



Tópicos Avanzados de Base de Datos.

UNIVERSIDAD DIEGO PORTALES.

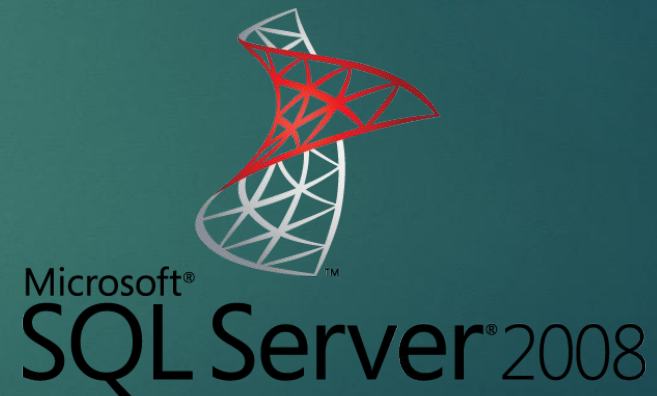
Ayudantía N° 2.

Procedimientos para la creación de un ETL.

- ▶ Profesor: Alejandro Romero Frantzen.
- ▶ Curso: Tópicos Avanzados de Base de Datos.
- ▶ Ayudante: Alvaro Cáceres Aravena.
- ▶ Correo: ayudantiaudp.bi@gmail.com

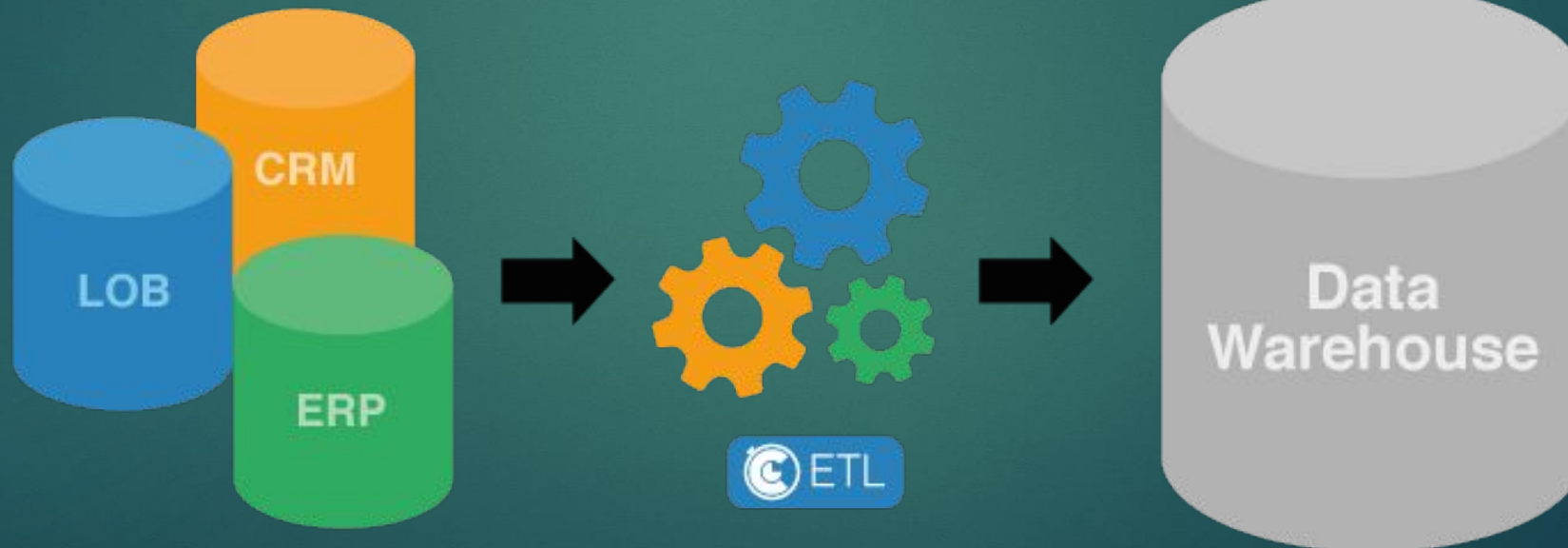
Plataforma TI.

- ▶ Para la realización del proceso ETL (extract, transform, load) se utilizarán las siguientes plataformas tecnológicas.

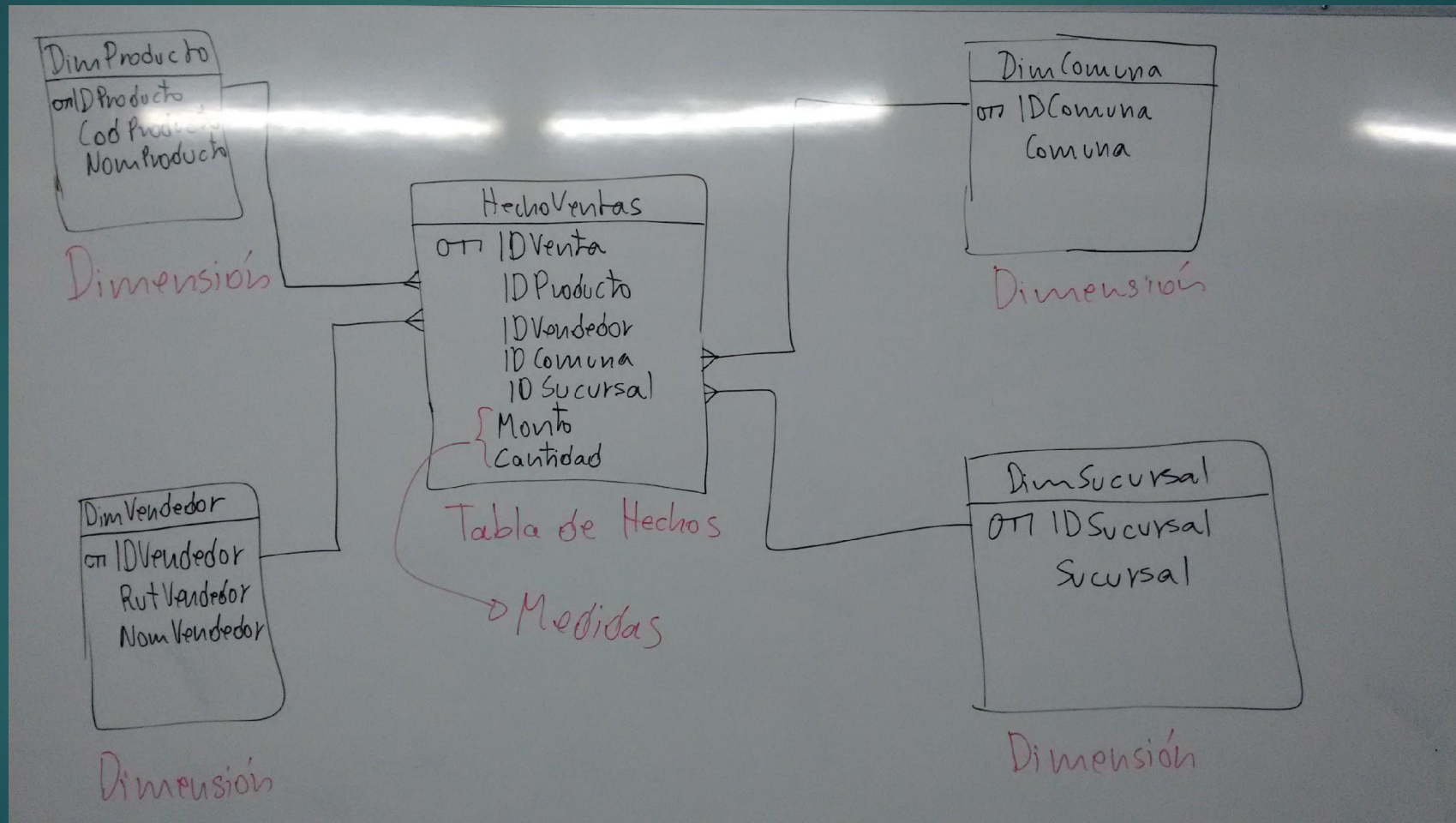


¿ Qué es ETL?

- ▶ Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, **Data Mart**, o **Data Warehouse** para analizar en otro sistema operacional para apoyar un proceso de negocio.



ETL (Ventas_AlejandroRomero).



Dimensiones.

1. Ejecutar SQL Server Business Intelligence Development Studio.
2. En el menú **Archivo** (File) seleccione **Nuevo** luego **Proyecto**(Project).
3. Seleccione **Integration Service Project** y colocar nombre del proyecto.
4. Desde Tool Box arrastre al panel Control Flow la herramienta **Data Flow Task** (con un click puede cambiar el nombre).
5. Al realizar doble click en Data Flow Task se direcciona al panel Data Flow. Desde Tool Box arrastre OLE DB Source.
6. Realice doble click en **OLE DB Source** para abrir **OLE DB connection manager**.
7. Presione New, en la nueva ventana nuevamente presione New, en el rectangulo Server name escriba **serverbd**(localhost).
8. Seleccione la base de datos (Ej: Ventas AlejandroRomero), presione **Test Connection** y finalmete ok.

Dimensiones.

9. En Data Access Mode seleccione **SQL command** y finalmente presione **Build Query**.
10. En la ventana **Query Builder** presione **Add Table**, seleccione la tabla Ventas (add).
11. Seleccione los campos codProducto, NomProducto (según los campos solicitados). No olvide colocar en la consulta SQL la sintaxis **DISTINCT**.
12. Al regresar a la ventana OLE DB Source Editor , presione Columns para confirmar los campos seleccionados y finalmente presione ok.
13. Ahora desde Tool Box arrastre al panel Data Flow **OLE DB Destination**.
14. No olvide vincular la flecha verde desde **OLE DB Source** a **OLE DB Destination**.

Dimensiones.

15. Realizar Doble click en **OLE DB Destination Editor**.
16. Importante!. Recuerde que en **OLE DB connection manager** debe estar direccionado a nuestro **Data Warehouse**.
17. Ejecute SQL Server y cree una base de datos con el nombre DW_NombreAlumno.
18. Desde Visual Studio Presione New y configure la conexion al Data Warehouse.
19. En Server name escriba **serverbd** (localhost) y seleccione la base de datos DW_NombreAlumno.
20. Realice un Test Connection y presione ok.
21. En la ventana principal del editor , en **Data access mode** seleccione **Table or view – Fast Load**, presione new.

Dimensiones

22. Se abrirá la ventana **Create Table** para la creacion de la dimension.
23. En Create Table coloque el nombre de la tabla dimension. (**Create Table [DimProducto]**).
24. Instancie las llaves para cada Dimension como IDProducto numeric (18,0) identity (1,1) . Esto se realiza para cada Dimension.
25. Presione ok , luego en el menú **Mapping** confirme que los campos direccionados desde el **OLE DB Source** lleguen a la tabla Dim y presione OK.
26. Regrese al panel **Control Flow** y desde ToolBox arrastre la herramienta Execute SQL Task.
27. Vincule la flecha de color verde desde **Execute SQL Task** a su **ETL**.

Dimensiones.

28. Realice doble click en **Execute SQL Task.**
29. En Connection Seleccione su Data warehouse y en **SQL Stament** escriba la query **delete from DimProducto.**
30. Guarde los cambios y ejecute su ETL desde Visual Studio.
31. Revise en **SQL Server** su Base de Datos **DW_NombreAlumno** para verificar que los datos se hayan cargado.
32. Este procedimiento se repite para la creación de cada una de las dimensiones.

Tabla Temporal.

1. Arraste desde ToolBox la herramienta **Data Flow Task**. Cambie el nombre **(ETL TempVentas)** y presione doble click para abrir el Panel Data Flow.
2. Arrastre OLE DB Source, y repita los pasos anteriores sin embargo en esta etapa debe copiar por completo la tabla Ventas. En Columns verifique los campos.
3. **OLE DB Destination** en **Create Table** solo cambie el nombre de la Tabla a Temp Ventas y finalmente desde **Mappings** confirme los campos de llegada a la tabla TempVentas y presione ok.
4. Desde **SQL Server** confirme si esta creada la table **TempVentas**.
5. No olvide escribir el **delete from TempVentas** desde **Execute SQL Task**.

Table de Hecho.

1. Arraste desde ToolBox la herramienta **Data Flow Task**. Cambie el nombre (ETL_HechoVentas) y presione doble click para abrir el Panel Data Flow.
2. Arrastre **OLE DB Source**, y repita los pasos anteriores sin embargo en esta etapa debe copiar por completo la tabla **TempVentas** desde el Data warehouse.
3. En **Query Builder** seleccione todas las tablas Dim y TempVentas.
4. Los campos que identificar a cada dimension (no son los ID) debe vincularlos a la tabla **TempVentas**. Seleccione los ID de cada tabla dimension y las medidas nativas.

Tabla de Hecho.

5. Revise los campos desde Columns para verificar su orden.
6. **OLE DB Destination** en **Create Table** solo cambie el nombre de la Tabla a HechoVentas. Incorpore el IDVentas a la tabla de Hecho y finalmente desde **Mappings** confirme los campos de llegada a la tabla HechoVentas y presione ok.
7. Desde **SQL Server** confirme si esta creada la table **HechoVentas**.
8. No olvide escribir el delete from HechoVentas desde Execute SQL Task.
9. Guardamos y Ejecutamos el Proyecto.
10. Por lo tanto usted ya ha creado su primer ETL.