

**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, 20 de noviembre 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
Sistema de Carga	32
Sistema ETL	32
Capa de Datos	32
Sistema de Visualización	32
Sistema de Aprendizaje Automático	32
Sistema de Logging y Backup	32
Sistema de Autenticación	32
8.4. Diagramas de interacción	33
9. Arquitectura y Pila Tecnológica	36
9.1. Aspectos Arquitecturales del Sistema	36
Suposiciones y Limitaciones	36
Decisiones de Diseño	36
9.2. Herramientas y Frameworks de Software	37
10. Modelado de Datos	39
10.1. Diccionario de Datos	39
Diccionario rotación	39
Geovictoria día	44
Geovictoria semana	45
Resumen anual de capacitación	46
Participantes de las capacitaciones	48
Perfiles de trabajo	49
Proceso de contratación/selección	51
10.2. Base de datos Staging	54
10.2.1. Data Warehouse	57
Objetivo del Data Warehouse	57
Decisiones de diseño	57
BI vs ML (Data Lake VS Data Marts)	57
Conformed Dimensions	57
Formulación de las Tablas de Hechos	58
1. Rotación de Empleados (Transaccional)	58
2. (De apoyo) Headcount mensual (Snapshot Periodico)	58
3. Asistencia Diaria (Transaccional)	59
4. Participación en Capacitación (Transaccional)	59

5. Proceso de Selección (Snapshot Acumulativo)	60
Dimensiones	61
Dimensión Empleado	61
Dimensión Tiempo	62
Dimensión Cargo	63
Dimensión Capacitación	64
Mini Dimensiones	65
Dimensión Clase Medida (Rotación)	65
Dimensión Motivo Medida (Rotación)	65
Dimensión Empresa (Rotación)	65
Dimensión Tipo Permiso (Asistencia)	66
Dimensión Tipo Turno (Asistencia)	66
Tablas de Hechos	67
Rotacion de Empleados	67
Dotación de Empleados	67
Asistencia Diaria	68
Participación en Capacitaciones	68
Procesos de Selección	69
Modelo Multidimensional Graficado	71
11. Diseño de Interfaz de Usuario y Prototipo	72
11.1. Mockups	72
12. Conclusiones	74
13. Bibliografía	75
14. Anexo	75

índice de figuras y tablas

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

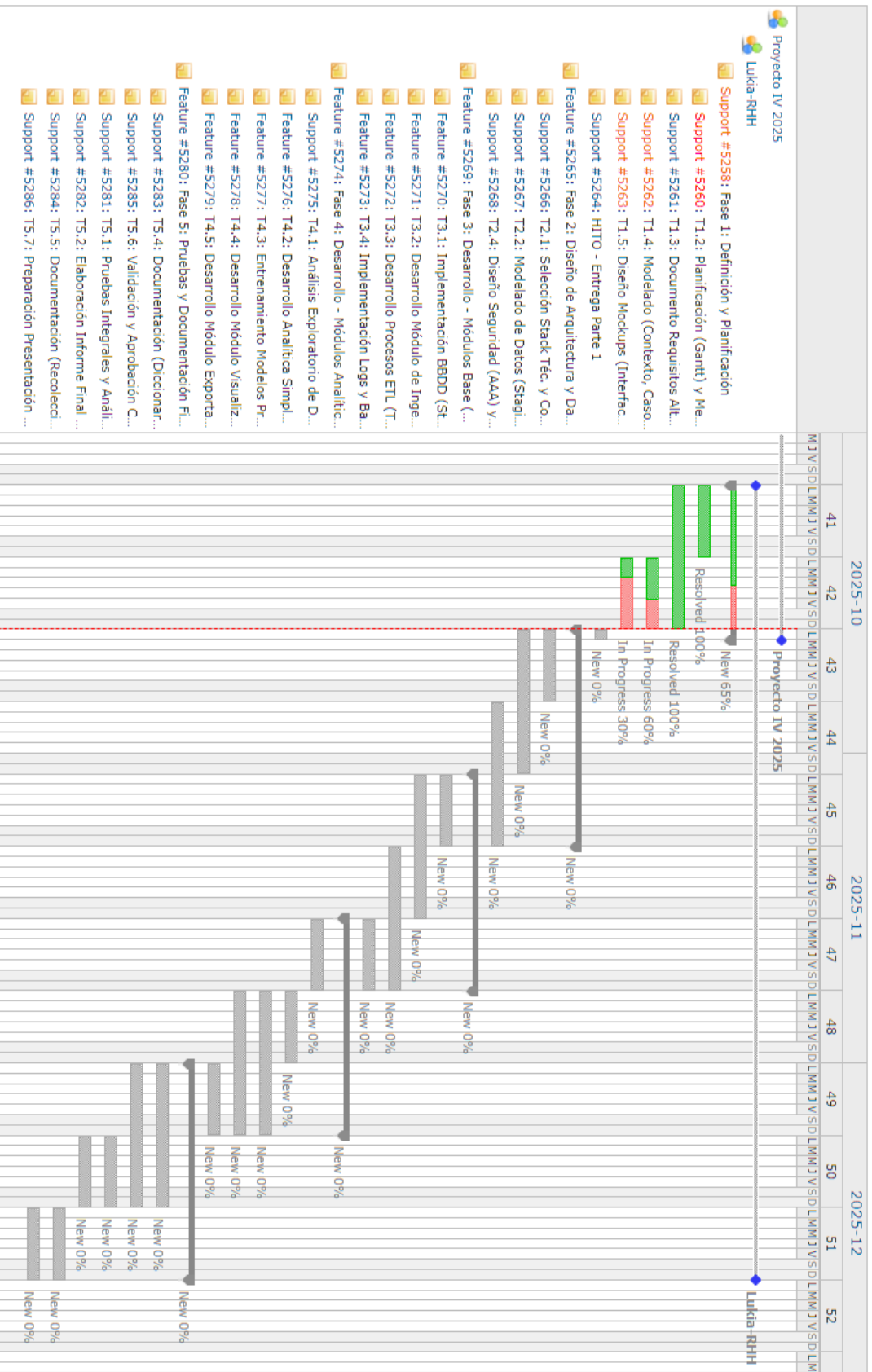
Implementar una plataforma analítica de datos para el área de RRHH de Casino Luckia Arica, que automatice la ingesta y procesamiento de información, permitiendo el análisis de métricas clave y la generación de modelos predictivos para facilitar la toma de decisiones proactiva sobre la gestión del personal.

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

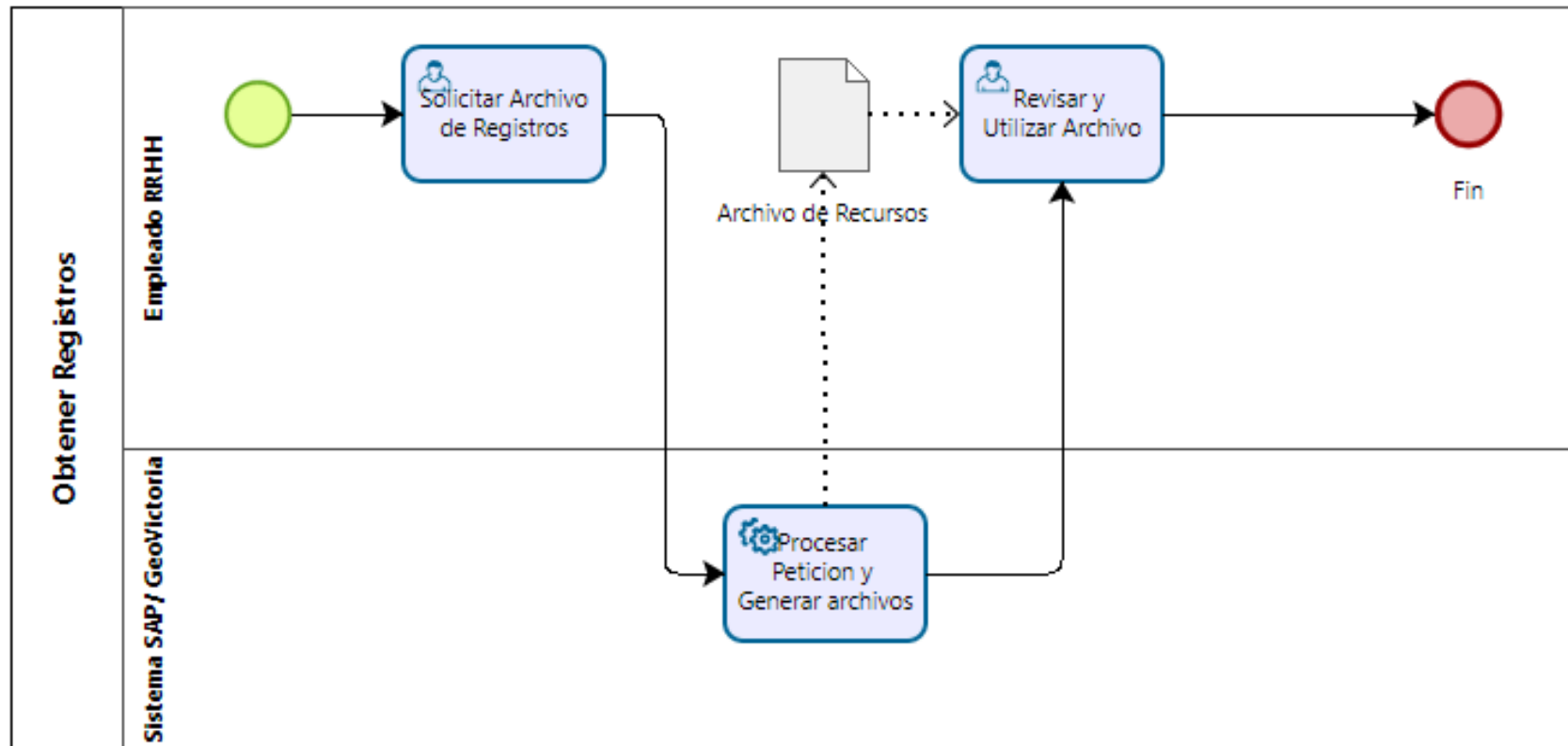
Descripción: Tareas de empaquetado del proyecto, creación de manuales y planes para la entrega final.

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

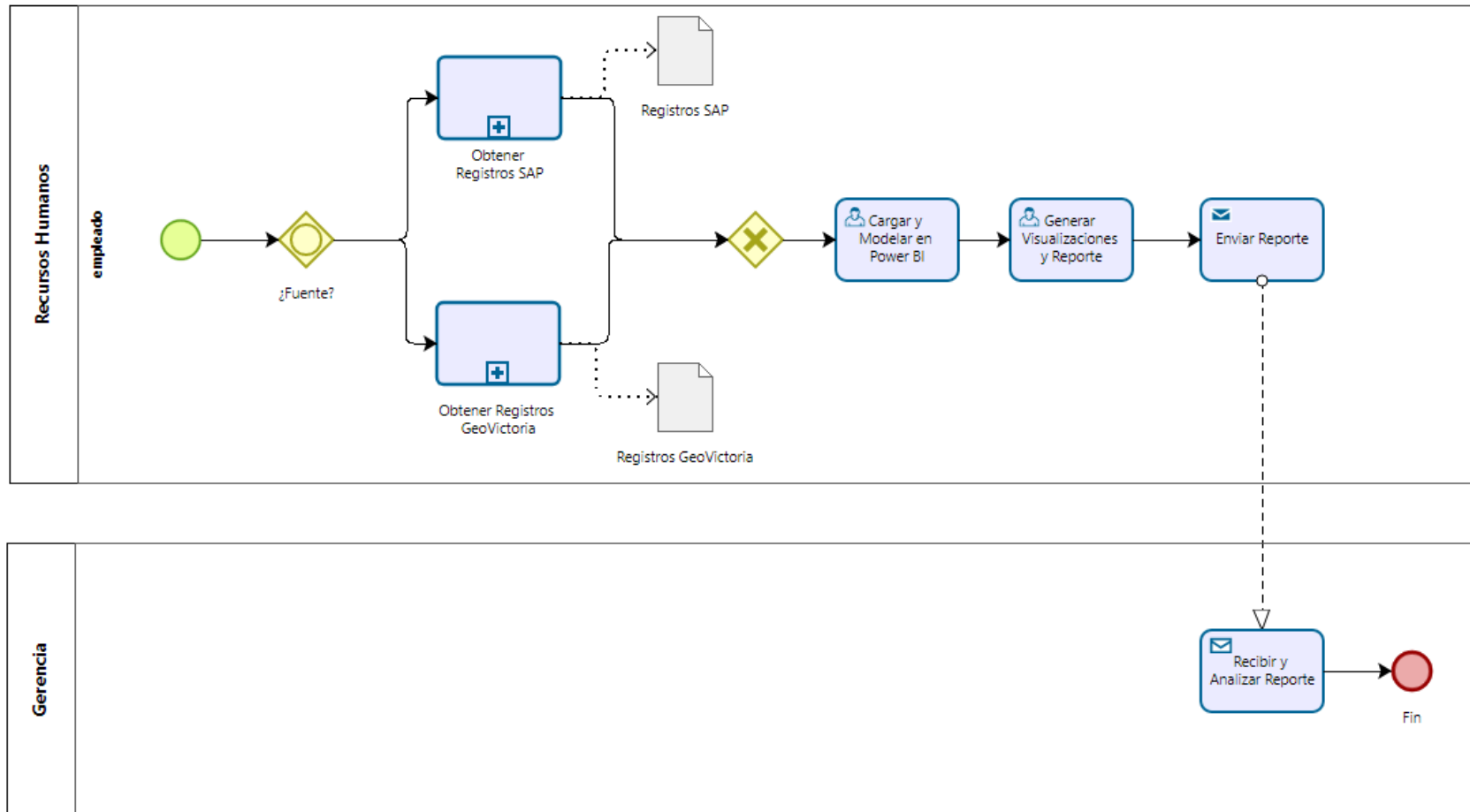
7. Modelado del Procesos de Negocio

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.

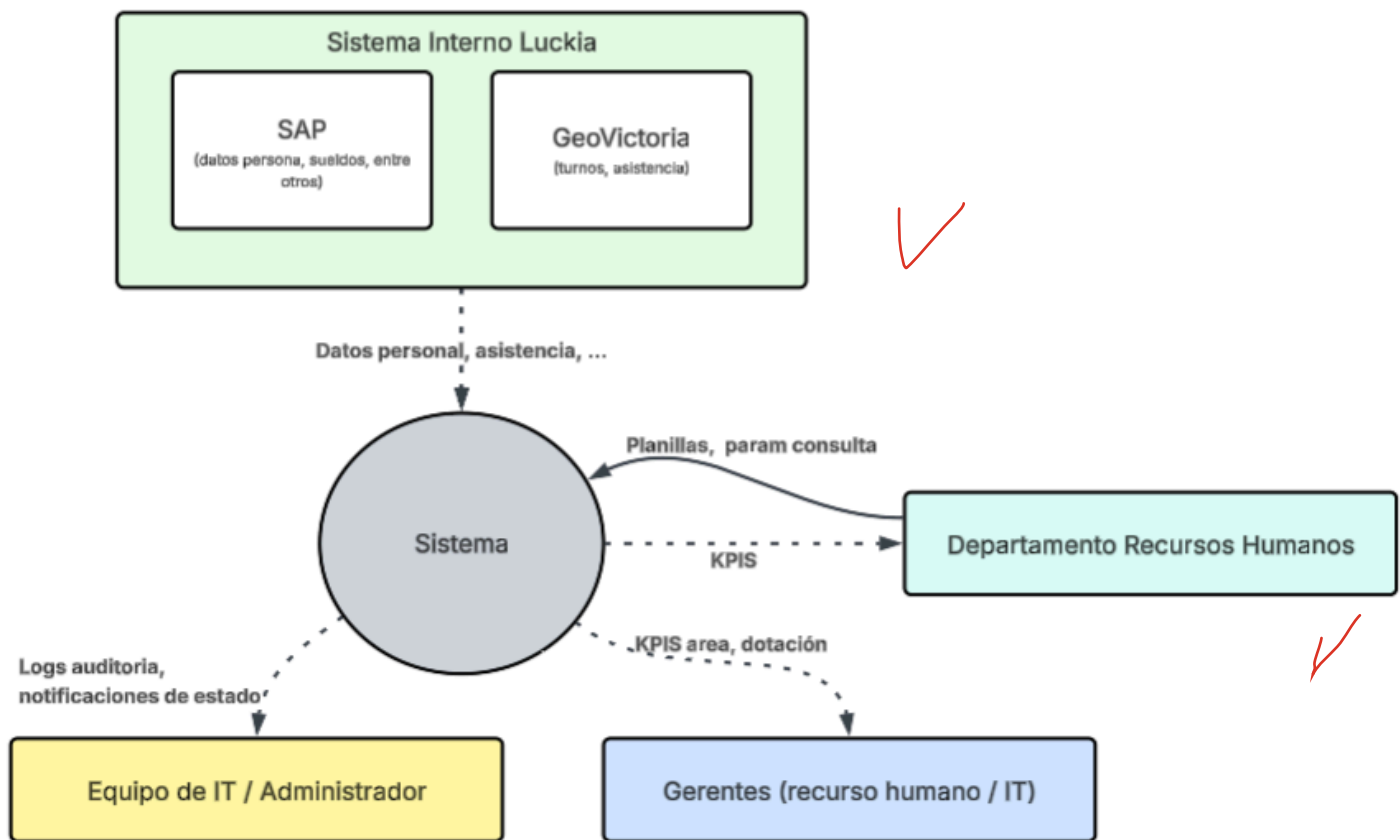


8. Modelado del sistema

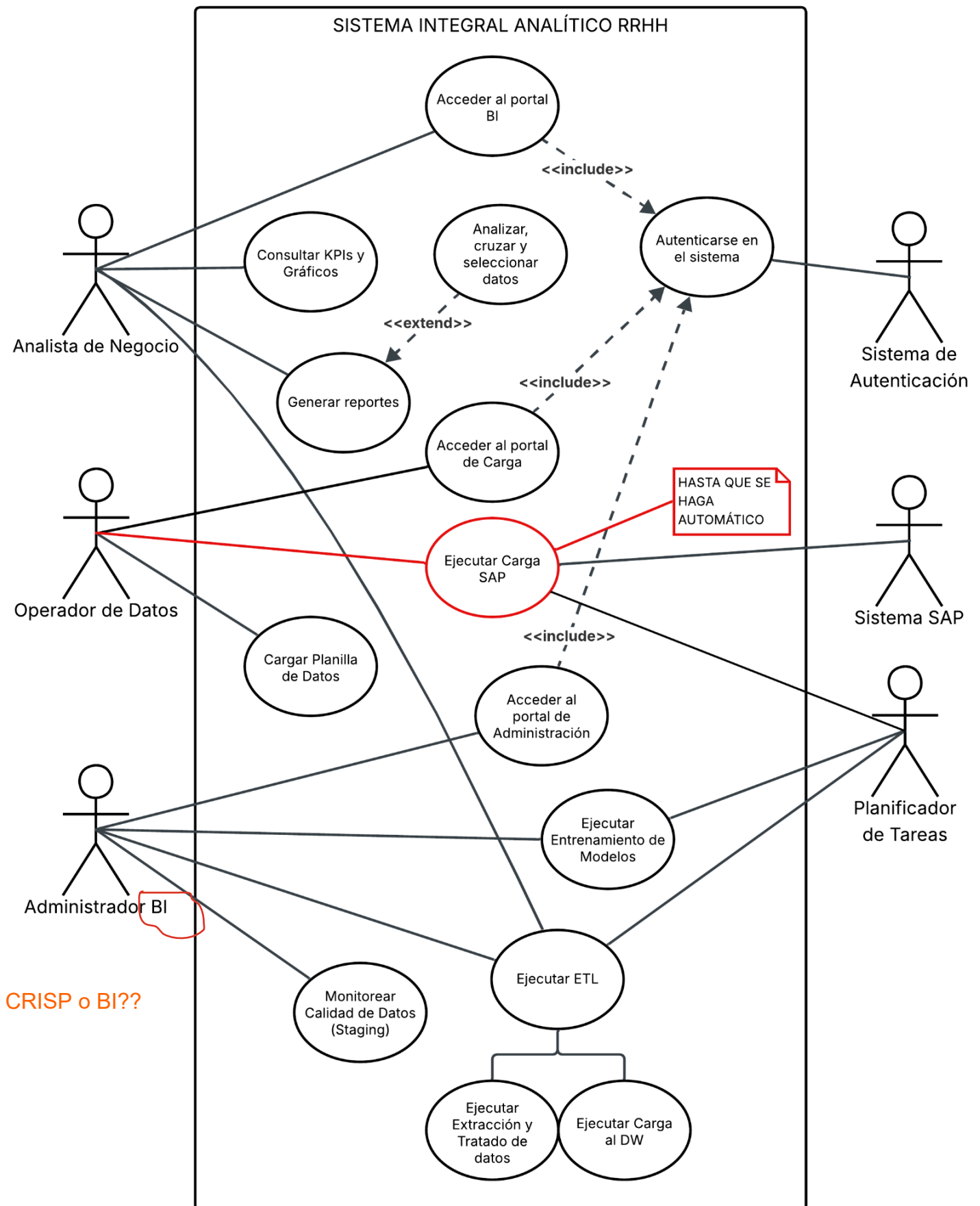
8.1. Modelo de contexto

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.

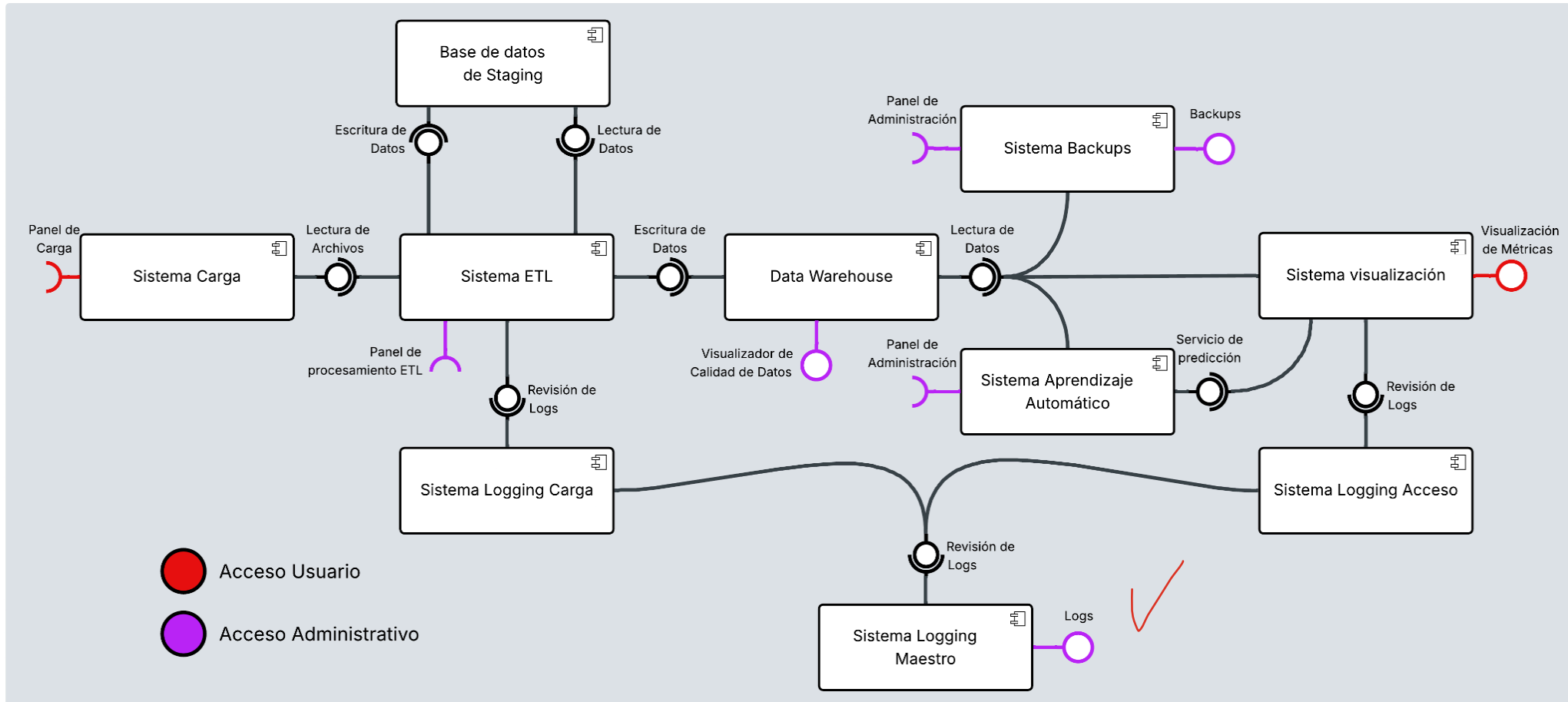
4. Actores: Sistemas Externos

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:

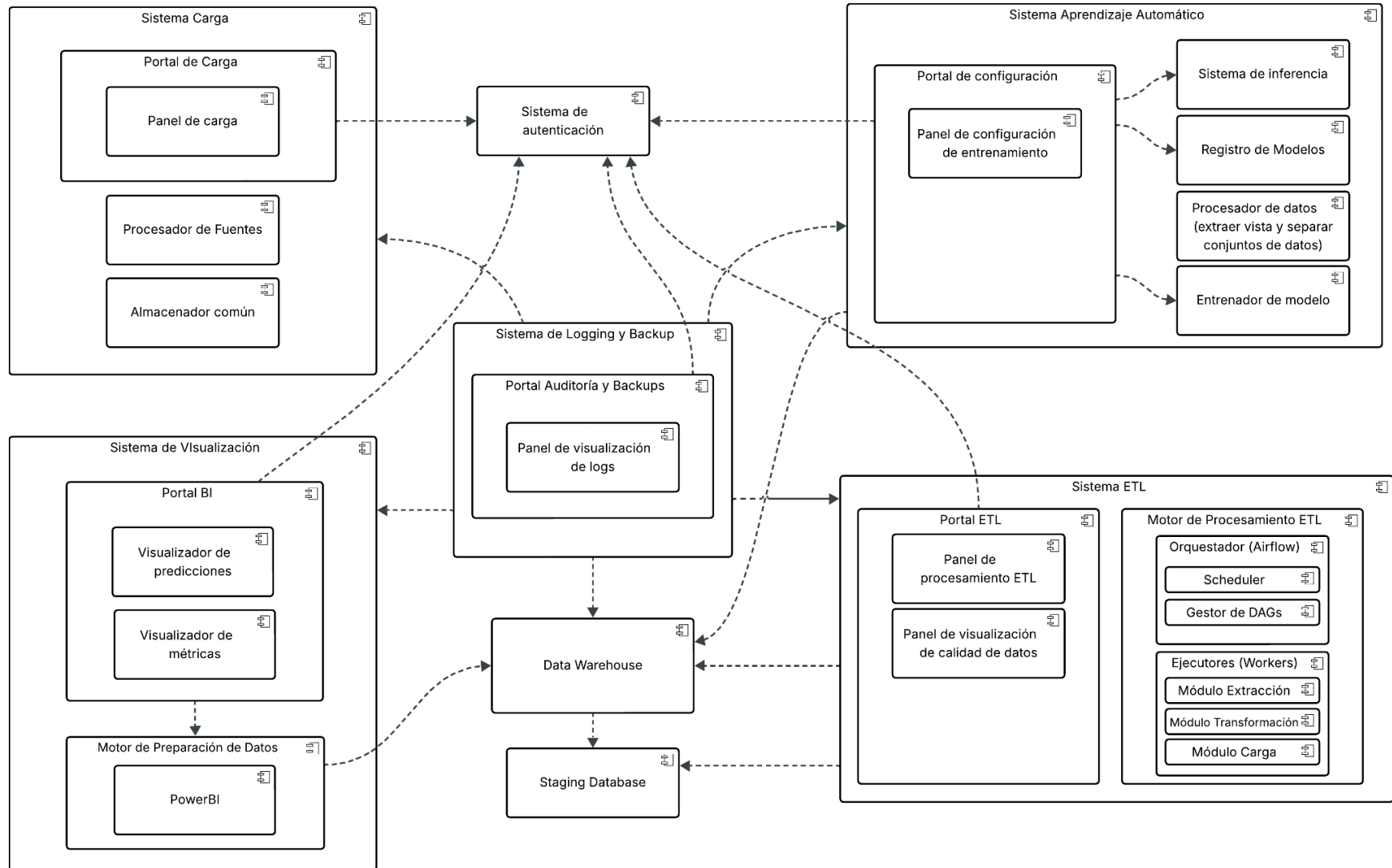
- **Sistema de Carga:** Recibe datos crudos, realiza un procesamiento mínimo de estandarización y los prepara para su recolección periódica (mensual o anual) por el ETL.
- **Sistema ETL:** Integrado por el Planificador de Tareas y el Ejecutor de Workflows, transforma exhaustivamente los datos preprocesados 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

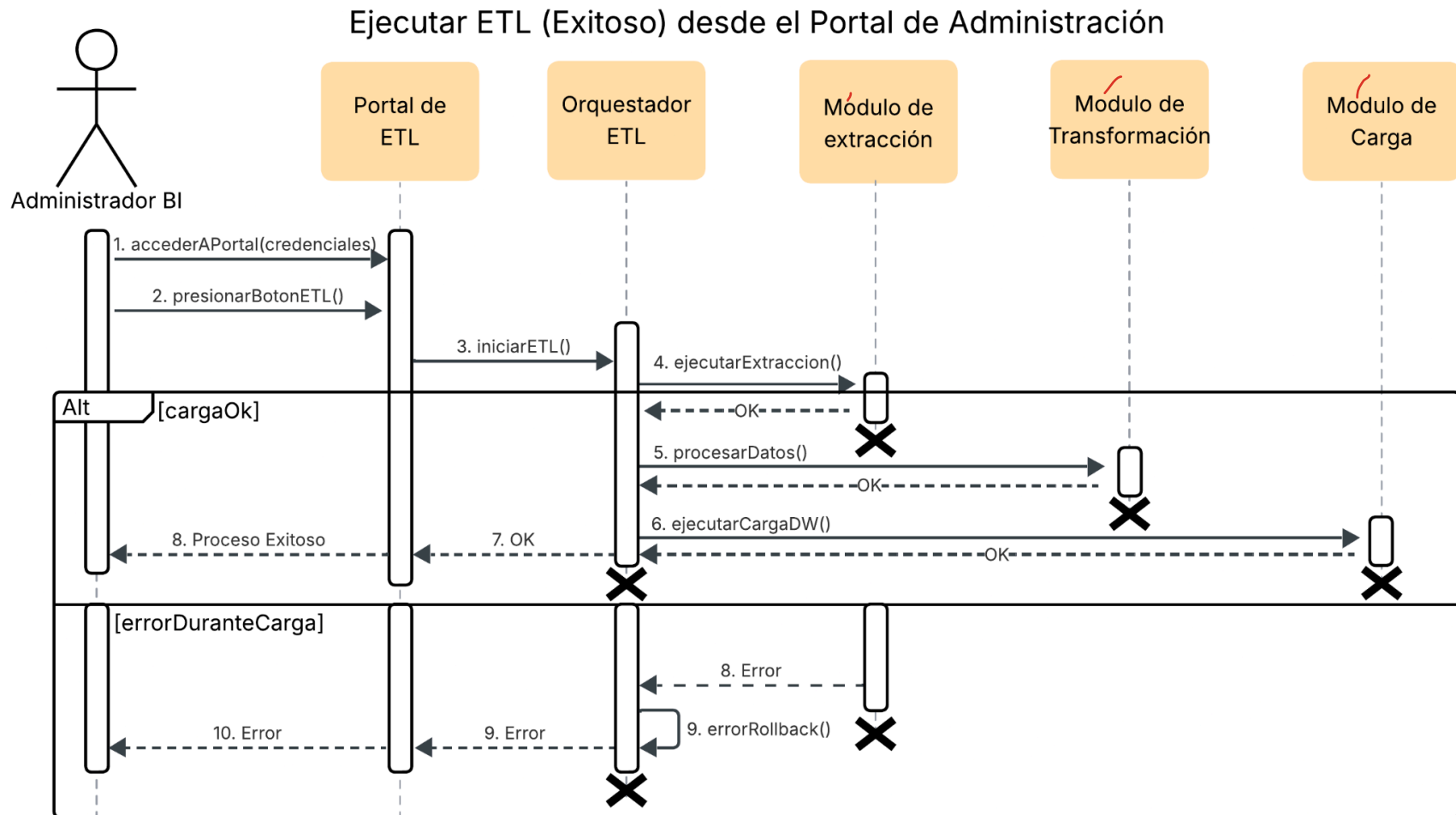


Arquitectura del Sistema de Componentes

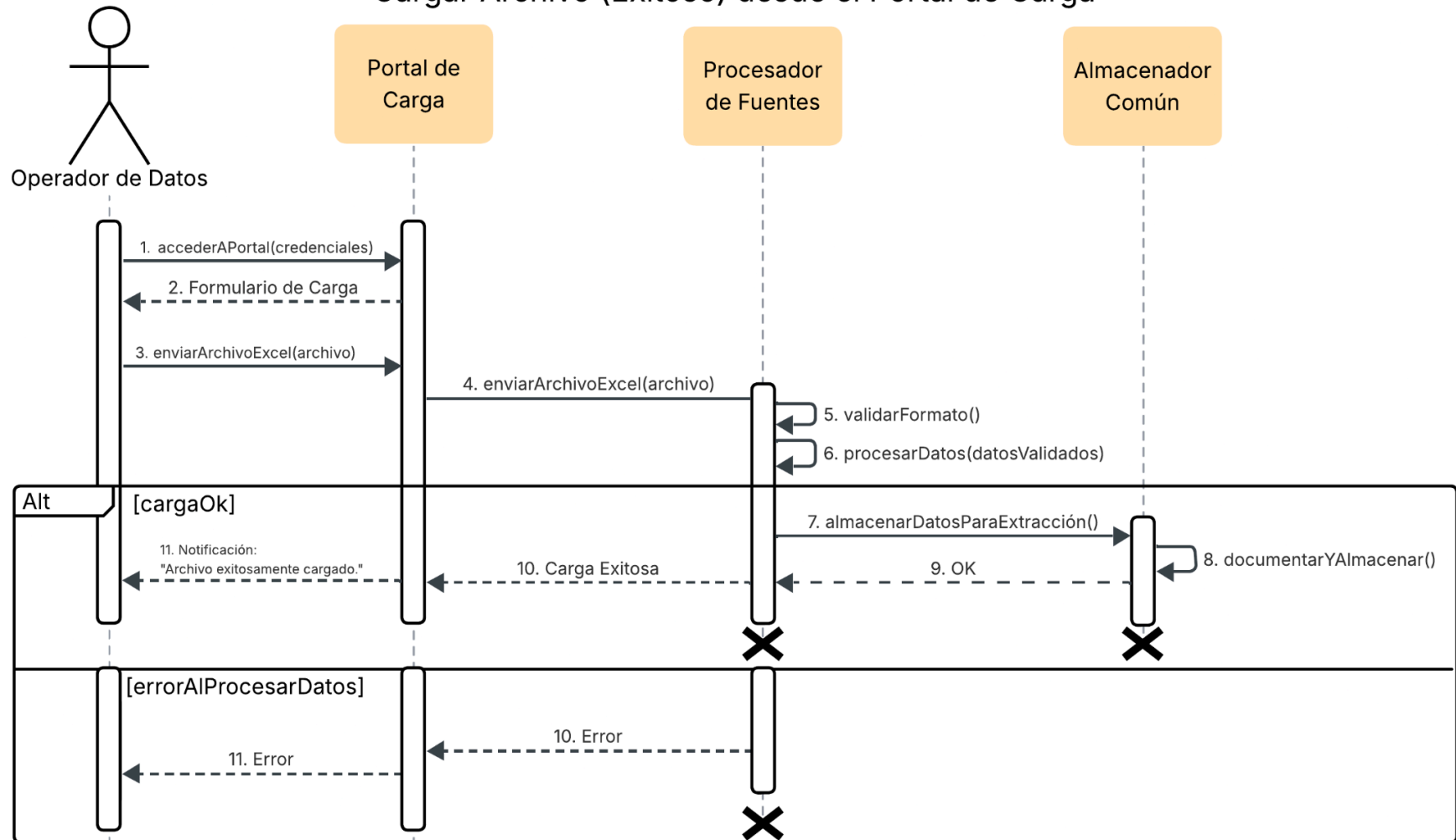
Sistema de Carga	Gestiona la ingesta de los datos. A través del Portal de Carga, el usuario interactúa con el Panel de Carga para subir y validar planillas según estándares predefinidos (si es que aplica). El Procesador de Fuentes convierte estos datos a un formato intermedio y finalmente el Almacenador 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.

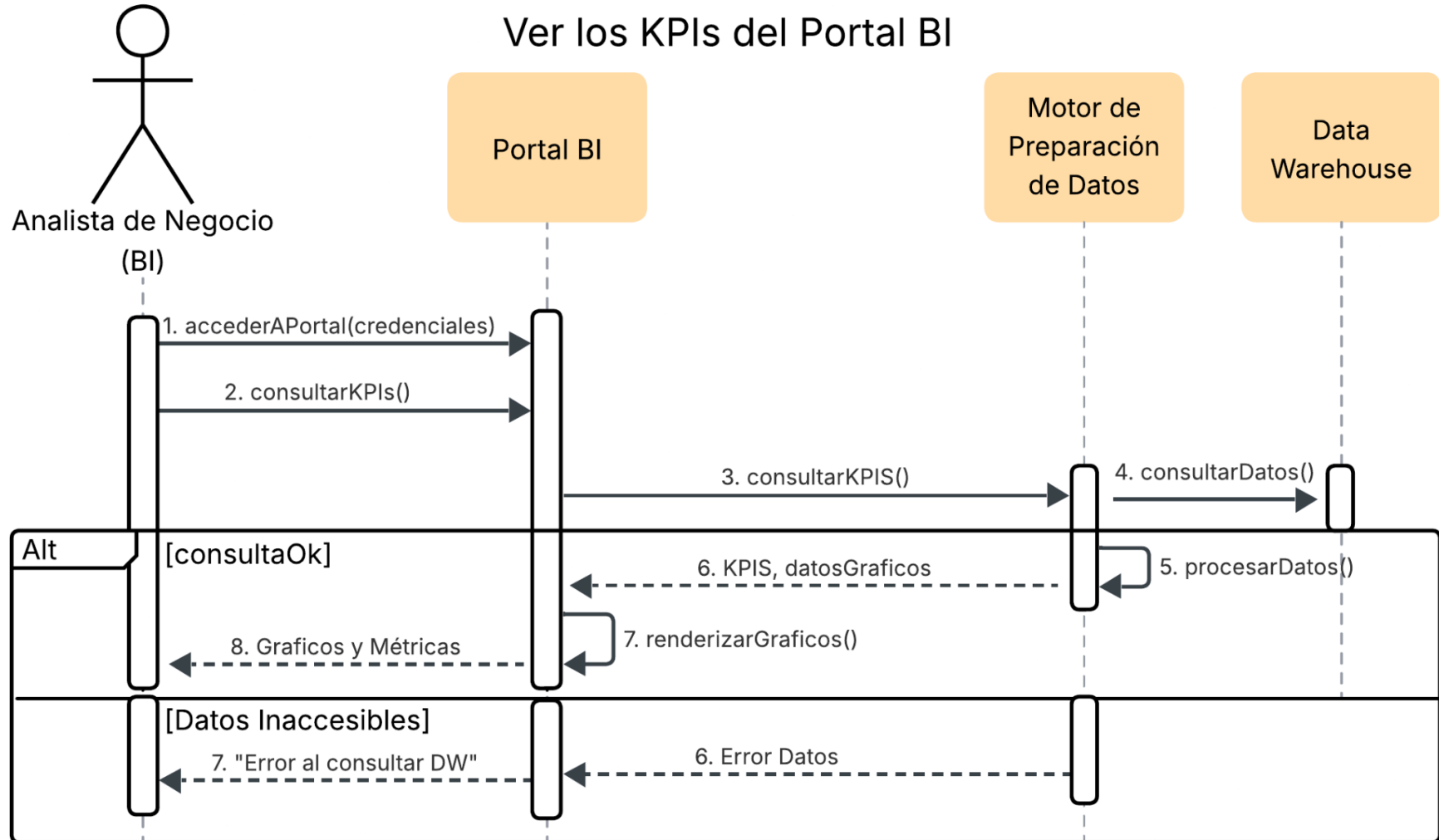


8.4. Diagramas de interacción



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. Diccionario de Datos

Diccionario rotación

obs	Campo Procesado	Campo	Tipo de dato sugerido	Tipo de información	Convención	Valores Posibles	Descripción
Detalles del empleado	id_empleado	Nº pers.	entero	id	Numérico entero (101318), sin separadores.	>100000	ID SAP
Detalles del empleado	nombre	Número de personal	cadena (varchar)	texto	APELLIDOS, NOMBRES (todo en mayúsculas)	CARPENTER SOSA, SABRINA JIMENA	NOMBRE
Detalles de la rotación	id_empresa	Soc.	entero	id	Tres números	837 841	codigo empresa
Detalles de la rotación	empresa	Nombre de la empresa	cadena (varchar)	texto	Texto en mayúsculas con razón social completa.	SIDESA CHILE S.A.	EMPRESA
Detalles de la rotación	tipo_empleo	Denom. área personal	cadena (varchar)	Categoría	Texto capitalizado.	Empleado Temporales Jubilado parcial	TIPO EMPLEADO (nuevo)
Detalles de la rotación	desde1	Desde	fecha	Timestamp (D)	Fecha Serial de Excel (Días desde 1/01/1900). DD-MM-AAAA	41399	debe ser una fecha referencial del empleo
Detalles de la rotación	hasta1	Hasta	fecha	Timestamp (D)	Fecha Serial de Excel (Días desde 1/01/1900). DD-MM-AAAA	41399	debe ser una fecha referencial del empleo
Detalles de la rotación	area	Denominación de unidad organiz	cadena (varchar)	Categoría	Texto en mayúscula. Puede tener símbolos.	AA&BB BARES BOVEDA CAJA EVENTUALES	AREA



						HOUSEKEEPING	
Detalles de la rotación	cargo	Denominación de posiciones	cadena (varchar)	Categoría	Texto capitalizado, puede tener símbolos.	Integración: Posición por defecto Bartender Cajero Mucama	CARGO
Detalles de la rotación	jornada	Relación laboral	cadena (varchar)	Categoría	Texto capitalizado, puede tener números.	Part time Full time Temporal Artículo 22	JORNADA
Detalles de la rotación	ant_puesto	Ant_puesto	fecha	Timestamp (D)	Fecha Serial de Excel (Días desde 1/01/1900). DD-MM-AAAA	41399	nuevo
Detalles de la rotación	ceco	Denominación	cadena (varchar)	Categoría	Texto en mayúsculas. Puede tener símbolos (puntos, asteriscos?)	ARICA CENTER BARES ARICA CENTER RESTAUR ARICA CENTER HOSPEDA S.AREA ARICA CENTER	DEN CECO
Detalles del empleado	fecha_nacimiento	Fe.nacim.	fecha	Timestamp (D)	Fecha Serial de Excel (Días desde 1/01/1900). DD-MM-AAAA	41399	FECHA NACIMIENTO
Detalles del empleado	edad	Edad del empleado	entero	Edad			EDAD
Detalles del empleado	pais_nacimiento	País de nacimiento	cadena (varchar)	Categoría	Texto capitalizado. Países.	Chile Bolivia Venezuela Colombia	PAIS NACIMIENTO



Detalles del empleado	lugar_nacimiento	Lugar de nacimiento	cadena (varchar)	Categoría	Texto en mayúsculas. Ciudades.	arica SANTIAGO SANTA CRUZ BOLIVIA	LUGAR NACIMIENTO
Detalles del empleado	nacionalidad	Nacionalidad	cadena (varchar)	Categoría	Texto en minúsculas.	chilena venezolana boliviana peruana	NACIONALIDAD
Detalles del empleado	estado_civil	Clave para el estado civil	cadena (varchar)	Categoría	Texto en minúsculas, abreviado.	solter casado divorc separa	ESTADO CIVIL
Detalles del empleado	nro_hijos	Nº de hijos	entero	Cantidad		0 1 2 3	NRO HIJOS
Detalles del empleado	sexo	Texto sexo	cadena (varchar)	Categoría	Texto en minúsculas.	masculino femenino	SEXO
Detalles de la rotación	desde2	Desde	fecha	Timestamp (D)	Fecha Serial de Excel (Días desde 1/01/1900). DD-MM-AAAA	41399	Debe ser la fecha inicio del intervalo de la clase de fecha
Detalles de la rotación	hasta2	Hasta	fecha	Timestamp (D)	Fecha Serial de Excel (Días desde 1/01/1900). DD-MM-AAAA	41399	Debe ser la fecha de término del intervalo de la clase de fecha
Detalles de la rotación	clase_fecha	Clase de fecha	cadena (varchar)	Categoría	Texto capitalizado, puede tener símbolos (puntos, números).	Fecha de alta tec. Alta Grupo LUCKIA Cese AFC Fecha ref.vacaci. Vacaciones Progres. Cambio antigüedad Electeur College 1	



						Fecha fin contrato Inicio bonificación	
Detalles de la rotación	fecha	Fecha	fecha	Timestamp (D)	Fecha Serial de Excel (Días desde 1/01/1900). DD-MM-AAAA		Si las fechas anteriores no corresponden a la clase de fecha, porque son de otra cosa o porque sirven de versionado para los registros (modificación), sería esta la fecha de la clase de fecha.
Detalles de la rotación	clase_prestamo	Clase de préstamo	cadena (varchar)	Categoría	Texto capitalizado.	Préstamo Empresa	Quizá indica las personas que han pedido préstamos a la empresa
Detalles de la rotación	movilidad_geografica	Movilidad geográfica			Sin datos.		Quizá indica si las personas han sido movilizadas geográficamente
Detalles del empleado	experiencia_profesional	Experiencia Profesional			Sin datos.		Idem, pero el formato se desconoce
Detalles de la rotación	desde3	Inicio	fecha	Timestamp (D)	Fecha Serial de Excel (Días desde 1/01/1900). DD-MM-AAAA	41399	
Detalles de la rotación	hasta3	Hasta	fecha	Timestamp (D)	Fecha Serial de Excel (Días desde 1/01/1900). DD-MM-AAAA	41399	

Detalles de la rotación	clase_medida	Denominación de la clase de me	cadena (varchar)	Categoría	Texto capitalizado. Puede contener símbolos (paréntesis, puntos).	Baja Contratación Cambio de puesto de trabajo Cambio de contrato Reingreso empresa TEMPORALES Reingreso a la empresa Reingreso a la empresa (ONB) Cambio de centro Jubilación parcial Cambio centro (LIQ.PAGA EXTRA)	
Detalles de la rotación	motivo_medida	Denominación del motivo de med	cadena (varchar)	Categoría	Texto capitalizado. Puede contener símbolos (guiones, puntos, paréntesis)	Art.159-1-Mutuo acuerdo Partes Reemplazo Baja voluntaria Art.159-4-Término Plazo Conve	
Detalles de la rotación	alta	Alta	fecha	Timestamp (D)	Fecha Serial de Excel (Días desde 1/01/1900). DD-MM-AAAA	41399	Alta de la persona
Detalles de la rotación	baja	Baja	fecha	Timestamp (D)	Fecha Serial de Excel (Días desde 1/01/1900). DD-MM-AAAA	41399	Cuando se dio de baja
Detalles de la rotación	encargado_superior	Nombre del superior (GO)	cadena (varchar)	nombre	APELLIDOS, NOMBRES (todo en mayúsculas)	CARPENTER SOSA, SABRINA JIMENA	



Geovictoria día

obs	Campo Procesado	Campo	Tipo de dato sugerido	Tipo de información	Convención	Valores posibles	Descripción
Detalles del empleado	grupo	Grupo	cadena (varchar)	Categoría	Texto en mayúsculas (ADMINISTRACION, etc.).	ADMINISTRACION Arica - Sidesa AVI	Nombre del grupo o área organizacional del trabajador.
Detalles de la asistencia	asistio_en	Fecha	fecha	Timestamp (D)	Día corto + DD-MM-YYYY (ej. Lun 29-04-2024).		Fecha del registro de asistencia.
Detalles de la asistencia	tipo_permiso	Permiso	cadena (varchar)	Categoría	Texto con primera letra en mayúscula (Ninguno, Vacaciones, etc.).	Permiso Ninguno Vacaciones	Indica si el trabajador posee algún permiso o licencia ese día.
Detalles del empleado	tipo_turno	Turno	cadena (varchar)	Texto descriptivo	Formato HH:MM - HH:MM(x mins) o Descanso.	HH:MM - HH:MM(mm "mins") Descanso No Planificado	Horario teórico de trabajo asignado al trabajador.
Detalles de la asistencia	hora_ingreso	Entró	hora	Timestamp (H)	HH:MM (24 horas).		Hora real en que el trabajador marcó su ingreso.
Detalles de la asistencia	atraso	Atraso	hora	Duración (hrs)	H:MM, con 0:00 si no hubo atraso.		Minutos u horas de atraso reales comparados con el turno.
Detalles de la asistencia	hora_salida	Salió	hora	Timestamp (H)	HH:MM (24 horas).		Hora real de salida registrada.
Detalles de la asistencia	adelanto	Adelanto	hora	Duración (hrs)	H:MM, con 0:00 si no hubo adelanto.		Minutos u horas que el trabajador salió antes del horario teórico.
Detalles de la asistencia	total_horas	Horas Totales	hora	Duración (hrs)	H:MM.		Cantidad total de horas efectivamente trabajadas ese día.
Detalles del empleado	id_empleado	Cargo	entero	ID	Numérico, sin separadores (101318).	≈320 unicos	Código o identificador interno del cargo del trabajador. Equivalente al Código SAP

Geovictoria semana

obs	Campo Procesado	Campo	Tipo de dato sugerido	Tipo de información	Convención	Valores posibles	Descripción
Detalles del empleado	grupo	Grupo	cadena (varchar)	Categoría	Texto en mayúsculas, exceptuando "Arica - Sidesa".	ADMINISTRACION Arica - Sidesa AVI BARES	Grupo o área organizacional del trabajador.
Detalles de la asistencia agregada	inicio_semana	Fecha Inicio	fecha	Timestamp (D)	DD-MM-YYYY.		Primer día del período semanal considerado.
Detalles de la asistencia agregada	fin_semana	Fecha Fin	fecha	Timestamp (D)	DD-MM-YYYY.		Último día del período semanal considerado.
Detalles de la asistencia agregada	horas_trabajadas	Horas Trabajadas	hora	Duración (hrs)	H:MM.		Total de horas efectivas trabajadas durante la semana.
Detalles de la asistencia agregada	horas_extra	Horas Extra	hora	Duración (hrs)	H:MM.		Total de horas trabajadas por sobre la jornada normal.
Detalles de la asistencia agregada	horas_no_trabajadas	Horas No Trabajadas	hora	Duración (hrs)	H:MM.		Total de horas no trabajadas dentro de la jornada semanal (ausencias, permisos, etc.).
Detalles del empleado	id_empleado	Cargo	entero	ID	Número, sin separadores (101318).	≈320 únicos	Código o identificador interno del cargo del trabajador.

Resumen anual de capacitación

obs	Campo Procesado	Campo	Tipo de dato sugerido	Tipo de la información	Convención	Valores Posibles	Descripción
Detalles de la realización de la capacitación	nro_capacitación	N°	entero	ID (no unica)	Número entero.	7, 39, ...	Identificador numérico único del registro de capacitación.
Detalles de la capacitación	mes	Mes	cadena (varchar)	Categoría	Texto en MAYÚSCULA S.	MARZO, MAYO, ...	Mes en que se reporta la capacitación.
Detalles de la capacitación	titulo	Título capacitación	cadena (varchar)	Categoría	Texto.	Ver lista de "CAPACITACIONES"	Nombre oficial del curso o programa.
Detalles de la capacitación	lugar	Lugar de impartición	cadena (varchar)	Categoría	Texto en MAYÚSCULA S.	Presencial, Online, Virtual	Modalidad de la capacitación.
Detalles de la capacitación	fecha_inicio	Fecha inicio	fecha	Timestamp (D)	Formato DD-mes (ej. 20-mar.).		Fecha de inicio del curso.
Detalles de la capacitación	fecha_fin	Fecha fin	fecha	Timestamp (D)	Formato DD-mes (ej. 20-mar.).		Fecha de finalización del curso.
Detalles de la capacitación	objetivo_area	Objetivo / Área temática	cadena (varchar)	Categoría	Texto.	Negocio, Cliente, Cultura, Digital, Liderazgo	Categoría principal a la que pertenece el curso.
Detalles de la capacitación	externo_interno	Externo / Interno	cadena (varchar)	Categoría	Texto en MAYÚSCULA S.	INTERNO, Externo, Externo / Interno	Indica si el formador es de la empresa o externo.
Detalles de la capacitación	tipo_curso	Tipo de curso	cadena (varchar)	Categoría	Texto en MAYÚSCULA S.	PRESENCIAL, ...	(Parece describir la modalidad, similar a 'Lugar de impartición').

Detalles de la realización de la capacitación	gerencia	Gerencia	cadena (varchar)	Categoría	Texto en MAYÚSCULA S.	Perú, Croacia, Copiapó, Arica	Gerencia que organiza o recibe la capacitación.
Detalles de la realización de la capacitación	formador_proveedor	Formador / Proveedor	cadena (varchar)	Texto	Texto.	RR.HH, Anglosaxon, ...	Nombre de la persona o entidad que imparte el curso.
Detalles de la realización de la capacitación	nro_asistentes	Asistentes	entero	Cantidad	Numérico entero.	15, 10, ...	Número de participantes en el curso.
Detalles de la realización de la capacitación	horas_persona	Horas	entero	Cantidad (hrs)	Numérico.	2, 12, ...	Duración en horas del curso por persona.
Detalles de la realización de la capacitación	total_horas	Horas totales	entero	Cantidad (hrs)	Numérico.	30, 120, ...	Horas multiplicadas por el número de asistentes.
Detalles de la realización de la capacitación	coste	Coste total	decimal	Cantidad (eur)	Formato XX,XX € (coma decimal).	69,82 €, ...	Costo total del curso en Euros.
Detalles de la realización de la capacitación	valoracion_formador	Valoración del formador	decimal	Rango (eur)	Numérico.	(Datos de ejemplo vacíos - €)	Calificación (ej. 1-5) del formador.
Detalles de la realización de la capacitación	indice_satisfaccion	Índice de satisfacción	decimal	Rango	Numérico (posiblemente 0-100).	(Datos de ejemplo vacíos)	Porcentaje de satisfacción de los asistentes.
Detalles de la realización de la capacitación	nps	NPS	entero	Rango	Numérico (ej. -100 a 100).	(Datos de ejemplo vacíos)	Net Promoter Score de la capacitación.

Participantes de las capacitaciones

obs	Campo Procesado	Campo	Tipo de dato sugerido	Tipo de la información	Convención	Valores Posibles	Descripción
Detalles de la realización de la capacitación	nro_capitacion	N°	entero	ID (no única)	Numérico entero.	1, ...	Identificador numérico único del registro de participante.
Detalles de la realización de la capacitación	mes	MES	cadena (varchar)	Categoría	Texto en MAYÚSCULAS.	MARZO, ...	Mes en que se realizó el curso.
Detalles de la realización de la capacitación	rut	Rut	cadena (varchar)	Texto único (rut)	Formato XXXXXXXX-X (sin puntos).	21149296-1, ...	RUT del participante (distinto al de la tabla inicial).
Detalles del empleado	id_empleado	N° EMPLEADO	entero	ID	Numérico entero.	103963, ...	ID del empleado (probablemente coincide con ID SAP).
Detalles del empleado	nombre	NOMBRE	cadena (varchar)	Texto	Texto en MAYÚSCULAS.	YEREZIM DIANA, ...	Nombres del participante.
Detalles del empleado	apellidos	APELLIDOS	cadena (varchar)	Texto	Texto en MAYÚSCULAS.	JIMENEZ HIDALGO, ...	Apellidos del participante.
Detalles del empleado	correo	CORREO	cadena (varchar)	Texto (correo)	Formato usuario@dominio.cl.	YHURTADO@antayhoteles.cl, ...	Email corporativo del participante.
Detalles del empleado	titulo	NOMBRE CURSO	cadena (varchar)	Categoría	Ver lista de " CAPACITACIONES "	Ver lista de "CAPACITACIONES"	Nombre del curso al que asistió.
Detalles del empleado	horas_persona	TOTAL HRS FORMACION	entero	Cantidad	Numérico.	2, ...	Total de horas de formación recibidas (coincide con 'Horas' de la tabla anterior).

Perfiles de trabajo

obs	Campo Procesado	Campo	Tipo de dato sugerido	Tipo de la información	Convención	Valores Posibles	Descripción
Detalles del Puesto de Trabajo	puesto	Puesto	cadena (varchar)	Texto	Texto	Ej: "Ejecutivo Comercial"	Nombre del puesto de trabajo.
Detalles del Puesto de Trabajo	categoria_puesto	Categoría	cadena (varchar)	Texto	Texto	Ej: "Técnico y Auxiliar"	Categoría laboral del puesto.
Detalles del Puesto de Trabajo	nombre_puesto	Nombre	cadena (varchar)	Texto	Texto	Puede estar vacío.	Nombre de quién documenta el puesto.
Detalles del Puesto de Trabajo	fecha_creacion	Fecha	fecha	Timestamp (D)	Formato de fecha (ej.DD/MM/AAAA)	N/A	Fecha de creación o actualización de la ficha.
Auto Referencia del Puesto de Trabajo	linea_asc	Línea ascendente	cadena (varchar)	Lista	Texto	N/A	Define de quién depende el puesto o a quién reporta.
Auto Referencia del Puesto de Trabajo	linea_desc	Línea Descendente	cadena (varchar)	Lista	Texto	N/A	Define si el puesto tiene colaboradores a cargo.
Detalles del Puesto de Trabajo	mision	Principal mision	texto	Texto	Texto libre	N/A	Párrafo que describe la misión principal del puesto.
Detalles del Puesto de Trabajo	responsabilidades	Responsabilidades	texto	Texto	Texto libre	N/A	Párrafo que describe las responsabilidades generales.
Detalles del Puesto de Trabajo	informa_y_reporta	Informa y Reporta que... a...	cadena (varchar)	Texto	Texto	N/A	Especifica los roles a los que reporta (similar a Línea Ascendente).
Detalles del Puesto de Trabajo	tareas_semanales	Tareas semanales	texto (lista)	Lista	Texto / Lista	N/A	Listado de tareas con frecuencia semanal.
Detalles del Puesto de Trabajo	tareas_mensuales	Tareas mensuales	texto (lista)	Lista	Texto / Lista	N/A	Listado de tareas con frecuencia mensual.

Detalles del Puesto de Trabajo	tareas_trimestrales	Tareas Trimestrales	texto (lista)	Lista	Texto / Lista	N/A	Listado de tareas con frecuencia trimestral.
Detalles del Puesto de Trabajo	tareas_semestrales	Tareas Semestrales	texto (lista)	Lista	Texto / Lista	N/A	Listado de tareas con frecuencia semestral.
Detalles del Puesto de Trabajo	tareas_anuales	Tareas Anuales	texto (lista)	Lista	Texto / Lista	N/A	Listado de tareas con frecuencia anual.
Detalles del Puesto de Trabajo	otras_tareas	Otras	texto (lista)	Lista	Texto / Lista	N/A	Tareas adicionales no cubiertas en otras frecuencias.
Detalles del Puesto de Trabajo	competencias	Competencias	texto (lista)	Lista	Texto / Lista	N/A	Listado de competencias conductuales requeridas.
Detalles del Puesto de Trabajo	habilidades_generales	Habilidades y conocimientos requeridos Generales	texto (lista)	Lista	Texto / Lista	N/A	Listado de habilidades y conocimientos generales.
Detalles del Puesto de Trabajo	habilidades_cognitivas	Habilidades y conocimientos requeridos Cognitivos	texto (lista)	Lista	Texto / Lista	N/A	Listado de habilidades y conocimientos cognitivos.
Detalles del Puesto de Trabajo	habilidades_fisicas	Habilidades y conocimientos requeridos Fisicos	texto (lista)	Lista	Texto / Lista	N/A	Listado de habilidades y conocimientos físicos.
Detalles del Puesto de Trabajo	habilidades_sensoriales	Habilidades y conocimientos requeridos Sensoriales	texto (lista)	Lista	Texto / Lista	N/A	Listado de habilidades y conocimientos sensoriales.
Detalles del Puesto de Trabajo	condiciones_trabajo	Condiciones de trabajo	texto (lista)	Lista	Texto / Lista	N/A	Descripción de las condiciones del entorno laboral.
Detalles del Puesto de Trabajo	apartado_legal	Apartado legal	texto	Texto / Firma	Texto y/o campo de firma	N/A	Párrafo legal de aceptación de términos y la firma.

Proceso de contratación/selección

obs	Campo Procesado	Campo	Tipo de dato sugerido	Tipo de la información	Convención	Valores Posibles	Descripción
Detalles del proceso de selección	fecha_cierre	FECHA CIERRE PROCESO	fecha	Timestamp (D)	Número (serial de fecha Excel).	45425, ...	Fecha de cierre. El número es un serial de Excel.
Detalles del Puesto de trabajo	cargo	PUESTO	cadena (varchar)	Categoría	Texto en MAYÚSCULAS.	ASISTENTE CONTABLE, ...	Nombre del cargo a cubrir.
Detalles del Puesto de trabajo	ceco	CENTRO AREA	cadena (varchar)	Categoría	Texto.	Casino, ...	Área o centro de costo solicitante.
Detalles del Puesto de trabajo	-	DETALLE PUESTO	cadena (varchar)	Categoría	Texto en MAYÚSCULAS.	ASISTENTE CONTABLE, ...	Descripción o nombre detallado del puesto.
Detalles del proceso de selección	detalle	SITUACION	cadena (varchar)	Texto	Texto largo.	Vamos a evaluar junto con..., ...	Estado o descripción actual del proceso.
Detalles del Puesto de trabajo	gerencia	GERENCIA	cadena (varchar)	Categoría	Texto en MAYÚSCULAS.	ARICA, ...	Gerencia solicitante.
Detalles del Puesto de trabajo	linea_negocio	LINEA NEGOCIO	cadena (varchar)	Texto	Texto.	Casino Hotel, ...	Unidad de negocio.
Detalles del Puesto de trabajo	grupo	GRUPO	cadena (varchar)	Categoría	Texto (Alfanumérico).	G6, ...	Clasificación o grupo salarial del puesto.
Detalles del proceso de selección	motivo	MOTIVO	cadena (varchar)	Categoría	Texto.	Despido (sustitución), ...	Causa de la vacante.
Detalles del proceso de selección	-	DETALLE MOTIVO	cadena (varchar)	Texto	Texto largo.	Búsqueda anticipada renuncia..., ...	Explicación del motivo de la vacante.
Detalles del proceso de selección	publicado_interno	PUBLICADO PORTAL INTERNO	booleano / cadena	Categoría			Indica si se publicó en la plataforma interna de la empresa.

Detalles del proceso de selección	-	FECHA ENVIO RESPONSABLE	fecha	Timestamp (D)	Numérico (serial de fecha Excel).	45594, ...	Cuando se le informa al responsable de la selección
Detalles del proceso de selección	-	FECHA ENTREVISTA RESPONSABLE	fecha	Timestamp (D)	Numérico (serial de fecha Excel).	45600, ...	Cuando el responsable entrevista
Detalles del proceso de selección	-	DURACION ENTREVISTA GERENTE	hora (o minutos)	Duración	Numérico.	0, ...	Duración en horas o días de la entrevista.
Detalles del proceso de selección	duracion_dias	DURACION DIAS TOTAL	hora (o minutos)	Duración	Numérico entero.	0, ...	Días totales que tomó el proceso.
Detalles del proceso de selección	fuentes_reclutamiento	FUENTE RECLUTAMIENTO	cadena (varchar)	Texto (o categoría)	Texto.	Feria Laboral, ...	Canal por el que se obtuvo al candidato.
Detalles del proceso de selección	-	DETALLE FUENTE	cadena (varchar)	Texto	Texto.		Detalle de la fuente por la cual se realizó el proceso de selección.
Detalles del empleado	nombre	PERSONA	cadena (varchar)	Categoría	Texto.	Mireya Carrillo, ...	Nombre del candidato seleccionado o finalista.
Detalles del empleado	sexo	SEXO	cadena (varchar)	Categoría	Texto.	Masculino, Femenino	Sexo del candidato.
Detalles del empleado	edad	EDAD	entero	Cantidad (edad)	Numérico entero.	32, ...	Edad del candidato.
Detalles del empleado	formacion	FORMACIÓN	cadena (varchar)	Categoría	Texto.	Bachillerato/Ed Media completa, ...	Nivel educacional del candidato.
Detalles del empleado	anos_experiencia	AÑOS EXPERIENCIA	entero	Cantidad (años)	Numérico.	12, ...	Años de experiencia laboral del candidato.
Detalles del empleado	sector_procedencia	SECTOR DE PROCEDENCIA	cadena (varchar)	Texto	Texto.	hotel, ...	Industria o sector del que proviene el candidato.

Detalles del empleado	activo	ACTIVO	booleano	Lógico	Texto.	si, no	Si se encuentra todavía trabajando la persona seleccionada.
Detalles del empleado	tiene_continuidad	CONTINUIDAD (+ DE 4 MESES)	booleano	Lógico			Si la persona seleccionada tiene continuidad dentro de la empresa (más de 4 meses).
Detalles del proceso de selección	-	OBSERVACIONES	cadena (varchar)	Texto	Texto largo.	Para compensar el equipo..., ...	Comentarios adicionales sobre el proceso.
Detalles del proceso de selección	-	COSTE PROCESO	decimal	Cantidad (eur)			Costo total del proceso de selección.
Detalles del proceso de selección	nro_cvs_recibidos	N CV RECIBIDOS	entero	Cantidad			Número de currículums recibidos.
Detalles del proceso de selección	nro_personas_entrevistadas_telefono	N PERSONAS ENTREVISTADAS TELEFONO	entero	Cantidad			Número de entrevistas telefónicas.
Detalles del proceso de selección	nro_personas_entrevistadas_presencial	N PERSONAS ENTREVISTADAS PRESENCIAL	entero	Cantidad			Número de entrevistas presenciales.
Detalles del proceso de selección	nro_personas_finalistas	N PERSONAS FINALISTAS	entero	Cantidad			Número de finalistas en el proceso.
Detalles del proceso de selección	-	ORIGEN	cadena (varchar)	Texto	Texto en MAYÚSCULAS.	ARICA	Origen geográfico del proceso.
Detalles del proceso de selección	-	TOTAL	entero	Cantidad	Numérico entero.	1, ...	Total de procesos de selección realizados
Detalles del proceso de selección	id_proceso	CONTROL INTERNO	entero	ID	Numérico entero.	6, ...	ID del proceso de selección

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 (Transaccional)

Esta tabla modela la cantidad de capacitaciones que se realizan y, crucialmente, la asistencia efectiva de los colaboradores a cada una de ellas. El fin es poder analizar tanto el impacto de las capacitaciones sobre la población total como a nivel individual del empleado.

Responde preguntas sobre la participación, el costo y la efectividad de la capacitación.

Ejemplos de preguntas:


- "¿Cuántas *horas_ppersona* de capacitación recibió el area de 'AA&BB BARES'?"
- "¿Cuál es el *indice_satisfaccion promedio* (desde *DimCapacitacion*) para los cursos de *objetivo_area* = 'Negocio'?"
- "¿Qué empleados no han recibido ninguna capacitación (*horas_ppersona*) en el último año?"
- "¿Cuál es el *coste por hora* de capacitación para los cursos de *objetivo_area* = 'Liderazgo'?"
- "¿Hay una correlación entre la *valoracion_formador* y el *indice_satisfaccion* de los asistentes?"
- "¿Qué *formador_proveedor* genera el NPS más alto?"

5. 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

Se contemplan estas dimensiones para el modelado multidimensional:

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	Tipo de SCD
Empleado_SK	INTEGER	Clave primaria sustituta (generada por ETL).	N/A
id_empleado_NK	VARCHAR	ID SAP / ID del empleado (clave natural).	SCD 1
id_proceso_NK	VARCHAR	ID del proceso de selección (clave natural de candidato).	SCD 1
rut_NK	VARCHAR	RUT del participante (clave natural de capacitación).	SCD 1
nombre	VARCHAR	Nombre completo del empleado o candidato.	SCD 1
correo	VARCHAR	Email corporativo del participante.	SCD 1
sexo	VARCHAR	Sexo (masculino/femenino).	SCD 1
fecha_nacimiento	DATE	Fecha de nacimiento del empleado.	SCD 1
pais_nacimiento	VARCHAR	País de nacimiento del empleado.	SCD 1
nacionalidad	VARCHAR	Nacionalidad del empleado.	SCD 1
estado_civil	VARCHAR	Estado civil del empleado.	SCD 1
nro_hijos	INTEGER	Número de hijos del empleado.	SCD 1
formacion	VARCHAR	Nivel educacional del candidato seleccionado.	SCD 1
agnos_experiencia	INTEGER	Años de experiencia laboral del candidato seleccionado.	SCD 1
cargo	VARCHAR	Cargo actual o histórico del empleado.	SCD 2

area	VARCHAR	Área actual o histórica del empleado.	SCD 2
ceco	VARCHAR	Centro de Coste actual o histórico.	SCD 2
gerencia	VARCHAR	Gerencia organizacional actual o histórica.	SCD 2
grupo	VARCHAR	Grupo o área organizacional actual o histórica.	SCD 2
empresa	VARCHAR	Razón social de la empresa actual o histórica.	SCD 2
tipo_empleo	VARCHAR	Tipo de empleado (Temporal, Empleado, etc.).	SCD 2
jornada	VARCHAR	Jornada laboral (Part time, Full time, etc.).	SCD 2
SCD_StartDate	DATETIME	Fecha de inicio de vigencia del registro.	N/A
SCD_EndDate	DATETIME	Fecha de fin de vigencia del registro.	N/A
SCD_IsCurrent	BOOLEAN	Indicador de registro actual (vigente).	N/A

Dimensión Tiempo

Dimensión estática (SCD 0) para llevar registro del tiempo.

Nombre de Columna	Tipo de Dato	Descripcion
Tiempo_SK	INTEGER	Clave primaria (Formato AAAAMMDD).
Fecha_Completa	DATE	Fecha completa (ej. 2024-05-15).
Dia_Semana_Nombre	VARCHAR	Nombre del día (ej. 'Miércoles').
Dia_Mes_Num	SMALLINT	Número del día en el mes (ej. 15).
Semana_Año_Num	SMALLINT	Número de la semana en el año (ej. 20).
Mes_Nombre	VARCHAR	Nombre del mes (ej. 'Mayo').
Mes_Año_Num	SMALLINT	Número del mes en el año (ej. 5).
Trimestre_Nombre	VARCHAR	Nombre del trimestre (ej. 'Q2-2024').
Año_Num	SMALLINT	Año (ej. 2024).
Indicador_FinDeMes	BOOLEAN	Indicador de fin de mes (True/False).
Indicador_DiaLaboral	BOOLEAN	Indicador de día laboral (True/False).

Dimensión Cargo

Nombre de Columna	Tipo de Dato	Descripcion	Tipo de SCD
Cargo_SK	INTEGER	Clave primaria sustituta.	N/A
puesto	VARCHAR	Clave natural (Nombre del puesto).	SCD 1
categoria_puesto	VARCHAR	Categoría laboral del puesto.	SCD 1
mision	TEXT	Misión principal del puesto.	SCD 1
responsabilidades	TEXT	Responsabilidades generales.	SCD 1
informa_y_reporta	VARCHAR	Roles a los que reporta.	SCD 1
tareas_semanales	TEXT	Listado de tareas semanales.	SCD 1
competencias	TEXT	Listado de competencias conductuales.	SCD 1
habilidades_generales	TEXT	Listado de habilidades generales.	SCD 1
habilidades_cognitivas	TEXT	Listado de habilidades cognitivas.	SCD 1
habilidades_fisicas	TEXT	Listado de habilidades físicas.	SCD 1
habilidades_sensoriales	TEXT	Listado de habilidades sensoriales.	SCD 1
condiciones_trabajo	TEXT	Descripción de las condiciones laborales.	SCD 1

Dimensión Capacitación

Nombre de Columna	Tipo de Dato	Descripción	Tipo de SCD
Capacitacion_SK	INTEGER	Clave primaria sustituta.	N/A
nro_capacitacion_NK	INTEGER	Clave natural (ID del evento).	SCD 1
titulo	VARCHAR	Nombre oficial del curso.	SCD 1
mes	VARCHAR	Mes en que se reporta.	SCD 1
lugar	VARCHAR	Modalidad (Presencial, Online, etc.).	SCD 1
fecha_inicio	DATE	Fecha de inicio del curso.	SCD 1
fecha_fin	DATE	Fecha de finalización del curso.	SCD 1
objetivo_area	VARCHAR	Categoría principal (Negocio, Cliente, etc.).	SCD 1
externo_interno	VARCHAR	Origen del formador (Interno/Externo).	SCD 1
tipo_curso	VARCHAR	Tipo de curso (Presencial, etc.).	SCD 1
gerencia	VARCHAR	Gerencia que organiza o recibe.	SCD 1
formador_proveedor	VARCHAR	Nombre del formador o proveedor.	SCD 1
nro_asistentes	INTEGER	Número de participantes (atributo del evento).	SCD 1
total_horas	INTEGER	Horas x Asistentes (atributo del evento).	SCD 1
coste	DECIMAL	Costo total del curso (atributo del evento).	SCD 1
valoracion_formador	DECIMAL	Calificación del formador (atributo del evento).	SCD 1
indice_satisfaccion	DECIMAL	Satisfacción de asistentes (atributo del evento).	SCD 1
nps	INTEGER	NPS de la capacitación (atributo del evento).	SCD 1

Mini Dimensiones

Separadas debido a su longitud mediana.

Dimensión Clase Medida (Rotación)

SCD Tipo 1.

Nombre de Columna	Tipo de Dato	Descripcion
Clase_Medida_SK	INTEGER	Clave primaria sustituta.
clase_medida	VARCHAR	Valor de la clase de medida (ej. 'Baja', 'Contratación').

Dimensión Motivo Medida (Rotación)

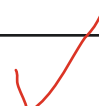
SCD Tipo 1.

Nombre de Columna	Tipo de Dato	Descripcion
Motivo_Medida_SK	INTEGER	Clave primaria sustituta.
motivo_medida	VARCHAR	Valor del motivo de medida (ej. 'Renuncia voluntaria', 'Cambio de puesto').

Dimensión Empresa (Rotación)

SCD Tipo 1.

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



Dimensión Tipo Permiso (Asistencia)

SCD Tipo 1.

Nombre de Columna	Tipo de Dato	Descripcion
Permiso_SK	INTEGER	Clave primaria sustituta.
tipo_permiso	VARCHAR	Tipo del permiso utilizado en la justificación de inasistencia.

Dimensión Tipo Turno (Asistencia)

SCD Tipo 1.

Nombre de Columna	Tipo de Dato	Descripcion
Turno_SK	INTEGER	Clave primaria sustituta.
es_descanso	BOOLEAN	Si la persona pidió descanso el día.
es_no_planificado	BOOLEAN	Si el turno no fue planificado para el día.
hora_inicio_planificada	TIME	Hora de inicio del turno si fue planificado.
hora_fin_planificada	TIME	Hora de fin del turno si fue planificado.
minutos_descanso_planificados	INTEGER	Minutos de descanso planificados para el turno.
duracion_turno_horas	DECIMAL	Duración calculada del turno.

Tablas de Hechos

Rotacion de Empleados

Granularidad: Una fila por evento de acción de RRHH (clase_medida SAP).

Tipo de Hecho: Transaccional Factless. Registra eventos puntuales en el tiempo.

Nombre de Columna	Tipo de Dato	Descripcion	Rol
Tiempo_SK	INTEGER	Clave foránea a DimTiempo (fecha del evento).	Clave Foránea
Empleado_SK	INTEGER	Clave foránea a DimEmpleado (versión post-evento).	Clave Foránea
Clase_Medida_SK	INTEGER	Clave foránea a DimClaseMedida.	Clave Foránea
Motivo_Medida_SK	INTEGER	Clave foránea a DimMotivoMedida.	Clave Foránea
Empresa_SK	INTEGER	Clave foránea a DimEmpresa.	Clave Foránea

Dotación de Empleados

Granularidad: Una fila por empleado, por mes.

Tipo de Hecho: Snapshot Periódico. Toma una "foto" del estado al final de cada mes.

Nombre de Columna	Tipo de Dato	Descripcion	Rol
Mes_SK	INTEGER	Clave foránea a DimTiempo (último día del mes).	Clave Foránea
Empleado_SK	INTEGER	Clave foránea a DimEmpleado (versión vigente).	Clave Foránea
Headcount	INTEGER	Medida (siempre 1) para conteo de personal.	Medida (Semi-aditiva)
Antigüedad_Meses	INTEGER	Medida de antigüedad en meses a la fecha del snapshot.	Medida (Semi-aditiva)

Asistencia Diaria

Granularidad: Una fila por empleado, por día.

Tipo de Hecho: Transaccional. Registra el evento de asistencia de un día.

Nombre de Columna	Tipo de Dato	Descripcion	Rol
Tiempo_SK	INTEGER	Clave foránea a DimTiempo (día de asistencia).	Clave Foránea
Empleado_SK	INTEGER	Clave foránea a DimEmpleado (versión vigente).	Clave Foránea
Permiso_SK	INTEGER	Clave foránea a DimPermiso.	Clave Foránea
Turno_SK	INTEGER	Clave foránea a DimTurno.	Clave Foránea
atraso	DECIMAL	Medida de minutos/horas de atraso.	Medida
adelanto	DECIMAL	Medida de minutos/horas de salida temprana.	Medida
total_horas	DECIMAL	Medida de horas totales trabajadas.	Medida

Participación en Capacitaciones

Granularidad: Una fila por empleado, por evento de capacitación.

Tipo de Hecho: Transaccional.

Nombre de Columna	Tipo de Dato	Descripcion	Rol
Empleado_SK	INTEGER	Clave foránea a DimEmpleado (asistente).	Clave Foránea
Capacitacion_SK	INTEGER	Clave foránea a DimCapacitacion (evento).	Clave Foránea
Tiempo_SK	INTEGER	Clave foránea a DimTiempo (fecha fin del curso).	Clave Foránea
horas_ppersona	INTEGER	Medida de horas de capacitación recibidas.	Medida

✓

Procesos 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.

Existen dimensiones degeneradas 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

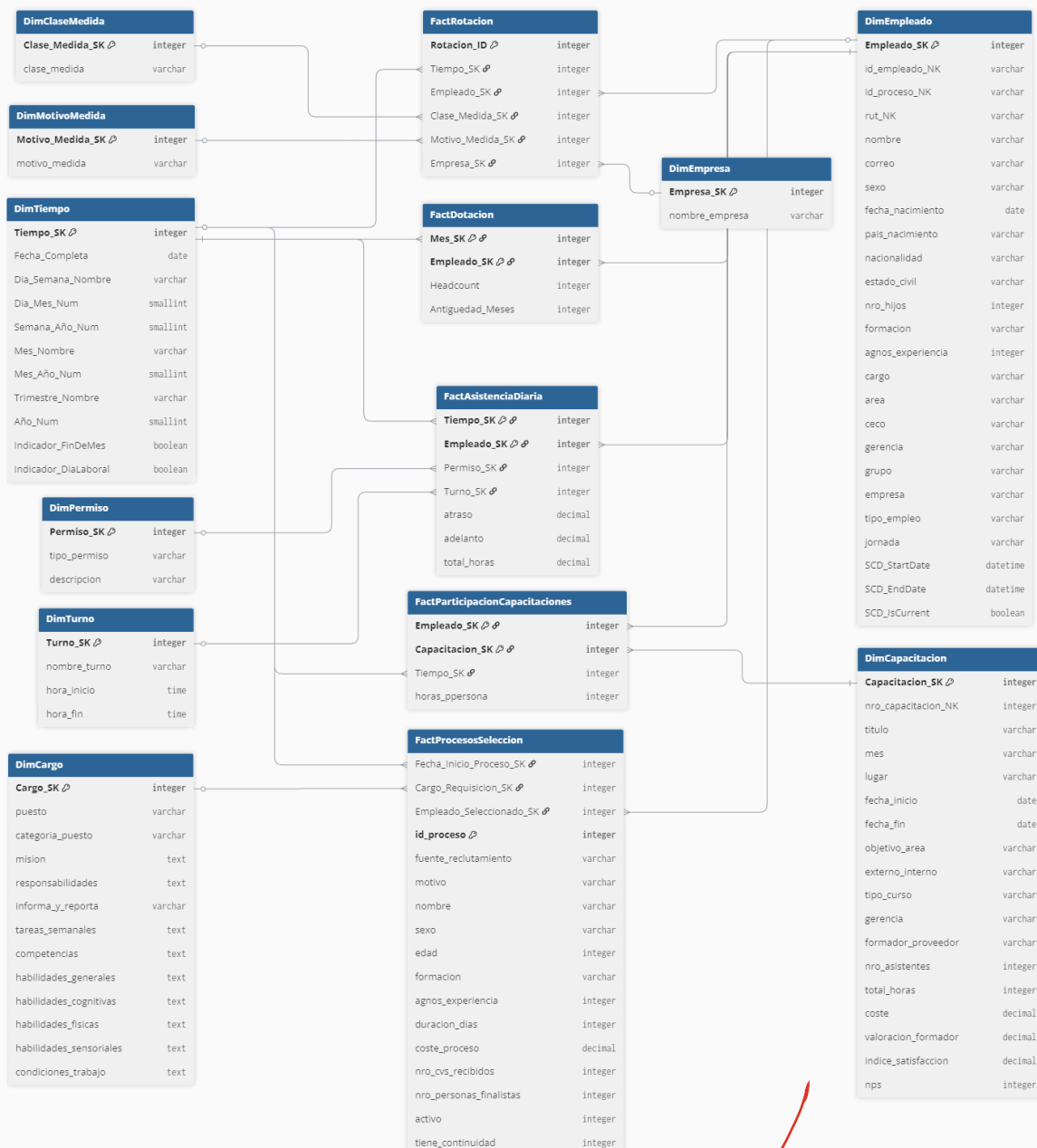
Nombre de Columna	Tipo de Dato	Descripcion	Rol
Fecha_Inicio_Proceso_SK	INTEGER	Clave foránea a DimTiempo (fecha inicio).	Clave Foránea
Cargo_Requisicion_SK	INTEGER	Clave foránea a DimCargo (cargo solicitado).	Clave Foránea
Empleado_Seleccionado_SK	INTEGER	Clave foránea a DimEmpleado (poblada al final).	Clave Foránea
id_proceso	INTEGER	ID del proceso de selección.	Dimensión Degenerada
fuentes_reclutamiento	VARCHAR	Canal por el que se obtuvo al candidato.	Dimensión Degenerada
motivo	VARCHAR	Causa de la vacante.	Dimensión Degenerada
nombre	VARCHAR	Nombre del candidato seleccionado.	Dimensión Degenerada
sexo	VARCHAR	Sexo del candidato seleccionado.	Dimensión Degenerada
edad	INTEGER	Edad del candidato seleccionado.	Dimensión Degenerada
formacion	VARCHAR	Nivel educacional del candidato seleccionado.	Dimensión Degenerada
agnos_experiencia	INTEGER	Años de experiencia del candidato seleccionado.	Dimensión Degenerada
duracion_dias	INTEGER	Medida de días totales del proceso.	Medida
coste_proceso	DECIMAL	Medida de costo total del proceso.	Medida

nro_cvs_recibidos	INTEGER	Medida de CVs recibidos.	Medida
nro_personas_finalistas	INTEGER	Medida de finalistas.	Medida
activo	INTEGER	Medida (1/0) si el seleccionado sigue activo.	Medida
tiene_continuidad	INTEGER	Medida (1/0) si el seleccionado superó 4 meses.	Medida



Modelo Multidimensional Graficado

El modelo multidimensional presentado es el resultado de un riguroso proceso de abstracción y diseño, orientado específicamente a satisfacer las necesidades analíticas y predictivas del proyecto. Más que un producto final inamovible, este esquema constituye una arquitectura de datos base para los indicadores de gestión de RRHH. Su diseño contempla la realidad dinámica de los datos: se entiende que la estructura deberá ser revisada y ajustada en un futuro para alinearse con la evolución de la organización. El objetivo de esta versión es proporcionar un punto de partida robusto que permita realizar las inevitables adaptaciones de forma controlada y eficiente. En la sección del anexo se puede encontrar un enlace para visualizar el modelo con una mayor resolución.



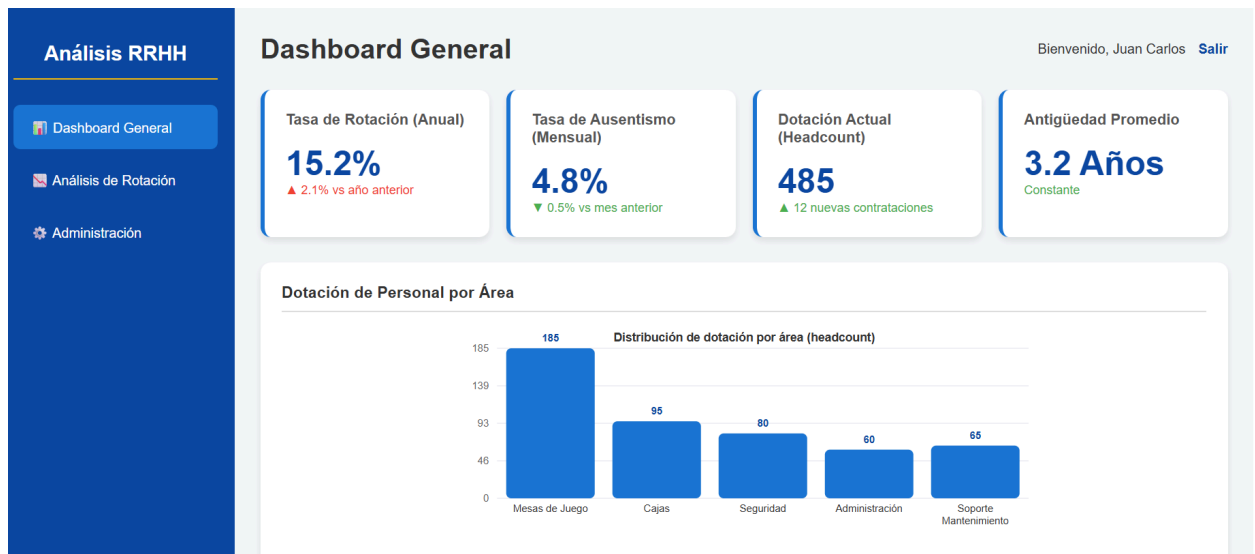
11. Diseño de Interfaz de Usuario y Prototipo

11.1. Mockups

Panel de Modelos Predictivos



Panel de KPIs



Panel de Carga y Administración

Análisis RRHH

Dashboard General

Análisis de Rotación

Administración

Administración del Sistema

Bienvenido, Juan Carlos [Salir](#)

Carga de Datos (ETL)

Sube los archivos Excel con los registros de asistencia, altas/bajas, etc.

Seleccionar Planilla Excel

Seleccionar archivo

Sin archivos seleccionados

Iniciar Carga y Transformación

Exportación de Datos

Exporta los datos consolidados del Data Warehouse a diferentes formatos.

Seleccionar Formato

CSV

Exportar Datos

Gestión de Acceso (AAA)

Administra los roles y permisos de los usuarios del sistema.

Usuario	Rol	Último Acceso	Acciones
jjampara	Administrador	2025-10-18 15:30	Editar Revocar
tsilva	Analista	2025-10-18 14:00	Editar Revocar
gerencia	Visualizador	2025-10-17 11:00	Editar Revocar



12. Conclusiones

Este documento ha detallado el plan y la arquitectura para abordar un problema tangible en el área de RRHH de Casino Luckia Arica: la fragmentación de sus datos. La situación actual se caracteriza por la dispersión de la información del personal, lo que deriva en un trabajo manual intensivo y limita la toma de decisiones a un plano reactivo. ✓

La solución propuesta es diseñar un Data Warehouse (un almacén de datos) para centralizar esta información. Se planea usar procesos ETL (Extracción, Transformación y Carga) para mover y limpiar los datos. El objetivo es que el área de RRHH pueda consultar métricas clave, como la rotación o el ausentismo, sin tener que armar reportes a mano cada vez.

Para gestionar el desarrollo, se eligió Kanban por su flexibilidad y menor carga administrativa, lo que parece adecuado para el tiempo ajustado del proyecto. El proceso de análisis de datos seguirá las fases de CRISP-DM, que es un estándar probado para proyectos de este tipo.

Si la ejecución de este plan es exitosa, se espera que la plataforma resultante reduzca significativamente la carga operativa. Más allá de la mera visualización de métricas en un dashboard, la verdadera prueba de éxito del proyecto residirá en la capacidad del modelo predictivo para anticipar la rotación de personal con un grado de precisión que resulte útil para la gestión. ✓

El plan es ambicioso. Su viabilidad dependerá críticamente de dos factores principales: primero, la calidad, completitud e integridad de los datos históricos que se logren extraer de los sistemas fuente (SAP y GeoVictoria) y, segundo, la disciplina en la ejecución de las tareas definidas en el backlog. Se espera que el diseño aquí presentado sienta una base técnica sólida para alcanzar dichos objetivos.

Para concretar esta visión, el presente diseño ha definido no solo la arquitectura, sino también la selección del stack tecnológico y el detalle exhaustivo del modelo de datos multidimensional. Se espera que esta definición sienta una base técnica sólida, permitiendo al equipo avanzar con confianza a la siguiente fase crucial: el desarrollo e implementación de los procesos ETL. ✓

Obs: Saquen el exceso, en todo caso es un muy buen informe.. hay aspecto que se contradicen.. en el uso de BI y Crisp-dm, usar BI está bien con el fin de limpiar los datos y Crisp-Dm se utiliza al final para predecir de acuerdo a su empleo fundamental..
Sacando el exceso y limpiando las contradicciones y pensando críticamente saquen la basura chatgpt que observo
Coloquen Índice de figuras y tablas y numeradas y citadas
Todo el documento justificado a ambos lados, hay partes que si otras que están a la izq.
Hay límites de los informes en la UTA son 50 páginas, yo acepto +-10%

13. 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. **hay recientes donde se convinan con bi..**
- [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.

14. 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