



Resumen de proyecto

Sistema de Análisis de RRHH para el Casino Luckia



Integrantes

- Tomás Silva
- Juan Yampara

Profesor a Cargo:

- Diego Aracena

Asignatura:

- Proyectos IV



Resumen del proyecto

Recursos Humanos en Casino Luckia Arica tiene fragmentación de sus datos, dificultando el análisis proactivo y la toma de decisiones estratégica.

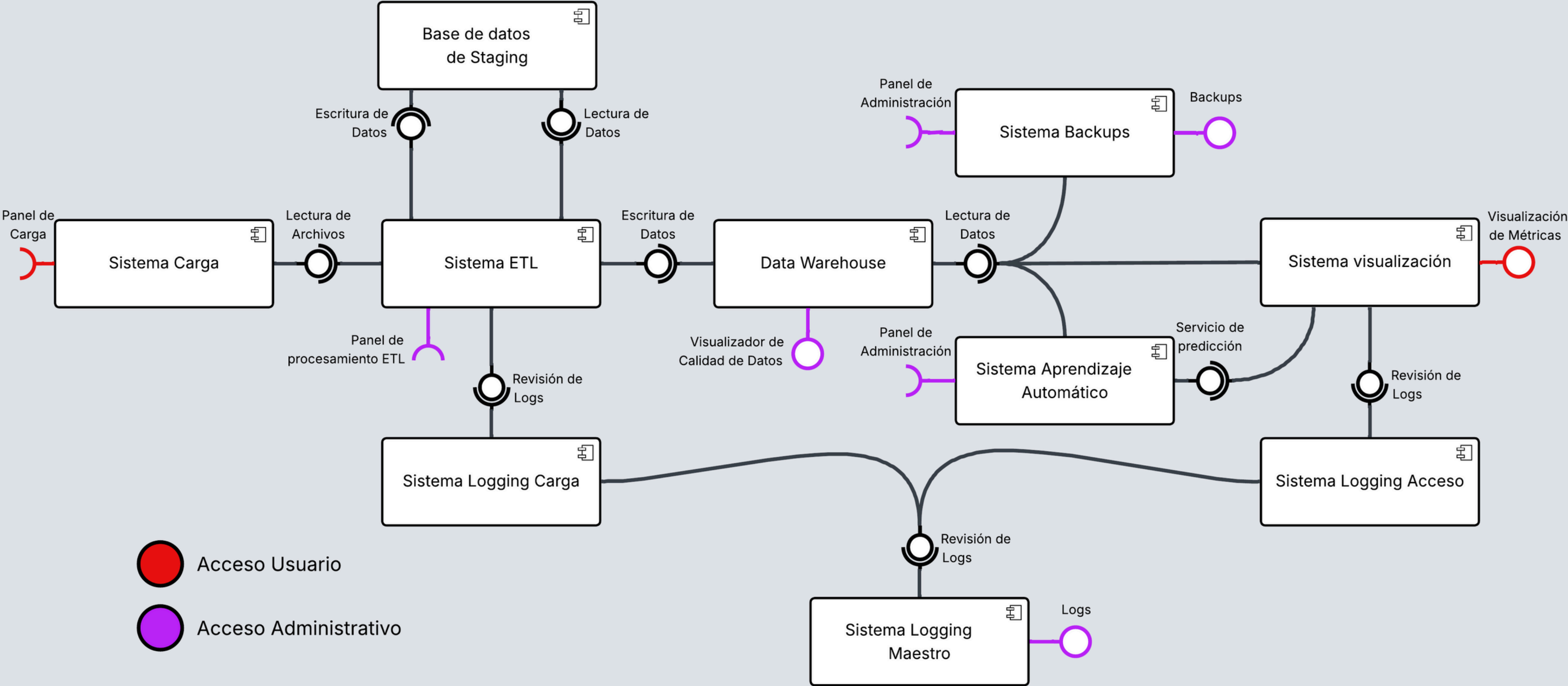
Se propone implementar una Plataforma Analítica para centralizar la información. Utilizando CRISP-DM y Kanban, se espera automatizar el cálculo de KPIs y presentarlos en dashboards interactivos, mejorando la gestión estratégica en RRHH.



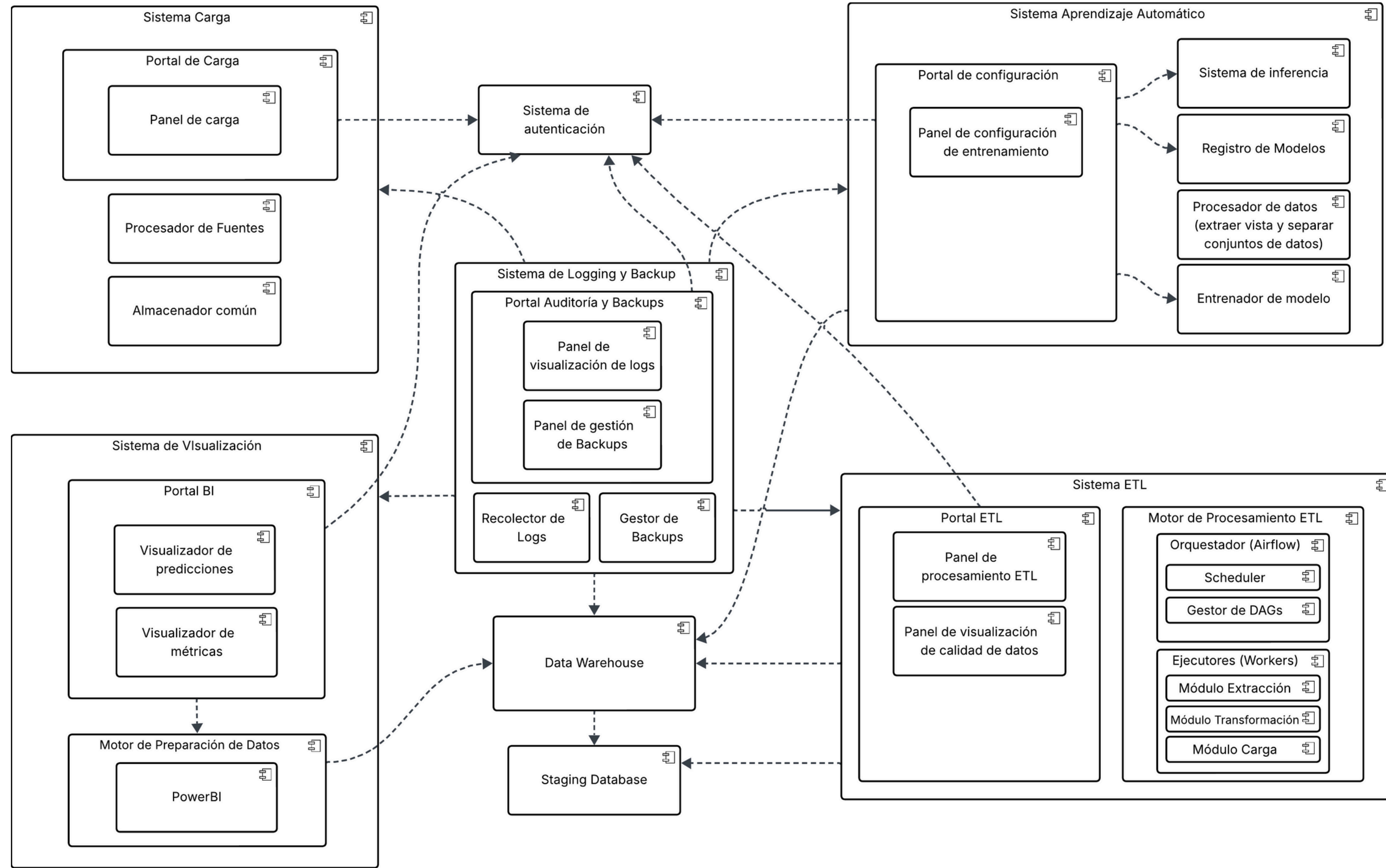
Modelo de Componentes lógicos

Modelo de Componentes lógicos

de Alto Nivel



Modelo de Componentes lógicos específico



Modelo de Componentes lógicos

1. Sistemas Primordiales

Sistema de Carga	Se encarga de recibir los datos en crudo y procesarlos hasta un nivel estándar mínimo
Sistema ETL	Ejecuta los Workflows, Integra los datos y los archiva en el Data Warehouse
Base de Datos	Provee el almacenaje temporal para el procesamiento del ETL
Data Warehouse	Provee el almacenaje de los datos integrados para su análisis y el entrenamiento de modelos de predicción
Sistema de Aprendizaje Automático	Es el sistema encargado de procesar los datos integrados para el entrenamiento de los modelos predictivos requeridos del sistema. Guarda los modelos entrenados en un Registro y expone su Servicio de Inferencia para permitir el acceso a la predicción
Sistema de Visualización	Toma tanto los datos del Data Warehouse como los resultados de la predicción de los modelos para mostrarlos en gráficos y en dashboards interactivos

Modelo de Componentes lógicos

1. Sistemas Secundarios

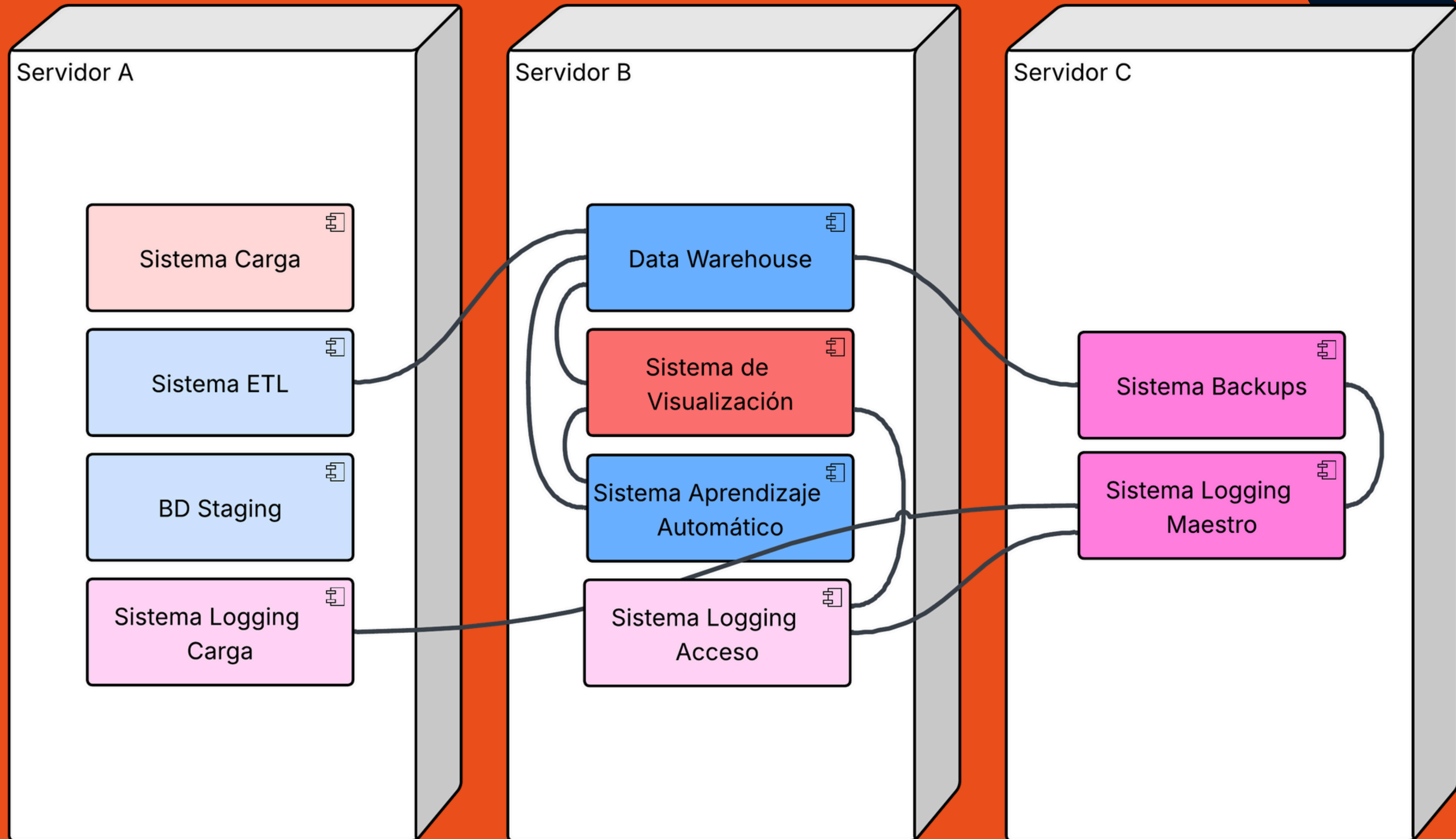
Sistemas de Logging
(Carga, Acceso y
Maestro)

Recopilan e integran los logs generados por los distintos subsistemas del macrosistema de análisis

Sistema de Backups

Gestiona las copias de seguridad periódicas de los datos procesados del Data Warehouse

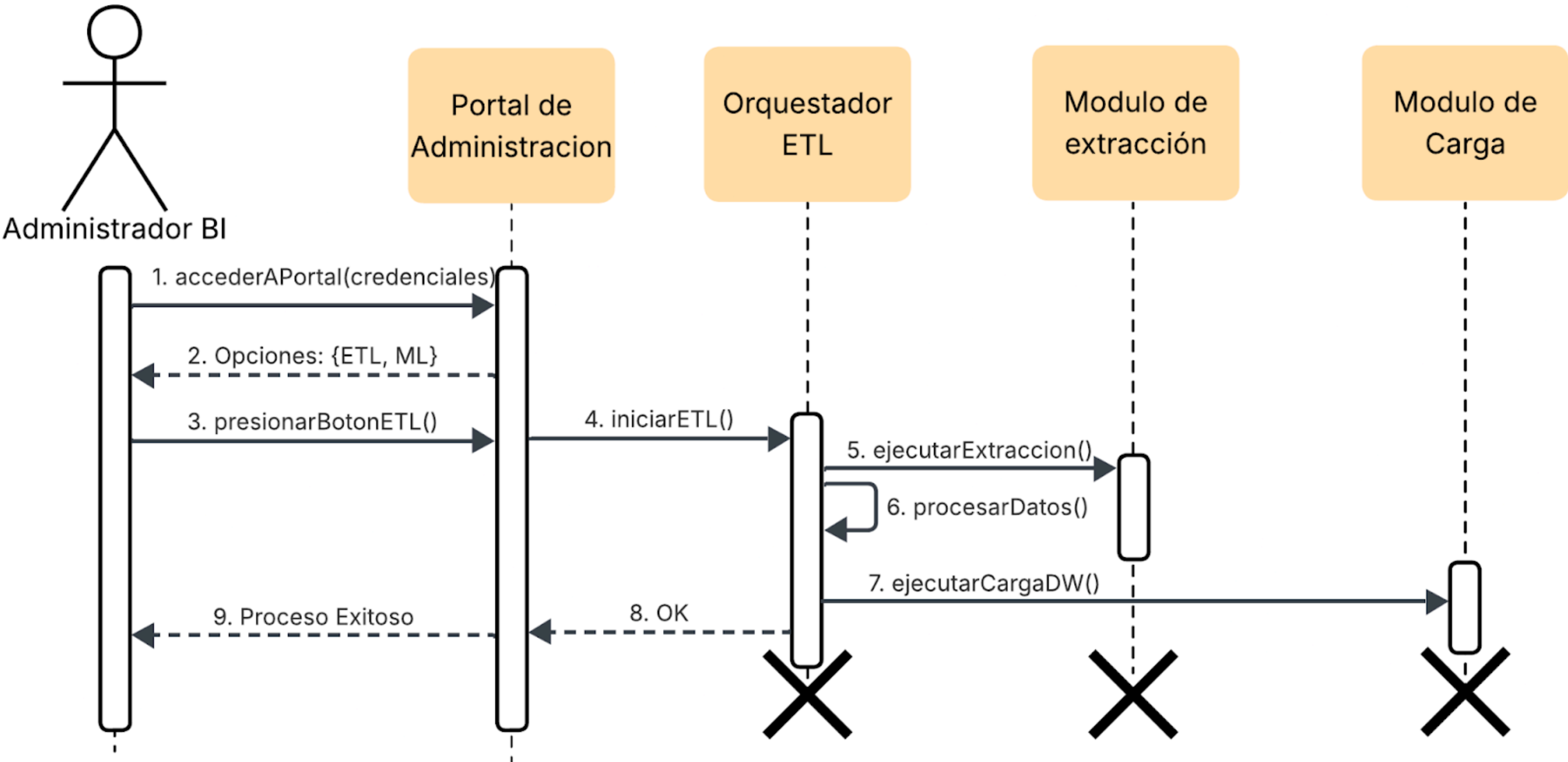
Arquitectura Física del sistema



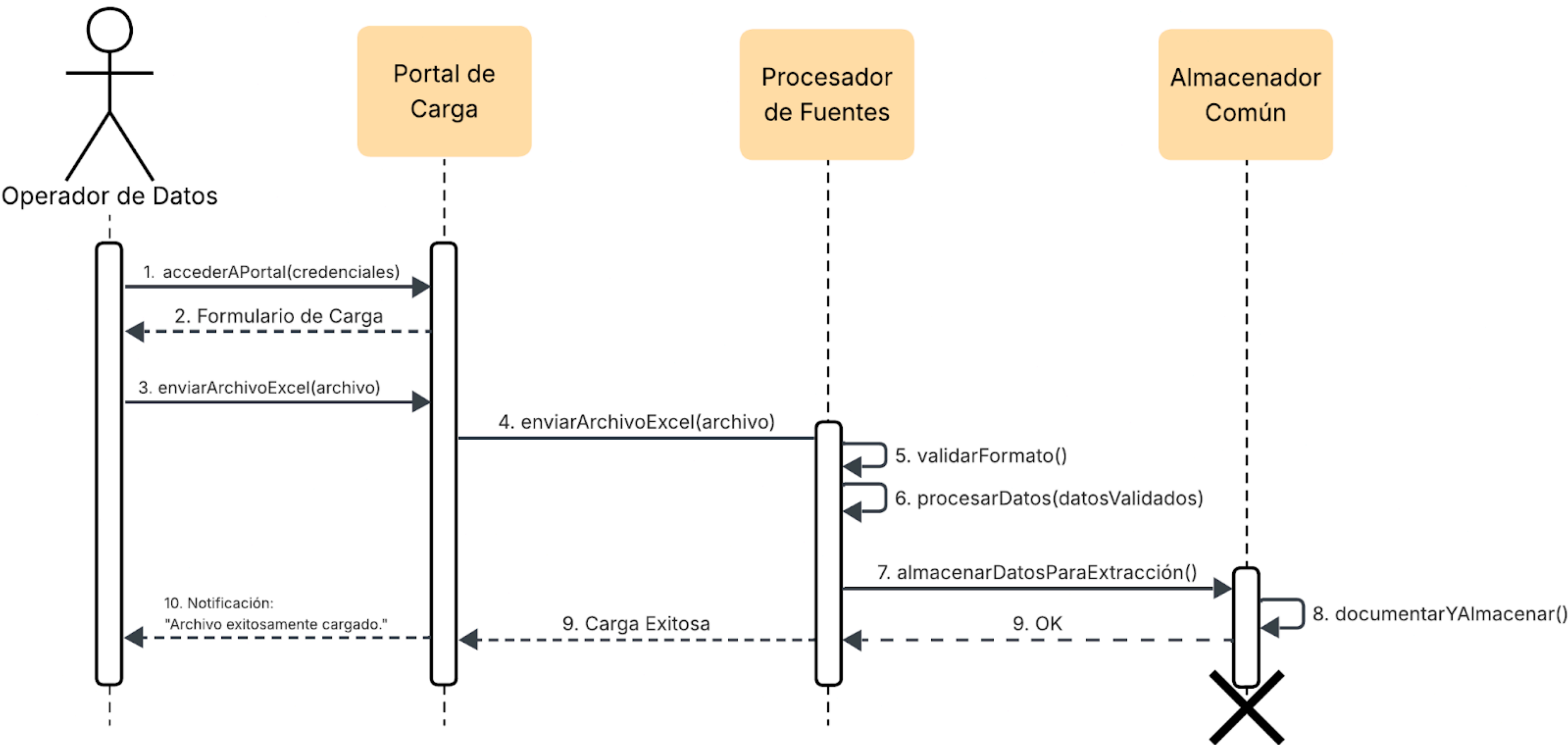


Diagramas de Interacción

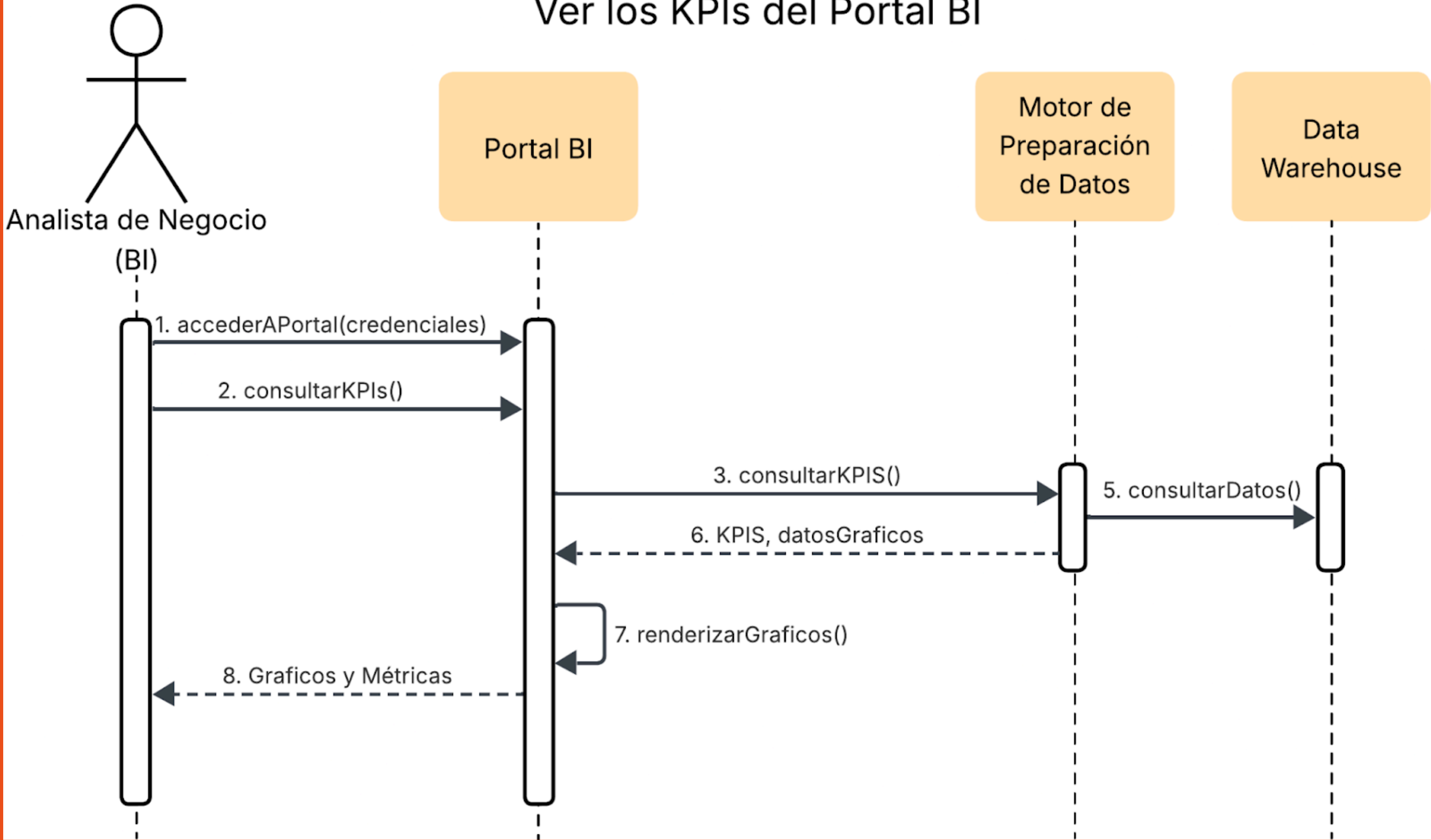
Ejecutar ETL (Exitoso) desde el Portal de Administración



Cargar Archivo (Exitoso) desde el Portal de Carga



Ver los KPIs del Portal BI





Arquitectura y Pila Tecnológica

Herramientas y Frameworks

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 usa 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.	2.149

Justificación

Nombre	Justificación
Python	Versátil para ETL, modelado predictivo y scripting, con ecosistema open-source que acelera el desarrollo.
FastApi	Alto rendimiento en APIs asíncronas para integración segura y escalable de módulos backend.
Scikit-learn	Procesamiento eficiente de datasets de RRHH con algoritmos robustos para predicciones precisas.
XGBoost	Maximiza precisión predictiva con bajo overhead, ideal para datasets estructurados en RRHH.
SvelteKit	Framework ligero en JavaScript para renderización reactiva, alineado con estándares internos de la empresa.
PostgreSQL	Garantiza integridad referencial y soporte OLAP para escalabilidad en staging y Data Warehouse.
Apache Airflow	Orquesta flujos ETL con programación y monitoreo automático para workflows confiables.
Docker	Proporciona contenerización para despliegue consistente y portable en todos los entornos.
PowerBI	Integra visualización interactiva de KPIs y dashboards con el DW para insights accionables.



Modelado de Datos

Base de Datos Staging

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
	agnos_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

Base de Datos Staging

stg_perfiles_trabajo		
pk	id_perfil	Integer
	puesto	Varchar
	categoria_puesto	Varchar
	nombre_puesto	Varchar
	fecha_creacion	Date
	linea_asc	Varchar
	linea_desc	Varchar
	mision	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

Data Warehouse

Objetivo del DW

Alimentar los análisis estadísticos (BI) y el entrenamiento de los modelos de predicción (Machine Learning).

Decisiones de Diseño

Para evitar duplicar esfuerzos de ETL entre BI y ML, priorizaremos los Data Marts (modelo estrella) como única fuente de verdad.

Las tablas anchas para ML se generarán mediante vistas sobre estos Marts, garantizando así la consistencia de los datos y eliminando la redundancia de un Data Lake intermedio.

Tablas de hechos

Se pueden extraer 5 tablas de hechos de las fuentes de datos de la empresa:

- Rotación Empleado
- Headcount Mensual
- Asistencia Diaria
- Participación en Capacitaciones
- Proceso de selección

Data Warehouse

Tablas de hechos

Descripción

Rotación Empleado

Describe los sucesos y cambios en el empleo (clase_medida) dentro de una persona dentro de la empresa. La estructura de los datos es transaccional, lo que significa que registra una fila por cada cambio en el estado de empleo de una persona (ej. contratación, baja, promoción o cambio de puesto) y los intervalos en los que estos cambios son válidos.

Headcount mensual

La tabla transaccional de rotación no permite un cálculo fácil de métricas como la "tasa de rotación", ya que no están en un formato fácil de procesar. 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 recuperar la información de dotación fácilmente y delega el esfuerzo de cómputo al proceso ETL, lo cual es ideal a largo plazo.

Data Warehouse

Tablas de hechos

Descripción

Asistencia Diaria

Aquí se busca modelar la asistencia efectiva de cada día para identificar patrones en los comportamientos de asistencia. 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.

Participación en Capacitación

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.

Data Warehouse

Tablas de hechos

Descripción

Proceso de Selección

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 flujo 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).

Data Warehouse

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'?"

Dimensiones

Principales

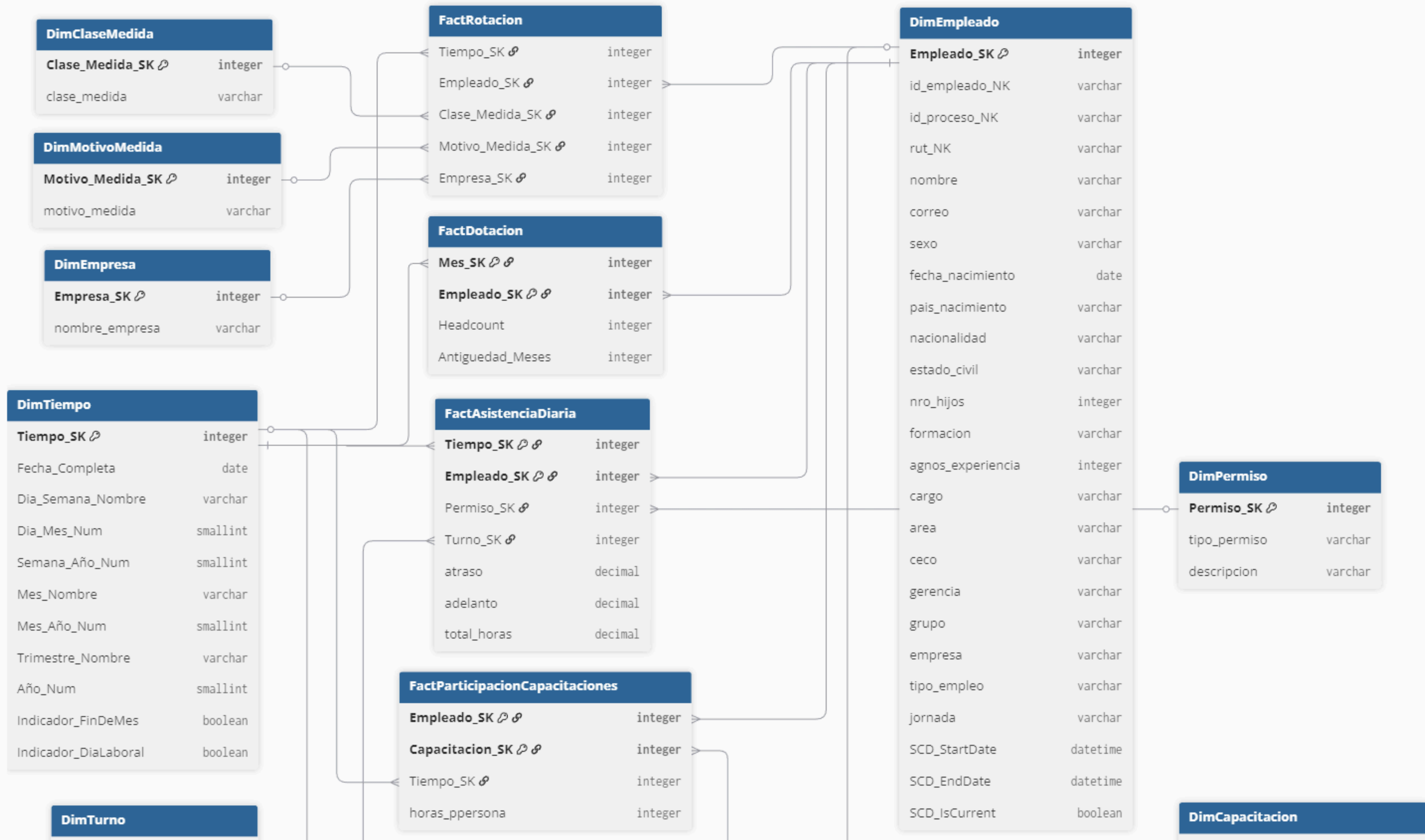
Empleado (SCD2)
Tiempo (SCD0)
Cargo (SCD1)
Capacitación (SCD1)

Secundarias (Mini Dimensiones)

Clase Medida (SCD1)
Motivo Medida (SCD1)
Empresa (SCD1)
Tipo Permiso (SCD1)
Tipo Turno (SCD1)



Modelo DW



o_Num	smallint
icador_FinDeMes	boolean
icador_DiaLaboral	boolean

FactParticipacionCapacitaciones	
Empleado_SK	integer
Capacitacion_SK	integer
Tiempo_SK	integer
horas_ppersona	integer

tipo_empleo	varchar
jornada	varchar
SCD_StartDate	datetime
SCD_EndDate	datetime
SCD_IsCurrent	boolean

DimTurno	
Turno_SK	integer
nombre_turno	varchar
hora_inicio	time
hora_fin	time

FactProcesosSeleccion	
Fecha_Inicio_Proceso_SK	integer
Cargo_Requisicion_SK	integer
Empleado_Seleccionado_SK	integer
id_proceso	integer
fuentes_reclutamiento	varchar
motivo	varchar
nombre	varchar
sexo	varchar
edad	integer
formacion	varchar
agnos_experiencia	integer
duracion_dias	integer
coste_proceso	decimal
nro_cvs_recibidos	integer
nro_personas_finalistas	integer
activo	integer
tiene_continuidad	integer

DimCargo	
Cargo_SK	integer
puesto	varchar
categoria_puesto	varchar
mision	text
responsabilidades	text
informa_y_reporta	varchar
tareas_semanales	text
competencias	text
habilidades_generales	text
habilidades_cognitivas	text
habilidades_fisicas	text
habilidades_sensoriales	text
condiciones_trabajo	text

DimCapacitacion	
Capacitacion_SK	integer
nro_capacitacion_NK	integer
titulo	varchar
mes	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
total_horas	integer
coste	decimal
valoracion_formador	decimal
indice_satisfaccion	decimal
nps	integer



Conclusión

La implementación de este Data Warehouse de RRHH **representa la transformación digital necesaria** para superar la fragmentación de datos y la gestión reactiva actual.

El éxito de esta arquitectura no solo **reducirá drásticamente la carga operativa manual**, sino que, a través de la analítica avanzada y la predicción de rotación, brindará una ventaja estratégica para **influir proactivamente en la retención de talento**.

La viabilidad técnica está establecida; ahora, el foco debe estar en **garantizar la integridad de los datos** y una ejecución rigurosa del plan Kanban para materializar este valor.



Resumen de proyecto

Sistema de Análisis de RRHH para el Casino Luckia



Integrantes

- Tomás Silva
- Juan Yampara

Profesor a Cargo:

- Diego Aracena

Asignatura:

- Proyectos IV