

**UNIVERSIDAD DE TARAPACÁ
FACULTAD DE INGENIERÍA**

**INGENIERÍA CIVIL EN COMPUTACIÓN E INFORMÁTICA ARICA
– CHILE**



Documento de requisitos de
Sistema de Análisis Integral
de RRHH para Casino Luckia

Equipo de Desarrollo:

- **Tomás Ignacio Silva Muñoz**
- **Juan Carlos Yampara Rojas**

Empresa o Unidad: Casino Luckia

Curso: Proyecto IV ICCI

Profesor: Diego Aracena Pizarro

Arica, 30 de Diciembre 2025

1. Índice

1. Índice	2
2. Resumen	5
3. Definición del Problema y Justificación	6
3.1. Contexto	6
3.2. Problema	6
3.3. Solución escogida	7
3.4. Otras soluciones	7
4. Alcance, Objetivos y Planificación	8
4.1. Alcance del proyecto	8
4.2. Objetivo General	9
4.3. Objetivos Específicos	9
4.4. Carta Gantt	10
5. Requerimientos del proyecto	12
5.1. De Alto nivel	12
Visión del sistema	12
Requisitos de alto nivel	12
5.2. Funcionales	13
5.3. No funcionales	14
5.4. De documentación	15
6. Metodología	16
6.1. Macro-Proceso Analítico: CRISP-DM	16
6.2. Gestión Ágil del Desarrollo: Kanban	17
6.3. Integración de la Metodología Híbrida	17
6.4. Definición de tareas	18
Sistema de Etiquetas (Tags)	18
Épica 1: Definición y Planificación (Entrega Parte 1)	18
Épica 2: Diseño de Arquitectura y Datos	19
Épica 3: Configuración de Infraestructura y BBDD	20
Épica 4: Desarrollo Módulo de Ingesta y ETL	20
Épica 5: Desarrollo Módulo de Seguridad y Operaciones	21
Épica 6: Análisis Exploratorio y Modelado Predictivo	21
Épica 7: Desarrollo Módulo Analítico y Visualización	22
Épica 8: Documentación Final y Cierre	22
7. Modelado del Procesos de Negocio	23
8. Modelado del sistema	25
8.1. Modelo de contexto	25
8.2. Modelo de casos de uso	26
1. Actor: Analista de Negocio	27
2. Actor: Operador de Datos	27

3. Actor: Administrador BI	27
4. Actores: Sistemas Externos	28
8.3. Modelo de Componentes Lógicos	29
Modelo de Alto Nivel	29
Componentes Principales:	30
Componentes Secundarios:	30
Modelado de nivel Medio	31
Arquitectura del Sistema de Componentes	32
Diagrama de despliegue	33
8.4. Diagramas de interacción	34
9. Arquitectura y Pila Tecnológica	37
9.1. Aspectos Arquitecturales del Sistema	37
Suposiciones y Limitaciones	37
Decisiones de Diseño	37
9.2. Herramientas y Frameworks de Software	38
10. Modelado de Datos	40
10.1. Fuentes de Datos	40
Diccionario rotación	40
GeoVictoria día y semana	40
Resumen anual de capacitación y Participantes de las capacitaciones	41
Proceso de selección (Contratación)	41
Perfiles de trabajo	41
10.2. Base de datos Staging	42
10.2.1. Data Warehouse	45
Objetivo del Data Warehouse	45
Decisiones de diseño	45
BI vs ML (Data Lake VS Data Marts)	45
Conformed Dimensions	45
Formulación de las Tablas de Hechos	46
1. Rotación de Empleados (Transaccional)	46
2. (De apoyo) Headcount mensual (Snapshot Periodico)	46
3. Asistencia Diaria (Transaccional)	47
4. Participación en Capacitación (Factless / Transactional)	47
6. Proceso de Selección (Snapshot Acumulativo)	48
Dimensiones y Tablas de Hechos	49
Dimensiones Conformadas	49
Dimensión Empleado	49
Dimensión Tiempo	50
Dimensión Cargo	50
Dimensión Empresa	50
Dimensión Gerencia	51

Dimensión Centro Costo	51
Dimensión Modalidad Contrato	51
Área de Selección	52
Hechos de Selección	52
Área de Asistencia	53
Dimensión Turno	53
Dimensión Permiso	53
Hechos de Asistencia	54
Área de Rotación y Dotación	55
Dimensión Medida Aplicada	55
Hechos de Rotación	55
Hechos de Dotación	56
Área de Capacitación	57
Dimensión Curso	57
Dimensión Proveedor	57
Dimensión Lugar	57
Hechos de Realización de Capacitaciones	58
Hechos de Participación en Capacitaciones	59
Modelo Multidimensional Graficado	60
11. Diseño de Interfaz de Usuario y Prototipo	61
11.1. Mockups	61
12. Conclusiones	63
13. Bibliografía	64
14. Anexo	64

2. Resumen

La gestión de Recursos Humanos en Casino Luckia Arica se enfrenta actualmente a desafíos significativos debido a la fragmentación de sus datos, los cuales residen en sistemas dispares y archivos manuales como SAP, GeoVictoria y planillas Excel. Esta desconexión impide un análisis estratégico y fomenta una toma de decisiones reactiva ante problemas críticos como la rotación de personal, el ausentismo y la optimización de la dotación.

Este proyecto propone el diseño y la futura implementación de una Plataforma Analítica Integral para abordar dicho problema. El objetivo principal es planificar la centralización de esta información a través de procesos automatizados de Extracción, Transformación y Carga (ETL) en un Data Warehouse (DW) unificado.

Para guiar la ejecución, se utilizará una metodología híbrida: CRISP-DM estructurará el proceso de minería de datos, desde la comprensión del negocio hasta el eventual despliegue de modelos predictivos; mientras que Kanban se empleará para la gestión ágil del desarrollo de software (ETLs, BBDD, Dashboards).

El resultado esperado de este plan es una solución funcional que automatizará el cálculo de KPIs (Tasa de Rotación, Ausentismo) y los presentará en dashboards interactivos, con el fin de dotar al área de RRHH de una herramienta que facilite la migración de una gestión operativa reactiva a una gestión estratégica y basada en datos.

3. Definición del Problema y Justificación

3.1. Contexto

Actualmente, el equipo de Recursos Humanos de Casino Luckia Arica gestiona información esencial del personal (como ingresos, asistencia, turnos y licencias) a través de múltiples archivos y sistemas separados. Esta información se consolida de forma manual y fragmentada, lo que limita la capacidad del área para anticipar y tomar decisiones estratégicas sobre la rotación de colaboradores, sobre el ausentismo por área o turno, y sobre la dotación óptima requerida para la operación del casino.

3.2. Problema

La gestión fragmentada y manual de los datos de personal impide tener una visión unificada y rápida de la plantilla, lo cual dificulta el análisis proactivo de temas críticos como la rotación y el ausentismo. Esta falta de integración obliga a tomar decisiones reactivas, lo que se traduce en un mayor riesgo de sobrecostos operativos (por ejemplo, horas extra no planificadas) y una potencial afectación a la calidad del servicio que se ofrece a los clientes.

3.3. Solución escogida

Se propone implementar una Plataforma Analítica Integral que unifique automáticamente todas las fuentes de datos de Recursos Humanos en un almacén de datos centralizado y permitir desde ahí el procesamiento analítico de los datos. Esta plataforma permitirá calcular métricas clave de rotación, ausentismo y dotación; además de permitir desarrollar modelos predictivos para anticipar el comportamiento del personal, y visualizar toda la información de manera sencilla a través de paneles de control interactivos. Esto transformará la gestión de personas en un proceso estratégico y proactivo. Este proyecto también contempla la identificación y recolección de nuevos datos de relevancia analítica, garantizando un ciclo de mejora continua donde los modelos se realimentarán con más información para aumentar progresivamente su precisión e impacto en las decisiones estratégicas de la empresa.

3.4. Otras soluciones

Antes de decidir la arquitectura final, se consideró una alternativa más simple: usar una herramienta como Power BI para conectarla directamente a las fuentes de datos (reportes de SAP, planillas Excel, etc.), sin construir un Data Warehouse (DW) intermedio.

Esta opción se analizó por ser, en teoría, mucho más rápida de implementar. Sin embargo, se descartó porque traía problemas importantes a futuro:

1. **Mantenimiento Complicado:** Toda la lógica para limpiar y unificar los datos tendría que vivir dentro de Power BI. Si una fuente de datos cambia, arreglar los reportes sería muy difícil.
2. **Dificultad para Modelos Predictivos:** Entrenar un modelo (como el de predicción de rotación) requiere datos limpios, históricos y consolidados. Hacer esto "al vuelo" desde las fuentes originales es ineficiente y muy complejo.
3. **Poca Flexibilidad:** Si en el futuro se quisieran añadir nuevas métricas o cruzar datos de formas no previstas, sería necesario rehacer gran parte del trabajo.

Se concluyó que, aunque la construcción de un Data Warehouse toma más tiempo al inicio, es la base necesaria para que el sistema sea mantenible, escalable y, sobre todo, para que los modelos predictivos puedan funcionar.

4. Alcance, Objetivos y Planificación

A continuación, se define el alcance y los objetivos del proyecto, integrando los requerimientos funcionales y técnicos.

4.1. Alcance del proyecto

El proyecto contempla el diseño, desarrollo e implementación de una plataforma analítica de datos para el área de RRHH de Casino Luckia Arica.

El alcance del sistema incluye:

1. **Ingesta de Datos:** Desarrollar un módulo que permita la carga de archivos (ej. planillas Excel) exportados desde los sistemas actuales (SAP y Geovictoria).
2. **Procesamiento y Almacenamiento:** Implementar procesos ETL (Extracción, Transformación y Carga) para limpiar, validar, transformar y unificar los datos en un Data Warehouse (DW) centralizado que almacenará el historial de la información.
3. **Analítica Descriptiva:** Calcular y exponer métricas clave de forma automatizada, incluyendo Tasa de Rotación, Tasa de Ausentismo y Dotación de personal por área/turno.
4. **Analítica Predictiva:** Entrenar un modelo predictivo inicial, basado *exclusivamente* en los datos históricos disponibles durante el proyecto, para estimar la probabilidad de rotación (baja) de colaboradores por área.
5. **Visualización y Reportes:** Disponer de una interfaz gráfica (dashboard) que permita a los usuarios finales consultar las métricas y los resultados del modelo de forma interactiva, así como generar reportes automatizados.
6. **Seguridad y Auditabilidad:** Aplicar controles de ciberseguridad para asegurar la confidencialidad, integridad y disponibilidad de los datos (ej. control de acceso, logs), además de permitir su revisión y gestión por los administradores de la herramienta.

Fuera del Alcance (Exclusiones):

1. El proyecto **no modificará los sistemas fuente** (SAP o Geovictoria).
2. El proyecto **no desarrollará nuevos módulos de captura de datos** (ej. formularios de ingreso); solo consumirá la información existente.
3. La implementación se realizará utilizando la infraestructura y tecnologías dispuestas o aprobadas por la organización.

4.2. Objetivo General

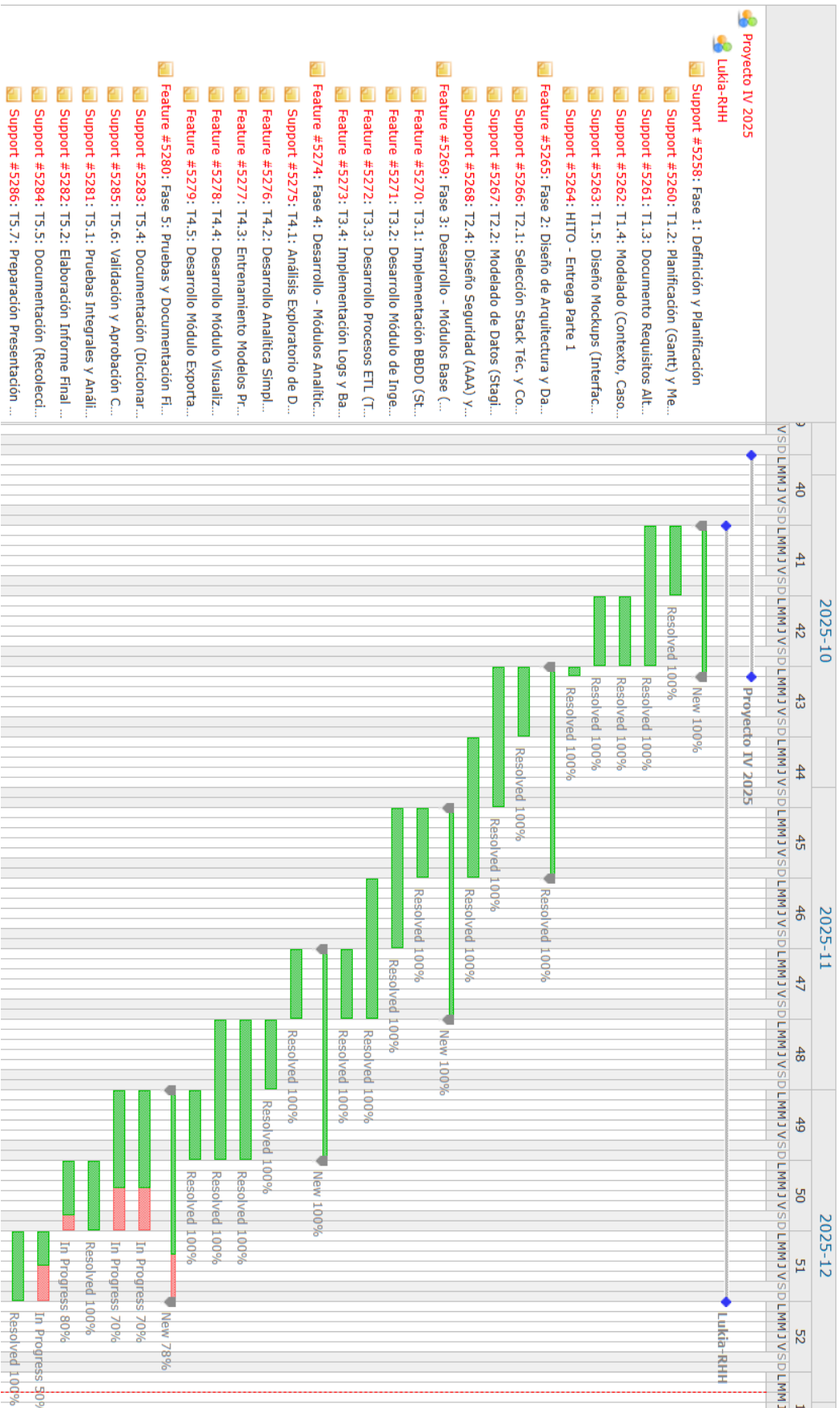
Implementar una plataforma analítica de datos para el área de RRHH de Casino Luckia Arica, que automatice la ingesta y procesamiento de información, permitiendo el análisis de métricas clave y la generación de modelos predictivos para facilitar la toma de decisiones proactiva sobre la gestión del personal.

4.3. Objetivos Específicos

1. **Diseñar y construir** un proceso ETL (Extracción, Transformación y Carga) capaz de ingestar, validar y unificar datos de múltiples fuentes (SAP y Geovictoria).
2. **Implementar** un Data Warehouse (DW) centralizado para consolidar y almacenar de forma histórica y optimizada la información del personal.
3. **Analizar** los datos históricos para desarrollar y entrenar un modelo predictivo inicial, enfocado en la probabilidad de rotación de personal por área.
4. **Desarrollar** un módulo de visualización con reportes automatizados y dashboards interactivos que muestren las métricas clave (rotación, ausentismo, dotación).
5. **Garantizar** la seguridad y privacidad de la información mediante la aplicación de principios de ciberseguridad (ej. control de acceso, logs, comunicaciones seguras) en la arquitectura de la solución.
6. **Documentar** la arquitectura del sistema, el diccionario de datos y los flujos ETL para asegurar su mantenibilidad y facilitar la futura captación de nuevos datos.

4.4. Carta Gantt

Título	Descripción	Semana (Relativa)	Fecha Inicio	Fecha Término
Fase 1: Definición y Planificación	Análisis de requisitos, alcance, modelado (Contexto, Casos de Uso, BPM) y mockups para la primera entrega.	1 - 2	06-oct-25	17-oct-25
HITO 1: Entrega 1	Entrega formal de la documentación de la Fase 1.	3	20-oct-25	20-oct-25
Fase 2: Diseño de Arquitectura y Datos	Selección de stack, configuración de Git, modelado de BBDD (Staging/DW) y diseño de procesos ETL y seguridad (AAA).	3 - 6	21-oct-25	14-nov-25
Fase 3: Desarrollo - Módulos Base (ETL/DW)	Implementación de BBDD, desarrollo de ingesta (Excel), procesos ETL (transformación/carga) y sistemas de logs/backups.	5 - 8	03-nov-25	21-nov-25
HITO 2: Entrega 2	<i>Presentación de avances de implementación y diseño detallado.</i>	<i>8 (Aprox)</i>	<i>(Sin Fecha)</i>	<i>(Sin Fecha)</i>
Fase 4: Desarrollo - Módulos Analíticos	Análisis exploratorio (EDA), desarrollo de KPIs (Rotación, etc.), entrenamiento de modelos predictivos y dashboards.	8 - 11	24-nov-25	16-dic-25
Fase 5: Pruebas y Documentación Final	Pruebas integrales, redacción de informes, manuales (Usuario, ETL, RPO/RTO) y validación con el cliente.	10 - 12	10-dic-25	23-dic-25
HITO 3: Entrega 3	<i>Entrega del sistema implementado y el informe final/pruebas.</i>	<i>11 (Aprox)</i>	<i>(Sin Fecha)</i>	<i>(Sin Fecha)</i>
HITO 4: Entrega Final	Presentación final del producto y entrega de manuales aprobados.	12	24-dic-25	24-dic-25



5. Requerimientos del proyecto

5.1. De Alto nivel

Visión del sistema

El sistema consiste en una **Plataforma Integral de Analítica de Personas** (asociados a la rama de estudio de People Analytics) diseñada para centralizar, procesar y explotar los datos de Recursos Humanos de la organización. Su propósito principal es transformar fuentes de información dispersas y no estructuradas en **una fuente única de verdad** que alimente tanto tableros de control históricos (Business Intelligence) como modelos predictivos de comportamiento laboral (Data Mining, Machine Learning). La solución garantiza la seguridad, privacidad y disponibilidad de la información sensible, proporcionando herramientas de toma de decisiones accesibles tanto para perfiles técnicos como gerenciales.

Requisitos de alto nivel

Requisito	Descripción
Gestión e Integración Centralizada de Datos	El sistema debe actuar como un repositorio central unificado capaz de ingerir, limpiar y estructurar datos provenientes de fuentes heterogéneas para garantizar la calidad y consistencia de la información organizacional.
Inteligencia de Negocios y Capacidad Predictiva	El sistema debe proveer capacidades analíticas avanzadas que permitan visualizar el estado actual de la organización (descriptivo) y anticipar comportamientos futuros de los colaboradores (predictivo) para apoyar la toma de decisiones estratégicas.
Gobernanza, Seguridad y Auditoría	El sistema debe garantizar la protección integral de los datos sensibles , asegurando el cumplimiento normativo, el control de acceso estricto y la trazabilidad de todas las operaciones realizadas en la plataforma.
Disponibilidad y Experiencia de Usuario (UX/Ops)	El sistema debe ofrecer una arquitectura robusta, escalable y fácil de mantener , con una interfaz que permita la explotación de datos y la administración del sistema de forma intuitiva para distintos perfiles de usuario.

5.2. Funcionales

ID	Título	Descripción
RF-01	Plataforma de integración de datos	El sistema debe contar con una plataforma para subir datos de distintas fuentes no relacionales (principalmente Planillas Excel), transformándolos en un formato común y trabajable.
RF-02	Base de datos de staging	El sistema debe contar con una base de datos separada para almacenar temporalmente los datos extraídos antes de su carga final en el Almacén de Datos.
RF-03	Almacén de Datos (DW)	<p>El sistema debe disponer de un Data Warehouse estructurado que permita almacenar los datos consolidados de forma histórica y optimizada para consulta.</p> <p>El sistema debe permitir la integración de datos desde múltiples fuentes mediante procesos ETL, incluyendo la validación, limpieza y transformación de la información antes de su carga en el Almacén de Datos.</p>
RF-04	Analítica simple de los datos	El sistema debe mostrar métricas simples calculadas automáticamente con los datos disponibles. Debe mostrar métricas como: Tasa de rotación, Tasa de Ausentismo, Dotación de personal, entre otros.
RF-05	Modelos predictivos	El sistema debe implementar modelos predictivos entrenados con los datos recopilados. Debe predecir: Probabilidad de rotación,
RF-06	Visualización de métricas y KPIs	El sistema debe mostrar gráficamente métricas relevantes obtenidas de los modelos y otros cruces de información complejos.
RF-07	Registro de logs	El sistema debe registrar eventos relevantes, como accesos, ejecuciones de procesos ETL, fallos y cambios de configuración.
RF-08	Backups automáticos	El sistema debe realizar copias de seguridad periódicas de las bases de datos y configuraciones del sistema.
RF-09	Control de acceso AAA	El sistema debe implementar autenticación, autorización y auditoría de usuarios (AAA) con distintos roles de acceso.
RF-10	Exportación de datos	El sistema debe permitir exportar los datos transformados o consolidados en formatos como CSV, Excel o JSON.

5.3. No funcionales

ID	Título	Descripción
RNF-01	Comunicaciones seguras	Toda comunicación entre módulos o usuarios debe realizarse mediante protocolos cifrados (TLS/SSL).
RNF-02	Disponibilidad	El sistema debe mantenerse disponible y operativo de forma continua, considerando réplicas y ventanas de mantenimiento programadas.
RNF-03	Escalabilidad	El sistema debe poder ampliar su capacidad de procesamiento y almacenamiento sin rediseñar su arquitectura base.
RNF-04	Mantenibilidad	El sistema debe diseñarse de forma modular para facilitar actualizaciones, pruebas y resolución de errores.
RNF-05	Compatibilidad	El sistema debe ser compatible con las tecnologías y herramientas de la organización (Microsoft, PostgreSQL, Power BI, etc.).
RNF-06	Confiabilidad	El sistema debe garantizar la integridad de los datos y la continuidad operativa ante fallos.
RNF-07	Eficiencia	El sistema debe procesar y cargar los datos en tiempos razonables según el volumen esperado.
RNF-08	Legalidad y privacidad	El sistema debe cumplir con la normativa vigente sobre protección de datos personales y confidencialidad laboral.
RNF-09	Auditabilidad	Todas las acciones del sistema deben poder rastrearse para fines de control y cumplimiento normativo.
RNF-10	Usabilidad dual	El sistema debe presentar una interfaz sencilla para usuarios no técnicos y funcionalidades avanzadas para expertos.

5.4. De documentación

ID	Título	Descripción
RD-01	Diccionario de datos	Se debe elaborar un diccionario que describa el significado, formato, dominio y relaciones de cada campo del sistema.
RD-02	Documentación de procesos ETL	Se debe documentar cada flujo de extracción, transformación y carga, incluyendo fuentes, reglas y errores posibles.
RD-03	Informe de Análisis Exploratorio de Datos (EDA)	Debe elaborarse un informe técnico que resuma la completitud, consistencia, representatividad y calidad de los datos disponibles; con el objetivo de analizar la viabilidad de los modelos predictivos y observar la necesidad de la captación de nuevos datos.
RD-05	Plan de recolección de nuevos datos	Debe elaborarse un documento con estrategias para mejorar la cobertura y calidad de datos.
RD-06	Manual de usuario administrativo	Debe documentarse el uso de la plataforma, incluyendo carga, validación, reportes, backups y administración de usuarios.
RD-07	Plan de recuperación ante fallos (RPO/RT0)	Se debe definir un plan para restaurar datos y servicios dentro de los tiempos objetivos de recuperación (RPO y RT0).
RD-08	Plan de adopción	Se debe elaborar un plan que facilite la incorporación del sistema en la organización, incluyendo capacitación y soporte inicial.

6. Metodología

Para abordar la naturaleza dual de este proyecto, que combina el **desarrollo de un producto de software** (la plataforma) con un proceso de **descubrimiento de conocimiento** (la analítica de datos), se implementará una metodología híbrida.

Esta metodología combina el macro-proceso **CRISP-DM** para guiar el proyecto de datos, con la agilidad y simplicidad de **Kanban** para la gestión diaria del desarrollo y las tareas.

6.1. Macro-Proceso Analítico: CRISP-DM

Para el componente de analítica y modelos predictivos, se utilizará el **Proceso Estándar Inter-industria para Minería de Datos (CRISP-DM)**. Este modelo es el estándar de la industria porque es iterativo y se centra en cómo los datos pueden responder a los problemas del negocio.

Las fases de CRISP-DM se aplicarán de la siguiente manera:

1. **Comprensión del Negocio (Business Understanding):** Completada. Corresponde a la "Definición del Problema" y "Objetivos", donde se identificó la necesidad de analizar la rotación (RF-05) y el ausentismo (RF-04).
2. **Comprensión de los Datos (Data Understanding):** Se realizará un Análisis Exploratorio de Datos (EDA) (RD-03) para identificar la calidad, completitud y potencial de los datos de SAP y Geovictoria.
3. **Preparación de los Datos (Data Preparation):** Es la fase más intensiva en desarrollo. Implica construir los flujos ETL (RF-03) para limpiar, transformar y consolidar los datos en el Data Warehouse (RF-02, RF-03).
4. **Modelado (Modeling):** Se seleccionarán y se usarán diferentes técnicas y algoritmos de minería de datos para intentar predecir la rotación de personal (RF-05), utilizando los datos preparados.
5. **Evaluación (Evaluation):** Se medirán los modelos para asegurar que su precisión y rendimiento sean aceptables y que respondan correctamente a la pregunta de negocio.
6. **Despliegue (Deployment):** El modelo predictivo final se integrará en la plataforma para ser consumido a través de los dashboards y reportes (RF-06).

6.2. Gestión Ágil del Desarrollo: Kanban

Para la gestión diaria del proyecto y el desarrollo de software (ETLs, BBDD, App Web, Dashboards), se utilizará **Kanban**. Esta metodología ágil se prefiere sobre Scrum por su simplicidad y flexibilidad, eliminando ceremonias complejas y centrándose en el flujo de trabajo visual.

Nuestra implementación de Kanban se basará en tres pilares:

1. **Visualizar el Flujo (Tablero Kanban):** Se utilizará un tablero digital simple para visualizar todas las tareas. Las columnas básicas serán:
 - **Backlog (Por Hacer):** Lista priorizada de todas las tareas.
 - **En Proceso (Doing):** Tareas que se están trabajando activamente. Se limitará el "Trabajo en Progreso" (WIP) para mantener el foco y simplificar el proceso.
 - **En Pruebas / Validación:** Tareas terminadas que están siendo probadas por el equipo o validadas por el cliente.
 - **Hecho (Done):** Tareas completadas y aprobadas.
2. **Cadencias (Reuniones Ágiles y Ligeras):** Para mantener la comunicación y la cadencia de desarrollo, adoptaremos solo las ceremonias más valiosas y simples:
 - **Reunión Diaria (Daily Stand-up):** Una reunión de 15 minutos cada mañana frente al tablero Kanban para sincronizar al equipo. Cada miembro responde: ¿Qué hice ayer? ¿Qué haré hoy? ¿Qué impedimentos tengo?
 - **Revisión de Entrega (Review/Demo):** Al finalizar un hito o una funcionalidad importante (ej. "Módulo de Reportes terminado"), se realizará una demo al cliente (Profesor y/o Luckia) para obtener feedback temprano, en lugar de esperar al final de un Sprint.
3. **Gestión del Flujo (Mejora Continua):** En lugar de planificar en Sprints, el equipo se guiará tomando la siguiente tarea de mayor prioridad del Backlog, tan pronto como termine una tarea actual.

6.3. Integración de la Metodología Híbrida

Ambas metodologías funcionan en conjunto:

- **CRISP-DM** define las **grandes fases analíticas** del proyecto, las mencionadas anteriormente.
- **Kanban** se usa para **desglosar y ejecutar** esas fases en tareas atómicas y manejables.

6.4. Definición de tareas

Para gestionar el trabajo siguiendo la metodología Kanban, el proyecto se desglosa en un **Backlog** (lista de trabajo pendiente). Este backlog se estructura jerárquicamente en **Épicas** y **Tareas**.

1. **Épicas (Epics)**: Son las grandes unidades de trabajo o módulos del proyecto (ej. "Módulo de Ingesta y ETL"). Una Épica es demasiado grande para trabajarla directamente, pero sirve para agrupar tareas relacionadas.
2. **Tareas (Tasks)**: Son las unidades de trabajo individuales y accionables que se moverán por el tablero Kanban (ej. "Crear tabla de Staging"). Cada tarea nace de uno o más requisitos (RF, RNF, RD).

Sistema de Etiquetas (Tags)

Cada tarea en el backlog se clasifica con un sistema de etiquetas simple para entenderla a simple vista dentro del mismo:

- **Fase (#fase[1-5])**: Indica a qué fase de la Carta Gantt pertenece (ej. #fase3).
- **Tipo (#[tipo])**: Define la naturaleza del trabajo (ej. #doc para documentación, #feature para funcionalidad, #config para infraestructura, #analisis para investigación).
- **Componente (#[componente])**: El módulo técnico que afecta (ej. #etl, #dw, #ui, #seguridad).
- **Referencia (#[ref])**: El requisito principal que cumple (ej. #rf01, #rnf02).

A continuación, se presenta el backlog inicial del proyecto.

Épica 1: Definición y Planificación (Entrega Parte 1)

Agrupar todas las tareas documentales, de análisis y de diseño de alto nivel requeridas para cumplir con la primera entrega del 20 de octubre.

Nombre Tarea	Descripción	Tags	Prioridad
Redactar Doc. Problema/Solución	Definir y justificar el problema y la solución escogida.	#fase1 #doc #proyecto	Crítica
Redactar Doc. Alcance y Objetivos	Detallar el alcance, objetivos (Gral/Específicos) del proyecto.	#fase1 #doc #proyecto	Crítica
Crear Carta Gantt	Planificar las fases, tareas e hitos del proyecto (06-Oct al 24-Dic).	#fase1 #doc #proyecto	Crítica
Redactar Doc. Metodología	Explicar el uso de CRISP-DM y Kanban (Sección 5).	#fase1 #doc #proyecto	Crítica

Crear Doc. Requisitos Alto Nivel	Enumerar las características principales y requisitos funcionales.	#fase1 #doc #proyecto #rf-all	Crítica
Desarrollar Modelo de Contexto	Crear diagrama de contexto (Sistema vs. Entorno externo).	#fase1 #doc #diseño	Crítica
Desarrollar Modelo Casos de Uso	Crear diagrama y descripción de Casos de Uso (Admin, Analista).	#fase1 #doc #diseño	Crítica
Desarrollar Diagrama BPMN	Mapear el proceso de negocio propuesto (flujo de datos y análisis).	#fase1 #doc #diseño	Crítica
Crear Mockups de Interfaces	Diseñar un prototipo de baja fidelidad de la UI (Carga y Dashboard).	#fase1 #doc #ui #rnf10	Crítica

Épica 2: Diseño de Arquitectura y Datos

Define los "planos" técnicos detallados del sistema, las bases de datos y la seguridad antes de construir.

Nombre Tarea	Descripción	Tags	Prioridad
Documentar Stack Tecnológico	Decidir y documentar SO, BBDD, Lenguajes, Frameworks.	#fase2 #doc #arquitectura #rnf05	Crítica
Diseñar Topología de Red	Crear diagrama de componentes (Servidor, BBDD, App Web).	#fase2 #doc #arquitectura #rnf01	Media
Modelar BBDD Staging	Diseñar el esquema de tablas para la BBDD de Staging.	#fase2 #doc #bd #rf02 #rd01	Crítica
Modelar BBDD Data Warehouse	Diseñar el modelo dimensional (tablas de Hechos y Dimensiones).	#fase2 #doc #dw #rf03 #rd01	Crítica
Mapear Flujo ETL	Documentar el mapeo de campos (Origen -> Staging -> DW).	#fase2 #doc #etl #rd02	Crítica
Definir Reglas de Limpieza	Especificar las reglas de negocio para transformar y validar datos.	#fase2 #doc #etl #rd02 #rnf06	Crítica
Definir Roles y Permisos	Detallar qué puede hacer cada rol (Admin, Analista, Visitante).	#fase2 #doc #seguridad #rf09	Alta
Definir Estrategia de Backup	Especificar frecuencia, retención y lugar de almacenamiento.	#fase2 #doc #backup #rf08	Baja

Épica 3: Configuración de Infraestructura y BBDD

Tareas para "levantar" el entorno de desarrollo, instalar servicios y desplegar las bases de datos vacías.

Nombre Tarea	Descripción	Tags	Prioridad
Instalar y Configurar Servidor	Instalar Aplicaciones y Servicios para soportar el sistema (BBDD, Git, Python, Docker).	#fase3 #config #servidor	Crítica
Configurar Comunicaciones Seguras	Implementar certificados TLS/SSL autofirmados o Let's Encrypt.	#fase3 #config #seguridad #rnf01	Baja
Desplegar BBDD Staging	Crear y ejecutar script SQL para crear la base de datos de Staging.	#fase3 #feature #bd #rf02	Alta
Desplegar BBDD Data Warehouse	Crear y ejecutar script SQL para crear el Data Warehouse (esquema).	#fase3 #feature #dw #rf03	Alta

Épica 4: Desarrollo Módulo de Ingesta y ETL

Construcción del "corazón" del sistema: la plataforma para cargar archivos y los scripts que mueven y transforman los datos.

Nombre Tarea	Descripción	Tags	Prioridad
Desarrollar UI Carga de Archivos	Crear la interfaz web para que el usuario suba planillas Excel.	#fase3 #feature #ui #rf01 #rnf10	Alta
Desarrollar API Carga (Excel -> Staging)	Crear el backend que recibe el archivo y lo inserta en BBDD Staging.	#fase3 #feature #api #rf01	Crítica
Desarrollar Script ETL (Staging -> DW)	Crear el proceso (Transform/Load) que aplica reglas y mueve datos al DW.	#fase3 #feature #etl #rf03	Crítica
Orquestar Flujo ETL	Crear el script principal que llama a la Carga y al T-L en orden.	#fase3 #feature #etl #rf03 #rnf07	Crítica

Épica 5: Desarrollo Módulo de Seguridad y Operaciones

Implementación de los sistemas de acceso, auditoría (logs) y los respaldos automáticos.

Nombre Tarea	Descripción	Tags	Prioridad
Desarrollar Módulo de Autenticación	Crear o integrar (SSO) la lógica de Login/Logout.	#fase3 #feature #seguridad #rf09	Media
Desarrollar Módulo de Autorización	Implementar la lógica de Roles y Permisos en la UI/API.	#fase3 #feature #seguridad #rf09	Media
Implementar Sistema de Logs	Configurar logging para accesos, errores y eventos ETL.	#fase3 #feature #auditoria #rf07 #rnf09	Alta
Automatizar Backups BBDD	Crear y programar (ej. cronjob) el script de backup automático.	#fase3 #config #backup #rf08 #rnf06	Baja

Épica 6: Análisis Exploratorio y Modelado Predictivo

Tareas de Ciencia de Datos (CRISP-DM) para entender los datos y construir el modelo de predicción de rotación.

Nombre Tarea	Descripción	Tags	Prioridad
Ejecutar EDA (Análisis Exploratorio)	Analizar la calidad, distribución y completitud de los datos históricos.	#fase4 #analisis #datos #rd03	Alta
Preparar Dataset para Modelo	Crear el dataset de entrenamiento (Feature Engineering, limpieza).	#fase4 #analisis #modelo #rf05	Media
Entrenar y Evaluar Modelo	Probar algoritmos (ej. Regresión Logística, Random Forest) y evaluarlos.	#fase4 #analisis #modelo #rf05	Media
Desplegar Modelo (Job/API)	Implementar el modelo entrenado para generar predicciones.	#fase4 #feature #modelo #rf05	Media

Épica 7: Desarrollo Módulo Analítico y Visualización

Descripción: Construcción de la interfaz de usuario final (Dashboards) para mostrar KPIs y predicciones.

Nombre Tarea	Descripción	Tags	Prioridad
Implementar Cálculos de KPIs	Crear Vistas o Procedimientos Almacenados para KPIs.	#fase4 #feature #dw #rf04	Alta
Desarrollar Dashboard Sección de KPIs	Crear la UI para visualizar la Analítica Simple (Gráficos).	#fase4 #feature #ui #rf04 #rf06	Alta
Desarrollar Dashboard Sección Predictivo	Crear la UI para visualizar los resultados del modelo de rotación.	#fase4 #feature #ui #rf05 #rf06	Media
Implementar Módulo de Exportación	Desarrollar la función para exportar datos (CSV/Excel).	#fase4 #feature #api #ui #rf10	Baja

Épica 8: Documentación Final y Cierre

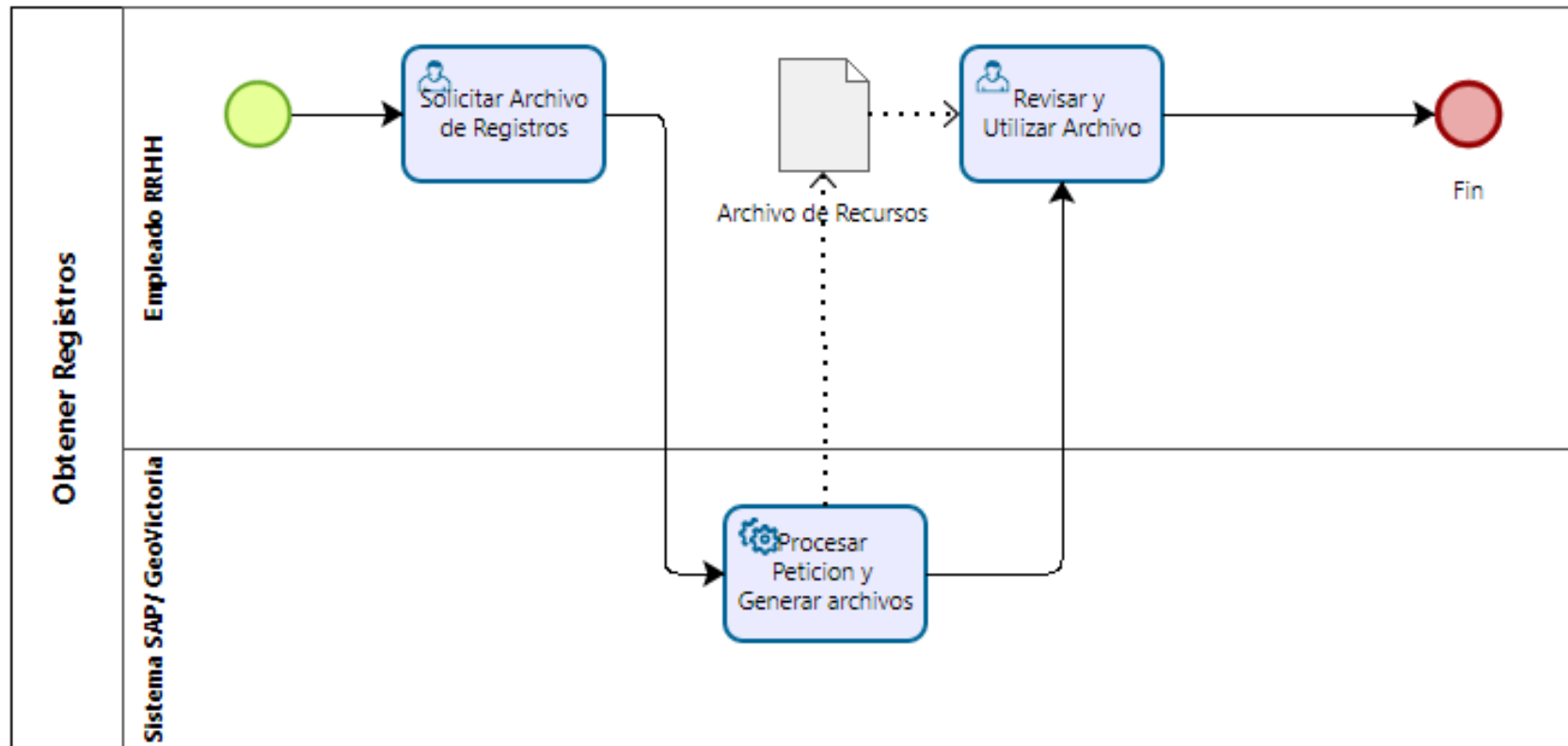
Descripción: Tareas de empaquetado del proyecto, creación de manuales y planes para la entrega final.

Nombre Tarea	Descripción	Tags	Prioridad
Redactar Diccionario de Datos	Documentar el esquema final del DW (tablas, campos, tipos).	#fase5 #doc #bd #rd01	Alta
Documentar Flujos ETL	Crear los diagramas y descripción final de los flujos ETL.	#fase5 #doc #etl #rd02	Media
Redactar Manual de Usuario/Admin	Crear guía de uso de la plataforma (Carga, Reportes, Usuarios).	#fase5 #doc #usuario #rd06	Media
Redactar Plan de Recuperación (RPO/RTO)	Documentar el plan de recuperación ante desastres.	#fase5 #doc #riesgos #rd07	Baja
Redactar Plan de Recolección de Datos	Proponer estrategias para capturar nuevos datos a futuro.	#fase5 #doc #analisis #rd05	Baja
Redactar Plan de Adopción	Documentar el plan de capacitación y transferencia.	#fase5 #doc #usuario #rd08	Baja
Pruebas Finales y Validación Cliente	Ejecutar pruebas integrales y obtener la aprobación final.	#fase5 #qa #cliente	Media

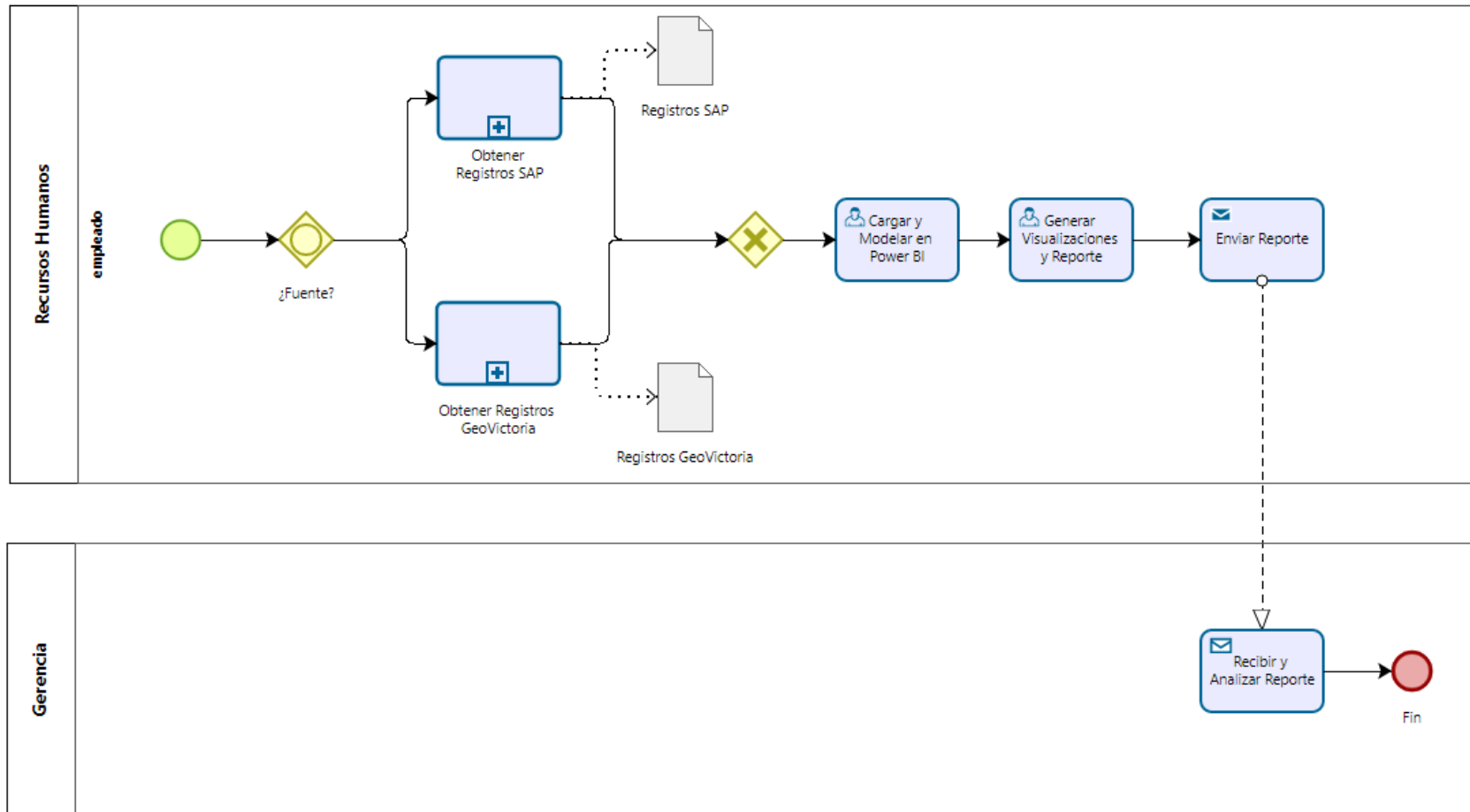
7. Modelado del Procesos de Negocio

El modelado de procesos de negocio es bastante sencillo comparado al sistema propuesto. Previo al sistema, los datos se extraen manualmente y las métricas se calculan a pulso.

Extracción manual de los datos de los principales sistemas presentes en la empresa (SAP-RRHH y GeoVictoria-Asistencia)



Unificación, modelamiento y cálculo de métricas desde los datos, para presentarse en gerencia y toma de decisiones.

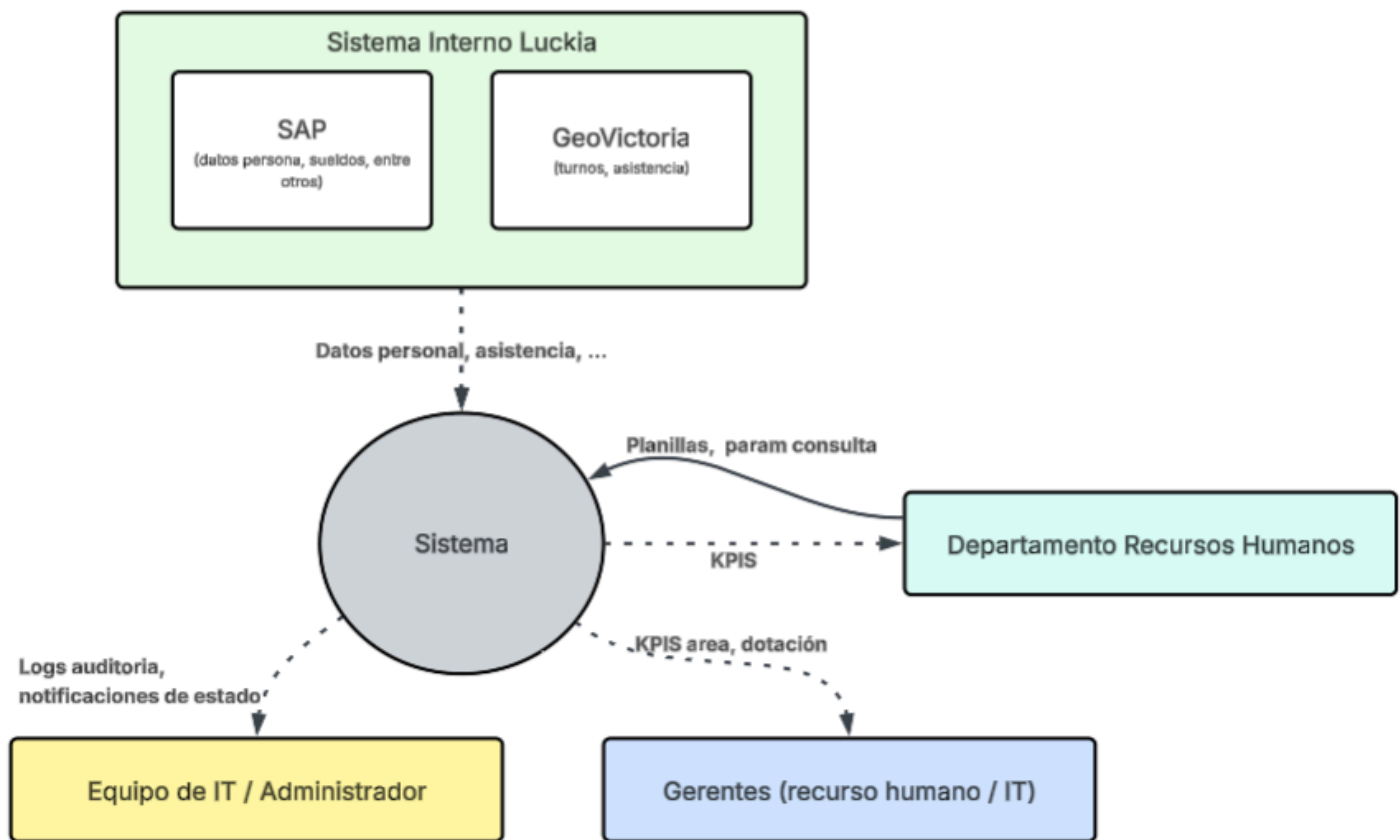


8. Modelado del sistema

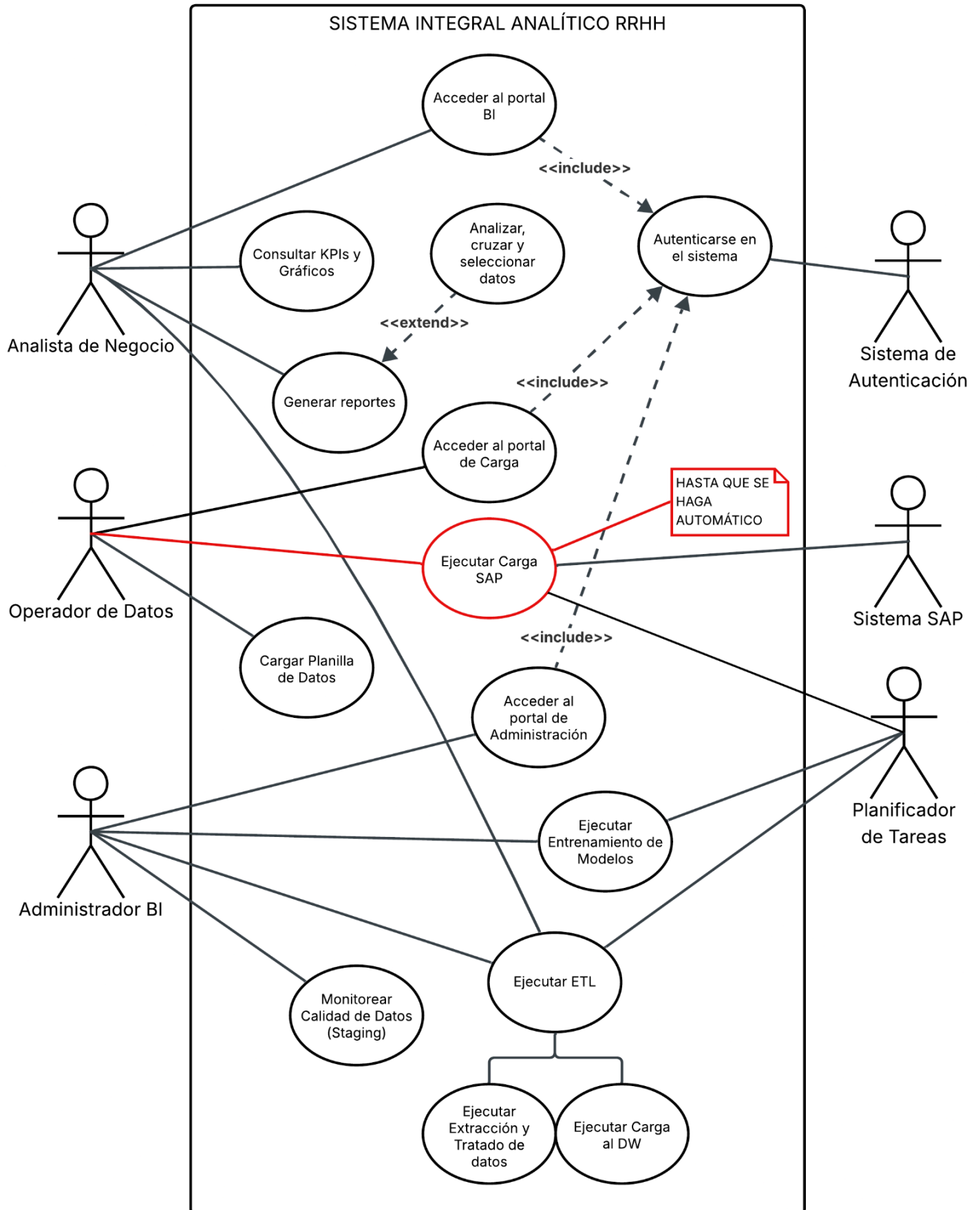
8.1. Modelo de contexto

El diagrama de contexto ilustra la plataforma analítica, denominada "Sistema", como la entidad central del flujo de información. Este sistema se alimenta desde dos canales principales: por un lado, ingiere los datos provenientes de los sistemas transaccionales de la empresa (SAP y GeoVictoria), los cuales proveen la información maestra del personal, como sus detalles contractuales, sueldos, turnos y registros de asistencia. Por otro lado, también recibe entradas de datos manuales, como planillas Excel, directamente desde el Departamento de Recursos Humanos.

Una vez que el sistema procesa esta información, la distribuye a sus usuarios a través de dos canales principales. Primero, provee la inteligencia de negocio (incluyendo indicadores de gestión (KPIs), métricas especializadas como la dotación por área y los resultados de los modelos de predicción) tanto al Departamento de Recursos Humanos como a los Gerentes. Segundo, reporta toda la información técnica, como los registros de auditoría y las notificaciones de estado del sistema, al Equipo de IT o Administrador para fines de monitoreo y control.



8.2. Modelo de casos de uso



1. Actor: Analista de Negocio

Es el consumidor principal de la información.

- **Acceder al portal BI:** El analista ingresa a la plataforma de visualización.
 - **(Include) Autenticarse en el sistema:** Para acceder, el sistema siempre le pedirá que inicie sesión.
- **Consultar KPIs y Gráficos:** Visualiza los reportes y dashboards predefinidos.
- **Generar reportes:** Utiliza la plataforma para crear reportes personalizados.
 - **(Extend) Analizar, cruzar y seleccionar datos:** Opcionalmente, durante la generación de reportes, puede realizar análisis más profundos y cruces de datos específicos.
- **Ejecutar ETL:** Tiene la potestad de iniciar un proceso de actualización de datos fuera del horario programado.

2. Actor: Operador de Datos

Es el encargado de la ingesta de los datos más manuales.

- **Acceder al portal de Carga:** Ingresa a la sección específica para subir archivos.
- **Cargar Planilla de Datos:** Sube los archivos (ej. Excel) que contienen datos que no se pueden obtener automáticamente.

3. Actor: Administrador BI

Es el rol técnico que mantiene y monitorea los datos del sistema analítico.

- **Acceder al portal de Administración:** Ingresa a la consola de administración del sistema.
- **Monitorear Calidad de Datos (Staging):** Revisa los datos en la base de datos temporal (staging) para asegurar que se cargaron correctamente antes de moverlos al DW.
- **Ejecutar ETL:** Tiene la potestad de iniciar un proceso de actualización de datos fuera del horario programado, ya sea para probar el sistema o depurar el proceso.
- **Ejecutar Entrenamiento de Modelos:** Inicia manualmente el proceso de re-entrenamiento del modelo predictivo.

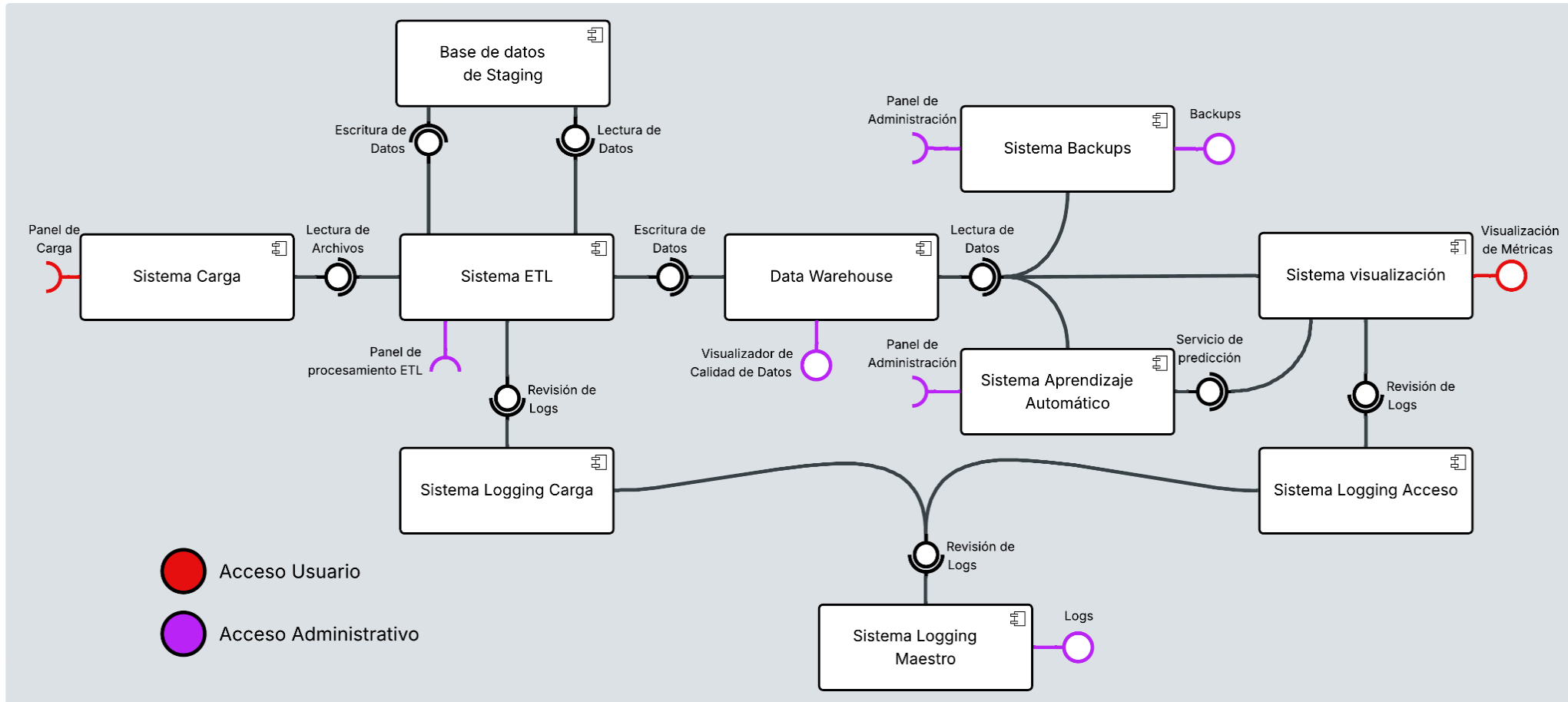
4. Actores: Sistemas Externos

Son sistemas automáticos que interactúan con la plataforma.

- **Sistema SAP (Actor):**
 - **Ejecutar Carga SAP:** El sistema SAP exporta sus datos.
- **Planificador de Tareas (Actor):**
 - **Ejecutar Carga SAP:** Invoca la exportación de datos al sistema SAP.
 - **Ejecutar ETL:** Inicia automáticamente el proceso de Extracción, Transformación y Carga (ETL) que mueve los datos de Staging al Data Warehouse.
 - **Ejecutar Entrenamiento de Modelos:** Inicia automáticamente el proceso de re-entrenamiento del modelo predictivo.
- **Sistema de Autenticación (Actor):**
 - **Autenticarse en el sistema:** Valida las credenciales de los usuarios.

8.3. Modelo de Componentes Lógicos

Modelo de Alto Nivel



A continuación, se describen los componentes principales y secundarios del sistema, representados en el diagrama de componentes lógicos. Estos elementos aseguran un flujo eficiente de datos desde la ingesta hasta el análisis y la visualización, alineados con los objetivos del proyecto.

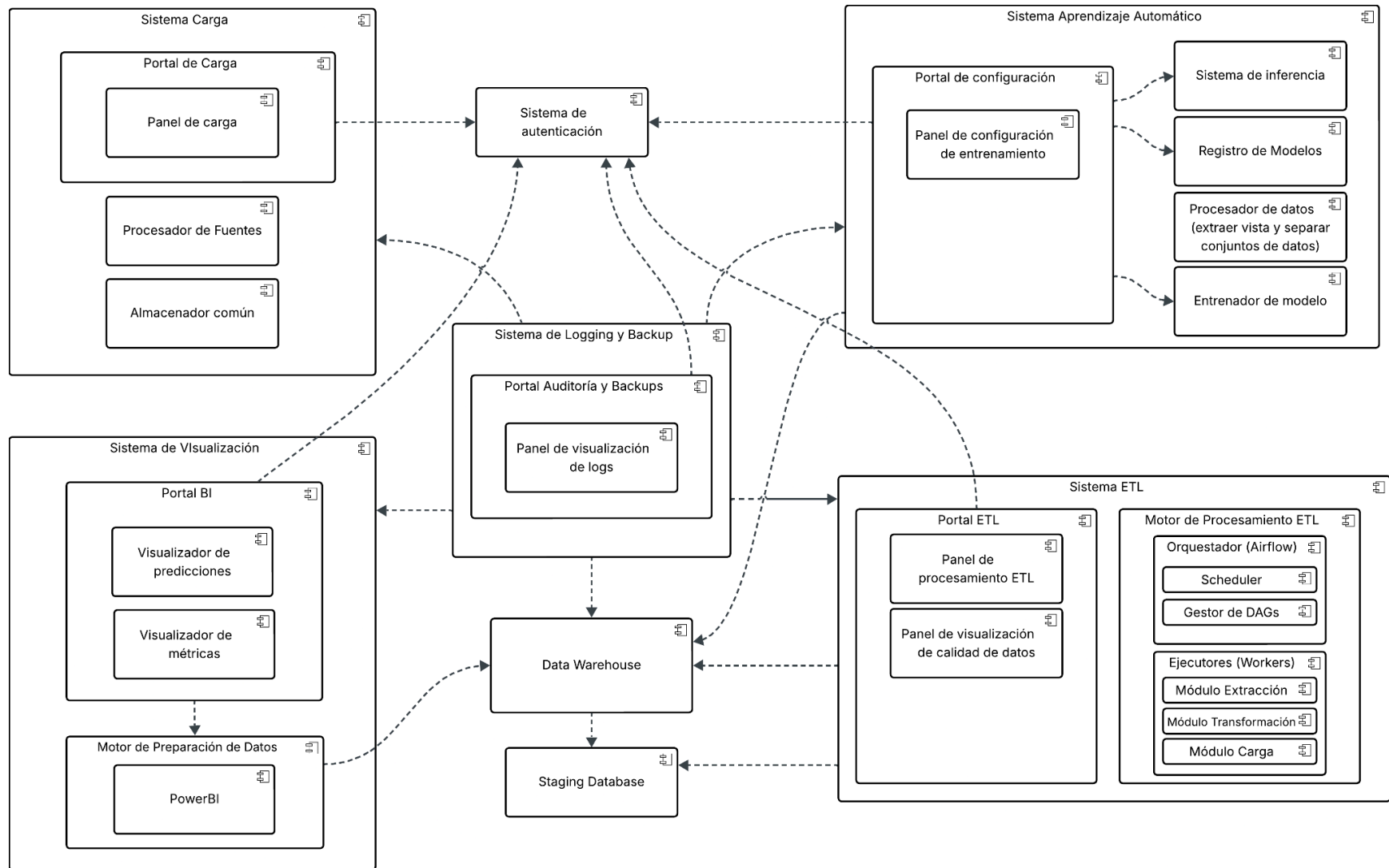
Componentes Principales:

- **Sistema de Carga:** Recibe datos crudos, realiza un procesamiento mínimo de estandarización y los prepara para su recolección periódica (mensual o anual) por el ETL.
- **Sistema ETL:** Integrado por el Planificador de Tareas y el Ejecutor de Workflows, transforma exhaustivamente los datos preprocesados para su integración óptima en el Data Warehouse.
- **Base de Datos de Staging:** Proporciona almacenamiento temporal para el procesamiento intensivo realizado por el ETL, facilitando la depuración y validación de datos.
- **Data Warehouse:** Ofrece almacenamiento persistente de datos integrados, optimizado para consultas analíticas y entrenamiento de modelos predictivos.
- **Sistema de Aprendizaje Automático:** Procesa datos del Data Warehouse para entrenar modelos predictivos, almacena los modelos en un registro y expone un servicio de inferencia para predicciones en tiempo real.
- **Sistema de Visualización:** Consulta el Data Warehouse y los resultados predictivos para generar gráficos y dashboards interactivos, apoyando la toma de decisiones.

Componentes Secundarios:

- **Sistemas de Logging (Carga, Acceso y Maestro):** Recopilan e integran logs de todos los subsistemas para monitoreo, auditoría y resolución de incidencias.
- **Sistema de Backups:** Gestiona copias de seguridad periódicas del Data Warehouse, garantizando la integridad y recuperación de datos en caso de fallos.

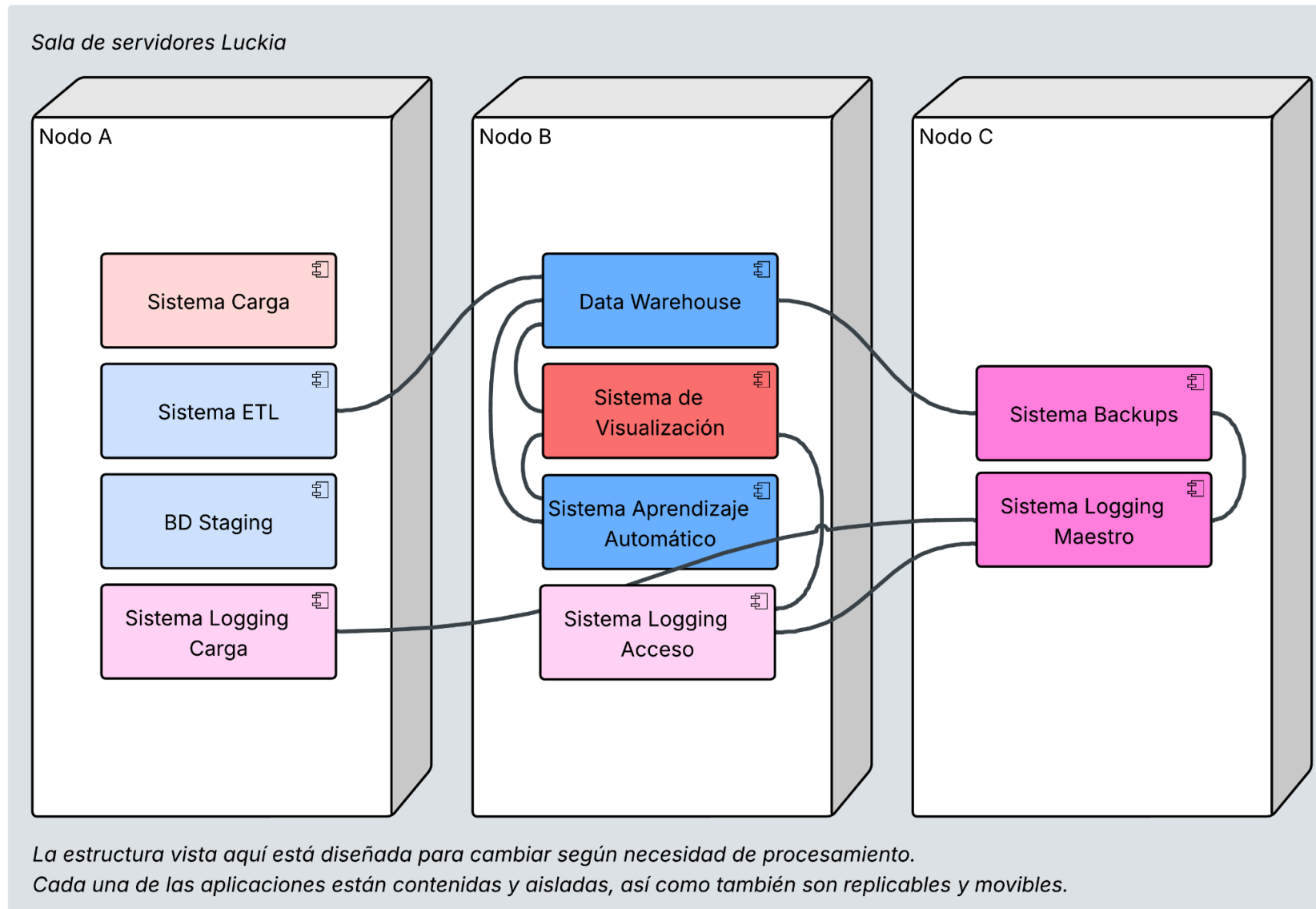
Modelado de nivel Medio



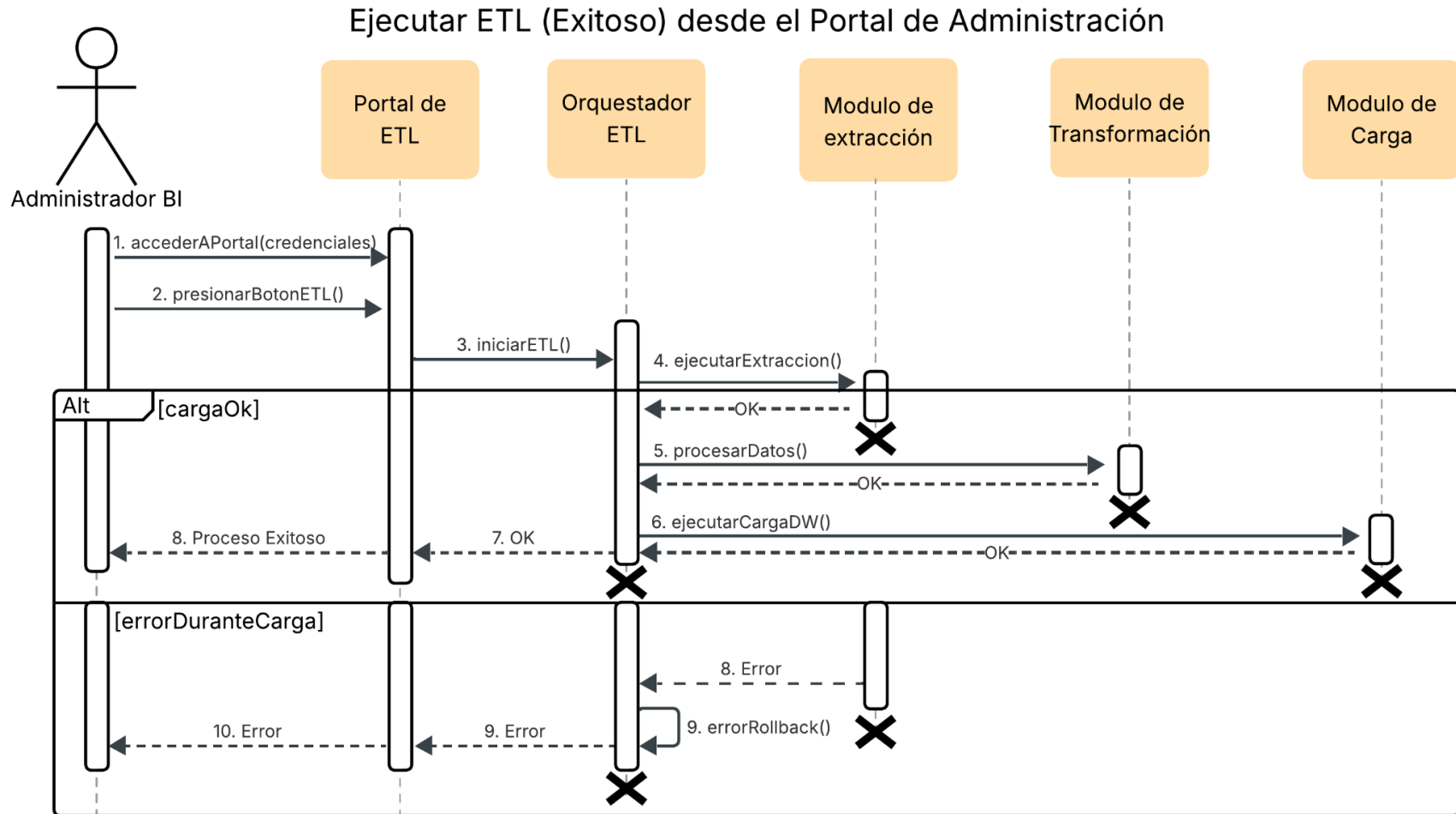
Arquitectura del Sistema de Componentes

Sistema de Carga	Gestiona la ingesta de los datos. A través del Portal de Carga, el usuario interactúa con el Panel de Carga para subir y validar planillas según estándares predefinidos (si es que aplica). El Procesador de Fuentes convierte estos datos a un formato intermedio y finalmente el Almacenador Común los almacena y categoriza. Esta última parte actúa como zona de aterrizaje intermedia para los archivos antes de ser procesados por el ETL.
Sistema ETL	Es el componente principal de integración de los datos. El Portal ETL permite la gestión operativa mediante el Panel de Procesamiento y la supervisión de datos con el Panel de Calidad. A nivel de backend, el Motor de Procesamiento ETL separa el control de la acción: el Orquestador (Airflow) utiliza un Planificador (Scheduler) y un Gestor de DAGs (Workflows) para dictar cuándo y cómo se ejecutan las tareas, mientras que la Ejecución (Workers) delimita las partes funcionales independientes, de cada Módulo de Extracción, Transformación y Carga. Estos finalmente son ejecutados por el Orquestador y mueven la información desde el Almacenador Común hacia el DW.
Capa de Datos	Constituye la base de almacenamiento del sistema. Los datos pasan primero por la Base de Datos de Staging, utilizada para conversiones de formato rápidas y temporales. El destino final es el Data Warehouse, que resguarda la información integrada y definitiva, alimentando tanto a los sistemas de visualización como a los de aprendizaje automático.
Sistema de Visualización	Se encarga de la presentación de datos consumiendo directamente del Almacén de Datos. Utiliza un Motor de Preparación de Datos (implementándose actualmente sobre PowerBI) como capa de abstracción de consultas. La información se despliega en el Portal BI, el cual contiene el Visualizador de Métricas para el análisis descriptivo y el Visualizador de Predicciones para exponer los resultados de los modelos de IA.
Sistema de Aprendizaje Automático	Administra el ciclo de vida de los modelos predictivos. El Procesador de Datos extrae conjuntos de entrenamiento ("vistas minables") del Almacén de Datos. Estos alimentan al Entrenador de Modelo, cuyos parámetros se definen previamente en el Panel de Configuración de Entrenamiento dentro del Portal de Configuración. Los modelos resultantes se guardan en el Registro de Modelos y son puestos en producción mediante el Sistema de Inferencia.
Sistema de Logging y Backup	Opera transversalmente para el monitoreo y resguardo del sistema. El Recolector de Logs centraliza la actividad de todos los sistemas, permitiendo su revisión en el Panel de Visualización de Logs del Portal Auditoría y Backups. Asimismo, el Gestor de Backups realiza copias de seguridad del Almacén de Datos, las cuales se administran y recuperan desde el Panel de Gestión de Backups.
Sistema de Autenticación	Servicio transversal que valida la identidad de los usuarios, controlando el acceso a los distintos portales y servicios del ecosistema (Carga, ETL, BI, etc.) de manera centralizada o individual.

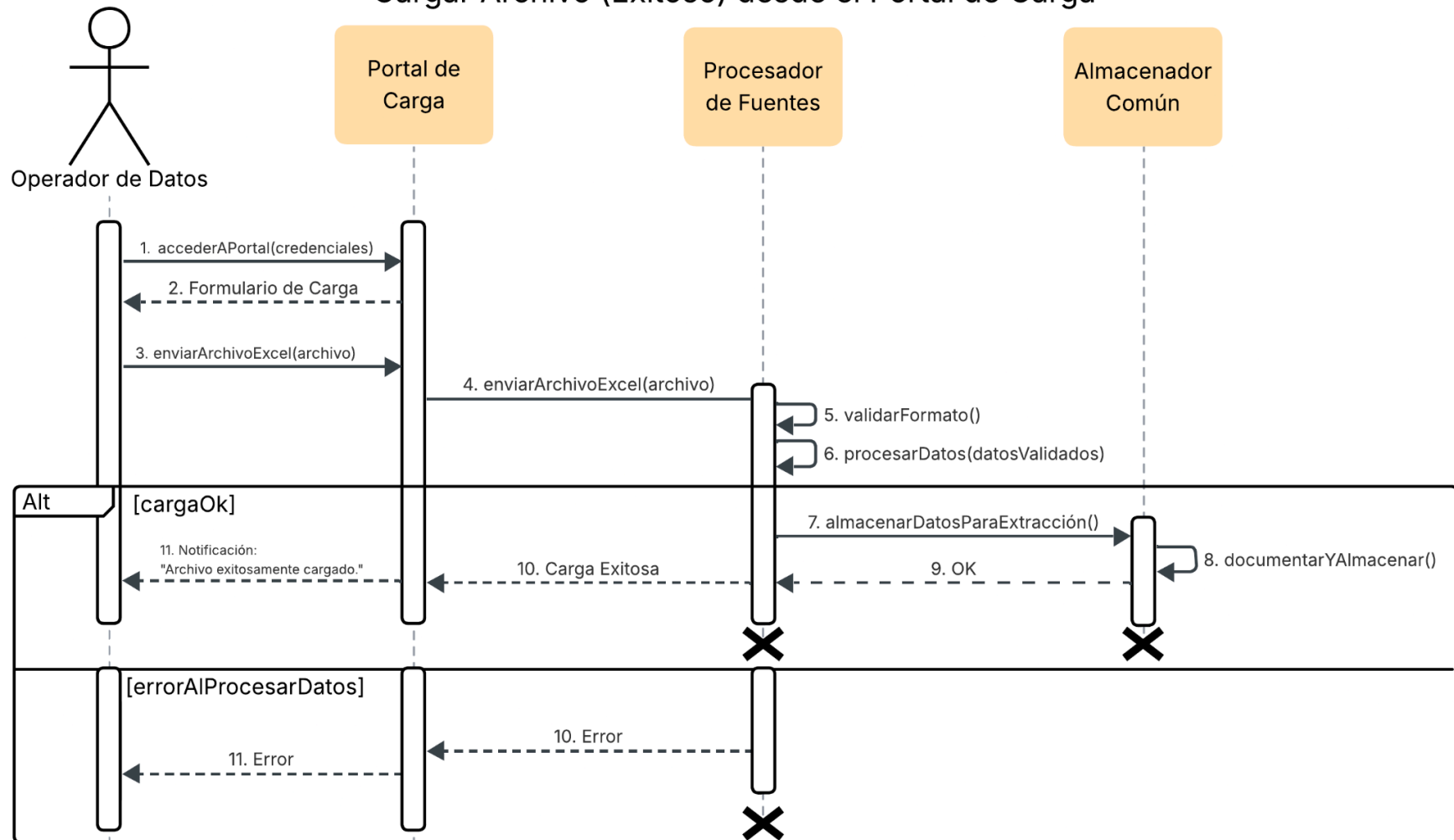
Diagrama de despliegue

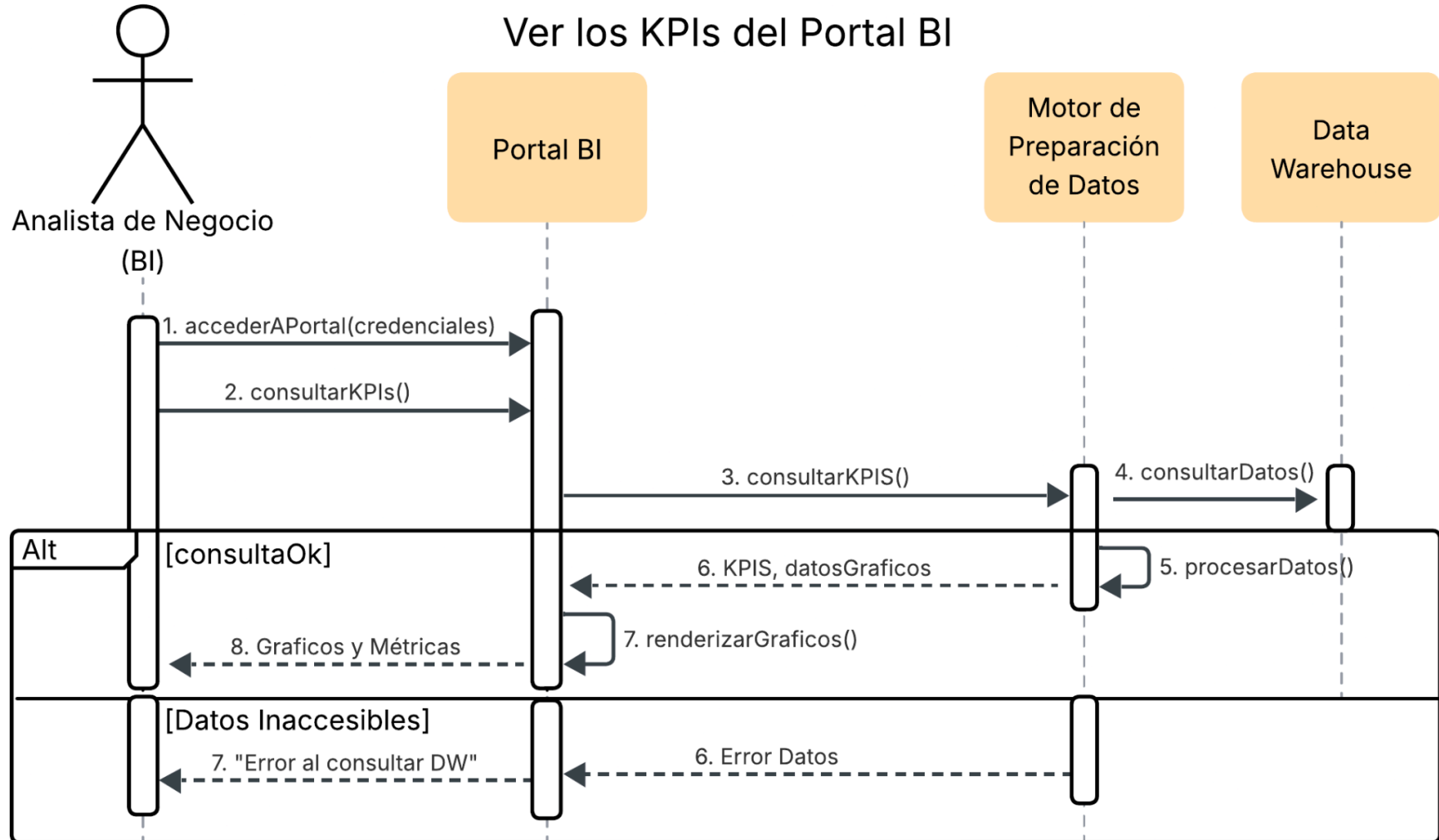


8.4. Diagramas de interacción



Cargar Archivo (Exitoso) desde el Portal de Carga





9. Arquitectura y Pila Tecnológica

9.1. Aspectos Arquitecturales del Sistema

Suposiciones y Limitaciones

- **Ausencia de Estándares de Fuentes de Datos:** Se parte de la base de que no existe un estándar definido para las fuentes de datos. En caso de existir, se asume que este es impuesto por un tercero (como Luckia España), es inmutable a nivel interno e intrínsecamente impredecible.
- **Fragilidad de la Implementación:** Dada la falta de estándares, se anticipa que la implementación resultante será extremadamente frágil. Se requerirá una adaptación y mantenimiento continuos ante cualquier cambio en los formatos de datos de origen.

Decisiones de Diseño

Las siguientes decisiones se han tomado en relación con la arquitectura y las herramientas del proyecto:

- **Bases de Datos:** Se implementarán dos bases de datos separadas (Staging y Data Warehouse) en ubicaciones físicas distintas para optimizar la integración de datos, el procesamiento y evitar sobrecargas.
- **Lenguaje de Programación:** Se seleccionó **Python** por su facilidad de uso e interoperabilidad con las distintas herramientas que se utilizarán (BI, Orquestador de Workflows, ML, Backends, etc.).
- **Carga de Datos:** Inicialmente, se utilizará un sistema de **carpetas compartidas** como mecanismo provisional para programar la carga de datos a la plataforma, pero luego, será reemplazada por una interfaz que al mismo tiempo de permitir cargar los archivos, integre los estándares necesarios una vez implantados.
- **Visualización:** Se priorizará el desarrollo de una visualización integrada utilizando **PowerBI** antes de considerar una solución de visualización personalizada.
- **Gestión de Logs:** Los logs se mantendrán en **texto plano**, dado que su gestión avanzada no constituye una prioridad dentro del proyecto.

9.2. Herramientas y Frameworks de Software

Para la implementación de este proyecto, se ha seleccionado un conjunto específico de tecnologías de código abierto y librerías que potenciarán las tareas clave de Extracción, Transformación y Carga (ETL), así como el desarrollo del modelo de Machine Learning.

A continuación, se detallan las tecnologías específicas seleccionadas para cada tarea:

Nombre	Uso	Versión
Python	Lenguaje para implementar la mayoría de cosas	>=3.13
FastApi	Para el Backend	0.121.1 (python 3.13)
Scikit-learn	Para el entrenamiento del modelo	1.7.2
XGBoost	Para el entrenamiento del modelo	3.1.1
SvelteKit	Framework robusto para frontends que usan JavaScript. Se utiliza debido a que es un framework común dentro de la empresa.	5
PostgreSQL	Para la Base de Datos Staging y Data Warehouse	18
Apache Airflow	Creación, planificación y seguimiento de flujos de trabajo (ETL).	3.1.0 (python 3.13)
Docker	Para la contenerización del proyecto.	28.4.0
PowerBI	Para la visualización y analítica de los datos.	>=2

La selección de la pila tecnológica para este proyecto se fundamenta en criterios de compatibilidad, escalabilidad, eficiencia y alineación con las prácticas estándar del sector de análisis de datos y desarrollo ágil. A continuación, se esquematiza la justificación de cada componente, destacando sus beneficios clave y razonamiento de elección:

Python (>=3.13): Elegido como lenguaje principal por su versatilidad en procesos ETL, modelado predictivo y scripting de datos. Ofrece una sintaxis clara y un ecosistema rico de bibliotecas open-source que aceleran el desarrollo, minimizando tiempos de implementación.

FastAPI (0.121.1, compatible con Python 3.13): Adoptado para el backend debido a su alto rendimiento en APIs asíncronas. Facilita la integración segura y escalable de módulos como ingesta de datos y operaciones, asegurando eficiencia en entornos de alto tráfico.

Scikit-learn (1.7.2): Integrado para el entrenamiento de modelos de machine learning. Es una herramienta consolidada que permite el procesamiento eficiente de datasets de RRHH, enfocada en algoritmos robustos para predicciones precisas sobre rotación de personal.

XGBoost (3.1.1): Complementa Scikit-learn en el entrenamiento de modelos. Maximiza la precisión predictiva con bajo overhead computacional, ideal para datasets estructurados como los de RRHH, mejorando la capacidad analítica del sistema.

SvelteKit (5): Seleccionado para el frontend por ser un framework ligero basado en JavaScript. Optimiza la renderización reactiva y se alinea con estándares internos de la empresa, promoviendo interfaces intuitivas, mantenibles y de rápida carga.

PostgreSQL (18): Soporta tanto la base de datos staging como el Data Warehouse. Garantiza integridad referencial, soporte para consultas OLAP y escalabilidad horizontal, esencial para manejar volúmenes de datos en entornos analíticos.

Apache Airflow (3.1.0, compatible con Python 3.13): Orquesta los flujos ETL permitiendo su creación utilizando únicamente código. Permite la programación, monitoreo y recuperación automática de workflows complejos, asegurando automatización confiable en la transformación y carga de datos.

Docker (28.4.0): Asegura la contenerización del proyecto. Proporciona despliegue consistente y portable, facilitando la replicabilidad en entornos de desarrollo, pruebas y producción.

Power BI (>=2): Se adoptó Power BI como estándar de visualización por su usabilidad e integración rápida y nativa con el Data Warehouse (DW), no priorizando soluciones personalizadas. Aunque su uso no es universal ni preferido por todos dentro de la empresa, funciona bien y permite la construcción ágil de métricas, KPIs y dashboards interactivos, transformando los datos en conocimiento accionable.

Estas decisiones minimizan dependencias externas, reducen tiempos de implementación y maximizan la reutilización de componentes, alineándose con los objetivos de transformación digital en Casino Luckia.

10. Modelado de Datos

10.1. Fuentes de Datos

Diccionario rotación

Describe los datos relacionados a la inscripción de los empleados dentro de la empresa, con los cuales se realiza la paga de los empleados. Son gestionados por la herramienta SAP interna, y la estructura la organiza en términos de “medidas” (atributos), que se “aplican” (se asignan y se les da un valor) sobre los empleados (ej: medida “cargo_empleado” = “Bartender”). Se estructura similar a un datawarehouse con modelo estrella, almacena la información aislada en infotipos (simil de dimensiones) que se pueden cruzar a discreción según necesidad.

Este archivo guarda registros de los empleados centrando la medida “clase_medida” que corresponde a los cambios ocurridos en el empleo en un punto del tiempo cualquiera. Es entonces que estructura la trayectoria del empleado en términos de intervalos en donde la medida aplicada es válida.

Cabe aclarar que, aunque el sistema SAP gestiona la nómina completa, esta fuente corresponde a una vista acotada de los datos enfocada en el análisis de la rotación y no en el proceso de pagos en sí. Para asegurar la integridad en el cruce de información, cada registro cuenta con identificadores únicos, específicamente el **ID SAP** y el **RUT** del empleado, permitiendo la trazabilidad histórica de sus movimientos.

GeoVictoria día y semana

Describe la asistencia registrada por el sistema GeoVictoria dentro del sistema. Es una planilla sencilla que almacena los turnos planificados día a día para cada empleado y la asistencia efectiva del mismo. Guarda información sobre el turno asignado, los permisos aplicados si es que existen, la referencia al empleado y las marcas de entrada y salida registradas. Incluye información agregada en una planilla adicional llamada “Semana”, que agrupa según el empleado la suma de horas trabajadas.

En cuanto al formato, tanto las marcas de tiempo como las duraciones se presentan en formato HH:MM.

Resumen anual de capacitación y Participantes de las capacitaciones

Esta fuente integra los datos de las capacitaciones realizadas dentro de la empresa. Incluye una vista por empleado y por capacitación, facilitando la vista agregada de la información. La vista más agregada (vista de resumen anual) guarda información sobre las fechas de capacitación, su tipo, el proveedor, valoraciones, entre otros. La vista por empleado (participantes) guarda información sobre los empleados que participan en dichas capacitaciones, desagregando algunas métricas.

Proceso de selección (Contratación)

Este archivo documenta los procesos de selección gestionados dentro de la empresa. Organiza la información de manera incremental, es decir, los datos se van actualizando conforme ocurran cambios en los procesos. Guarda información sobre las fechas relevantes del proceso, el cargo que se está buscando reemplazar, detalles de la publicación del proceso, detalles sobre la persona seleccionada y otras métricas como número de entrevistas realizadas y currículums recopilados.

Dentro de las variables capturadas se incluye la 'fuente de reclutamiento', que permite identificar el origen de la postulación. Es importante notar que el registro almacena datos exclusivamente sobre el candidato seleccionado; no se guarda información ni motivos de rechazo sobre los postulantes que no cerraron el proceso.

Perfiles de trabajo

Más que un solo archivo, es un directorio en el que se guardan los descriptores de funciones de trabajo (DFTs) de cada cargo de la empresa. Guarda información sobre la identificación del cargo, categorización, misión y responsabilidades, tareas a cumplir y las competencias y habilidades requeridas para ejercer el cargo.

Estos archivos se encuentran en formato Word (.docx) y no en planillas Excel como el resto de las fuentes. Dado que no existe una tabla maestra con IDs de cargo, la vinculación con las demás fuentes se realiza mediante el cruce textual del nombre completo del cargo, el cual se encuentra estandarizado en los registros de SAP.

10.2. Base de datos Staging

La base de datos staging es una BD creada con la finalidad de ser temporal, su finalidad es almacenar todos los datos que serán posteriormente procesados y enviados al destino final. En este sentido, se identificaron las siguientes tablas:

stg_rotacion_empleados		
pk	id_empleado	Integer
	nombre	Varchar
	id_empresa	Integer
	empresa	Varchar
	tipo_empleo	Varchar
	desde1	Date
	hasta1	Date
	area	Varchar
	cargo	Varchar
	jornada	Varchar
	ant_puesto	Date
	ceco	Varchar
	fecha_nacimiento	Date
	edad	Integer
	pais_nacimiento	Varchar
	lugar_nacimiento	Varchar
	nacionalidad	Varchar
	estado_civil	Varchar
	nro_hijos	Integer
	sexo	Varchar
	desde2	Date
	hasta2	Date
	clase_fecha	Date
	fecha	Varchar
	clase_prestamo	Varchar
	movilidad_geografica	Varchar
	experiencia_profesional	Varchar
	desde3	Date
	hasta3	Date
	clase_medida	Varchar
	motivo_medida	Varchar
	alta	Date
	baja	Date
	encargado_superior	Varchar

stg_resumen_anual_capacitaciones		
pk	nro_capacitacion	Integer
	mes	Varchar
	titulo	Varchar
	lugar	Varchar
	fecha_inicio	Date
	fecha_fin	Date
	objetivo_area	Varchar
	externo_interno	Varchar
	tipo_curso	Varchar
	gerencia	Varchar
	formador_proveedor	Varchar
	nro_asistentes	Integer
	horas_ppersona	Integer
	total_horas	Integer
	coste	Decimal
	valoracion_formador	Decimal
	indice_satisfaccion	Decimal
	nps	Integer

stg_proceso_seleccion		
pk	id_proceso	Integer
	fecha_cierre	Date
	cargo	Varchar
	ceco	Varchar
	detalle_puesto	Varchar
	detalle_situacion	Text
	gerencia	Varchar
	linea_negocio	Varchar
	grupo	Varchar
	motivo	Varchar
	detalles_motivo	Text
	duracion_dias	Integer
	fuentes_reclutamiento	Varchar
	nombre	Varchar
	sexo	Varchar
	edad	Integer
	formacion	Varchar
	anos_experiencia	Integer
	sector_procedencia	Varchar
	tiene_continuidad	Boolean
	nro_cvs_recibidos	Integer
	nro_personas_entrevistadas_telefono	Integer
	nro_personas_entrevistadas_presencial	Integer
	nro_personas_finalistas	Integer

stg_perfiles_trabajo		
pk	id_perfil	Integer
	puesto	Varchar
	categoria_puesto	Varchar
	nombre_puesto	Varchar
	fecha_creacion	Date
	linea_asc	Varchar
	linea_desc	Varchar
	misión	Text
	responsabilidades	Text
	inform_y_reporta	Varchar
	tareas_semanales	Text
	tareas_mensuales	Text
	tareas_trimestrales	Text
	tareas_semestrales	Text
	tareas_anuales	Text
	otras_tareas	Text
	competencias	Text
	habilidades_generales	Text
	habilidades_cognitivas	Text
	habilidades_fisicas	Text
	habilidades_sensoriales	Text
	condiciones_trabajo	Text
	apartado_legal	Text

stg_capacitaciones_participantes		
pk	nro_capacitacion	Integer
	mes	Varchar
	rut	Varchar
	id_empleado	Integer
	nombre	Varchar
	apellidos	Varchar
	correo	Varchar
	titulo	Varchar
	horas_ppersona	Integer

stg_asistencia_diaria_geovictoria		
pk	id_asistencia	Integer
	grupo	Varchar
	asistio_en	Date
	tipo_permiso	Varchar
	tipo_turno	Date
	hora_ingreso	Date
	atraso	Date
	hora_salida	Date
	adelanto	Date
	total_horas	Date
	id_empleado	Integer

stg_asistencia_semanal_geovictoria		
pk	id_asistencia	Integer
	grupo	Varchar
	inicio_semana	Date
	fin_semana	Date
	horas_trabajadas	Date
	horas_extra	Date
	horas_no_trabajadas	Date
	id_empleado	Integer

10.2.1. Data Warehouse

Objetivo del Data Warehouse

El objetivo del Data Warehouse (DW) es doble: alimentar los análisis estadísticos (BI) y el entrenamiento de los modelos de predicción (Machine Learning).

Aunque el diseño está impulsado por los datos (data-driven), identificando los procesos de negocio desde las fuentes, se implementará la **arquitectura de bus de Kimball**. Se prioriza el desarrollo de un modelo estrella por ser efectivo para ambas tareas (analítica y ML).

Los pilares del diseño son la **simplicidad, la reutilización de datos y la conexión** entre los distintos data marts mediante conformed dimensions (dimensiones conformadas). Esto facilitará el análisis y documentará el ciclo de vida del colaborador, enriqueciendo el contexto para los modelos de ML.

Decisiones de diseño

BI vs ML (Data Lake VS Data Marts)

Los objetivos de analítica (BI) y entrenamiento (ML) requieren estructuras de datos distintas: BI tiende a usar modelos estrella, mientras que ML prefiere una "tabla ancha" o *flat table*.

Ambos procesos exigen localizar, limpiar, estandarizar e integrar datos. Sin embargo, construir un Data Lake primero duplicaría este esfuerzo de ETL.

Por ello, la estrategia es:

1. Priorizar la construcción de los Data Marts (modelo estrella).
2. Generar las tablas anchas para ML a demanda, usando vistas (normales o materializadas) que consulten los Data Marts.

Este enfoque mantiene una única fuente de verdad (los Data Marts) y reduce los riesgos de redundancia.

Conformed Dimensions

Dada la alta interrelación de los datos de Recursos Humanos, el modelo usará dimensiones conformadas (compartidas entre data marts) para asegurar la eficiencia y la consistencia del análisis. Las dimensiones conformadas planificadas son:

- DimEmpleado (La persona que cumple un rol dentro de la empresa)
- DimTiempo (El tiempo en el que ocurren los sucesos)
- DimCargo (La descripción del rol que cumple el empleado)

Formulación de las Tablas de Hechos

Se identificaron las siguientes tablas de hechos necesarias:

1. Rotación de Empleados (Transaccional)

Esta tabla modela los eventos clave que ocurren durante el empleo de un colaborador dentro de la empresa. La estructura de los datos, manejada por SAP, presenta cualidades transaccionales, lo que significa que se registra una fila por cada cambio en el estado de empleo de una persona (ej. contratación, baja, promoción, cambio de puesto).

Fundamentalmente, esta tabla proporciona el **numerador** necesario para todas las métricas de rotación y está diseñada para responder preguntas basadas en eventos.

Ejemplos de preguntas:

- *"¿Cuántas 'Bajas' (clase_medida) ocurrieron en el Q1?"*
- *"¿Cuáles son los principales motivo_medida de baja en el area de 'CAJA'?"*
- *"¿Cuál es la tasa de promoción (clase_medida = 'Cambio de puesto') por area o gerencia?"*
- *"¿Qué motivo_medida (ej. 'Baja voluntaria') es más común entre los empleados con menos de un año de antigüedad (calculado desde DimEmpleado)?"*
- *"¿Hay estacionalidad en las 'Contrataciones' o 'Bajas'?"*

2. (De apoyo) Headcount mensual (Snapshot Periodico)

La tabla transaccional de rotación no permite un cálculo fácil de métricas como la "tasa de rotación", ya que estas necesitan un **denominador** (la cantidad de empleados en un momento específico). El cálculo de este denominador a partir de los eventos transaccionales es costoso a largo plazo y engorroso para los desarrolladores.

Por ello, se incorpora esta tabla de apoyo. Su función es guardar el antecedente de la dotación de empleados mensual a través de un **snapshot periódico**. Esto ayuda a cruzar la información y delega el esfuerzo de cómputo al proceso ETL, lo cual es ideal a largo plazo. Responde preguntas basadas en el estado o balance en un punto en el tiempo.

Ejemplos de preguntas:

- *"¿Cuál era nuestro headcount total en 'BOVEDA' al 31-Ene?"*
- *"¿Cuál es la antigüedad promedio por area?"*
- *"¿Cómo ha evolucionado nuestro headcount (Headcount) mes a mes durante el último año?"*
- *"¿Cuál es la distribución de la antigüedad (Antigüedad_Meses) por cargo?"*
- *"¿Cuál es el headcount por tipo_empleo (ej. 'Empleado' vs 'Temporales') al final de cada trimestre?"*

3. Asistencia Diaria (Transaccional)

Aquí se busca modelar la asistencia efectiva de cada día para identificar patrones en los comportamientos de asistencia. Si bien se usará primariamente para *machine learning* (para crear *features*), también tiene un alto interés analítico.

Esta tabla es la base para todos los análisis de ausentismo crónico, puntualidad y la comparación entre el total de horas trabajadas versus las programadas.

Ejemplos de preguntas:

- "¿Qué empleados (*DimEmpleado*) tienen el mayor promedio de atraso en el último mes?"
- "¿Existe una correlación entre el *tipo_turno* y el *total_horas* trabajadas?"
- "¿Cuáles son los días de la semana con mayor ausentismo (conteo de filas donde *total_horas* = 0 o *tipo_permiso* != 'Ninguno')?"

4. Participación en Capacitación (Factless / Transactional)

Esta tabla modela la vinculación individual de cada empleado con las instancias formativas, registrando quién asistió a qué curso y en qué momento. Su granularidad es de una fila por empleado por curso, permitiendo cruces directos con ***DimEmpleado***, ***DimTiempo*** y ***DimCurso***. El fin es analizar la cobertura, el cumplimiento de planes de formación y la carga horaria por persona o departamento. Responde preguntas sobre la asistencia, las horas invertidas por talento y la penetración de la capacitación en la estructura organizacional. **Ejemplos de preguntas:**

- "¿Cuántas horas_ppersona de capacitación acumuló el área de 'AA&BB BARES' este trimestre?"
- "¿Qué empleados no registran ninguna participación en cursos obligatorios?"
- "¿Cuál es el promedio de cursos tomados por los empleados con cargo 'Bartender'?"
- "¿Cuál es la tasa de deserción (inscritos vs. asistentes) por gerencia?"

5. Realización de Capacitación (Snapshot / Transaccional)

Esta tabla modela la ejecución de los eventos de capacitación desde la perspectiva logística y financiera, independientemente de quiénes asistieron específicamente. Consolida los datos de la instancia formativa, conectando con ***DimProveedor***, ***DimCurso*** y ***DimTiempo***. El fin es controlar la inversión, la gestión de proveedores y la oferta formativa disponible. Responde preguntas sobre el gasto, la calidad del proveedor y la distribución de la oferta de cursos. **Ejemplos de preguntas:**

- "¿Cuál fue el *costo_total* invertido en cursos con *objetivo_area* = 'Liderazgo'?"
- "¿Qué formador_proveedor tiene el *indice_satisfaccion* promedio más alto en sus realizaciones?"
- "¿Cuántas realizaciones de tipo 'Presencial' vs 'Online' se gestionaron el último año?"
- "¿Cuál es el costo promedio por hora de instrucción (*valor_hora*) de los proveedores externos?"
- "¿Existe correlación entre el *costo_curso* y la *valoracion_formador* obtenida?"

6. Proceso de Selección (Snapshot Acumulativo)

Esta tabla modela los procesos de selección de principio a fin. Está diseñada para analizar la calidad y eficiencia de dichos procesos e integrarlos con el ciclo de vida que terminan teniendo las personas en la empresa.

Utiliza un diseño de **snapshot acumulativo** para seguir el *pipeline* del proceso. Responde preguntas de conversión y, lo más importante, sobre la **Calidad de Contratación** (Quality of Hire), lo cual se logra al cruzar esta tabla con otros data marts (como Rotación o Asistencia).

Ejemplos de preguntas:

- "¿Cuál es nuestra *duracion_dias promedio* (Tiempo para Contratar) por gerencia?"
- "¿Cuál es nuestra tasa de conversión de *nro_cvs_recibidos* a *nro_personas_finalistas*?"
- "¿Cuánto nos cuesta (*coste_proceso*) contratar para un cargo específico?"
- "¿Qué *fuentes_reclutamiento* (ej. 'Feria Laboral') genera más contrataciones que *tiene_continuidad* (más de 4 meses)?"
- "¿Los empleados contratados a través de procesos con una *duracion_dias* muy larga (>60 días) tienen más probabilidad de baja (uniendo con *FactRotacionTransaccional*) en sus primeros 6 meses?"
- "¿Las personas que contratamos de la 'Feria Laboral' (*fuentes_reclutamiento*) tienen más atraso (uniendo con *FactAsistenciaDiaria*) que las que vienen por 'Recomendación'?"
- "¿Cuál es el *coste_proceso promedio* de los empleados que *SÍ tiene_continuidad* vs. los que no?"

Dimensiones y Tablas de Hechos

Se contemplan estas dimensiones para el modelado multidimensional:

Dimensiones Conformadas

Dimensión Empleado

Modela el ciclo de vida de la persona dentro de la empresa. Captura el estado del empleado en un determinado punto del tiempo, ya que atributos como su cargo o encargado pueden cambiar con el tiempo.

Su granularidad es: “Una fila por versión histórica de una persona (candidato o empleado)”.

Se implementa como una SCD 2 para conservar la historia del trayecto. Se integra en las tablas de hechos tanto de rotación como de selección, para integrar el proceso completo y analizar el ciclo de vida del empleado.

Nombre de Columna	Tipo de Dato	Descripcion
Empleado_SK	INTEGER	Clave primaria sustituta (generada por ETL).
Empleado_ID_NK	VARCHAR	ID SAP / ID del empleado (clave natural).
Rut	VARCHAR	RUT del participante (clave natural).
Nombre_Completo	VARCHAR	Nombre completo del empleado o candidato.
Sexo	VARCHAR	Sexo (masculino/femenino).
Fecha_Nacimiento	DATE	Fecha de nacimiento del empleado.
Nacionalidad	VARCHAR	Nacionalidad del empleado.
Estado_Civil	VARCHAR	Estado civil del empleado.
Formación	VARCHAR	Nivel educacional del candidato seleccionado/empleado.
Estado_Laboral_Activo	BOOLEAN	Si el empleado está ejerciendo sus labores actualmente en la empresa.
SCD_Fecha_Inicio_Vigencia	DATE	Fecha de inicio de vigencia del registro.
SCD_Fecha_Fin_Vigencia	DATE	Fecha de fin de vigencia del registro.
SCD_Es_Actual	BOOLEAN	Indicador de registro actual (vigente).

Dimensión Tiempo

Dimensión estática que sirve como eje temporal maestro para todo el modelo. SCD 0.

Nombre de Columna	Tipo de Dato	Descripcion
Tiempo_SK	INTEGER	Clave primaria (Formato AAAAMMDD).
Fecha	DATE	Fecha completa (ej. 2024-05-15).
Anio	SMALLINT	Año (ej. 2024).
Mes_Nombre	VARCHAR	Nombre del mes (ej. 'Mayo').
Mes_Numero	SMALLINT	Número del mes en el año (ej. 5).
Trimestre	VARCHAR	Nombre del trimestre (ej. 'Q2-2024').
Dia_Semana	VARCHAR	Nombre del día (ej. 'Miércoles').
Es_Fin_De_Semana	BOOLEAN	Indica si es fin de semana.
Es_Feriado	BOOLEAN	Indica si es feriado.
Semana_Anio	SMALLINT	Número de la semana en el año (ej. 20).

Dimensión Cargo

Catálogo maestro de puestos de trabajo, agrupados por familias y áreas funcionales. SCD 1.

Nombre de Columna	Tipo de Dato	Descripcion
Cargo_SK	INTEGER	Clave primaria sustituta.
Nombre_Cargo	VARCHAR	Clave natural (Nombre del puesto).
Familia_Puesto	VARCHAR	Indica la clasificación familiar del puesto de trabajo.
Area_Funcional	VARCHAR	Indica la área funcional de la empresa relacionada con el trabajo

Dimensión Empresa

Identifica la entidad legal o razón social a la que está vinculado el empleado. SCD 1.

Nombre de Columna	Tipo de Dato	Descripcion
Empresa_SK	INTEGER	Clave primaria sustituta.
Codigo	VARCHAR	Código interno de la empresa.
Nombre	VARCHAR	Nombre de la empresa.

Dimensión Gerencia

Estructura organizacional de alto nivel (geográfica o funcional). SCD 1.

Nombre de Columna	Tipo de Dato	Descripcion
Gerencia_SK	INTEGER	Clave primaria sustituta.
Nombre_Gerencia	VARCHAR	Gerencia organizacional actual o histórica.

Dimensión Centro Costo

Unidad financiera para la imputación de gastos y presupuestos. SCD 1.

Nombre de Columna	Tipo de Dato	Descripcion
Ceco_SK	INTEGER	Clave primaria sustituta.
Nombre_Ceco	VARCHAR	Centro de Coste actual o histórico.

Dimensión Modalidad Contrato

Combina los atributos de vinculación legal (Indefinido, Plazo Fijo) y el régimen horario (Full-time, Part-time). Se modela como una Junk Dimension para evitar la proliferación de tablas pequeñas, agrupando combinaciones de atributos relacionados de baja cardinalidad. SCD 1.

Nombre de Columna	Tipo de Dato	Descripcion
Modalidad_SK	INTEGER	Clave primaria sustituta.
Tipo_Vinculo_Legal	VARCHAR	Tipo de empleado (Temporal, Empleado, etc.).
Regimen_Horario	VARCHAR	Jornada laboral (Part time, Full time, etc.).
FTE_Estándar	DECIMAL	Valor teórico por defecto según el régimen (ej. 1.0 Full, 0.5 Part).
Horas_Capacidad_Mensual_Estándar	INTEGER	Horas base teóricas definidas por regla de la modalidad contractual.

Área de Selección

Hechos de Selección

Granularidad: Una fila por proceso de selección.

Tipo de Hecho: Snapshot Acumulativo. Una única fila que se actualiza a medida que el proceso avanza.

Métricas:

- **Eficiencia:** *Duración del Proceso (Time to Fill) y Costo del Proceso.*
- **Calidad (Quality of Hire):** *Tiene Continuidad*, una métrica diferida que indica si el candidato seleccionado superó los 4 meses de permanencia, crucial para validar el éxito de la selección.

Existen dimensiones degeneradas (en café) debido a la forma que toma el proceso de la creación de la fuente de datos (varios de ellos no están disponibles durante la creación del registro), además del tipo de hecho que se está registrando. Apoya también al hecho de que los datos no son confiables y de que carecen de un estándar riguroso.

Nombre de Columna	Tipo de Dato	Descripcion
Seleccion_ID	INTEGER	Clave primaria de la tabla.
Fecha_Apertura_SK	INTEGER	Clave foránea a DimTiempo (fecha inicio).
Fecha_Cierre_SK	INTEGER	Clave foránea a DimTiempo (fecha fin/cierre).
Cargo_Solicitado_SK	INTEGER	Clave foránea a DimCargo (cargo solicitado).
Empleado_Seleccionado_SK	INTEGER	Clave foránea a DimEmpleado (poblada al final).
Gerencia_Solicitante_SK	INTEGER	FK a la gerencia que pidió la vacante.
ID_Solicitud_NK	VARCHAR	ID único del proceso en el sistema origen.
Fuente_Reclutamiento	VARCHAR	Canal de origen del candidato (LinkedIn, Portal, Referido).
Recruiter_Responsable	VARCHAR	Nombre del analista de selección a cargo.
Estado_Final_Proceso	VARCHAR	Resultado del cierre (Cubierta, Cancelada, Desierta).
Duracion_Proceso_Dias	INTEGER	Días transcurridos entre apertura y cierre.
Costo_Proceso	DECIMAL	Gasto total directo (publicaciones, agencias, tests).
Cantidad_Candidatos	INTEGER	Número total de postulantes recibidos.
Cantidad_Entrevistas	INTEGER	Número total de entrevistas realizadas.
Tiene_Continuidad	INTEGER	Si el contratado superó los 4 meses (Quality of Hire).
Motivo_Fallo_Continuidad	VARCHAR	Causa de la baja si ocurrió antes de los 4 meses.

Área de Asistencia

Dimensión Turno

Define los esquemas horarios teóricos, estableciendo las horas de entrada, salida y descansos esperados. SCD 1.

Nombre de Columna	Tipo de Dato	Descripción
Turno_SK	INTEGER	Clave primaria sustituta.
Nombre_Turno	VARCHAR	Nombre del turno en formato condensado.
Hora_Inicio_Teorica	TIME	Hora de inicio del turno si fue planificado.
Hora_Fin_Teorica	TIME	Hora de fin del turno si fue planificado.
Descanso_Min	INTEGER	Minutos de descanso planificados para el turno.
Horas_Jornada	DECIMAL	Duración calculada del turno.

Dimensión Permiso

Clasifica los motivos de ausencia o justificación de asistencia (Vacaciones, Licencias, Fallas), permitiendo distinguir entre ausentismo planificado y no planificado. SCD 1.

Nombre de Columna	Tipo de Dato	Descripción
Permiso_SK	INTEGER	Clave primaria sustituta.
Codigo_Permiso_NK	VARCHAR	Código interno del permiso aplicado.
Descripcion	VARCHAR	Tipo del permiso utilizado en la justificación.
Categoria_Analitica	VARCHAR	Categoría derivada del tipo de permiso aplicado, para permitir un análisis más profundo.

Hechos de Asistencia

Granularidad: Una fila por empleado, por día.

Tipo de Hecho: Transaccional. Registra el evento diario de cumplimiento de jornada.

Métricas:

- **Aditivas:** *Horas Trabajadas* para cálculos de productividad.
- **Desviaciones:** *Minutos de Atraso* y *Salida Anticipada*, calculados comparando las marcas reales contra la definición teórica de **DimTurno**. Permite el análisis granular de puntualidad y ausentismo.

Nombre de Columna	Tipo de Dato	Descripcion
Asistencia_ID	INTEGER	Clave primaria de la tabla.
Fecha_SK	INTEGER	Clave foránea a DimTiempo (día de asistencia).
Permiso_SK	INTEGER	Clave foránea a DimPermiso.
Empleado_SK	INTEGER	Clave foránea a DimEmpleado (versión vigente).
Turno_Planificado_SK	INTEGER	Clave foránea a DimTurno.
Hora_Entrada_Real	DATETIME	Hora de entrada real registrada.
Hora_Salida_Real	DATETIME	Hora de salida real registrada.
Horas_Trabajadas	DECIMAL	Medida de horas totales trabajadas.
Minutos_Atraso	INTEGER	Medida de minutos de atraso.
Minutos_Adelanto_Salida	INTEGER	Medida de minutos de salida temprana.
Tolerancia_Aplicada_Min	INTEGER	Tolerancia aplicada para evaluar si es atraso o no.
Permiso_Aplicado	STRING	Indica si se aplicó un permiso o no.
Es_Atraso	INTEGER	Indica si el registro contempla un atraso o no.
Es_Salida_Anticipada	INTEGER	Indica si el registro contempla una salida anticipada o no.
Es_Ausencia	INTEGER	Indica si el registro contempla una ausencia o no.

Área de Rotación y Dotación

Dimensión Medida Aplicada

Catálogo de acciones de personal (Altas, Bajas, Transferencias) que explican la naturaleza de un movimiento en la tabla de rotación. SCD 1.

Nombre de Columna	Tipo de Dato	Descripcion
Medida_SK	INTEGER	Clave primaria sustituta.
Tipo_Movimiento	VARCHAR	Valor de la clase de medida (ej. 'Baja', 'Contratación').
Razon_Detallada	VARCHAR	Valor del motivo de medida (ej. 'Renuncia', 'Cambio').
Es_Voluntario	BOOLEAN	Indica si el cambio realizado corresponde a un cambio negociado/acordado con el empleador o no.

Hechos de Rotación

Granularidad: Una fila por evento de movimiento de personal (Alta, Baja, Cambio de puesto).

Tipo de Hecho: Transaccional (Eventos).

Métricas:

- **Flujo:** *Variación Headcount* (+1 para entradas, -1 para salidas, 0 para cambios internos). Es la base para calcular índices de rotación.
- **Salarial:** *Sueldo Base Intervalo*, captura la remuneración en el momento exacto del movimiento.

Nombre de Columna	Tipo de Dato	Descripcion
Rotacion_ID	INTEGER	Clave primaria de la tabla.
Fecha_Inicio_Vigencia_SK	INTEGER	Clave foránea a DimTiempo (fecha evento).
Fecha_Fin_Vigencia_SK	INTEGER	Clave foránea a DimTiempo.
Modalidad_SK	INTEGER	Clave foránea a DimModalidadContrato.
Empleado_SK	INTEGER	Clave foránea a DimEmpleado.
Empresa_SK	INTEGER	Clave foránea a DimEmpresa.
Ceco_SK	INTEGER	Clave foránea a DimCentroCosto.
Cargo_SK	INTEGER	Clave foránea a DimCargo.
Medida_SK	INTEGER	Clave foránea a DimMedidaAplicada.
Sueldo_Base_Intervalo	DECIMAL	Sueldo base de la persona durante el intervalo medido.
Variacion_Headcount	INTEGER	Variación producida en el headcount (-1,0,+1).

Hechos de Dotación

Granularidad: Una fila por empleado activo, al último día de cada mes.

Tipo de Hecho: Snapshot Periódico. "Foto fija" del estado de la plantilla.

Métricas:

- **Stock:** *Headcount* y *Antigüedad en Meses*.
- **Capacidad:** *FTE Real* (Full Time Equivalent), métrica semi-aditiva vital para normalizar la fuerza laboral considerando jornadas parciales (ej. 2 personas de media jornada = 1.0 FTE). Fundamental para reportes de evolución histórica de plantilla sin tener que reconstruir estados desde los eventos de rotación.

Nombre de Columna	Tipo de Dato	Descripción
Mes_Cierre_SK	INTEGER	Clave foránea a DimTiempo (último día del mes).
Empleado_SK	INTEGER	Clave foránea a DimEmpleado (versión vigente).
Cargo_SK	INTEGER	Clave foránea a DimCargo.
Empresa_SK	INTEGER	Clave foránea a DimEmpresa.
Modalidad_SK	INTEGER	Clave foránea a DimModalidadContrato.
Headcount	INTEGER	Medida (siempre 1) para conteo de personal.
FTE_Real	DECIMAL	Equivalente full-time del cargo real aplicado.
Horas_Capacidad_Mensual	INTEGER	Total de horas teóricas que el empleado debe cumplir en el mes según su jornada contractual y calendario.
Sueldo_Base_Mensual	DECIMAL	Remuneración bruta fija vigente para el empleado al cierre del mes (excluye bonos o variables).
Antigüedad_Meses	INTEGER	Medida de antigüedad en meses a la fecha del snapshot.

Área de Capacitación

Dimensión Curso

Catálogo de la oferta formativa, incluyendo temáticas y modalidades. SCD 1.

Nombre de Columna	Tipo de Dato	Descripción
Curso_SK	INTEGER	Clave primaria sustituta.
Nombre_Curso	VARCHAR	Nombre oficial del curso.
Categoria_Tematica	VARCHAR	Categoría principal (Negocio, Cliente, etc.).
Modalidad	VARCHAR	Modalidad (Presencial, Online, etc.).

Dimensión Proveedor

Registro de entidades o instructores (internos/externos) que imparten la capacitación. SCD 1.

Nombre de Columna	Tipo de Dato	Descripción
Proveedor_SK	INTEGER	Clave primaria sustituta.
Nombre_Proveedor	VARCHAR	Nombre del formador o proveedor.
Tipo_Proveedor	VARCHAR	Origen del formador (Interno/Externo).

Dimensión Lugar

Modalidad y Ubicación geográfica específica donde ocurren los eventos de capacitación. SCD 1.

Nombre de Columna	Tipo de Dato	Descripción
Lugar_SK	INTEGER	Clave primaria sustituta.
Lugar	STRING	Modalidad de la capacitación (PRESENCIAL, ONLINE, VIRTUAL)
Región	STRING	Región geográfica de la capacitación.

Hechos de Realización de Capacitaciones

Granularidad: Una fila por edición o evento de un curso realizado (Oferta).

Tipo de Hecho: Transaccional.

Métricas:

- **Gestión:** Costo Total y Duración.
- **Calidad Agregada:** NPS y Satisfacción Promedio. Estas métricas resumen el feedback de todos los asistentes al evento, permitiendo evaluar la calidad del proveedor o instructor.

Nombre de Columna	Tipo de Dato	Descripción
Realizacion_ID	INTEGER	Clave primaria de la tabla.
Fecha_Inicio_SK	INTEGER	Clave foránea a DimTiempo (fecha inicio).
Fecha_Fin_SK	INTEGER	Clave foránea a DimTiempo (fecha fin).
Curso_SK	INTEGER	Clave foránea a DimCurso.
Proveedor_SK	INTEGER	Clave foránea a DimProveedor.
Gerencia_Organizadora_SK	INTEGER	Gerencia que organiza o recibe.
Lugar_Realización_SK	INTEGER	Clave foránea a DimLugarRealización.
Costo_Total_Curso	DECIMAL	Costo total del curso.
Duracion_Horas_Total	INTEGER	Horas por persona * Número total de asistentes.
Dias_Extension	INTEGER	Extensión en días de la capacitación.
Total_Asistentes	INTEGER	Número de participantes totales.
NPS_Score	INTEGER	NPS de la capacitación.
Satisfaccion_Promedio	DECIMAL	Satisfacción registrada promedio de los asistentes.
Valoracion_Formador	DECIMAL	Calificación otorgada al formador.

Hechos de Participación en Capacitaciones

Granularidad: Una fila por empleado, por curso (Demanda).

Tipo de Hecho: Transaccional.

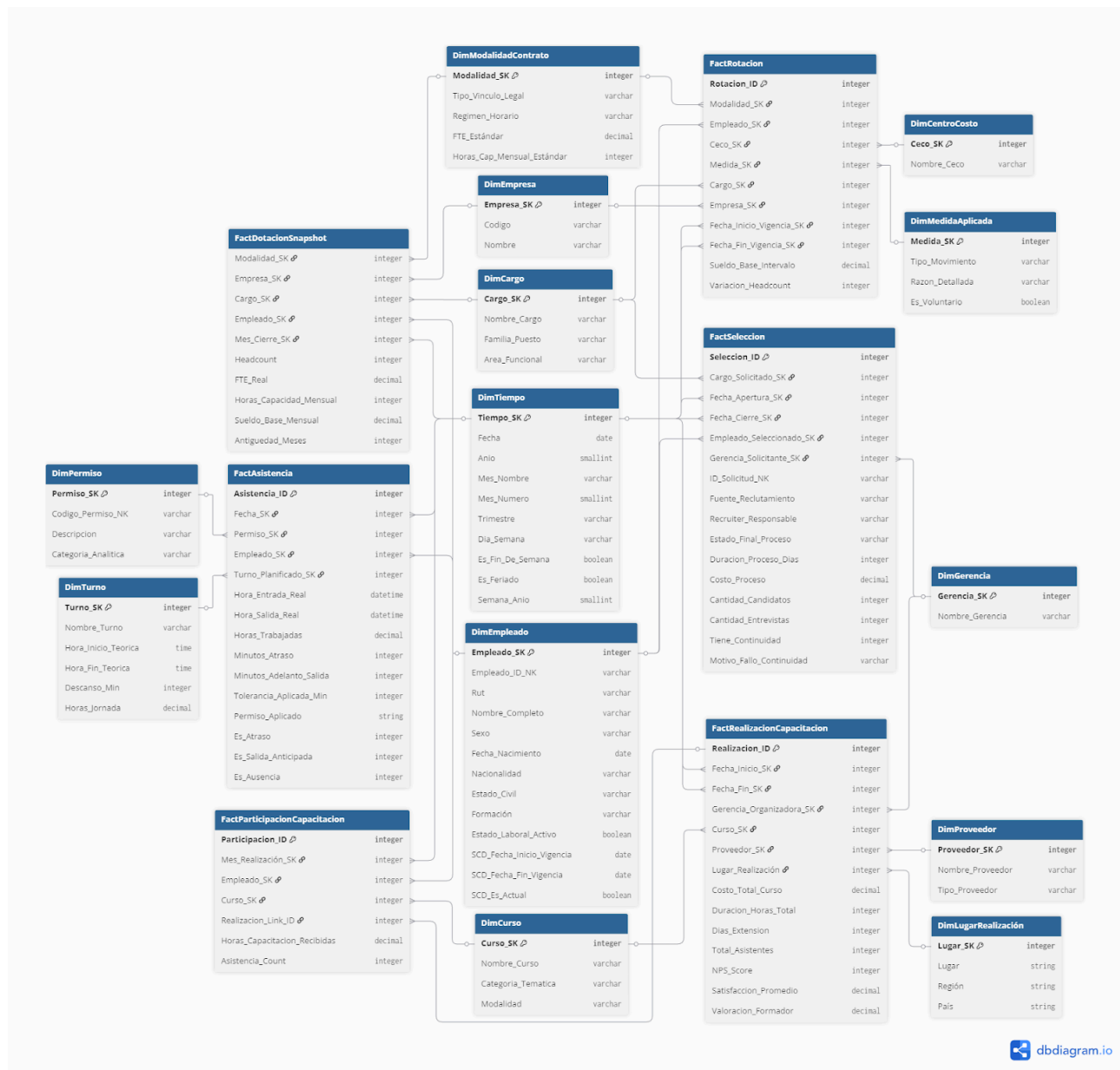
Métricas:

- **Consumo:** Horas de Capacitación Recibidas. Permite analizar la cobertura de formación por empleado o área (quién se capacitó y cuánto), vinculándose a la tabla de realización para obtener costos prorrateados si fuera necesario.

Nombre de Columna	Tipo de Dato	Descripcion
Participacion_ID	INTEGER	Clave primaria de la tabla.
Mes_Realización_SK	INTEGER	Clave foránea a DimTiempo.
Empleado_SK	INTEGER	Clave foránea a DimEmpleado (asistente).
Curso_SK	INTEGER	Clave foránea a DimCurso.
Realizacion_Link_ID	INTEGER	Clave foránea a FactRealizacionCapacitacion.
Horas_Capacitacion_Recibidas	DECIMAL	Medida de horas de capacitación recibidas.
Asistencia_Count	INTEGER	Cuenta fija (=1) para facilitar cálculos.

Modelo Multidimensional Graficado

Esta segunda versión del modelo multidimensional representa una evolución significativa respecto al diseño anterior, consolidando una arquitectura en esquema de estrella (Star Schema) optimizada para el rendimiento analítico y la eficiencia de las consultas. La estructura se articula en torno a **6 tablas de hechos y 13 dimensiones**, permitiendo una granularidad más fina y la incorporación de nuevas medidas críticas que no existían previamente, tales como indicadores de Quality of Hire (continuidad), desviaciones exactas de asistencia y de capacidad de la dotación (FTE real, etc.). Este diseño robustece las capacidades del proyecto, abarcando integralmente desde la adquisición de talento hasta el desarrollo y retención, proporcionando una base sólida para la inteligencia de negocios.



11. Procesamiento ETL

Para garantizar que los modelos de Machine Learning cuenten con datos íntegros, actualizados y de alta calidad, se ha implementado una capa de procesamiento basada en el paradigma de **Orquestación de Flujos de Datos**. Esta sección detalla el uso de **Apache Airflow** como el núcleo que gobierna el ciclo de vida de los datos, desde su ingesta en bruto hasta su transformación en tablas de hechos y dimensiones.

11.1. Apache Airflow: El Motor de Orquestación

Apache Airflow es una plataforma de código abierto utilizada para programar y monitorear flujos de trabajo (workflows). En el contexto de esta aplicación de RRHH, Airflow actúa como el "director de orquesta", asegurando que cada paso del proceso ETL (Extract, Transform, Load) se ejecute en el orden correcto y gestione los errores de manera automática.

Conceptos Clave

DAG (Directed Acyclic Graph)	Es la pieza central de Airflow. Un DAG es una colección de todas las tareas que se desean ejecutar, organizadas de manera que reflejen sus relaciones y dependencias. Al ser "acíclico", garantiza que el flujo de datos no entre en bucles infinitos.
Tasks (Tareas)	Representan una unidad de trabajo definida (por ejemplo, ejecutar un script SQL o cargar un archivo Excel).
Operators (Operadores)	Son las plantillas que definen qué hace realmente una tarea. En este proyecto, se utilizan principalmente para ejecutar procedimientos en PostgreSQL y lógica de negocio en Python.

11.2. Estrategia y Arquitectura de ETL

La arquitectura se ha diseñado bajo un enfoque de **desacoplamiento funcional**, separando la lógica de orquestación de la lógica de transformación de datos.

Pilares de la implementación:

1. **Modularización de SQL:** Las transformaciones complejas residen en archivos .sql o en scripts de código independientes. Esto permite que los analistas de datos puedan modificar la lógica de negocio sin alterar el código de los DAGs.
2. **Arquitectura de Tareas (Tasks):** Se utiliza una capa intermedia de scripts de Python (/tasks) que actúan como "wrappers". Estas tareas invocan los scripts SQL, permitiendo la reutilización de código y facilitando las pruebas unitarias.
3. **Infraestructura Contenerizada:** Todo el ecosistema de Airflow (Webserver, Scheduler, DAG Processor y Base de Datos) se despliega mediante **Docker**, de manera aislada pero manteniendo la conectividad entre contenedores.

11.3. Descripción de los Flujos de Trabajo (DAGs)

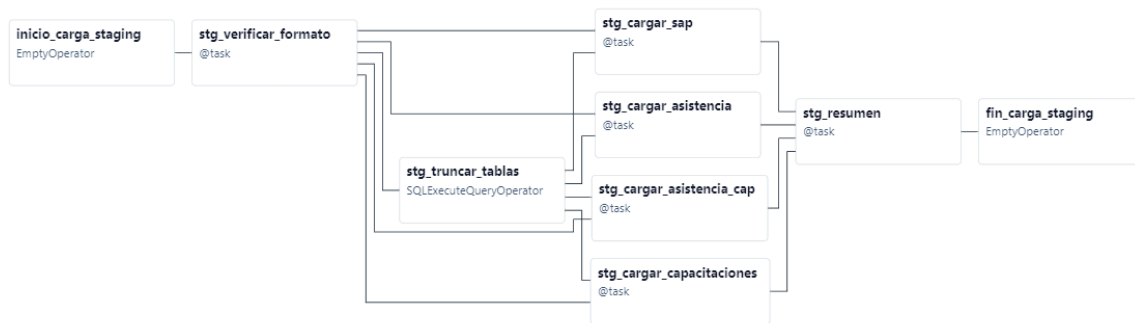
El sistema se divide en **6 DAGs atómicos** y un **DAG Maestro**, lo que permite una ejecución granular para correcciones puntuales o una ejecución integral para procesos de carga masiva.

DAG	Descripción	Entidades y Tablas Impactadas
<i>dag_carga_staging</i>	Ingesta Técnica	Esquema staging (Carga cruda de archivos Excel).
<i>dag_conformed</i>	Dimensiones Maestras	DimEmpleado, DimCargo, DimEmpresa, DimGerencia, DimCentroCosto, DimTiempo.
<i>dag_asistencia</i>	Gestión de Tiempos	FactAsistencia, DimTurno, DimPermiso.
<i>dag_capacitacion</i>	Formación y Desarrollo	FactRealizacionCapacitacion, FactParticipacionCapacitacion, DimCurso, DimProveedor.
<i>dag_rotacion</i>	Movimientos y Dotación	FactRotacion, FactDotacionSnapshot, DimMedidaAplicada, DimModalidadContrato.
<i>dag_seleccion</i>	Talent Acquisition	FactSeleccion.
<i>dag_maestro_completo</i>	Orquestación Global	Ejecución secuencial de todas las áreas anteriores.

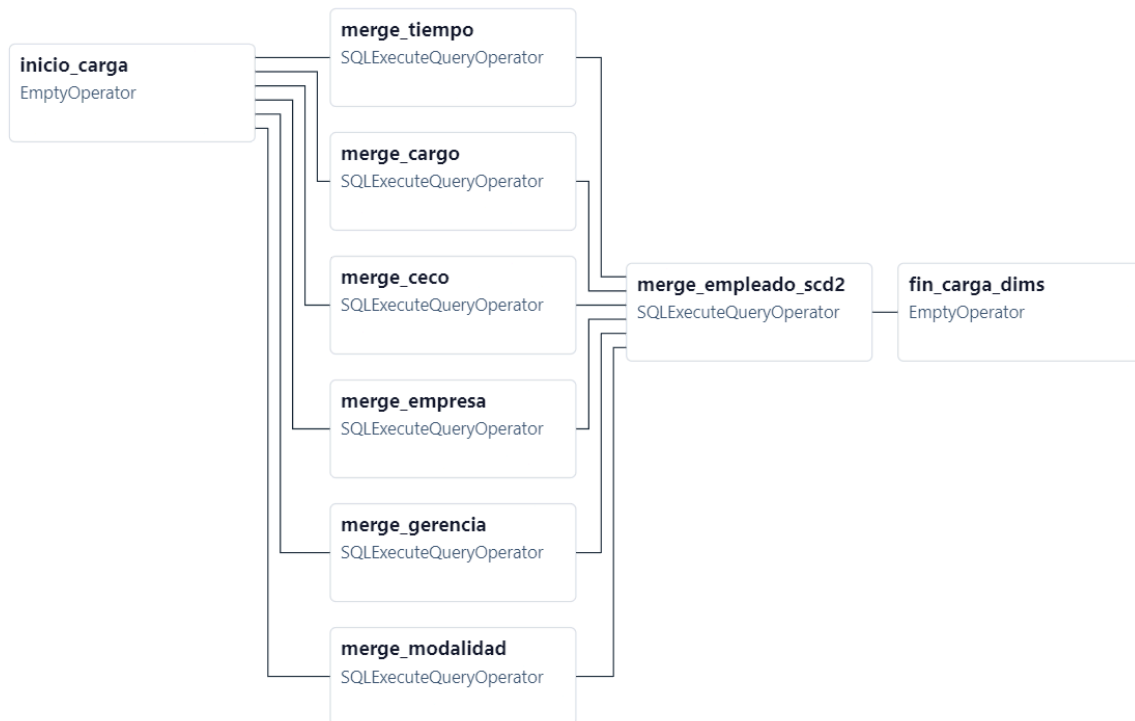
11.4. Visualización de los Grafos de Dependencia

A continuación, se presentan las representaciones gráficas de los flujos de trabajo, donde se observa la jerarquía de ejecución y las dependencias entre las tareas de extracción y carga de dimensiones/hechos.

00 dag_carga_staging



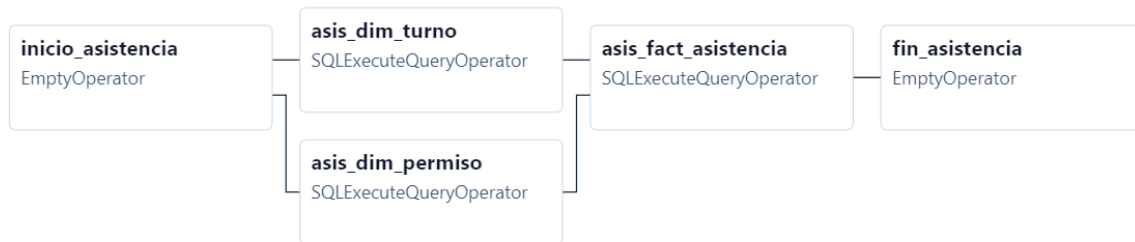
01 dag_conformed



02 dag_rotacion



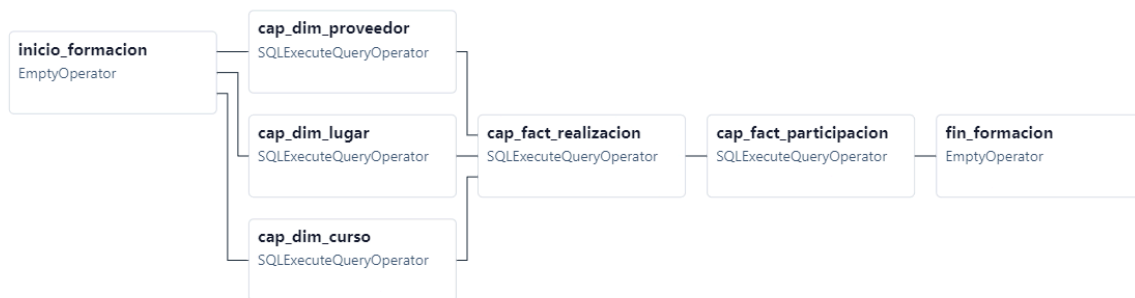
03 dag_asistencia



04 dag_seleccion



05 dag_capacitacion



99 dag_maestro_completo

Este DAG es un compendio del resto de DAGs mostrados en esta sección. Como se dijo anteriormente, solo concatena la ejecución de cada uno para facilitar la ejecución completa del ciclo ETL.

12. Diseño de Interfaz de Usuario y Prototipo

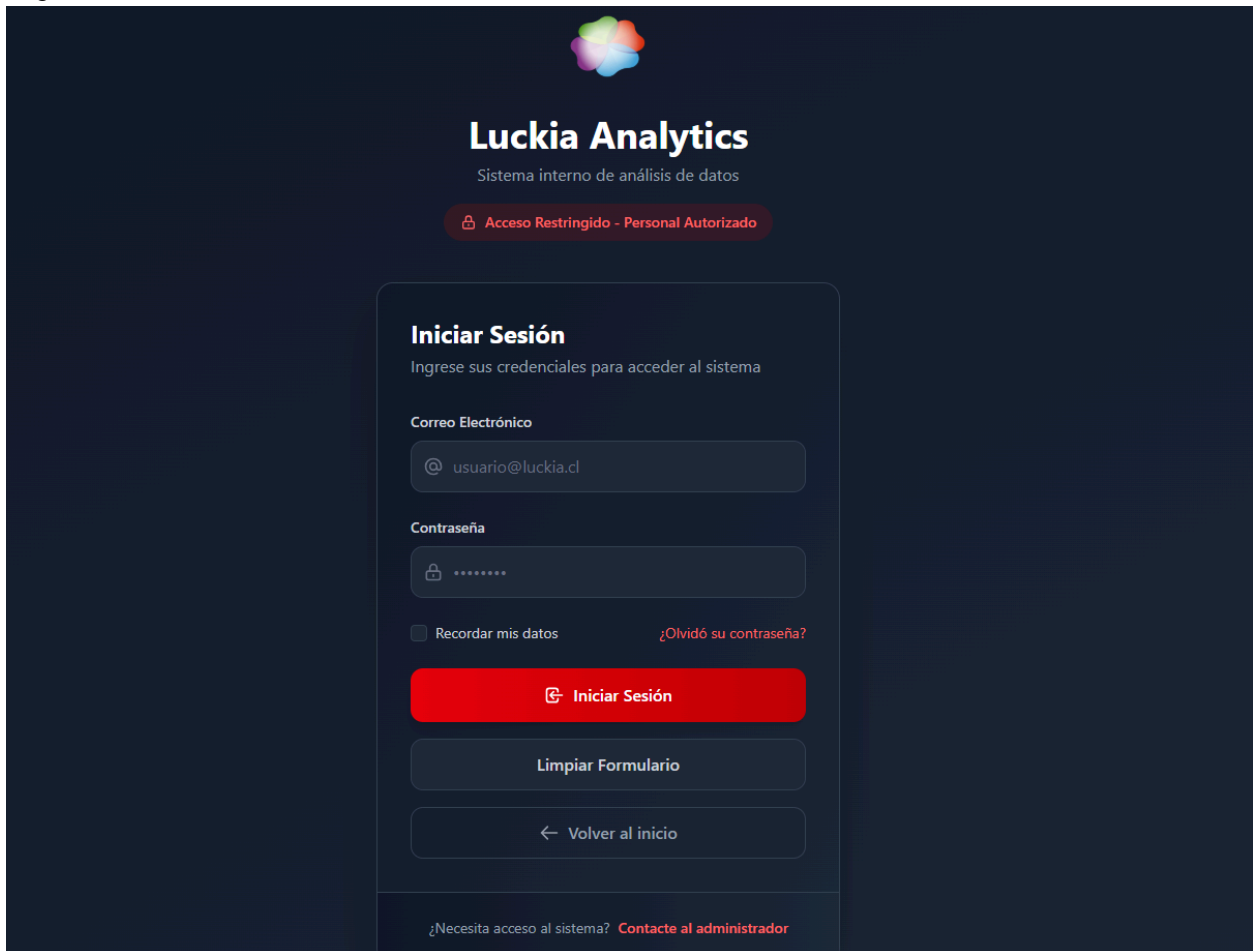
Inicialmente se propusieron unos diseños bastante sencillos, por lo que, pensando en una mejor visualización para el cliente, se dió un toque más moderno a la interfaz.

12.1. Mockups

Página Inicial



Login



The login page for Luckia Analytics features a dark blue background. At the top center is a logo consisting of five overlapping circles in green, yellow, orange, red, and blue. Below the logo, the text "Luckia Analytics" is displayed in a large, white, sans-serif font, followed by "Sistema interno de análisis de datos" in a smaller, lighter font. A red banner with white text and a lock icon reads "Acceso Restringido - Personal Autorizado". The main login form is a light blue rounded rectangle. It has a title "Iniciar Sesión" and a subtitle "Ingresa sus credenciales para acceder al sistema". It contains two input fields: "Correo Electrónico" with the placeholder "usuario@luckia.cl" and "Contraseña" with a masked password "*****". There is a checkbox for "Recordar mis datos" and a link "¿Olvidó su contraseña?". A prominent red button with a white lock icon and the text "Iniciar Sesión" is below the password field. Below this are two light blue buttons: "Limpiar Formulario" and "← Volver al inicio". At the bottom of the form, a link "¿Necesita acceso al sistema? Contacte al administrador" is displayed.

Luckia Analytics
Sistema interno de análisis de datos

Acceso Restringido - Personal Autorizado

Iniciar Sesión

Ingresa sus credenciales para acceder al sistema

Correo Electrónico

@ usuario@luckia.cl

Contraseña

☐ Recordar mis datos [¿Olvidó su contraseña?](#)

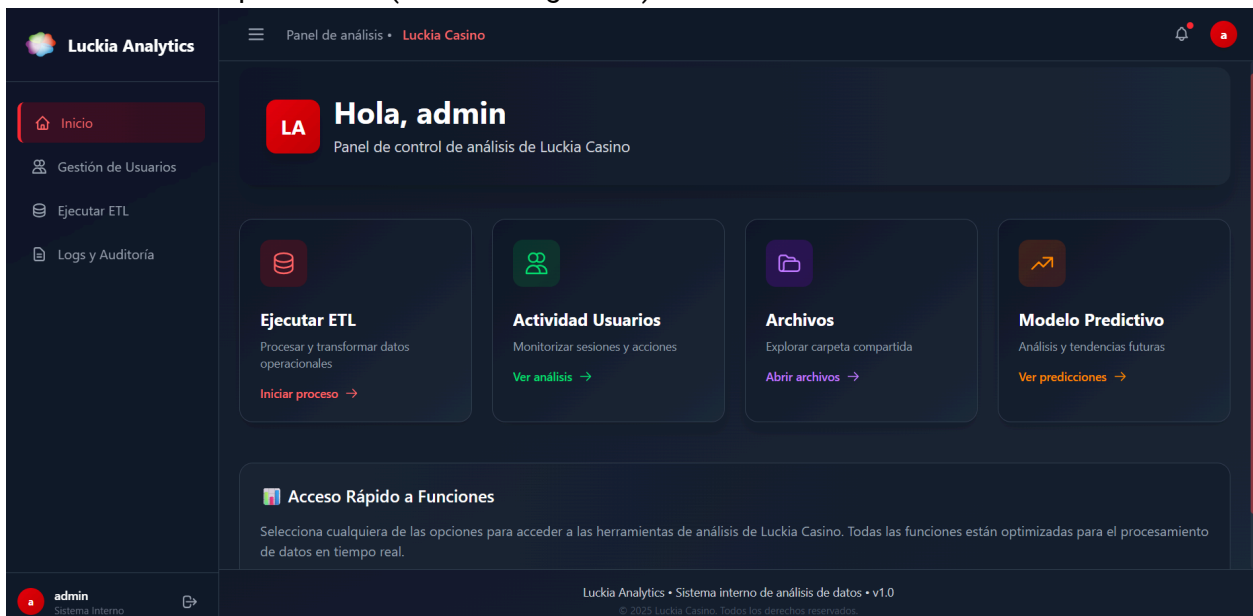
[Iniciar Sesión](#)

[Limpiar Formulario](#)

[← Volver al inicio](#)

¿Necesita acceso al sistema? [Contacte al administrador](#)

Panel Central de operaciones (dashboard general)



The dashboard has a dark blue theme. On the left is a sidebar with the "Luckia Analytics" logo and four menu items: "Inicio" (highlighted), "Gestión de Usuarios", "Ejecutar ETL", and "Logs y Auditoría". The main area has a header with a hamburger menu, "Panel de análisis • Luckia Casino", and a notification bell. Below the header is a large card with the "LA" logo, the greeting "Hola, admin", and the subtitle "Panel de control de análisis de Luckia Casino". The dashboard features four main functional cards: "Ejecutar ETL" (Process and transform operational data, with an "Iniciar proceso" link), "Actividad Usuarios" (Monitor sessions and actions, with a "Ver análisis" link), "Archivos" (Explore shared folder, with an "Abrir archivos" link), and "Modelo Predictivo" (Analyze and future trends, with a "Ver predicciones" link). At the bottom is a section titled "Acceso Rápido a Funciones" with a brief description. The footer includes a user profile "admin" and system info "Luckia Analytics • Sistema interno de análisis de datos • v1.0" and "© 2025 Luckia Casino. Todos los derechos reservados."

Luckia Analytics

Inicio

Gestión de Usuarios

Ejecutar ETL

Logs y Auditoría

Panel de análisis • Luckia Casino

Hola, admin

Panel de control de análisis de Luckia Casino

Ejecutar ETL

Procesar y transformar datos operacionales

[Iniciar proceso →](#)

Actividad Usuarios

Monitorizar sesiones y acciones

[Ver análisis →](#)

Archivos

Explorar carpeta compartida

[Abrir archivos →](#)

Modelo Predictivo

Análisis y tendencias futuras

[Ver predicciones →](#)

Acceso Rápido a Funciones

Selecciona cualquiera de las opciones para acceder a las herramientas de análisis de Luckia Casino. Todas las funciones están optimizadas para el procesamiento de datos en tiempo real.

admin
Sistema Interno

Luckia Analytics • Sistema interno de análisis de datos • v1.0
© 2025 Luckia Casino. Todos los derechos reservados.

Gestión de Usuarios

Administración de accesos, roles y permisos del sistema - Control AAA (Autenticación, Autorización, Auditoría)

[+ Nuevo Usuario](#)

Total Usuarios
1

Activos
1

Administradores
1

Último acceso
31-12-2025

Buscar usuario
Nombre, email o área...


Filtrar por rol
Todos los roles

Filtrar por estado
Todos los estados

Usuario	Rol	Estado	Último Acceso	Acciones
<div>A</div> <div>Administrador</div> <div>admin@sail.cl</div> <div>N/A</div>	Administrador	Activo	31-12-2025, 11:53:36 a. m. Creado: N/A	✎ ✖ 🔑

Luckia Analytics • Sistema interno de análisis de datos • v1.0
© 2025 Luckia Casino. Todos los derechos reservados.

Ejecución del Proceso ETL



Ejecución del ETL

Inicia el proceso de Extracción, Transformación y Carga de datos operacionales.

[▶ Iniciar Proceso](#)

Historial

Últimas 5 ejecuciones

✓

2023-10-25 14:30

Exitoso

2m 15s

✗

2023-10-24 09:15

Fallido

45s

✓

2023-10-23 18:00

Exitoso

2m 18s

✓

2023-10-22 10:00

Exitoso

2m 28s

✓

2023-10-21 15:45

Exitoso

2m 12s

[Ver historial completo →](#)

Historial Completo de Ejecuciones

<

Historial de Ejecuciones

Registro completo de actividades del motor ETL

TOTAL EJECUCIONES

154

TASA DE ÉXITO

92%

DURACIÓN PROMEDIO

2m 10s

Q

Buscar por usuario o detalle...

Todos

Exitoso

Fallido

ESTADO	FECHA / HORA	DURACIÓN	REGISTROS	USUARIO	DETALLES
Exitoso	2025-10-25 14:30	1m 15s	244 filas	Sistema (Auto)	Carga completada OK
Exitoso	2025-10-25 15:31	2m 45s	254 filas	jyampara	Carga completada OK
Fallido	2025-10-24 16:32	3m 8s	112 filas	jyampara	Error de conexión SAP
Exitoso	2025-10-24 17:33	2m 51s	460 filas	Sistema (Auto)	Carga completada OK
Exitoso	2025-10-23 18:34	3m 13s	522 filas	jyampara	Carga completada OK
Exitoso	2025-10-23 14:35	1m 39s	497 filas	jyampara	Carga completada OK
Exitoso	2025-10-22 15:36	3m 34s	336 filas	Sistema (Auto)	Carga completada OK
Exitoso	2025-10-22 16:37	2m 29s	563 filas	jyampara	Carga completada OK
Fallido	2025-10-21 17:38	3m 19s	168 filas	jyampara	Carga completada OK

Sistema de Logs y Auditoria

Sistema de Logs y Auditoría

Monitoreo centralizado de eventos del sistema, accesos y acciones de usuarios (RF-07)

Sin conexión en tiempo real

Exportar CSV

Exportar Local

Total Logs

15

Errores Críticos

2

Eventos de Auditoría

3

Último Evento

18-10-2025

Auto

Buscar en logs

Texto en mensaje, usuario, IP...

Nivel

Todos los niveles

Módulo

Todos los módulos

Usuario

Todos los usuarios

Limpiar Filtros

Eventos del Sistema

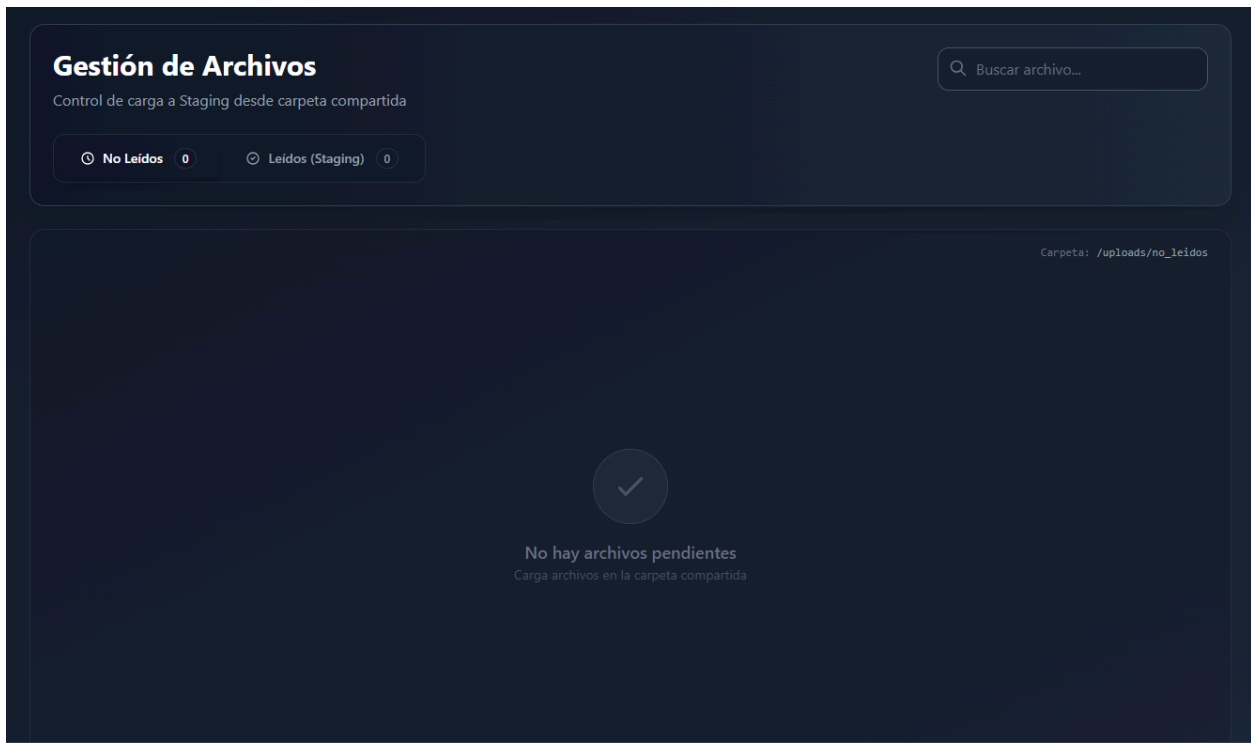
Mostrando 15 de 15 logs • Última actualización: 9:52:24 a. m.

Limpiar Logs Antiguos

Timestamp	Nivel	Módulo	Usuario	Mensaje	Acciones
18-10-2025, 3:39:23 p. m.	INFO	AUTH	juan.carlos 192.168.1.100	Usuario juan.carlos inició sesión exitosamente	Detalles
18-10-2025, 3:40:12 p. m.	SUCCESS	ETL	sistema	Proceso ETL completado exitosamente	Detalles

69

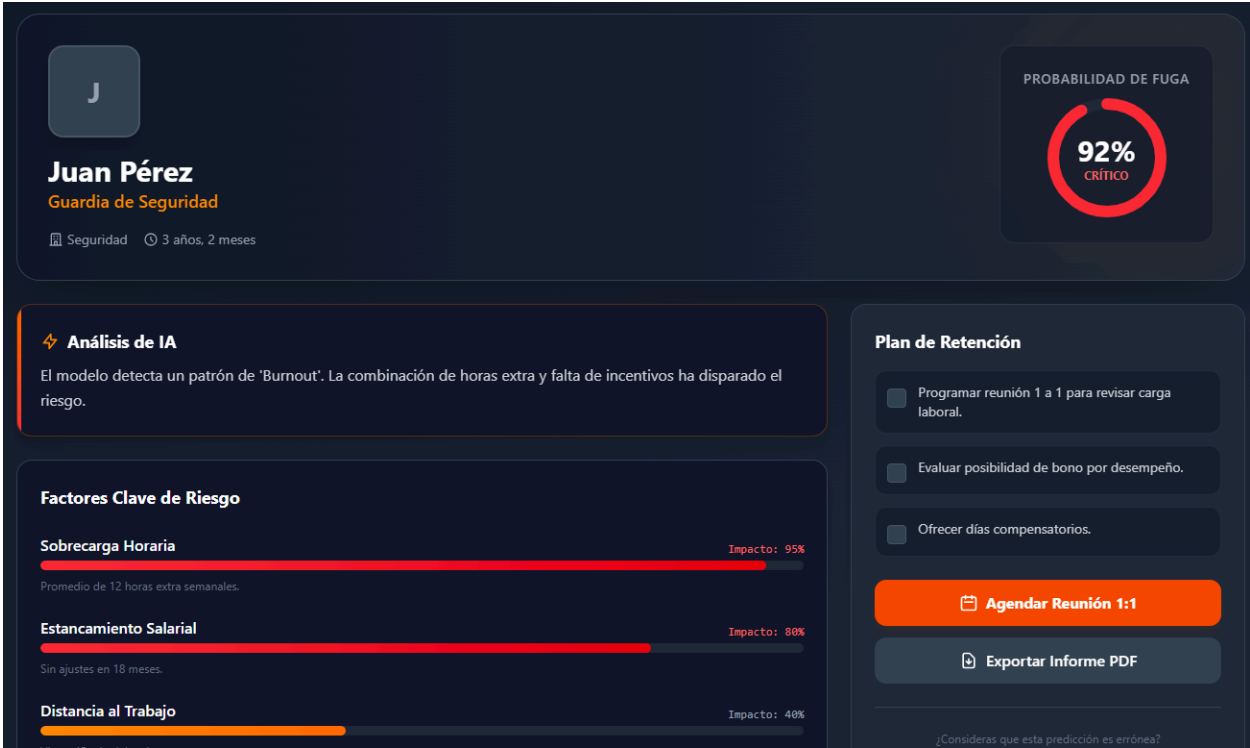
Gestión de Archivos



Interfaz del Modelo predictivo + Métricas provenientes del DW



Detalle del empleado (Predicción de riesgo)



13. Conclusiones

Este documento ha presentado la implementación de una plataforma web integral de análisis de datos diseñada para abordar las necesidades operativas del área de Recursos Humanos de Casino Luckia Arica. El sistema desarrollado materializa técnicamente la arquitectura previamente diseñada, transformando el plan conceptual de Data Warehouse en una aplicación funcional y accesible.

La solución consiste en una aplicación web desarrollada en SvelteKit 5 que centraliza el acceso a la información. Se implementaron módulos especializados para gestión de usuarios con control de accesos basado en roles (RBAC), visualización de predicciones de rotación de personal, ejecución controlada de procesos ETL orquestados por Apache Airflow, y auditoría completa de eventos del sistema. El diseño del frontend permite que los usuarios inicien operaciones ETL, pero el procesamiento y transformación de los datos es gestionada por Apache Airflow y el backend desarrollado con FastAPI, manteniendo una clara separación de responsabilidades.

Si la adopción del sistema es exitosa, se espera que la plataforma reduzca significativamente el tiempo que el equipo de RRHH dedica a tareas operativas repetitivas. Más allá de la visualización de métricas en dashboards, la verdadera prueba de éxito residirá en la utilidad práctica del módulo de predicciones para anticipar la rotación de personal, permitiendo acciones preventivas de retención antes de que los empleados tomen la decisión de abandonar la empresa. Lo último dependerá a su vez de la población de datos que se vaya formando en el DW.

La viabilidad del sistema depende críticamente de dos factores: primero, la continuidad y confiabilidad de los procesos ETL orquestados por Airflow que alimentan el Data Warehouse con datos actualizados de SAP y GeoVictoria principalmente; y segundo, la adopción efectiva de la plataforma por parte de los usuarios finales del área de RRHH (Gerencia). Se espera que la interfaz intuitiva y las funcionalidades implementadas faciliten esta adopción.

La arquitectura modular implementada, con separación clara entre presentación (SvelteKit), lógica de negocio (FastAPI), orquestación (Airflow) y almacenamiento (PostgreSQL), sienta una base técnica sólida que permite la extensión futura del sistema. El código está estructurado para facilitar la adición de nuevos módulos según surjan necesidades adicionales, ya sea ampliación de métricas, integración de nuevas fuentes de datos o refinamiento continuo del modelo predictivo basado en retroalimentación del equipo de RRHH.

14. Bibliografía

- [1] D. J. Anderson, *Kanban: Successful Evolutionary Change for Your Technology Business*. Sequim, WA: Blue Hole Press, 2010.
- [2] R. Wirth and J. Hipp, "CRISP-DM: Towards a standard process model for data mining," in *Proc. 4th Int. Conf. Practical Application of Knowledge Discovery and Data Mining*, Manchester, UK, 2000, pp. 29–39.
- [3] S. K. Wagh, A. A. Andhale, K. S. Wagh, J. R. Pansare, S. P. Ambadekar, and S. H. Gawande, "Customer churn prediction in telecom sector using machine learning techniques," *Results Control Optim.*, vol. 14, p. 100342, Mar. 2024, doi: 10.1016/j.rico.2023.100342.
- [4] R. Kimball and M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3rd ed. Hoboken, NJ: John Wiley & Sons, 2013.
- [5] E. Turban, R. Sharda, and D. Delen, *Business Intelligence: A Managerial Approach*, 2nd ed. Upper Saddle River, NJ: Pearson, 2014.
- [6] F. Diez, M. Bussin, and V. Lee, *Fundamentals of HR Analytics: A Manual on Becoming HR Analytical*. Bingley, UK: Emerald Publishing Limited, 2019.

15. Anexo

Repositorio de GitHub: <https://github.com/dyri0n/sail>

Modelo multidimensional:

- Fuente: <https://dbdiagram.io/d/691b29046735e1117031dd2e>
- Resumen:
https://drive.google.com/file/d/1yXgh9l5j8nxovqi31wP941dAgSadb0Po/view?usp=s_haring