

**MAESTRÍA EN CIENCIA DE DATOS**

**COHORTE 2024-2025**

**BUENOS AIRES**

**INTRODUCCIÓN A DATA WAREHOUSING**

**TRABAJO PRÁCTICO FINAL**

Docentes: - Mg. Eduardo A. Poggi - CC. Esteban J. Alonso - Dr. Hernán Etiennot

**Alumnos:**

* **GALLARDO, Ezequiel**
* **FILIPUZZI, Juan Manuel**

Contenido

[TPG01 – Flujo de datos en un DWA 3](#_Toc197291818)

[Objetivos de trabajo y Entregables 3](#_Toc197291819)

[1. Arquitectura de la solución 5](#_Toc197291820)

[2. Carga de Schema TMP 6](#_Toc197291821)

[2.1 Proceso Schema TMP 1: CATEGORIES 6](#_Toc197291822)

[2.2 Proceso Schema TMP 2: REGIONS 8](#_Toc197291823)

[2.3 Proceso Schema TMP 3: SHIPPERS 10](#_Toc197291824)

[2.4 Proceso Schema TMP 4: TERRITORIES 12](#_Toc197291825)

[2.5 Proceso Schema TMP 5: CUSTOMERS 15](#_Toc197291826)

[2.6 Proceso Schema TMP 6: EMPLOYEE TERRITORIES 17](#_Toc197291827)

[2.7 Proceso Schema TMP 7: EMPLOYEE 19](#_Toc197291828)

[2.8 Proceso Schema TMP 8: EMPLOYEE TERRITORIES 21](#_Toc197291829)

[2.9 Proceso Schema TMP 9: ORDERS 23](#_Toc197291830)

[2.10 Proceso Schema TMP 10: ORDERS 25](#_Toc197291831)

[2.11 Proceso Schema TMP 11: SUPPLIERS 27](#_Toc197291832)

[2.12 Proceso Schema TMP 11: COUNTRIES 30](#_Toc197291833)

# TPG01 – Flujo de datos en un DWA

## Objetivos de trabajo y Entregables

El objetivo de este trabajo práctico es desarrollar todas las capas de datos y ejecutar los procesos correspondientes del flujo end-to-end en un DWA (Data Warehouse Analítico), desde la ingesta hasta la publicación y la explotación.

El material básico para la elaboración del presente trabajo se encuentra publicado en la plataforma del curso, además de lo expuesto en clase.

**Aclaraciones**

• Este trabajo debe elaborarse por equipos según los grupos establecidos para la materia. Los grupos de más de tres integrantes serán penalizados.

• La entrega de este TP consiste en publicar un documento resumiendo lo realizado según se especifica más abajo, además de los componentes desarrollados.

• Cada grupo deberá exponer en clase una síntesis del trabajo realizado con una duración máxima de 10’. Podría reemplazarse con un video.

• Las fechas de publicación y presentación serán indicadas en la plataforma

• Incluyan en los archivos a entregar la lista de los integrantes. Se recomienda considerar una carátula en donde se identifique el posgrado, la materia, el título del informe, los integrantes del equipo y la fecha.

• La evaluación se realizará según la rúbrica descrita más abajo.

• Los integrantes de cada grupo obtendrán la misma calificación.

• Los docentes evalúan el trabajo realizado por lo que se manifiesta en la presentación y en los documentos entregados, por lo tanto se recomienda una elaboración cuidada y comentada. El contenido debe transmitir las tareas realizadas con la especificidad suficiente para comprenderlas pero sin entrar en detalles irrelevantes. No copien textos externos, si fuera necesario, citen la fuente.

**Informe y presentación exponiendo:**

1. Entrega de un informe y/o presentación (.PDF/.PPTX) con un resumen de lo realizado. Esto permitirá evaluar el resultado sin necesariamente abrir ningún entorno de base de datos.

2. Se deben incluir como anexos todos los scripts desarrollados, los DER y estructuras correspondientes.

3. Entregar como .ZIP la base resultante con todos los componentes (.db, .sql, etc. y los tableros) para verificación de autoría si fuera necesario.

4. Entregar el tablero desarrollado (por ejemplo, Tablero.PBIX).

5. En la presentación en clase deberán ejecutar los tableros desarrollados.

6. Salvo el informe que debe ser publicado en el aula viryual, los demás objetos pueden ser publicados en un drive con libre acceso.

## Arquitectura de la solución

La arquitectura de la solución, será la siguiente

## Carga de Schema TMP

En ésta sección de trabajo, tenemos como objetivo alimentar todas las tablas del schema TEMPORAL a partir de los files de INGESTA. Tomamos como patrón de diseño lo siguiente:

1. El input serán los flat files del directorio de INGESTA
2. Se crea un registro de Log propio de la entidad a correr.
3. Las Tablas Temporales se truncan al momento inicial.
4. Se envían todos los registros del archivo de INPUT, hacia la tabla en el schema TMP.
   1. Como premisa, todos los registros y sus respectivos campos, no reciben manipulación, para tener una copia exacta del transaccional; Motivo por el cual, todos los campos se almacenan como VARCHAR.
   2. En el caso de ser rejectada alguna fila; se almacena en un registro de Auditoría.
5. Se registra nivel auditoría el status del proceso.

Correremos todos los procesos con los FILES DE INGESTA 1.

### 2.1 Proceso Schema TMP 1: CATEGORIES

|  |  |
| --- | --- |
| JobName | TMP\_Categories.dtsx |
| Input |  |
| Control Flow |  |
| Data Flow |  |
| Data |  |

### 2.2 Proceso Schema TMP 2: REGIONS

|  |  |
| --- | --- |
| JobName | TMP\_Regions.dtsx |
| Input |  |
| Control Flow |  |
| Data Flow |  |
| Data |  |

### 2.3 Proceso Schema TMP 3: SHIPPERS

|  |  |
| --- | --- |
| JobName | TMP\_Shippers.dtsx |
| Input |  |
| Control Flow |  |
| Data Flow |  |
| Data |  |

### 2.4 Proceso Schema TMP 4: TERRITORIES

|  |  |
| --- | --- |
| JobName | TMP\_Territories.dtsx |
| Input |  |
| Control Flow |  |
| Data Flow |  |
| Data |  |

### 2.5 Proceso Schema TMP 5: CUSTOMERS

|  |  |
| --- | --- |
| JobName | TMP\_Customers.dtsx |
| Input |  |
| Control Flow |  |
| Data Flow |  |
| Data |  |

### 2.6 Proceso Schema TMP 6: EMPLOYEE TERRITORIES

|  |  |
| --- | --- |
| JobName | TMP\_Employee\_Territories.dtsx |
| Input |  |
| Control Flow |  |
| Data Flow |  |
| Data |  |

### 2.7 Proceso Schema TMP 7: EMPLOYEE

|  |  |
| --- | --- |
| JobName | TMP\_Employee.dtsx |
| Input |  |
| Control Flow |  |
| Data Flow |  |
| Data |  |

### 2.8 Proceso Schema TMP 8: EMPLOYEE TERRITORIES

|  |  |
| --- | --- |
| JobName | TMP\_Order\_Details.dtsx |
| Input |  |
| Control Flow |  |
| Data Flow |  |
| Data |  |

### 2.9 Proceso Schema TMP 9: ORDERS

|  |  |
| --- | --- |
| JobName | TMP\_Orders.dtsx |
| Input |  |
| Control Flow |  |
| Data Flow |  |
| Data |  |

### 2.10 Proceso Schema TMP 10: ORDERS

|  |  |
| --- | --- |
| JobName | TMP\_Orders.dtsx |
| Input |  |
| Control Flow |  |
| Data Flow |  |
| Data |  |

### 2.11 Proceso Schema TMP 11: SUPPLIERS

|  |  |
| --- | --- |
| JobName | TMP\_Suppliers.dtsx |
| Input |  |
| Control Flow |  |
| Data Flow |  |
| Data |  |

### 2.12 Proceso Schema TMP 11: COUNTRIES

|  |  |
| --- | --- |
| JobName | TMP\_Countries.dtsx |
| Input |  |
| Control Flow |  |
| Data Flow |  |
| Data |  |