



FABRICa tu plataforma de datos

Guía del alumno

Verne Academy



ÍNDICE

1	Introducción	1
2	Escenario.....	2
3	Implementando una arquitectura medallón	3
3.1	Laboratorio 1: Bronce.....	3
3.1.1	Tareas.....	4
3.2	Laboratorio 2: Plata.....	5
3.2.1	Tareas.....	5
3.3	Laboratorio 3: Oro.....	7
3.3.1	Tareas.....	7
3.4	Laboratorio 4: Modelo semántico.....	9
3.4.1	Tareas.....	9



1 INTRODUCCIÓN

Mediante este taller introductorio, presentaremos las diversas experiencias disponibles en Fabric. Al final de este, habremos aprendido a utilizar Lakehouses, Warehouses, Dataflows Gen2 o Data Pipelines entre otros.

2 ESCENARIO

Somos analistas de datos de la empresa World Wide Importers.

Wide World Importers (WWI) es un importador y distribuidor mayorista de artículos novedosos que opera desde el área de la bahía de San Francisco.

Como mayorista, los clientes de WWI son principalmente empresas que revenden a particulares. WWI vende a clientes minoristas de todo Estados Unidos, incluidas tiendas especializadas, supermercados, tiendas de informática, tiendas de atracciones turísticas y algunos particulares. WWI también vende a otros mayoristas a través de una red de agentes que promocionan los productos en nombre de WWI.

Recientemente, WWI se ha planteado la necesidad de llevar a cabo un análisis exhaustivo de aquellos productos más vendidos, averiguar qué clientes tenían un volumen de facturación elevado históricamente, pero han dejado de comprar, con la intención de lanzar campañas de venta que puedan volver a atraerlos.

3 IMPLEMENTANDO UNA ARQUITECTURA MEDALLÓN

La arquitectura de medallón describe una serie de capas de datos que denotan la calidad de estos. Garantiza la atomicidad, la coherencia, el aislamiento y la durabilidad a medida que los datos pasan por varias capas de validaciones y transformaciones antes de almacenarse en un diseño optimizado para un análisis eficiente.

Los términos **bronce (sin procesar)**, **plata (validado)** y **oro (enriquecido)** describen la calidad de los datos en cada una de estas capas.

Para nuestro caso de uso, seguiremos el siguiente esquema:



- Nuestro origen de datos será un conjunto de ficheros CSV como resultado de una extracción realizada sobre la base de datos de World Wide Importers (WWI).
- Los datos en crudo serán almacenados en la capa bronce.
- Mediante el uso de cuadernos, realizaremos las validaciones necesarias para llevar los datos hasta la capa plata.
- Seguidamente, consolidaremos la información desde la capa plata hasta la capa oro, donde estableceremos las relaciones necesarias y las métricas pertinentes para crear el modelo semántico de la solución.

3.1 Laboratorio 1: Bronce

Como se ha mencionado anteriormente, la capa bronce contiene datos que no están validados.

En nuestro caso, extraeremos los ficheros CSV alojados en el repositorio de GitHub para almacenarlos en el lago de datos de la capa bronce.

3.1.1 Tareas

- **Tarea 1:** Habilitar una capacidad de prueba de Microsoft Fabric
 - **Portal Power BI:** <https://app.powerbi.com/>
 - Emplea el usuario y contraseña facilitados en el entorno del laboratorio.
- **Tarea 2:** Creación de un área de trabajo
 - **Nombre del área de trabajo:** Verne Academy Summit 2024
- **Tarea 3:** Creación de la capa bronce
 - **Nombre de la carpeta:** bronze
 - **Nombre del Lakehouse:** bronze_WWI
- **Tarea 4:** Ingesta de datos
 - **Nombre de la subcarpeta en el Lakehouse:** wwi
 - **Nombre de la nueva canalización:** bronze_etl_WWI
 - **Actividad Copiar datos:**
 - **General**
 - **Nombre:** cpactivity_Customers
 - **Reintentos:** 3
 - **Origen**
 - **Conexión**
 - **Tipo:** HTTP
 - **URL:** <https://raw.githubusercontent.com/javendia/verne-academy-summit-2024/main/sources/Customers.csv>
 - **URL relativa:** Customers.csv
 - **Formato de archivo:** DelimitedText
 - **Destino**
 - **Lakehouse:** bronze_WWI
 - **Carpeta raíz:** Archivos
 - **Ruta de acceso del archivo:** wwi / Customers.csv
- **Tarea 5:** Ingesta de datos automatizada
 - **Variables**
 - Files
 - **Tipo:** Array
 - **Valor:**

```
[ "Colors.csv", "Customers.csv", "CustomerCategories.csv", "DeliveryMethods.csv", "Invoices.csv", "Orders.csv", "People.csv", "StockItems.csv" ]
```

- **Actividad ForEach**
 - **General**
 - **Nombre:** cpactivities
 - **Configuración**

- **Elementos:**

```
@variables('files')
```

- **Actividad interna Copiar datos**
 - **General**
 - **Nombre:** cpactivity_file
 - **Reintentar:** 3
 - **Origen**
 - **Conexión:** Definida anteriormente
 - **Tipo de conexión:** HTTP
 - **Dirección URL relativa:**

```
@item()
```

- **Formato de archivo:** DelimitedText
- **Destino**
 - **Lakehouse:** bronze_WWI
 - **Carpeta raíz:** Archivos
 - **Ruta de acceso del archivo:** wwi / @item()
 - **Formato de archivo:** DelimitedText

3.2 Laboratorio 2: Plata

Como hemos comentado en el punto anterior, si la capa bronce contiene todo el historial de datos en un estado casi sin procesar, la capa de plata representa una versión validada y enriquecida de los datos.

3.2.1 Tareas

- **Tarea 1:** Creación de la capa plata.
 - **Nombre de la carpeta:** silver
 - **Nombre del Lakehouse:** silver_WWI
 - Nuevo acceso directo a la carpeta wwi del Lakehouse bronze_WWI
- **Tarea 2:** Creación de cuadernos de transformación
 - **Nombre de la subcarpeta en silver:** notebooks
 - Importación de notebooks
 - **Lakehouse:** silver_WWI
- **Tarea 3:** Orquestación de notebooks
 - Importación del cuaderno fill_md_extract_bronze_silver
 - Reemplazo de identificadores GUID para cada uno de los cuadernos
 - Ejecución del cuaderno fill_md_extract_bronze_silver

- **Tarea 4:** Orquestación de recarga de la capa de plata
 - **Nombre de la nueva canalización:** silver_et_WWI
 - **Variables**
 - notebooks
 - **Tipo:** Array
 - notebook_name
 - **Tipo:** String
 - **Actividad Búsqueda**
 - **General**
 - **Nombre:** get_metadata
 - **Reintentar:** 3
 - **Configuración**
 - **Lakehouse:** silver_WWI
 - **Carpeta raíz:** Tablas
 - **Nombre de la tabla:** md_extract_bronze_silver
 - **Actividad Establecer variable**
 - **General**
 - **Nombre:** list_files
 - **Configuración**
 - **Nombre:** notebooks
 - **Valor**

```
@activity('get_metadata').output.value
```

- **Actividad ForEach**
 - **General**
 - **Nombre:** call_notebooks
 - **Configuración**
 - **Secuencial:** Sí
 - **Elementos**

```
@variables('notebooks')
```

- **Actividad interna Establecer variable**
 - **General**
 - **Nombre:** set_notebook_name
 - **Configuración:**
 - **Nombre:** notebook_name
 - **Valor**

```
@item().notebook_id
```


- **Actividad interna Bloc de notas**
 - **General**
 - **Nombre:** run_notebook
 - **Reintentar:** 3
 - **Bloc de notas**

```
@variables('notebook_name')
```

3.3 Laboratorio 3: Oro

En esta capa de la arquitectura medallón, los datos de la capa plata son agregados y enriquecidos con el objetivo de convertirlos en información relevante.

3.3.1 Tareas

- **Tarea 1:** Creación de la capa oro
 - **Nombre de la carpeta:** gold
 - **Nombre del Warehouse:** gold_WWI
- **Tarea 2:** Creación de tablas delta en la capa oro
- **Tarea 3:** Creación de vistas sobre la capa plata
- **Tarea 4:** Creación de esquema etl y procedimientos almacenados
- **Tarea 5:** Creación de flujo de datos para la entidad *Calendario*
 - **Nombre del flujo de datos:** gold_etl_calendar_WWI
 - **Nombre de la consulta:** Calendario
- **Tarea 6:** Orquestación
 - **Nombre de la nueva canalización:** gold_etl_WWI
 - **Variables**
 - stored_procedures
 - **Nombre:** stored_procedures
 - **Tipo:** Array
 - **Valor:**

```
["[etl].[load_WWI_dim_customer]", "[etl].[load_WWI_dim_people]", "[etl].[load_WWI_dim_stockitem]", "[etl].[load_WWI_fact_invoices]", "[etl].[load_WWI_fact_orders]"]
```

- stored_procedure
 - **Nombre:** stored_procedure
 - **Tipo:** String
- **Actividad Flujo de datos**
 - **General**
 - **Nombre:** gold_etl_calendar_WWI
 - **Configuración**
 - **Flujo de datos:** gold_etl_calendar_WWI

- **Actividad ForEach**
 - **General**
 - **Nombre:** copy_activities
 - **Configuración**
 - **Secuencial:** Sí
 - **Elementos:**

```
@variables('stored_procedures')
```

- **Actividad interna Establecer variable**
 - **General**
 - **Nombre:** set_stored_procedure
 - **Configuración**
 - **Nombre:** stored_procedure
 - **Valor:**

```
@item()
```

- **Actividad interna Procedimiento almacenado**
 - **General**
 - **Nombre:** copy_data
 - **Configuración**
 - **Almacén:** gold_WWI
 - **Nombre del procedimiento almacenado:**

```
@variables('stored_procedure')
```

- **Tarea 7: Orquestación general**
 - **Nombre de la nueva canalización:** orchestrator_WWI
 - **Actividad Invocar canalización**
 - **General**
 - **Nombre:** bronze_etl_WWI
 - **Configuración**
 - **Canalización invocada:** bronze_etl_WWI
 - **Esperar a la finalización:** Sí
 - **Actividad Invocar canalización**
 - **General**
 - **Nombre:** silver_etl_WWI
 - **Configuración**
 - **Canalización invocada:** silver_etl_WWI
 - **Esperar a la finalización:** Sí
 - **Actividad Invocar canalización**
 - **General**
 - **Nombre:** gold_etl_WWI

- **Configuración**
 - **Canalización invocada:** gold_etl_WWI
 - **Esperar a la finalización:** Sí

3.4 Laboratorio 4: Modelo semántico

Empleando los datos consolidados en la capa oro, los ficheros iniciales se convierten en información lista para su análisis a través de la construcción de un modelo en estrella, donde las entidades se relacionan entre sí y los datos se agregan a través de la construcción de métricas.

3.4.1 Tareas

- **Tarea 1:** Administración de relaciones desde el servicio

TABLA 1	TABLA 2	CARDINALIDAD	DIRECCIÓN
StockItem[StockItemID]	Orders[StockItemID]	1 a muchos	Único
People[PersonID]	Orders[SalespersonPersonID]	1 a muchos	Único
Customer[CustomerID]	Orders[CustomerID]	1 a muchos	Único
Calendario[Date]	Orders[OrderDate]	1 a muchos	Único
StockItem[StockItemID]	Invoices[StockItemID]	1 a muchos	Único
People[PersonID]	Invoices [SalespersonPersonID]	1 a muchos	Único
Customer[CustomerID]	Invoices [CustomerID]	1 a muchos	Único
Calendario[Date]	Invoices [InvoiceDate]	1 a muchos	Único



- **Tarea 2:** Creación de métricas

MÉTRICA	DEFINICIÓN	FORMATO
Total Sales Quantity	Total Sales Quantity = SUM('Orders'[Quantity])	Número entero
Sales Amount	Sales Amount = SUMX('Orders', 'Orders'[Quantity] * 'Orders'[UnitPrice])	Moneda
Sales Tax Amount	Sales Tax Amount = SUMX('Orders', [Sales Amount] * DIVIDE('Orders'[TaxRate], 100))	Moneda
Sales Net Cost	Sales Net Cost = [Sales Amount] - [Sales Tax Amount]	Moneda
Total Invoice Quantity	Total Invoice Quantity = SUM(Invoices[Quantity])	Número entero
Invoice Amount	Invoice Amount = SUMX('Invoices', 'Invoices'[Quantity] * 'Invoices'[UnitPrice])	Moneda
Invoice Tax Amount	Invoice Tax Amount = SUMX('Invoices', [Invoice Amount] * DIVIDE('Invoices'[TaxRate], 100))	Moneda
Invoice Net Cost	Invoice Net Cost = [Invoice Amount] - [Invoice Tax Amount]	Moneda

The image features a light blue background. In the center, the text "Verne Academy" is displayed in a blue, sans-serif font. Surrounding the text are ten yellow lightbulbs arranged in a circular pattern. At the bottom of the image, two yellow hands are shown, palms up, as if holding or presenting the central text and lightbulbs.

Verne Academy