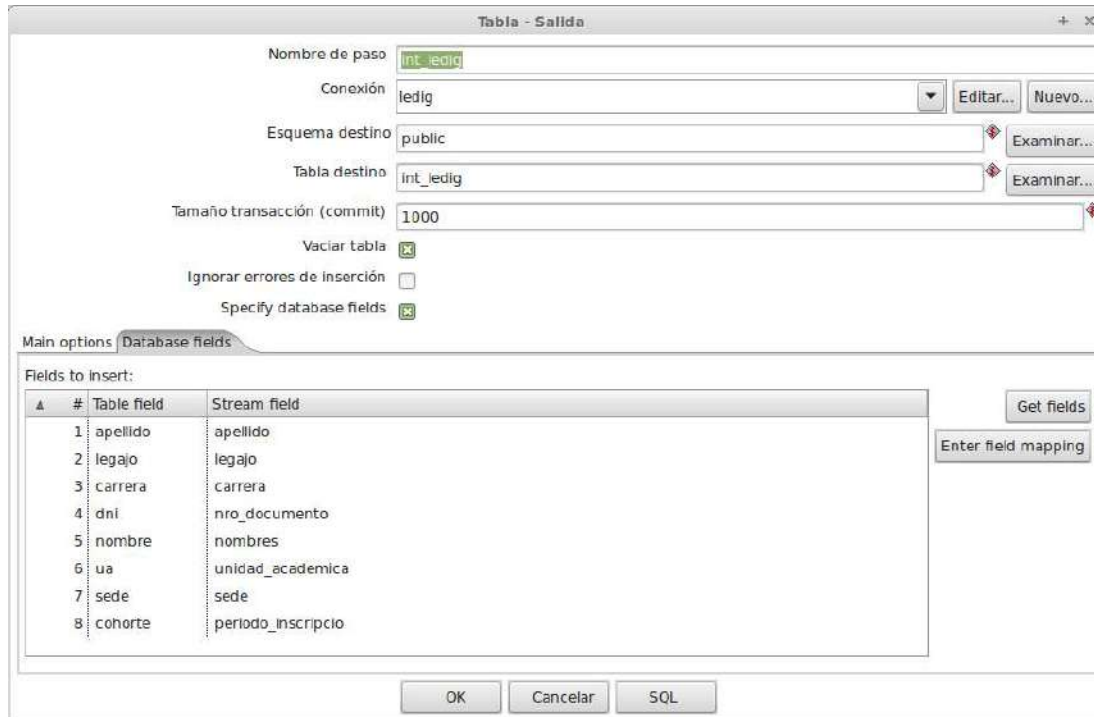


Figura 13. Correspondencia entre atributos de *informix* y *postgreSQL*

En esta pantalla también pueden configurarse algunas opciones interesantes como por ejemplo:

- Tamaño transacción (*commit*): permite configurar cada cuántas tuplas se realizará la ejecución de una sentencia *commit*.
- Vaciar TABLA: aquí se puede especificar si previa a la importación la tabla de destino será vaciada (sentencia *truncate*) o no.

Finalmente, para verificar el correcto funcionamiento de las conexiones entre los motores y de la migración en general es recomendable utilizar la instancia *DUMMY*, la cual permite realizar una simulación de todo lo desarrollado sin necesidad de manipular las bases de datos. También existe la posibilidad de ejecutar

la transformación desde el icono *play* de la barra de herramientas, desde la opción “Ejecutar” del menú *Action* o simplemente presionando la tecla “F9”.

Luego de esto ya se puede integrar la transformación a un flujo de trabajo (Figura 14).

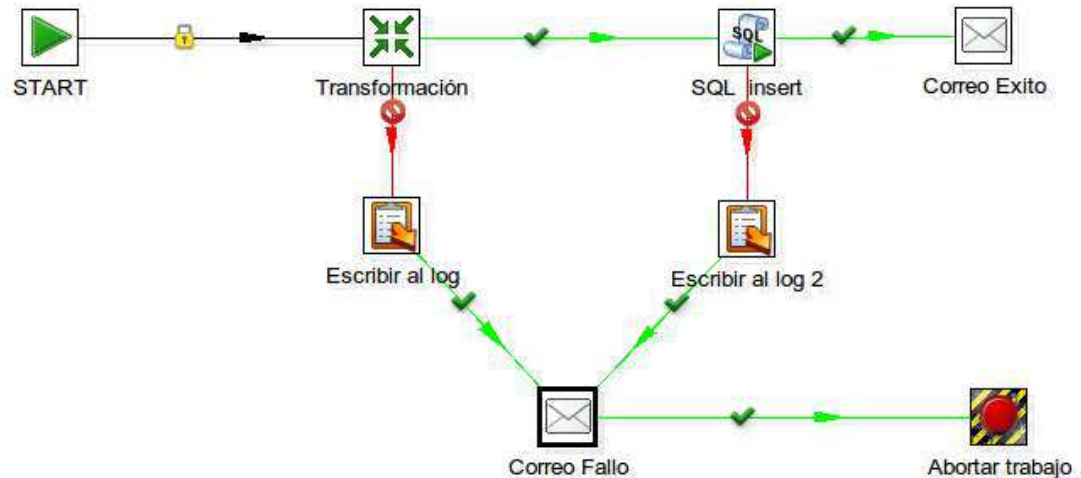


Figura 14. PDI - Flujo de trabajo interfaz SIU Guaraní y LeDig

El flujo de trabajo al iniciarse (instancia *STAR*) referencia y ejecuta la transformación que exporta los datos del SIU Guaraní y los importa al sistema LeDig (Figura 10). Paso seguido, en la instancia “*SQL insert*”, se ejecuta un *script* (Figura 15) que toma los datos depositados en la tabla *int_ledig* y los inserta en la tabla *ledig* para que luego sean utilizados como metadatos para referenciar la documentación digitalizada.

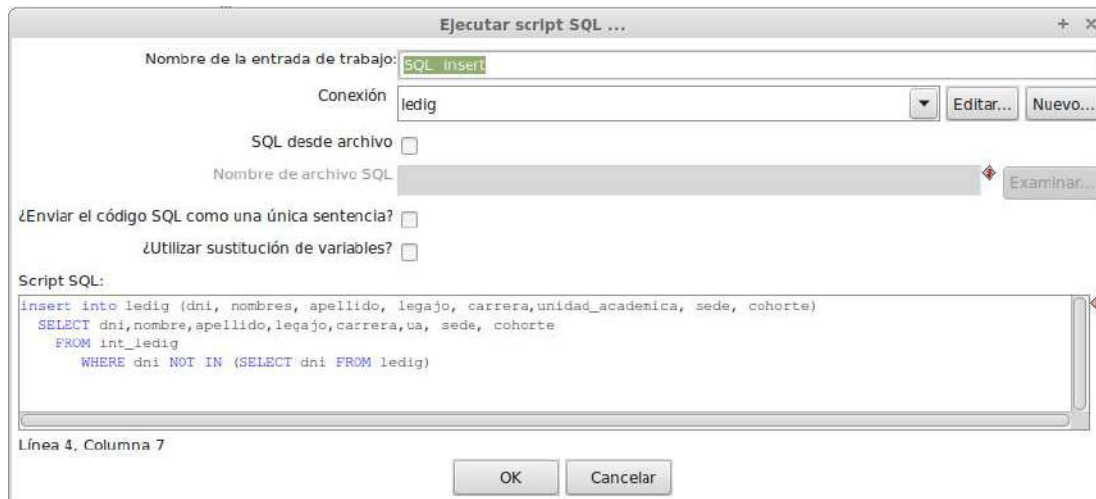


Figura 15. PDI - Operador "Script SQL"

Es importante añadir al *script* la cláusula *where* de la última línea, ya que esto evitará que los alumnos que registren inscripciones en distintos años académicos generen un conflicto de violación de clave primaria o duplicación de tupla en la tabla *ledig* a la hora de ser importados.

En el caso de que alguna instancia de ejecución falle, se genera un *log*, se envía un correo electrónico, preestablecido, al administrador de la base de datos y se aborta la ejecución.

Es recomendable ajustar el nivel de detalle de los *logs* de ejecución, ya que por defecto PDI no proveen mucho detalle. Esto se realiza en la pantalla previa a la ejecución de la transformación o flujo de trabajo como se exhibe en la Figura 16.

Ejecutar una transformación

Ejecución local, remota o clustered

☒ Ejecución local ☐ Ejecución remota

Servidor remoto: [dropdown]

☐ Pass export to remote server

Detalles

☐ Habilitar modo seguro

☒ Clear the log before execution

Nivel de registro: [dropdown: A nivel de fila (muy detallado)]

Fecha de Ejecución (yyyy/MM/dd HH:mm:ss): [text field]

Starting point of job: [dropdown]

Parameters

#	Parameter	Value	Default value
1	ciaventormix		
2	clavePostgres		
3	cohorte		
4	userInformix		dba
5	userPostgres		postgres

Variables

#	Variable	Valor
1		

Parámetros

Ejecutar Cancelar

Figura 16. PDI - Parametrización para ejecución del flujo de trabajo

En esta pantalla, en la sección “Parameters”, también se pueden asignar los valores necesarios a cada parámetro para realizar la simulación o ejecución del flujo de trabajo

Trabajar con esta estructura tiene como ventaja, entre otras opciones:

- Control de flujos en caso de fallos o éxitos
- Comunicación vía correo electrónico con el administrador de la base de datos
- Creación de *logs* en tiempo de ejecución
- Interpretación del lenguaje SQL en tiempo de ejecución
- La parametrización de todos los valores de la transformación

Esta última opción, posteriormente permite generar un comando de ejecución personalizado con las opciones necesarias, que puede ser ejecutado desde la línea de comandos del sistema operativo o programar su ejecución en el planificador de tareas.

Para definir estos parámetros se debe ejecutar la opción “configurar” del menú “editar” y luego acceder a la solapa “parameters”. Para el flujo de trabajo de la Figura 14, la definición de parámetros es la que se exhibe en la Figura 17.

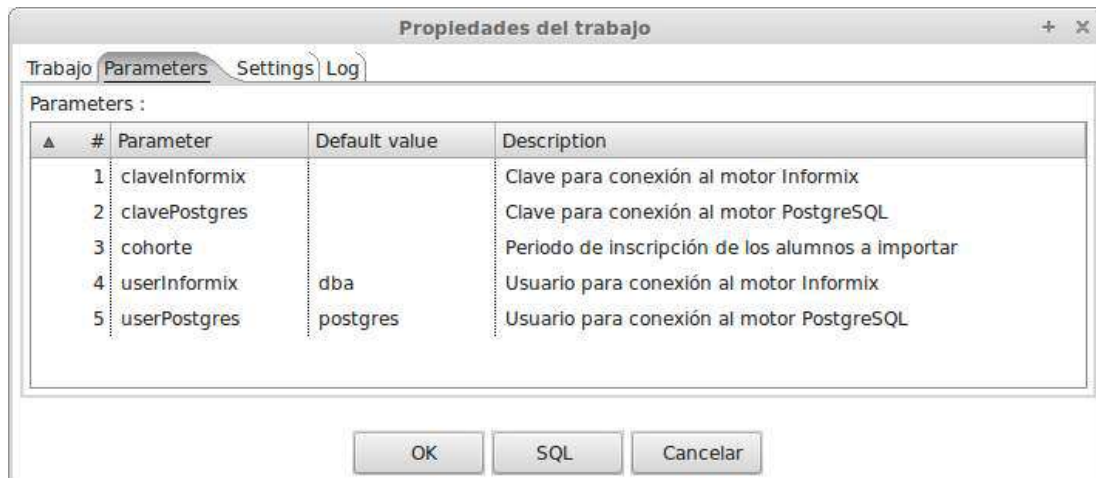


Figura 17. Parámetros definidos para el flujo de trabajo LeDig

Suponiendo que el flujo de trabajo desarrollado se guarde bajo el nombre de “jobLeDig.kjb” el comando de PDI que permite la ejecución desde la consola del sistema operativo se llama *kitchen* y quedaría conformado de la siguiente manera:

```
$ ./kitchen.sh -file = "jobLeDig.kjb" -param = "claveInformix = xxxxx" -  
param = "clavePostgres = xxxxx" -param = "cohorte = 2010"
```

La ejecución del anterior comando exportará del sistema SIU Guaraní todas las tuplas pertenecientes a los alumnos de todas las carreras cuyas inscripciones se hayan realizado en el periodo de inscripción dos mil diez (2010) y luego los importará al sistema LeDig para que puedan ser utilizadas como metadatos en la gestión de los legajos digitales.

Para este ejemplo se utilizaron solo estos parámetros bajo el supuesto de que los demás parámetros, *userInformix* y *userPostgres*, están definidos por defecto en la transformación, pero debe tenerse en cuenta que también pueden parametrizarse los datos de conexión a cada motor de base de datos aunque esto genera luego una línea de ejecución extensa que suele generar errores de sintaxis y/o tipeo al momento de su ejecución.

Por último es recomendable que la ejecución del comando *kitchen* se realice mediante el planificador de tareas del sistema operativo, en un horario que no afecte la performance de los OLTP de los sistemas SIU Guaraní y LeDig.