



UCSC

UNIVERSIDAD CATÓLICA DE LA SANTÍSIMA CONCEPCIÓN
FACULTAD DE INGENIERÍA
INGENIERÍA CIVIL INFORMÁTICA

**EVALUACIÓN E IMPLEMENTACIÓN DE LA ARQUITECTURA DE DATOS DE
MEDALLÓN
SOCIEDAD PESQUERA LANDES S.A.**

DERQUI ARIEL SANHUEZA BALBOA
Informe de Práctica Tutelada para optar al título de
INGENIERO CIVIL INFORMÁTICO

Supervisor: Alex Pacheco Segura
Profesor tutor: Braulio Quiero Hernández

Concepción, 03 de Marzo de 2025

Agradecimientos

Agradezco a Sociedad Pesquera Landes S.A. por brindarme la oportunidad de realizar esta práctica, permitiéndome aplicar mis conocimientos en un entorno profesional real. A mi supervisor Alex Pacheco Segura por su guía y enseñanzas, a mis colegas Omar, Roberto y Andru que me apoyaron y me enseñaron cosas valiosas del mundo laboral, así como a mi profesor tutor Braulio Quiero Hernández por su constante apoyo académico y dedicación. Además, quisiera agradecer a toda la comunidad universitaria, especialmente a profesores y profesoras que estuvieron ahí para enseñar y ayudarme a crecer como profesional.

Resumen ejecutivo

Durante la práctica, se logró realizar una evaluación del estado actual de la infraestructura de datos en Sociedad Pesquera Landes S.A. Este análisis detallado permitió identificar los distintos orígenes de datos, como también los métodos de obtención y almacenamiento de estos, especialmente en lo relacionado con el manejo y procesamiento de la información crítica para la empresa. A partir de esta evaluación, se comenzó a implementar la arquitectura de medallón de Microsoft, iniciando con los datos del área de recursos humanos los cuales con anterioridad se manejaban directamente en sistemas creados por Landes y al momento de la práctica estos eran manejados dentro de la plataforma de Talana, una empresa que provee varios servicios de RR.HH. Para lograr la implementación de la arquitectura de medallón se trabajó en Microsoft Fabric. Fue aquí donde se crearon y orquestaron las pipelines de datos que pudieran conectarse a distintos endpoints que entrega el servicio de Talana para la extracción de datos, luego, para la transformación de estos, fue necesario utilizar PySpark (Interfaz de Apache Spark en Python) y finalmente para almacenarlos se dispuso de Lakehouses (Almacenes de datos optimizados para la lectura, utilizan una mezcla de tecnologías de Delta Lake y Data Lake), estos almacenes actúan como bases de datos de tipo SQL Server, capaces de ejecutar código T-SQL. Esta arquitectura moderna y escalable permite el acceso a una fuente de datos centralizada, confiable y segura para los distintos departamentos dentro de la empresa.

Adicionalmente se comenzaron a desarrollar soluciones para optimizar y predecir resultados dentro de la planta, fue posible desarrollar una API con FastAPI de Python con un algoritmo de optimización llamado Simplex (desplegada en Microsoft Azure con Azure Functions) y una pequeña aplicación móvil (desarrollada en la plataforma de low-code de Microsoft PowerApps) para hacer uso de este algoritmo por personal responsable del proceso pertinente dentro de la planta de producción.

Índice de contenidos

Capítulo 1: Introducción	1
1.1 Sobre la organización	1
1.2 Unidad organizacional	2
1.3 Periodo y modalidad de trabajo	2
1.4 Objetivo general	3
1.5 Objetivos específicos	3
1.6 Metodología	4
Capítulo 2: Antecedentes generales de la organización	4
2.1 Situación tecnológica actual	4
2.2 Software principal	5
Capítulo 3: Descripción detallada de las actividades realizadas	6
Capítulo 4: Resultados	13
Capítulo 5: Reflexión	14
Capítulo 6: Conclusiones	15

Índice de figuras

Figura 1: Organigrama	2
Figura 2: Ubicación de datos	7
Figura 3: Arquitectura de medallón	9
Figura 4: Flujo de trabajo para las altas de la empresa	9

Índice de tablas

Tabla 1: Responsabilidades por tipo de servicio	5
Tabla 2: Lakehouse vs MySQL	11

Capítulo 1: Introducción

1.1 Sobre la organización

Sociedad Pesquera Landes, es una empresa chilena con una rica historia en la industria pesquera. Fundada en 1955, ha sido parte fundamental del crecimiento pesquero en Chile y ha ganado reconocimiento tanto a nivel nacional como internacional.

La empresa se dedica a la captura, procesamiento y comercialización de productos del mar. Entre sus principales productos se encuentran la sardina, anchoveta, jurel y merluza, además de harina y aceite de pescado utilizados en diversas industrias. Gracias a su propia flota y a plantas de procesamiento modernas ubicadas a lo largo de la costa chilena, Landes asegura que sus productos lleguen frescos y de calidad a sus clientes.

1.2 Unidad organizacional

A continuación, se muestra en la Figura 1 el organigrama de la empresa y a que unidad de la organización se me asignó.

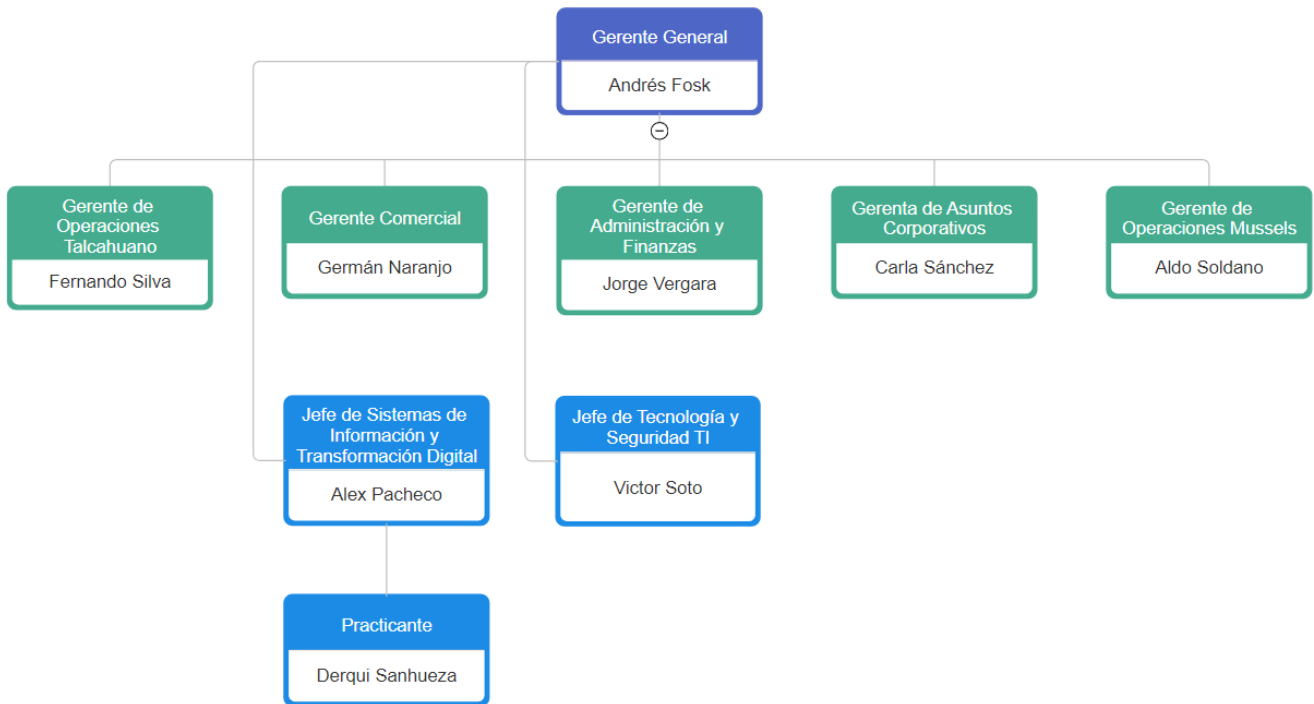


Figura 1: Organigrama
(Origen Landes, creación propia)

1.3 Periodo y modalidad de trabajo

El periodo de la práctica profesional estuvo comprendido entre el día 12 de agosto de 2024 y el día 20 de diciembre de 2024 sumando un total de 800 horas hábiles.

La modalidad de trabajo se definió como presencial, con la opción de algunas horas en remoto.

1.4 Objetivo general

La empresa se encuentra en un proceso de transformación tecnológica, lo que ha generado la necesidad de optimizar la gestión de datos para respaldar decisiones estratégicas y mejorar su desempeño.

Landes desea poder asegurar una fuente de datos centralizada, confiable y segura, que pueda adaptarse a las necesidades de los distintos departamentos dentro de la organización que necesiten de estos datos para la toma de decisiones.

Para lograr el objetivo de la compañía, se abordó como objetivo general de la práctica, la mejora de la infraestructura de datos de la compañía en todos sus ámbitos.

1.5 Objetivos específicos

Para lograr el objetivo anterior se consideró lo siguiente:

- Realizar una evaluación de la arquitectura de datos actual de la compañía.
- Proponer tecnologías y elementos de mejora para la arquitectura actual.
- Diseñar un plan de trabajo tomando en cuenta las mejores prácticas tanto técnicas como estratégicas.
- Implementar el plan de trabajo definido, lo cual incluye programación, pruebas y operación de sistemas.
- Capacitar y entrenar a los usuarios finales de esta solución de datos (analistas en distintos departamentos de la organización).

1.6 Metodología

El desarrollo de esta solución de datos se definió de forma incremental e iterativa, esto debido a que los requerimientos de datos por departamento pueden ir cambiando en el tiempo, es decir que esta solución de datos se tendrá que adaptar continuamente a las necesidades del negocio.

Capítulo 2: Antecedentes generales de la organización

2.1 Situación tecnológica actual

Landes actualmente se encuentra en una etapa de transformación tecnológica, no solamente en el aspecto de software si no que también en la infraestructura tecnológica física de la empresa.

Esta transformación tiene énfasis en:

- Datos y software
- Ciberseguridad y redes

La empresa depende principalmente de la contratación de servicios de terceros para la adquisición de software, aun así, la empresa también desarrolla software propio (es software puntual a una escala muy pequeña).

En cuanto a ciberseguridad y redes, se esta capacitando al personal constantemente sobre amenazas cibernéticas y se están adquiriendo controladores de red capaces de gestionar el tráfico dentro de la organización.

2.2 Software principal

En el contexto de la práctica profesional, se han utilizado distintos softwares para apoyar la etapa de transformación tecnológica de la organización. A continuación, se indican los softwares principales que fueron utilizados durante la realización de la práctica:

Microsoft Azure:

Plataforma principal en la nube que provee múltiples servicios de los tipos IaaS (Infraestructura como servicio), SaaS (Software como servicio), PaaS (Plataforma como servicio), luego están los servicios propios del cliente, en este caso On-prem (on premises o localmente en la empresa). A continuación, se muestra en la Tabla 1 la responsabilidad de cada parte al utilizar los distintos tipos de servicio.

Responsabilidad	SaaS	PaaS	IaaS	On-prem
Responsabilidad siempre retenida por el cliente				
Información y datos	Ciente	Ciente	Ciente	Ciente
Dispositivos (Móviles y PCs)	Ciente	Ciente	Ciente	Ciente
Cuentas e identidades	Ciente	Ciente	Ciente	Ciente
Responsabilidad varía según el tipo				
Infraestructura de identidad y directorios	Compartido	Compartido	Ciente	Ciente
Aplicaciones	Microsoft	Compartido	Ciente	Ciente
Controles de red	Microsoft	Compartido	Ciente	Ciente
Sistema operativo	Microsoft	Microsoft	Ciente	Ciente
Responsabilidad transferida al proveedor de la nube				
Hosts físicos	Microsoft	Microsoft	Microsoft	Ciente
Red física	Microsoft	Microsoft	Microsoft	Ciente
Centro de datos físico	Microsoft	Microsoft	Microsoft	Ciente

Tabla 1: Responsabilidades por tipos de servicio
(Extracción de la documentación de Microsoft)

Como se puede apreciar, IaaS nos permite tener el control de todo menos la parte física del servicio la cual es manejada por Microsoft, finalmente, si se tiene un servicio de tipo SaaS, solo nos debemos preocupar de la información que se genera, los dispositivos a los que se conecta y las cuentas que tienen acceso, reduciendo el esfuerzo que conlleva configurar el resto de los elementos.

Power Platform:

Power Platform es otro servicio de la nube de Microsoft, el cual proporciona otros 3 sub-servicios usados por la empresa, los cuales son **Power Automate** el cual se utiliza para crear flujos que ejecutan tareas de forma automática, **PowerApps** donde se nos permite crear aplicaciones de escritorio y móviles low-code como también almacenar datos y **Power BI** herramienta para la creación de reportes, el análisis de datos e inteligencia empresarial.

Existen otros softwares que se utilizan de forma interna en temas de producción y calidad de los productos clasificados para consumo humano, con los cuales (en esta ocasión) no interactué directamente.

Capítulo 3: Descripción detallada de las actividades realizadas

3.1 Evaluación de la arquitectura de datos actual de la compañía

La primera tarea dentro de la empresa fue realizar un diagnóstico para evaluar el estado actual de la infraestructura de datos de la empresa. Esto contempló identificar todos los orígenes de datos de la empresa ya sea en servidores físicos de la empresa (on-premises) o en la nube (cloud). También fue necesario identificar si estos datos pasaban por procesos de transformación y/o refinamiento previo a ser utilizados, finalmente se identificaron las aplicaciones y personas que hacían

uso de estos datos. En cuanto a los orígenes de datos, se descubrió que son variados, incluyendo desde datos capturados por dispositivos IoT, sensores de PLC en las plantas de proceso, datos generados por aplicaciones, hasta datos ingresados de forma manual en un archivo Excel. En la Figura 2, se ilustran distintas ubicaciones de datos, entre ellos están SQL Server Local, Data Lake Gen 2 en la nube de Azure, una REST API de parte de Talana, Microsoft Dataverse que es parte de Power Platform (una plataforma de desarrollo de software low-code) y Microsoft Dynamics 365 el cual está también ubicado en la nube y maneja los aspectos financieros de la empresa

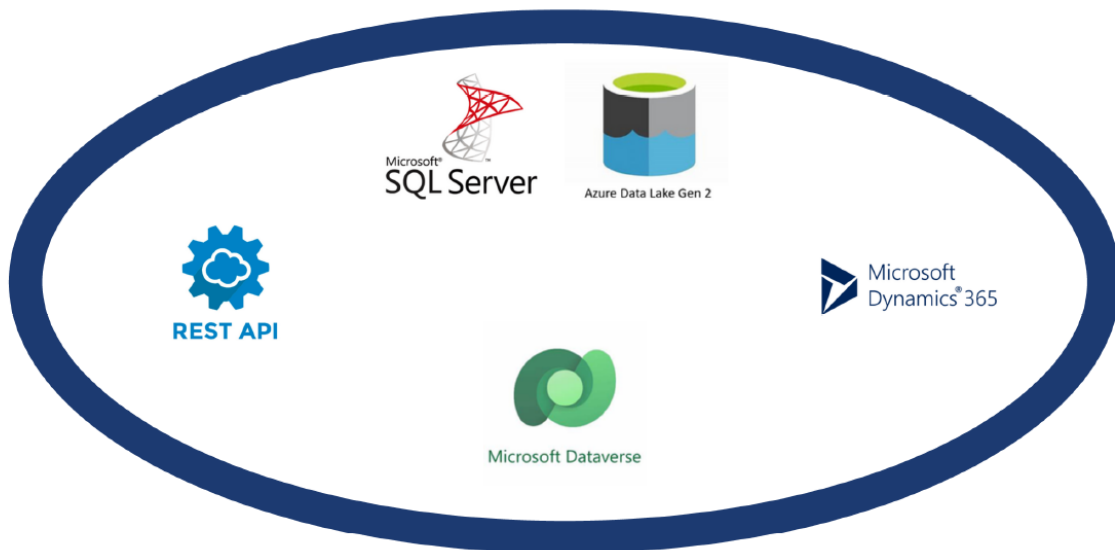


Figura 2: Ubicación de datos
(Creación propia)

Como existen diversos orígenes de datos, también existen diversos lugares donde estos se almacenan y procesan, para luego ser utilizados ya sea en análisis o reportería. Un análisis en profundidad reveló que los datos almacenados no seguían ninguna directiva después de ser almacenados, es decir, se modificaban en el momento y/o simplemente no se procesaban. Debido a esto surgió la necesidad de modificar la infraestructura de datos actual, esto con el objetivo de poder tener un lugar centralizado y bien estructurado para acceder a los datos necesarios por parte de los distintos departamentos de la empresa.

3.2 Propuesta de tecnologías y elementos de mejora para la arquitectura actual

Luego del diagnóstico, se tuvieron varias reuniones para definir cómo se debía mejorar y que tecnología usar, personalmente sugerí utilizar Microsoft Fabric, una plataforma relativamente nueva con meses de antigüedad que ofrecía solucionar el problema que la empresa estaba enfrentando, mi supervisor, también al tanto de esta tecnología dio el visto bueno para comenzar a implementarla dentro de la empresa.

3.3 Diseño e implementación de mejoras

El primer departamento de la empresa en comenzar la centralización de sus datos dentro de Fabric fue el departamento de recursos humanos (personas), este departamento, recientemente comenzó a utilizar Talana (software de rr.hh como servicio en la nube, capaz de manejar remuneraciones, asistencia.. etc.), Talana ofrece varios endpoints para obtener los datos que se manejan dentro del software, para poder comprender la respuesta de los endpoint y cómo hacer las respectivas consultas, se me proporcionó un archivo de 100 pag aproximadamente que contiene la documentación de Talana, ya con esto fue posible avanzar. El paso siguiente fue crear los primeros entornos de trabajo en Microsoft Fabric, estos entornos de trabajo estarían definidos por la arquitectura por capas de medallón de Microsoft. La arquitectura medallón consta de tres capas:

- **Bronce:** también conocida como *zona sin procesar*, esta primera capa almacena los datos de origen en su formato original. Los datos de esta capa suelen ser de solo anexión e inmutables.
- **Plata:** también conocida como *zona enriquecida*, esta capa almacena los datos procedentes de la capa de bronce. Los datos sin procesar se han limpiado y estandarizado, y ahora se estructuran como tablas (filas y columnas).

- **Oro:** también conocida como *zona AI ready*, esta capa final almacena los datos procedentes de la capa de plata. Los datos se refinan para satisfacer requisitos específicos de análisis y están listos para ser usados por modelos de inteligencia artificial.

A continuación, se muestra en la Figura 3 la arquitectura de medallón, cortesía de Microsoft

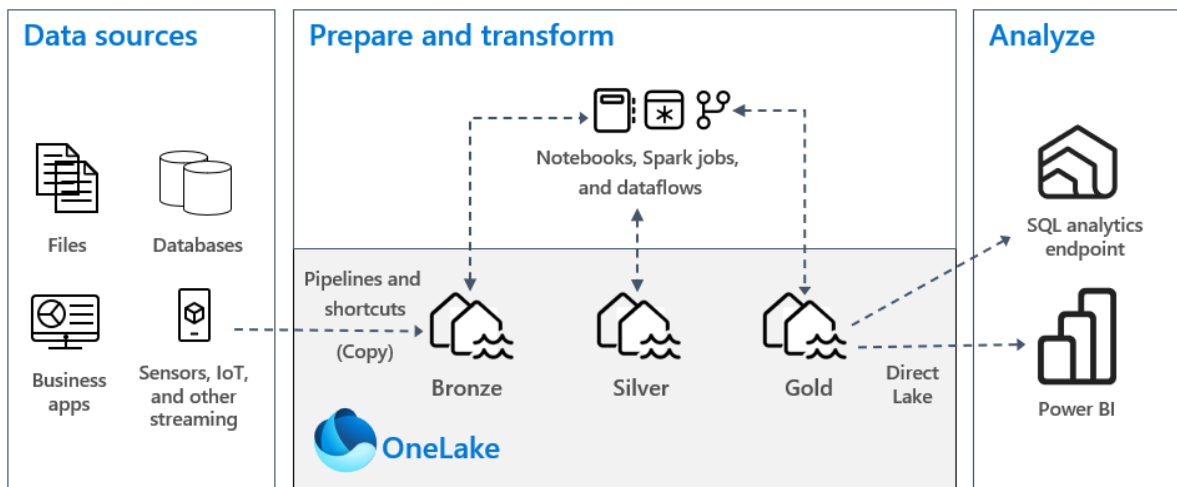


Figura 3: Arquitectura de medallón de Microsoft
(Sacado de la documentación de Microsoft)

Siguiendo la arquitectura de tipo medallón, se crearon 3 entornos de trabajo (Workspaces) uno para cada capa de la arquitectura, dentro de cada uno se creó un Lakehouse, un Lakehouse es un almacén de datos que combina tecnologías de Delta Lake (que utiliza archivos .parquet para guardar datos, archivos optimizados para la lectura) y Data Lake (almacén de datos que es capaz de almacenar datos no estructurados, semi- estructurados y estructurados.). En el Lakehouse de Bronze es imperativo que los datos lleguen tal cual los entrega el origen, en este caso un endpoint de Talana entrega una respuesta .JSON con los datos. Fue creada una pipeline por endpoint, esta pipeline fue configurada para conectarse al endpoint de Talana, recibir la respuesta de este y guardar la respuesta como archivo .JSON

dentro del Lakehouse. En la Tabla 2 se muestra la diferencia entre el Lakehouse de Microsoft y la base de datos MySQL.

Característica	Lakehouse (Parquet, Delta Lake, etc.)	MySQL (Relacional)
Modelo de datos	Archivos optimizados (Parquet, ORC, Delta)	Tablas relacionales
Estructura	Datos estructurados y semiestructurados	Principalmente estructurados
Escalabilidad	Alta (distribuido, Big Data)	Limitada (vertical/sharding)
Costo	Bajo (separación almacenamiento/cómputo)	Más alto (almacenamiento en disco)
Velocidad	Óptimo para análisis y Big Data	Rápido en transacciones pequeñas
Transaccionalidad	ACID con Delta Lake	ACID nativo
SQL	Compatible con SQL (Spark SQL, Trino)	SQL relacional estándar
Flexibilidad	Esquema flexible (Schema-on-read)	Esquema rígido (Schema-on-write)
Uso ideal	Análisis de datos, Data Lakes	Aplicaciones OLTP, transacciones rápidas

Tabla 2: Lakehouse vs MySQL
(Comparación utilizando la documentación de Microsoft)

A continuación, la Figura 4 muestra cómo las pipelines se conectan a los endpoints y generan los nuevos archivos en formato JSON

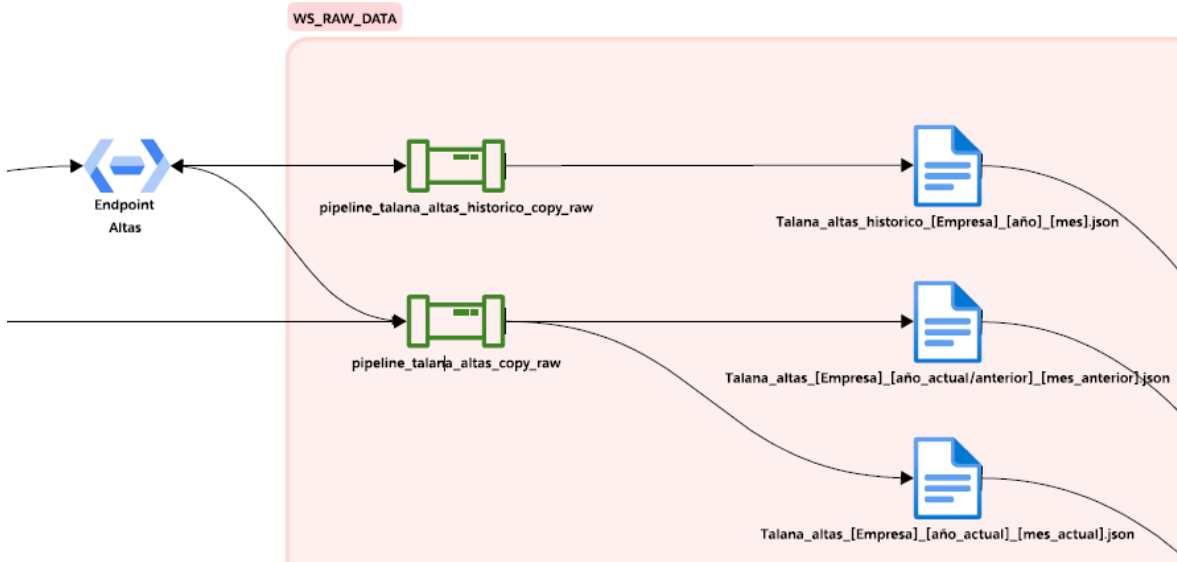


Figura 4: Flujo de trabajo para las altas de la empresa (Creación propia)

Este proceso se repitió para decenas de endpoints, asegurando que las respuestas de la API fueran validas antes de almacenar los archivos en el Lakehouse.

A continuación se llevó la respuesta .JSON a otro entorno de trabajo, en este caso correspondiente a la capa Silver (según el modelo medallón), ahora, este archivo generado a partir de la respuesta de la API sería convertido a una Tabla Delta (nombre que se le da a la tabla de un Lakehouse), para poder llevar la respuesta compleja de la API (dentro del archivo existían listas con otros .JSON dentro, arreglos, structs y muchos tipos de dato) a una tabla, fue necesario aplanar cada archivo .JSON, esto fue posible gracias a la creación de un elemento llamado “Notebook” dentro de Microsoft Fabric, el cual utiliza el framework Apache Spark para el procesamiento de datos, en este caso, existía la opción de utilizar Java (Scala) o Python (PySpark), personalmente elegí Python debido a mi experiencia

previa y conocimientos mas frescos al haber tenido los cursos del Profesor Pedro Gómez con Python. A pesar de que se trataba de Python, tuve que aprender a utilizar PySpark y descubrir todo el potencial oculto para mí en ese entonces. Gracias a la utilización de Python, se logro llevar lo que en el pasado era solo un archivo a una tabla.

Dentro de Microsoft Fabric, existe la posibilidad de crear “Modelos semánticos” (equivalentes a un modelo relacional) y relacionar tablas entre sí, tal como lo haríamos en una base de datos relacional, pero al tratarse de la capa de Silver, las tablas no deberían relacionarse aún, ya que la creación de modelos semánticos se lleva a cabo en la capa de oro, ya que estos últimos son los que el usuario final (analistas y científicos de datos) tendrá acceso para proceder al análisis de datos y creación de modelos de machine learning.

Finalmente se llega a la capa oro con otro proceso de transformación utilizando PySpark, en este caso, esta transformación dependerá de los requerimientos de cada departamento para las tablas finales, es decir, si las tablas tienen relación con recursos humanos, es ese departamento el que debe solicitar que datos quiere para sus respectivos análisis, además es posible satisfacer cualquier requerimiento extra que necesiten (como otras columnas calculadas, otras tablas, etc..).

Durante el periodo de práctica, se tuvieron varias reuniones para definir los requerimientos del departamento de recursos humanos y también se optimizó el código utilizado para realizar las transformaciones de datos necesaria.

A continuación, se comenzaría a dar marcha a la posibilidad de hacer análisis mas complejo y crear modelos predictivos, estos últimos por el tiempo no fueron capaces de ver la luz durante el periodo de práctica, aun así, fue posible la creación de una API con FastAPI de Python para optimizar un proceso interno, en conjunto con ingenieros industriales se determinaron las funciones objetivo, y la naturaleza del modelo, en este caso, lineal. Gracias a mi formación en Investigación de operaciones fui capaz de utilizar el método simplex para solucionar el problema con

Python, teniendo mucho cuidado con las restricciones, parámetros de entrada y salida a la hora de programar los endpoints de la API.

Capítulo 4: Resultados

Durante mi práctica en Sociedad Pesquera Landes S.A., pude ver de primera mano cómo una adecuada gestión de datos puede marcar la diferencia en la operación de una empresa. Lo que al principio parecía un simple proceso de recolección y almacenamiento de información, terminó convirtiéndose en un proyecto que impactó directamente en la eficiencia del área de Recursos Humanos y abrió las puertas para futuras optimizaciones en otras áreas.

Uno de los logros más satisfactorios fue la **automatización de procesos**. Diseñar y desarrollar pipelines de datos para la extracción y transformación de información desde Talana hacia la arquitectura de medallón fue un reto, pero el resultado final demostró que valía la pena. Ahora, la empresa tiene una estructura sólida donde la información fluye sin necesidad de intervención manual constante, minimizando errores y acelerando tiempos de consulta.

Además, la implementación de la estructura en **capas de Bronce, Plata y Oro** hizo que los datos estén organizados de manera más accesible, lo que facilita su análisis y uso por distintos equipos dentro de la empresa. Este fue un punto clave para mejorar la toma de decisiones basada en información confiable.

Otro gran avance fue la **optimización de procesos internos** mediante el desarrollo de una API con FastAPI y el método Simplex. Esta solución, junto con la aplicación móvil creada en PowerApps, permitió a los encargados de producción tomar decisiones más informadas y mejorar el rendimiento en ciertos procesos clave dentro de la planta.

En resumen, el trabajo realizado en esta práctica no solo mejoró la gestión de datos dentro de la empresa, sino que también sentó las bases para futuras mejoras y proyectos que podrían llevar la digitalización de Landes aún más lejos.

Capítulo 5: Reflexión

Si tuviera que resumir mi experiencia en esta práctica en una sola palabra, sería **aprendizaje**. Desde el primer día, me encontré con desafíos que me obligaron a salir de mi zona de confort y a buscar soluciones en áreas en las que tenía poco o ningún conocimiento previo.

Trabajar con **Microsoft Fabric y PySpark** fue un reto inesperado. Si bien tenía experiencia en Python, sumergirme en el mundo de los datos a gran escala y el procesamiento distribuido fue una curva de aprendizaje pronunciada. Sin embargo, al final del proceso, no solo logré adaptarme, sino que también adquirí habilidades que sin duda serán valiosas en mi carrera profesional.

Más allá de los aspectos técnicos, esta práctica también me ayudó a desarrollar **habilidades blandas esenciales**. La comunicación con diferentes áreas de la empresa, la capacidad de traducir necesidades de negocio en soluciones tecnológicas y la colaboración con colegas fueron aspectos que me ayudaron a crecer profesionalmente. Aprendí que, más allá del código y los algoritmos, lo más importante es cómo la tecnología puede impactar positivamente en el trabajo diario de las personas.

Esta experiencia también me permitió ver el valor de la innovación dentro de las empresas. En muchas ocasiones, las compañías no implementan nuevas tecnologías por desconocimiento o por temor al cambio. Haber podido aportar con ideas y herramientas que realmente hicieron una diferencia en Landes me dejó la satisfacción de haber contribuido con algo significativo.

Capítulo 6: Conclusiones

Mirando en retrospectiva, la práctica en Sociedad Pesquera Landes S.A. fue un punto de inflexión en mi formación profesional. No solo me permitió aplicar conocimientos adquiridos en la universidad, sino que me enfrentó a problemas reales que requerían soluciones creativas y bien fundamentadas.

Uno de los mayores aprendizajes fue la importancia de una **infraestructura de datos bien estructurada**. Comenzando con la evaluación de la infraestructura actual de datos en la compañía, se logró identificar varias falencias y puntos de mejoras que luego se abordarían en la implementación. Además, durante las reuniones fue posible ponerse de acuerdo en utilizar la tecnología de Microsoft Fabric.

También comprendí que la tecnología por sí sola no es suficiente. Es clave que las soluciones sean diseñadas **pensando en las personas que las utilizarán**. Por eso, fue necesario capacitar al personal, en específico, a un equipo de analistas en el departamento de mejora continua de la organización, los cuales solicitaron acceso para poder tomar decisiones.