



BUSINESS INTELLIGENCE PROFESSIONAL

Data Automatization

SESIÓN I

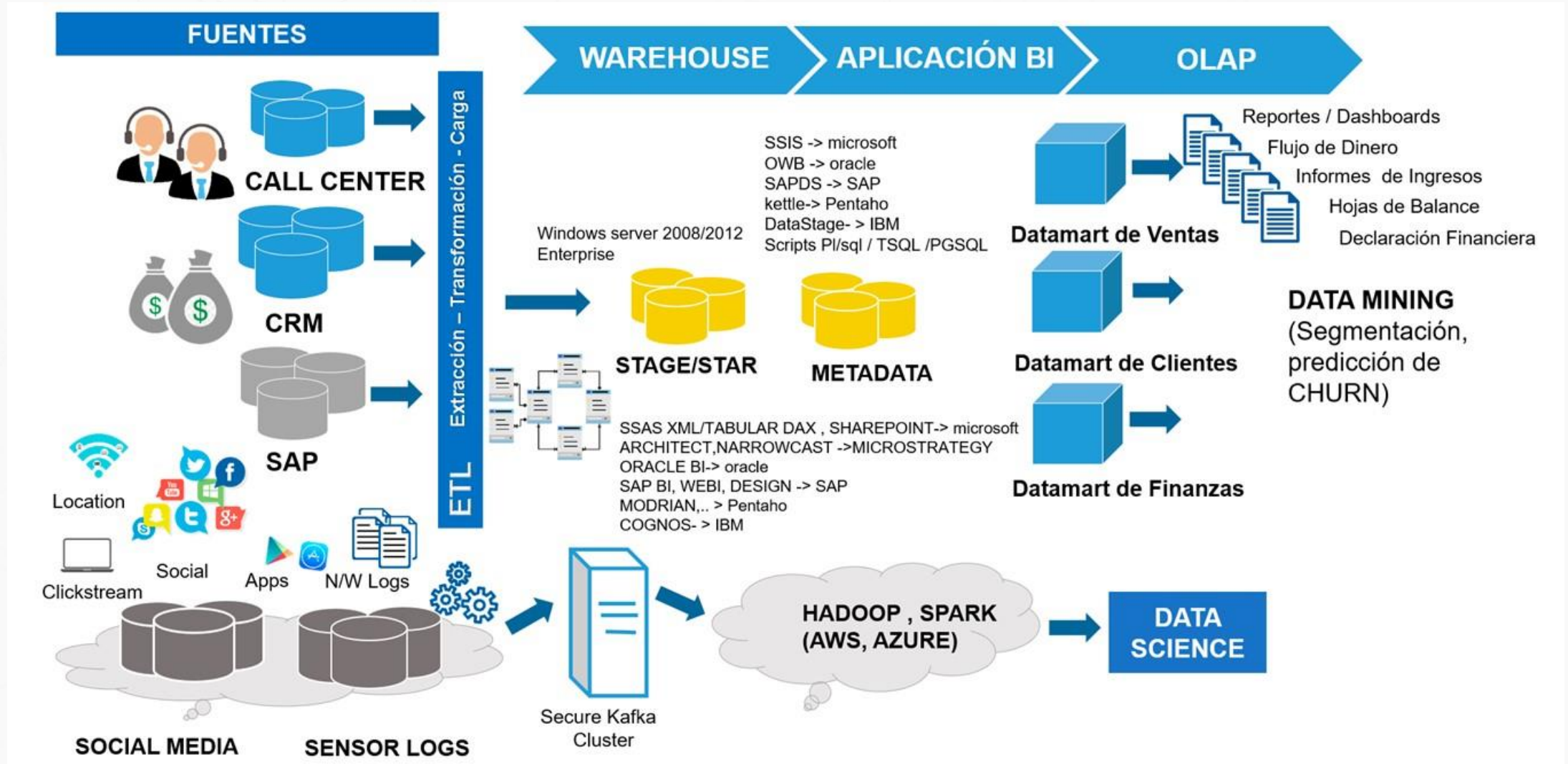
Docente: Victor Gutierrez
Data Architect

Agenda

Introducción a DWH y ETL (SSIS)

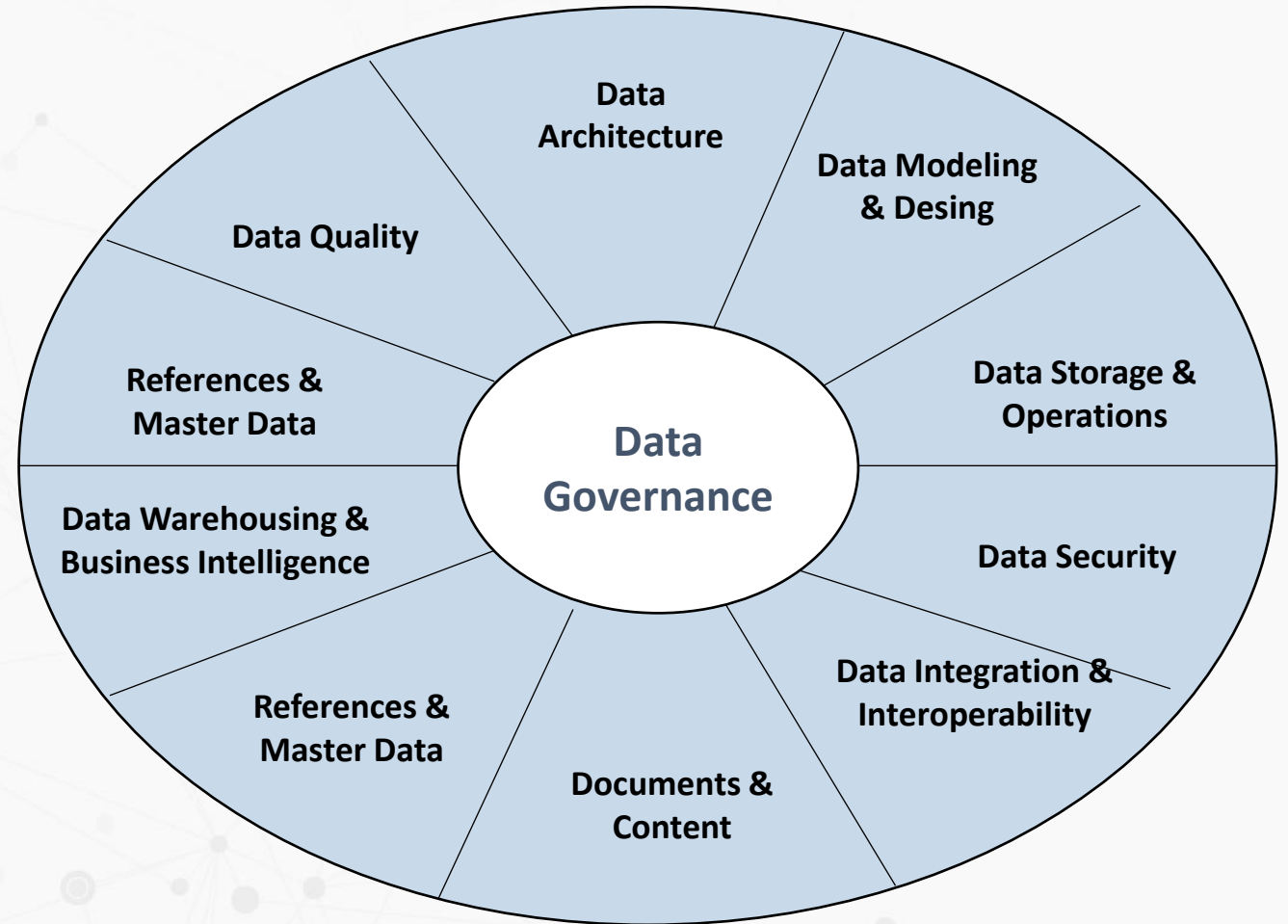
- Overview del Diseño de solución para proyectos DWH / ETL
- Conceptos fundamentales de la herramienta SSIS
- Arquitectura SSIS
- Herramientas de SSIS.
- Soluciones SSIS.
- Asistentes de importación y exportación
- Opciones para implementar procesos ETL

Diseño de solución para proyectos DWH / ETL



Datos Gobernados

- TODOS los datos deben estar gobernados
- Una analítica completa, precisa datos completos
- Los datos ya no son (y no volverán a serlo) propiedad de los departamentos técnicos



El origen de los datos y su transformación para atender los requerimientos operativos y consolidados del negocio

Fuentes:

Generan o complementan las operaciones del negocio

Ticket de venta



ETL

Capa ODS:

Enriquece la data del negocio, dando valor a la data operativa.

Reporte Semanal de Ventas								
Producto	Lunes	Martes	Miércoles	Jueves	Viernes	Total	Proyectado	Diferencia
A	\$ 10.00	\$ 15.00	\$ 12.00	\$ 11.00	\$ 13.00	\$ 61.00	\$ 100.00	-\$ 39.00
B	\$ 30.00	\$ 35.00	\$ 20.00	\$ 40.00	\$ 45.00	\$ 170.00	\$ 200.00	-\$ 30.00
C	\$ 48.00	\$ 49.00	\$ 52.00	\$ 67.00	\$ 87.00	\$ 303.00	\$ 300.00	\$ 3.00
D	\$ 12.00	\$ 15.00	\$ 24.00	\$ 26.00	\$ 28.00	\$ 105.00	\$ 100.00	\$ 5.00
E	\$ 80.00	\$ 89.00	\$ 85.00	\$ 72.00	\$ 83.00	\$ 409.00	\$ 400.00	\$ 9.00
F	\$ 60.00	\$ 59.00	\$ 41.00	\$ 58.00	\$ 50.00	\$ 268.00	\$ 250.00	\$ 18.00
G	\$ 10.00	\$ 14.00	\$ 15.00	\$ 18.00	\$ 20.00	\$ 77.00	\$ 50.00	\$ 27.00
H	\$ 32.00	\$ 35.00	\$ 38.00	\$ 39.00	\$ 37.00	\$ 181.00	\$ 150.00	\$ 31.00
I	\$ 20.00	\$ 22.00	\$ 18.00	\$ 19.00	\$ 20.00	\$ 99.00	\$ 100.00	-\$ 1.00
J	\$ 66.00	\$ 67.00	\$ 61.00	\$ 70.00	\$ 78.00	\$ 342.00	\$ 350.00	-\$ 8.00
K	\$ 10.00	\$ 19.00	\$ 18.00	\$ 17.00	\$ 15.00	\$ 79.00	\$ 50.00	\$ 29.00
Total	\$ 378.00	\$ 419.00	\$ 384.00	\$ 437.00	\$ 476.00	\$ 2,094.00	\$ 2,050.00	\$ 44.00

Capa BDS:

Consolida la información operativa, se busca validar los indicadores de la data del negocio

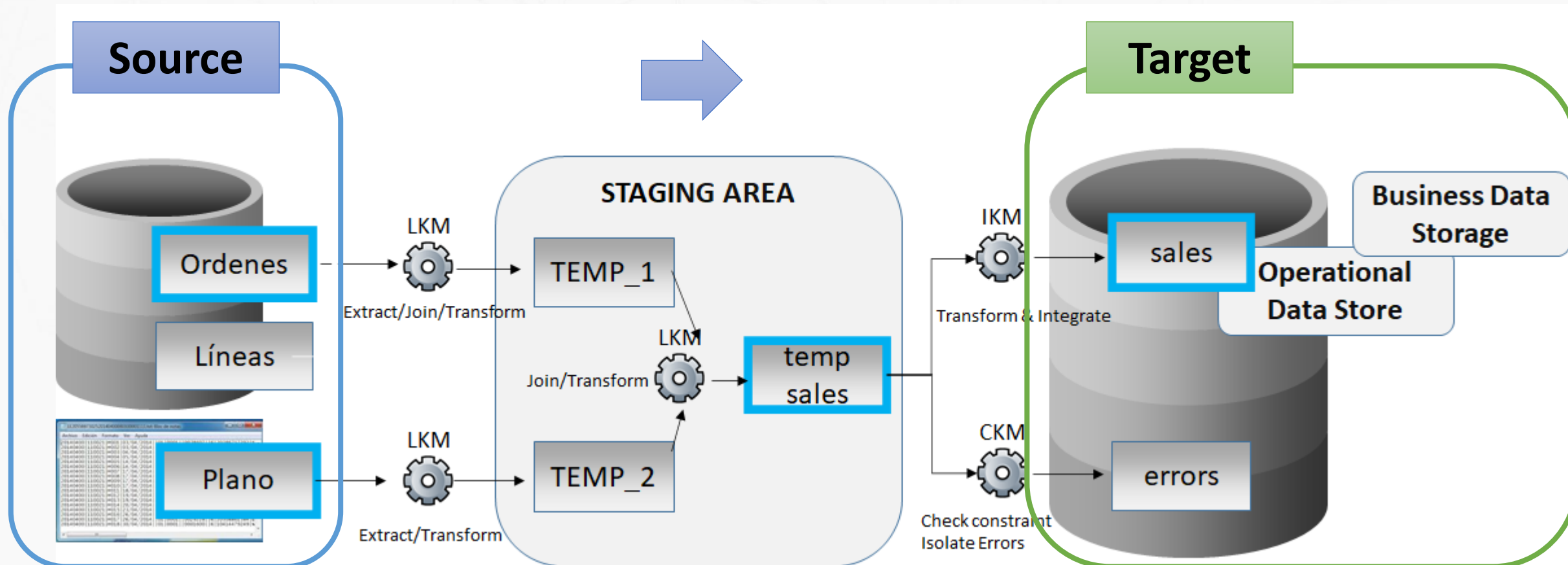


Información del Cliente



Desde la vista de diseño de la solución:

- El proceso de integración, busca homologar las fuentes heterogéneas.
- Diseñado para contener datos de bajo nivel o atómicos (indivisibles) (como transacciones y precios) con un historial limitado que se captura "en tiempo real" o "casi en tiempo real"



STG / Staging Area

Stage - Operational Data Storage que viene a ser una etapa previa al ambiente ODS. Esta es el punto de comunicación entre los sistemas fuente y el DWH. Aquí se realizan las tareas de limpieza de datos, normalización y transformación de la información para que entre de manera limpia al ambiente ODS. El ODSSTG es el punto de medición de la calidad de información que proviene de los aplicativos fuente.

ODS

Operational Data Storage. Dentro de las buenas prácticas de un DWH el ambiente ODS es usado para la integración de los aplicativos fuentes en un modelo valido para toda la compañía y se encuentra orientada por temas. Adicionalmente mantiene la información actualizada y tiene una naturaleza histórica. ODS tiene la mayor atomicidad en detalle de la información en todo el DWH

BDS

Bussines Data Storage. Forma parte de las buenas prácticas dentro de un DWH. El ambiente BDS es usado para contener la información resumida y agregada de lo que se tiene en ODS. Este ambiente es usado para la exploración analítica de lo que ocurra en el negocio pues ya se ha procesado y resumido la información operacional y sus estructuras de datos se encuentran preparadas para procesar consultas con un mayor performance que un modelo relacional

Especificaciones

Arquitectura

- Debe soportar Reproceso
- Dividir Procesos Diarios y Mensuales
- El proceso mensual deberá integrar las fuentes.

CRM

- Debe Categorizar los reclamos del cliente
- Debe Categorizar al cliente por el numero de líneas

Tráfico

- Debe acumular y categorizar el consumo de internet (Usar Data Entry)
- Debe acumular y aplicar los descuentos del consumo de voz según país destino.

Modelo ODS y BDS

CRM

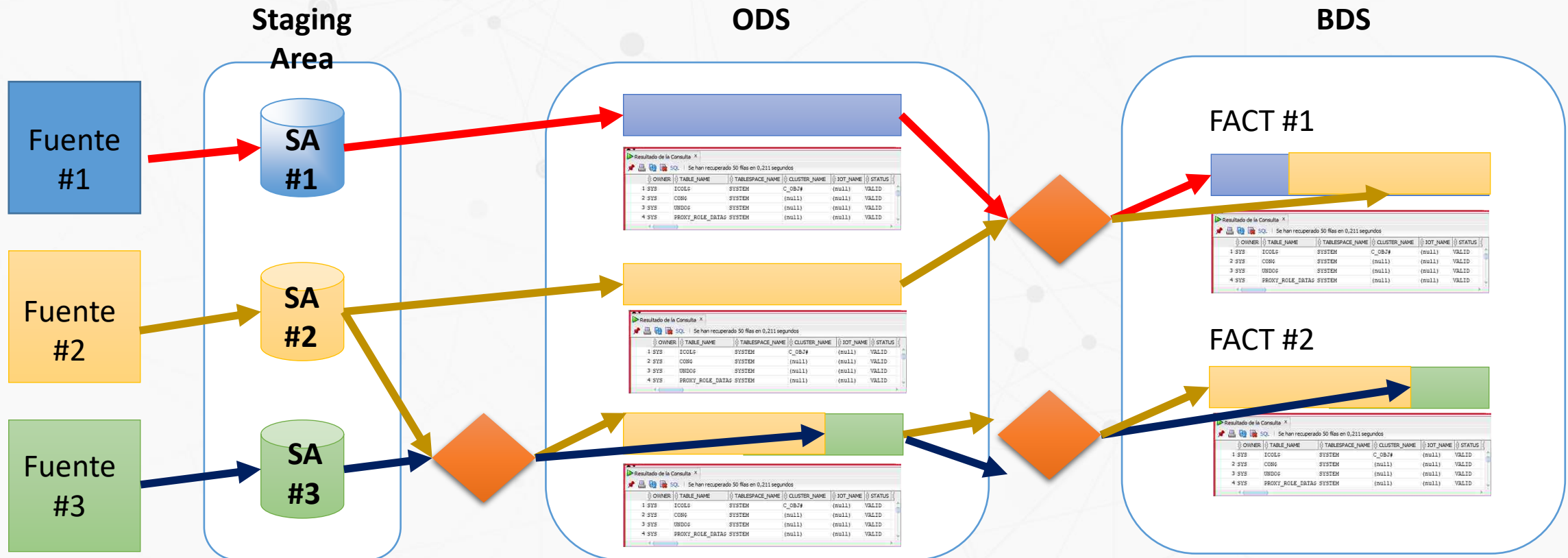
TRAFICO

DATA ENTRY



Diseño de Malla

- Es la secuencia ordenada de trazabilidad de los datos
- Implementación de la integración de datos.
- Deberá pensarse en ejecuciones por fuentes
- Deberá tener puntos de control



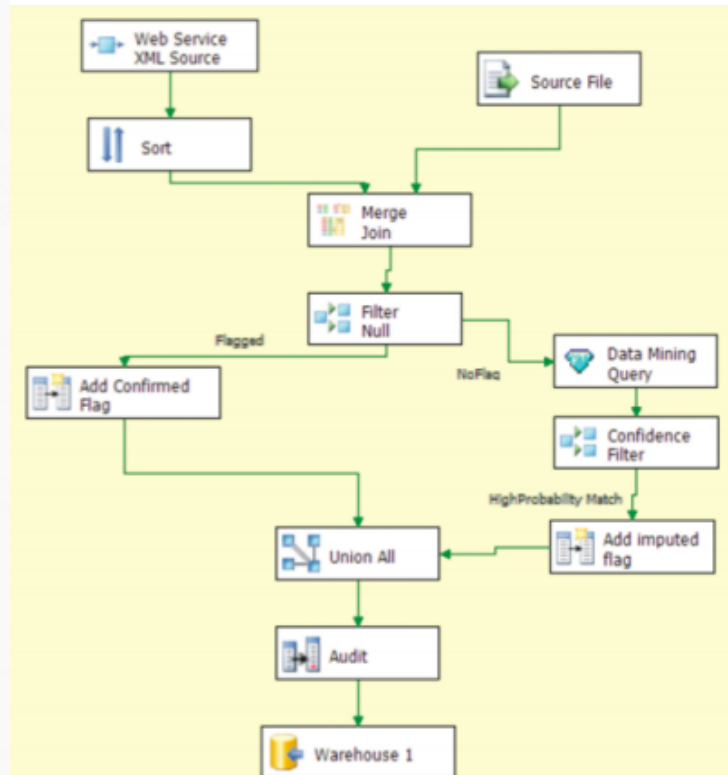
- Servicio de Integración de Datos de Microsoft BI.
- Herramienta basada en el concepto de Flujo de datos
- Herramienta para la **E**xtracción, **T**ransformación, carga de datos. (**L**oad)
- Herramienta para realizar operaciones DML/DDL,
- Herramienta para automatizar el mantenimiento y administración de base de datos

Conceptos herramienta SSIS

Road Map

Visual Studio

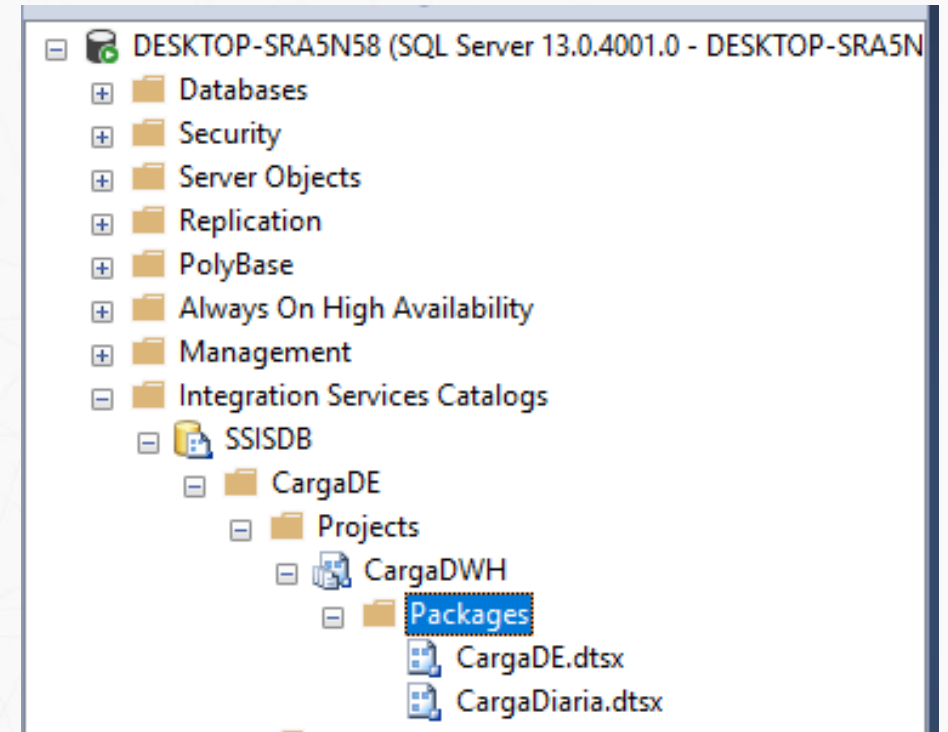
- Se construye el flujo de datos
- Se enlazan las tareas, que permite generar la trazabilidad



Despliegue

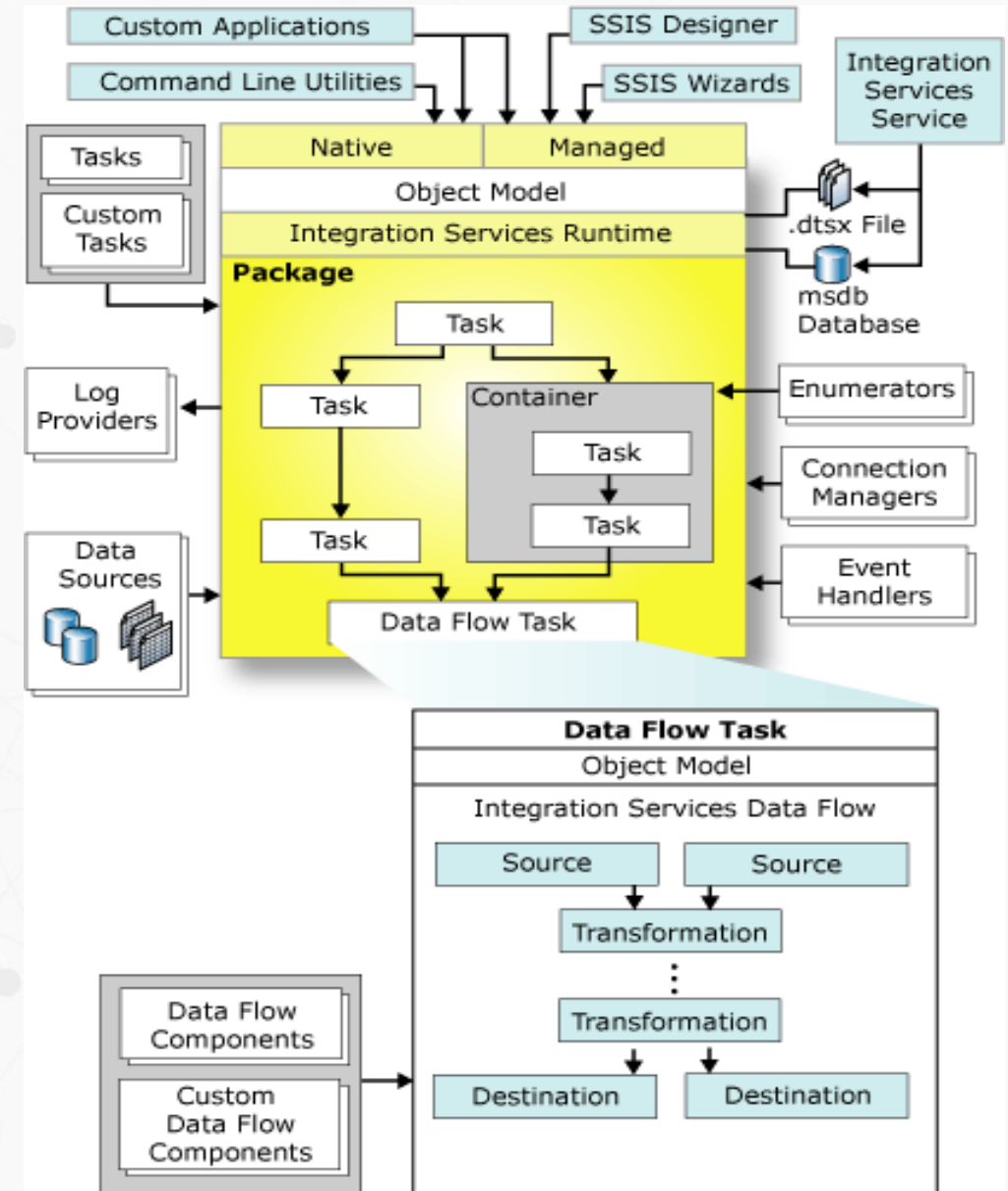
Catalogo de Servicios SSIS

- Al desplegar el proyecto, este refleja los packages
- SQL Agent, se procede a crear los jobs



Integration Services (SSIS) se compone de:

- SSIS Designer
- Integration Services Data Flow
- Integration Services Runtime
- Integration Services Object Model



Componente	Descripción
SSIS Designer	<ul style="list-style-type: none">SSIS es una herramienta gráfica que puede utilizar para crear y mantener paquetes de Integration Services.
Runtime Engine	<ul style="list-style-type: none">El runtime de Integration Services guarda el diseño de paquetes, corre paquetes, y proporciona apoyo a la auditoria, los puntos de interrupción, configuración, conexiones y transacciones.
Data Flow Engine	<ul style="list-style-type: none">La tarea Flujo de datos encapsula el motor de flujo de datos.El motor de flujo de datos proporciona los buffers de memoria en que se mueven los datos desde el origen al destino, y llama a las fuentes que extraen datos de archivos y bases de datos relacionales.El motor de flujo de datos también gestiona las transformaciones que modifican los datos, y los destinos que se cargan los datos o hacen que los datos disponibles para otros procesos.
Object Model	<ul style="list-style-type: none">El modelo de objetos de Integration Services incluye application programming interfaces (API) gestionadas para la creación de componentes personalizados para su uso en paquetes o aplicaciones personalizadas que crea, carga, ejecuta y administra paquetes.

Herramienta SSIS

Vista de un proyecto SSIS (Visual Studio)

Flujo de Control, Flujo de Datos, Controladores de Eventos y Explorador de Paquete aquí

Solution Explorer
Administre su proyecto SSIS de aquí

ToolBox
Seleccionar componentes pre-definidos de aquí

Connection Managers
Crear/administrar las conexiones de origen de datos para su proyecto

Properties
Cambiar la configuración de los componentes

The screenshot shows the Visual Studio interface for an SSIS project named 'calugas'. The main design area displays the 'Flujo de control' (Control Flow) task. The left sidebar contains the 'Cuadro de herramientas de SSIS' (SSIS Toolbox) with a search bar and a list of tasks like 'Execute SQL Task' and 'Tarea Flujo de datos'. The right sidebar shows the 'Explorador de soluciones' (Solution Explorer) with the project structure, including 'Project.params', 'Administradores de conexiones', 'Paquetes SSIS', and 'Varios'. Below the Solution Explorer is the 'Propiedades' (Properties) window for the selected 'calugas' component, showing the 'Ubicación' (Location) property with the value 'C:\Calugas\DE'. A callout points to the 'Flujo de control' task in the design area, indicating it is the selected view.

Package1.dtsx [Diseño]

Flujo de control | Flujo de datos | Parámetros | Controladores de eventos | Explorador de paquete

100%

Información... Arrastre un elemento del cuadro

calugas	
Ubicación	
Nombre	calugas
RutaDeAccesoCompleta	C:\Calugas\DE

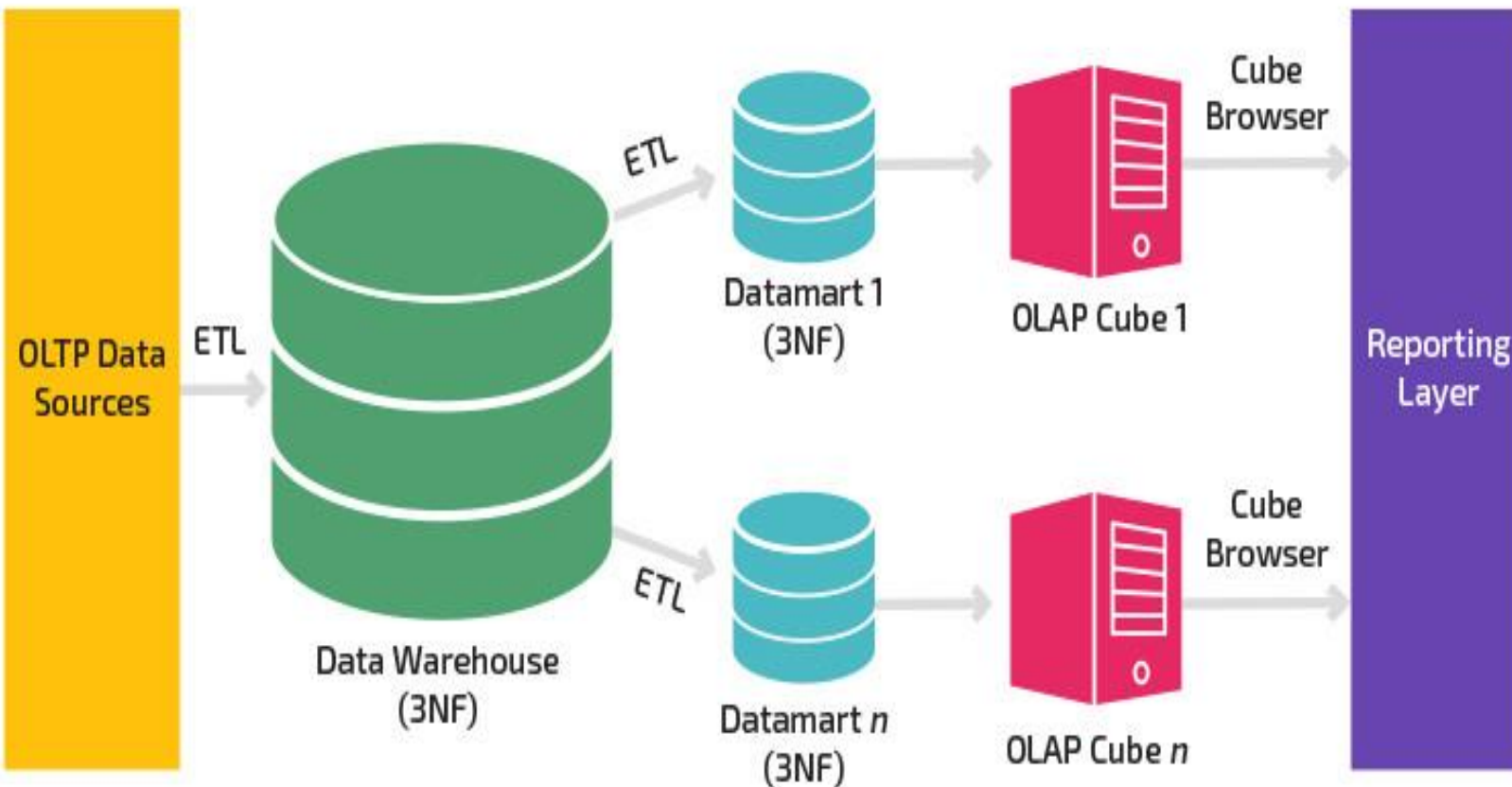
DEMO 1

Despliegue

Consideraciones para la infraestructura de un Data Warehouse



Inmon Model



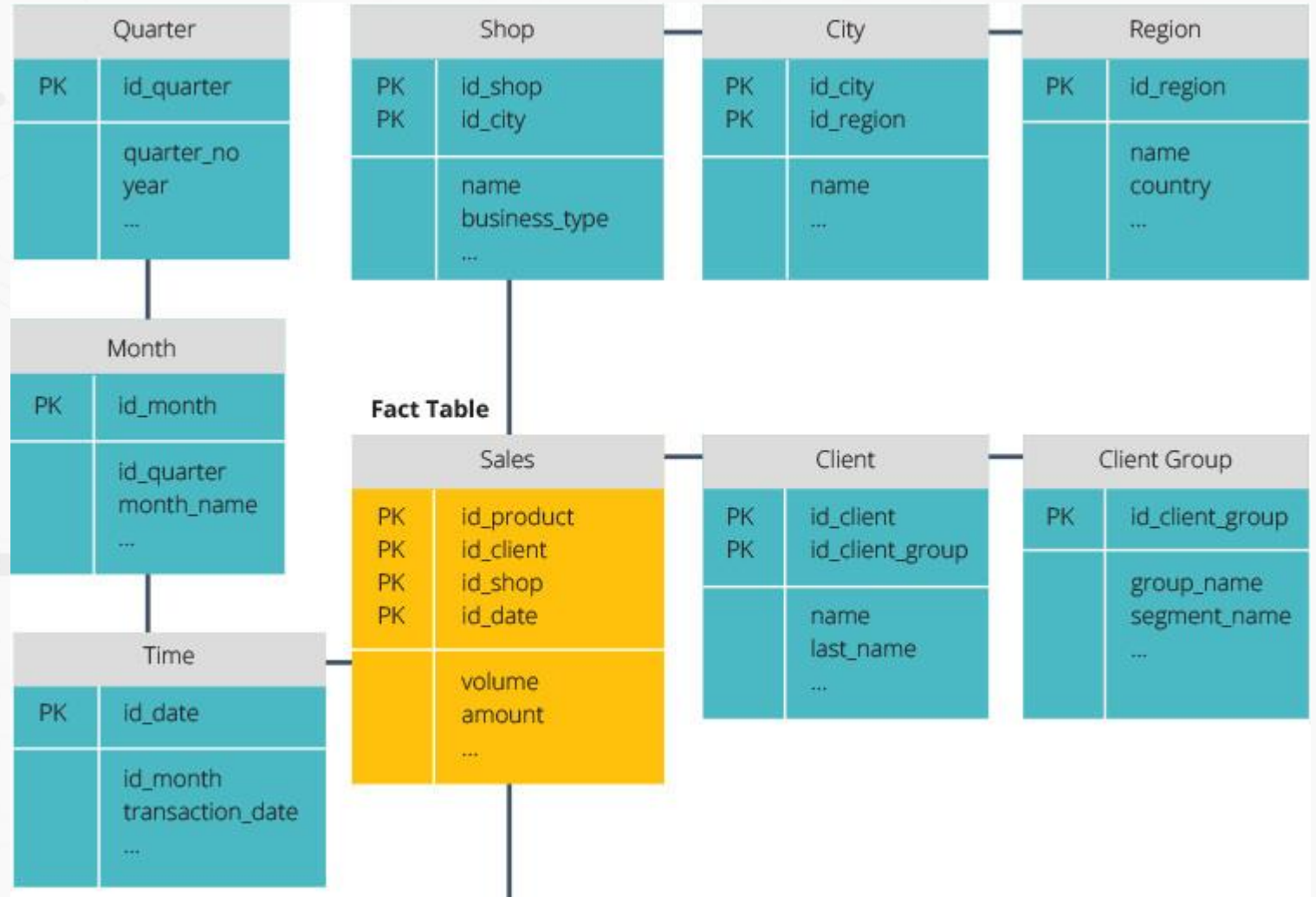
El enfoque de Bill Inmon favorece un diseño descendente en el que el DW es el depósito de datos centralizado y el componente más importante de los sistemas de datos de una organización.

El enfoque de Inmon primero construye el modelo centralizado de datos corporativos, y el almacén de datos se considera la representación física de este modelo. Los almacenes de datos dimensionales relacionados con líneas comerciales específicas se pueden crear desde el almacén de datos cuando se necesitan.

Opciones para implementar procesos ETL

El esquema de copo de nieve

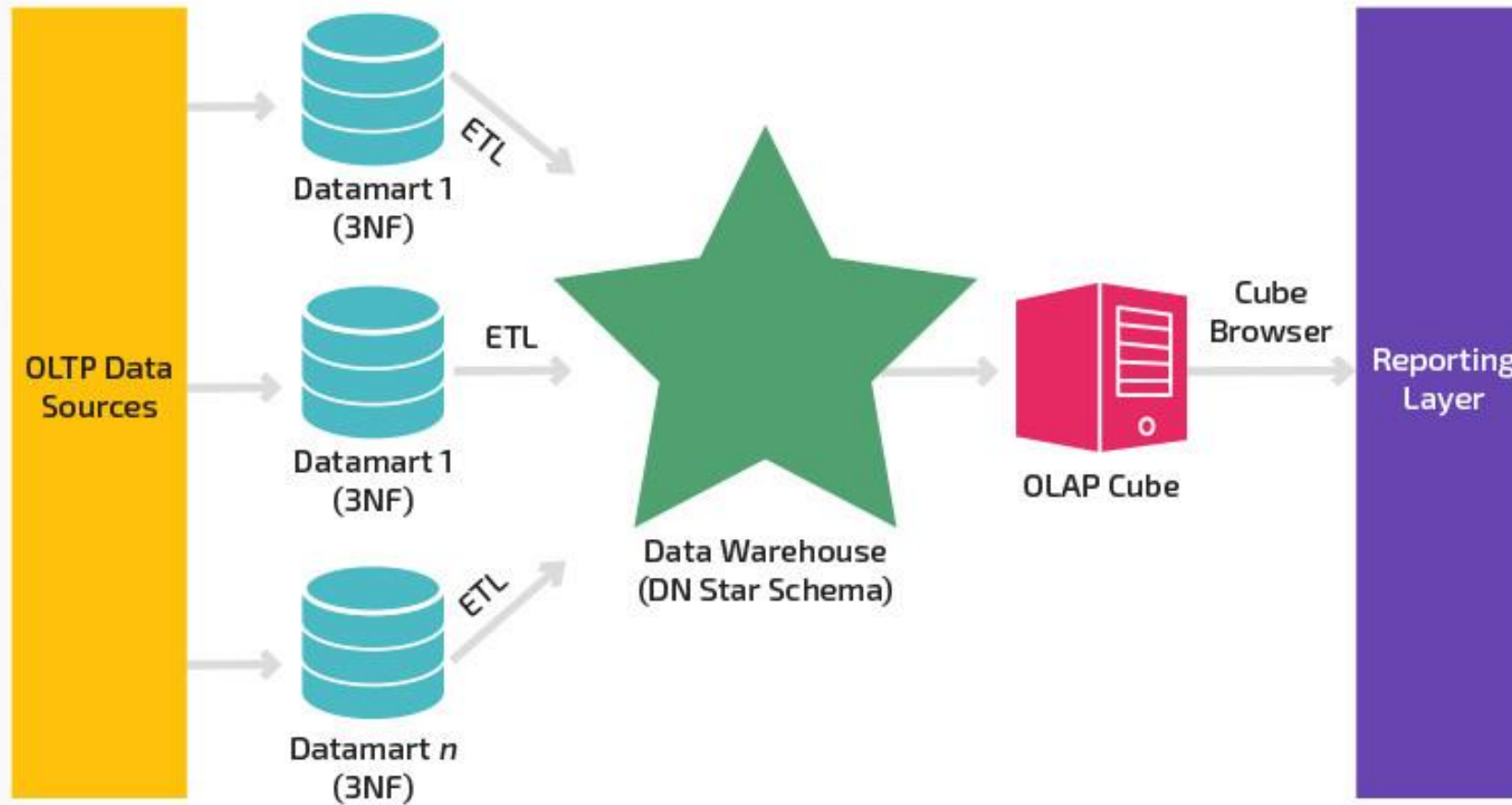
Divide la tabla de hechos en una serie de tablas de dimensiones normalizadas. La normalización crea más tablas de dimensiones con múltiples combinaciones y reduce los problemas de integridad de datos. Sin embargo, las consultas son más desafiantes usando el esquema de copo de nieve, porque las consultas deben profundizar más para acceder a los datos relevantes.



Opciones para implementar procesos ETL

Modelo de datos

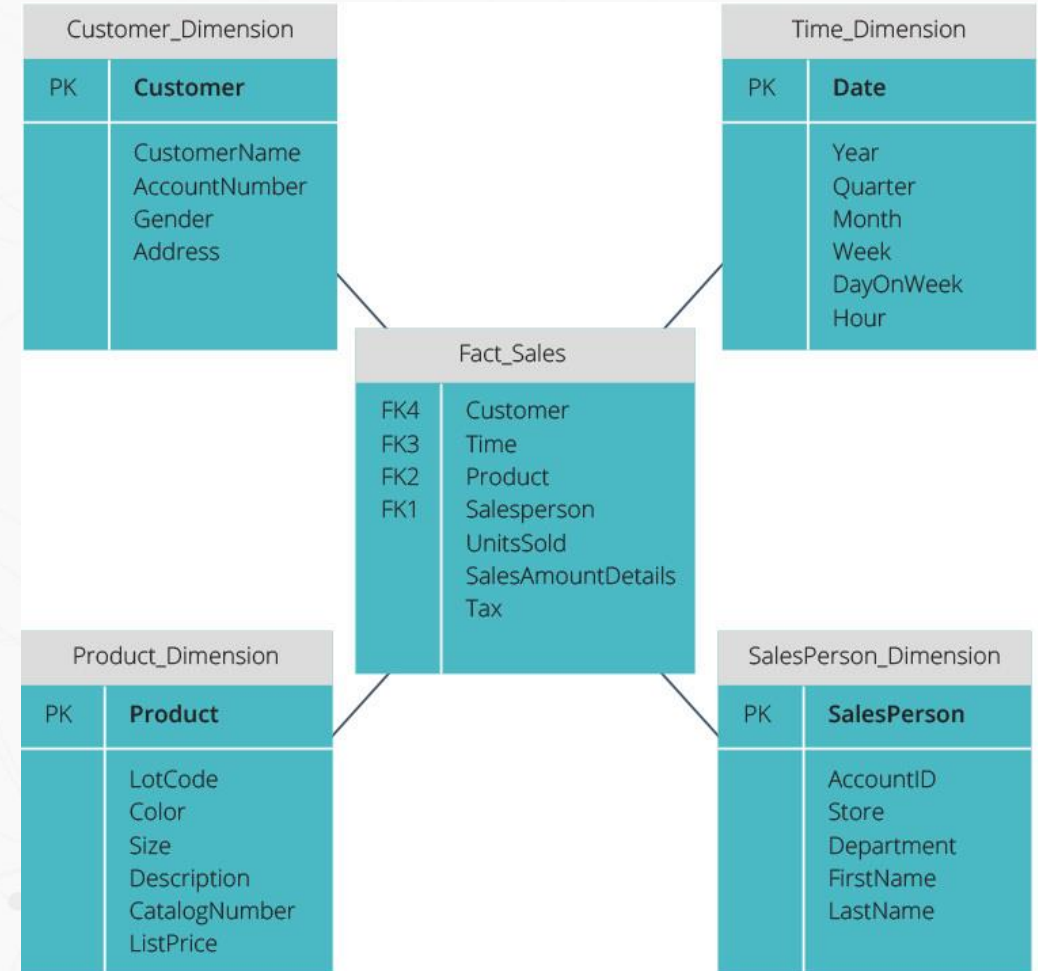
Kimball Model



El diseño DW de Ralph Kimball comienza con los procesos comerciales más importantes. En este enfoque, una organización crea DM que agregan datos relevantes en torno a áreas específicas del tema. El DW es la combinación de los DM individuales de la organización.

El esquema en estrella

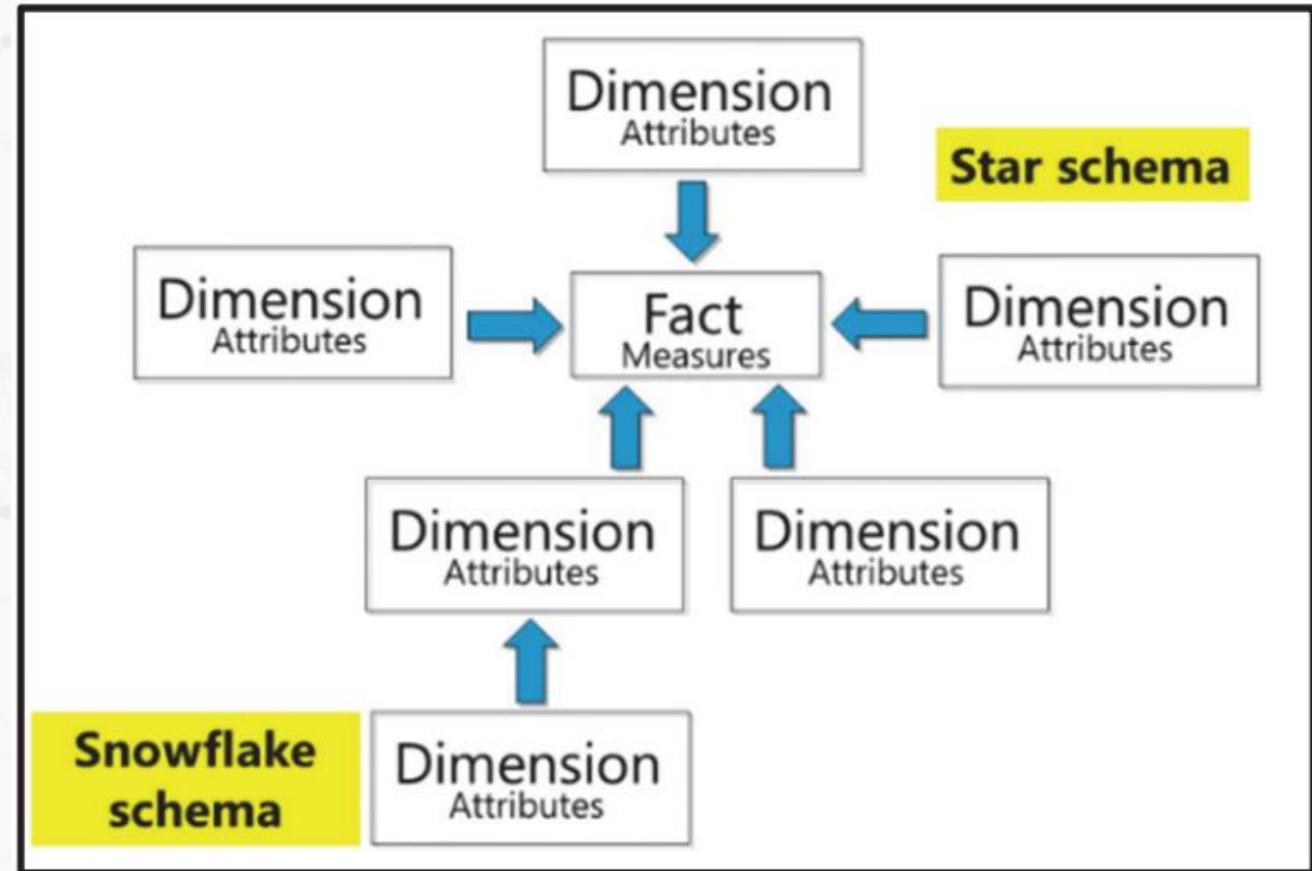
Toma la información de la tabla de hechos y la divide en tablas de dimensiones desnormalizadas. El énfasis para el esquema en estrella está en la velocidad de consulta. Solo se necesita una combinación para vincular tablas de hechos a cada dimensión, por lo que consultar cada tabla es fácil.



Legend: PK = primary key | FK = foreign key

Data Warehouse Design Overview

- Aplicar un modelo dimensional del DWH
- Aplicar un proceso para el diseño del DWH
- Utilice un enfoque basado en procesos de negocio para el modelado dimensional.
- Documentar diseños de modelos dimensionales.





The Data Warehouse Design Process

Aunque cada proyecto tiene sus consideraciones **singulares**, existe un proceso de común para *diseñar* un DWH

El método se basa en gran medida en el DWH. patrones de diseño identificados y documentados por Kimball

1. Determine analytical and reporting requirements
2. Identify the business processes that generate the required data
3. Examine the source data for those business processes
4. Conform dimensions across business processes
5. Prioritize processes and create a dimensional model for each
6. Document and refine the models to determine the database logical schema
7. Design the physical data structures for the database

Dimensional Modeling

Después de identificar los procesos comerciales y dimensiones, se pueden representar en una matriz

Esta técnica se basa en la matriz de procesos de Kimball.

Business Processes	Time	Product	Customer	Salesperson	Factory Line	Shipper	Account	Department	Warehouse
Manufacturing	x	x			x				
Order Processing	x	x	x	x					
Order Fulfillment	x		x			x			
Financial Accounting	x						x	x	
Inventory Management	x	x							x

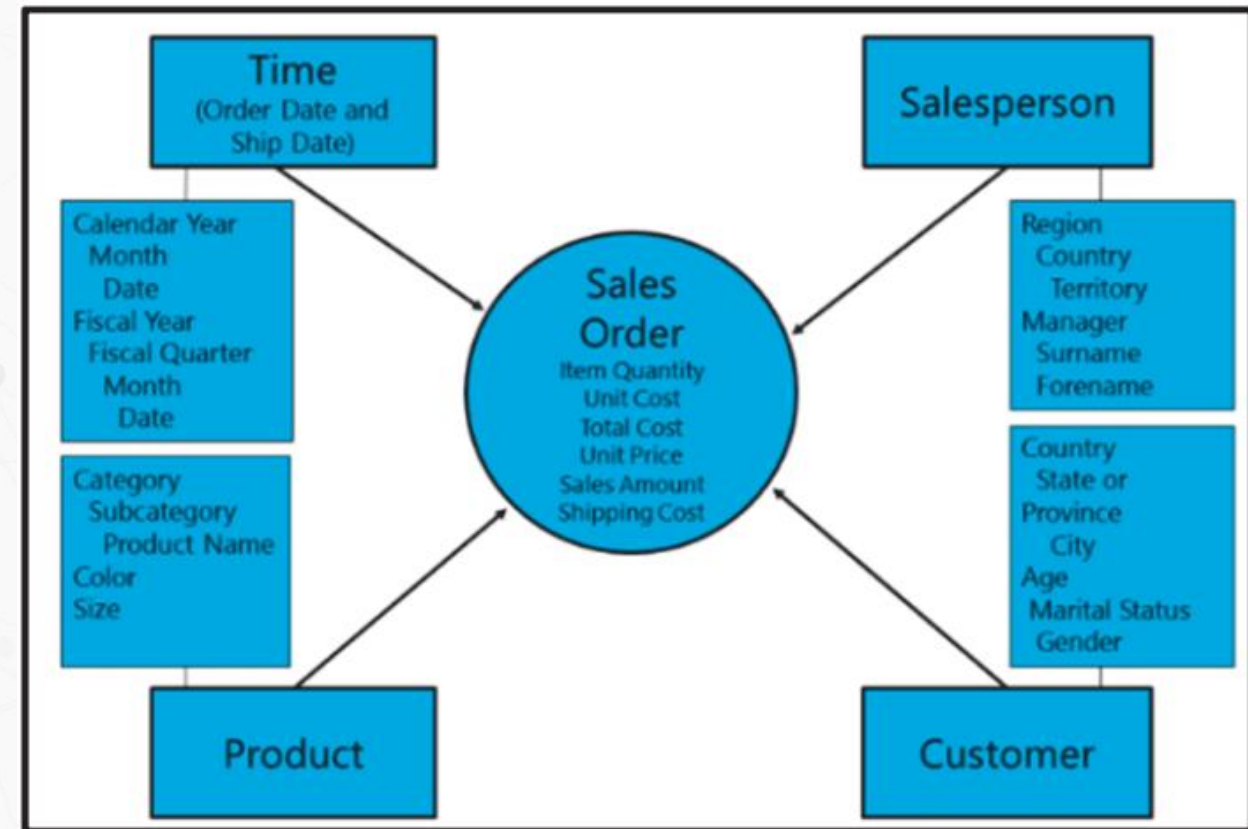
- **Grain:** 1 row per order item
- **Dimensions:** Time (order date and ship date), Product, Customer, Salesperson
- **Facts:** Item Quantity, Unit Cost, Total Cost, Unit Price, Sales Amount, Shipping Cost

Documenting Dimensional Models

Luego de diseñar los modelos dimensionales para cada *proceso empresarial*, cree un diagrama simple. Dejando al centro las tablas de hecho con las dimensiones a las que se relaciona.

A medida que refina el modelo dimensional, puede agregar más atributos

También puede tener una jerarquía para un análisis mas detallado, por ejemplo, ver agregaciones de la cantidad de ventas por año, mes y fecha o por país, estado y ciudad.



Types of Fact Table

Tablas de hechos de transacciones. El tipo más común de hechos es una tabla de hechos de "transacciones", en la que cada fila registra una transacción o evento

Tablas periódicas. Estos registros miden valores en un momento específico.

Tablas Acumuladas. Estos se pueden usar en escenarios en los que es posible que desee usar un hecho tabla para seguir el progreso de un proceso empresarial a través de múltiples etapas

• Transaction Fact Table

OrderDateKey	ProductKey	CustomerKey	OrderNo	Qty	Cost	SalesAmount
20160101	25	120	1000	1	125.00	350.99
20160101	99	120	1000	2	2.50	6.98
20160101	25	178	1001	2	250.00	701.98

• Periodic Snapshot Fact Table

DateKey	ProductKey	OpeningStock	UnitsIn	UnitsOut	ClosingStock
20160101	25	25	1	3	23
20160101	99	120	0	2	118

• Accumulating Snapshot Fact Table

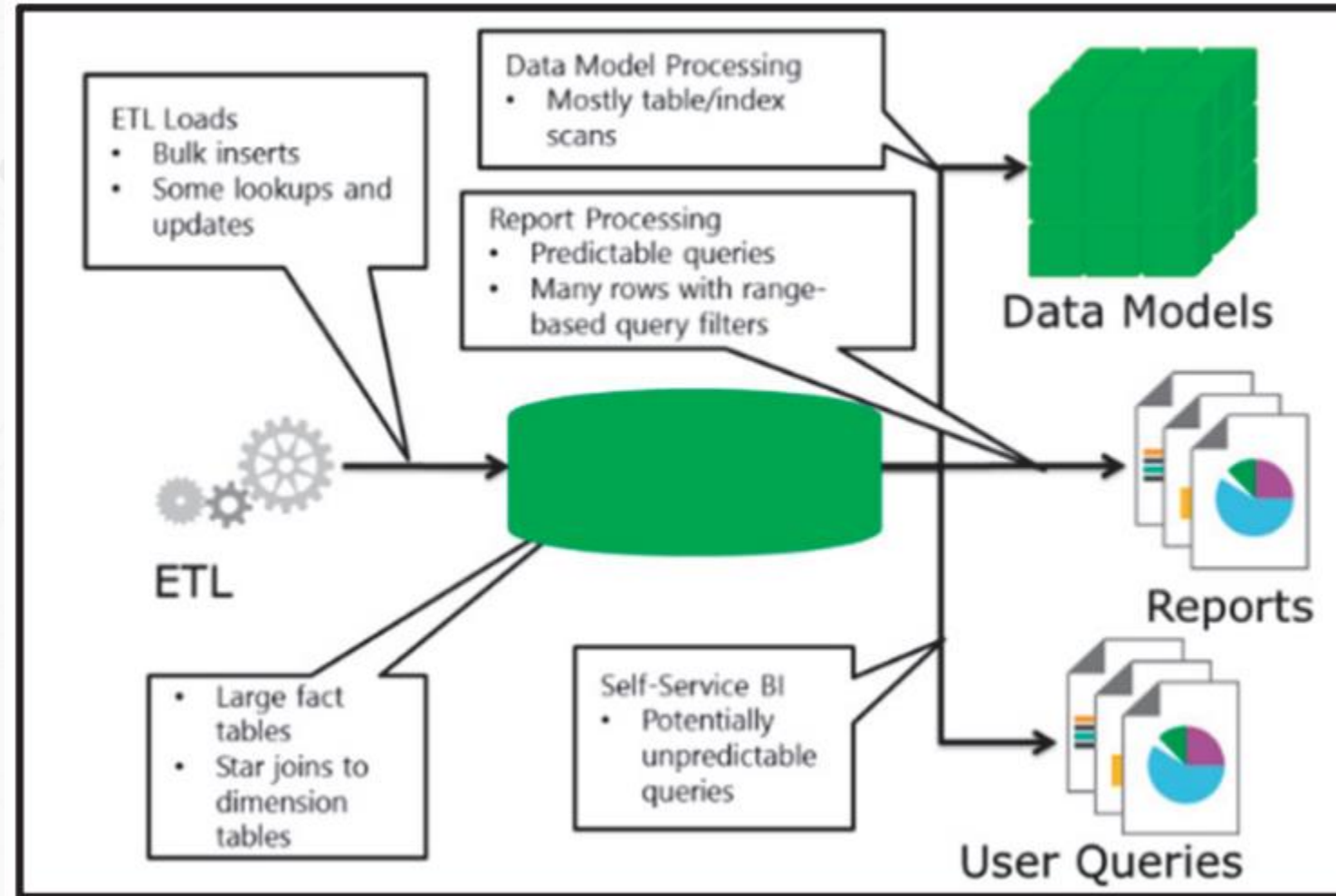
OrderNo	OrderDateKey	ShipDateKey	DeliveryDateKey
1000	20160101	20160102	20160105
1001	20160101	20160102	00000000
1002	20160102	00000000	00000000

Data Warehouse I/O Activity

Antes de diseñar la base de datos física para el DWH, es útil considerar los tipos de la carga de trabajo que debe soportar y *donde almacenarla*.

La base de datos debe almacenar potencialmente tablas grandes con millones de filas.

Normalmente, la actividad de I/O en la base de datos es generada por una de las cargas de trabajo o por operaciones de mantenimiento, como copias de seguridad.



Azure Synapse Analytics

<https://docs.microsoft.com/es-es/azure/sql-data-warehouse/sql-data-warehouse-overview-what-is>

SQL Server Integration Services

<https://docs.microsoft.com/en-us/sql/integration-services/sql-server-integration-services?view=sql-server-ver15>

SSIS How to Create an ETL Package

<https://docs.microsoft.com/en-us/sql/integration-services/ssis-how-to-create-an-etl-package?view=sql-server-ver15>



BUSINESS INTELLIGENCE PROFESSIONAL