



UNIVERSIDAD NACIONAL DE LA MATANZA

INTELIGENCIA DE NEGOCIOS

Estrategias de Integración y Procesamiento de Datos **Proceso ETL - Implementación**

Docentes: ING. LORENA R. MATTEO / LIC. MARCELO N. SAIED

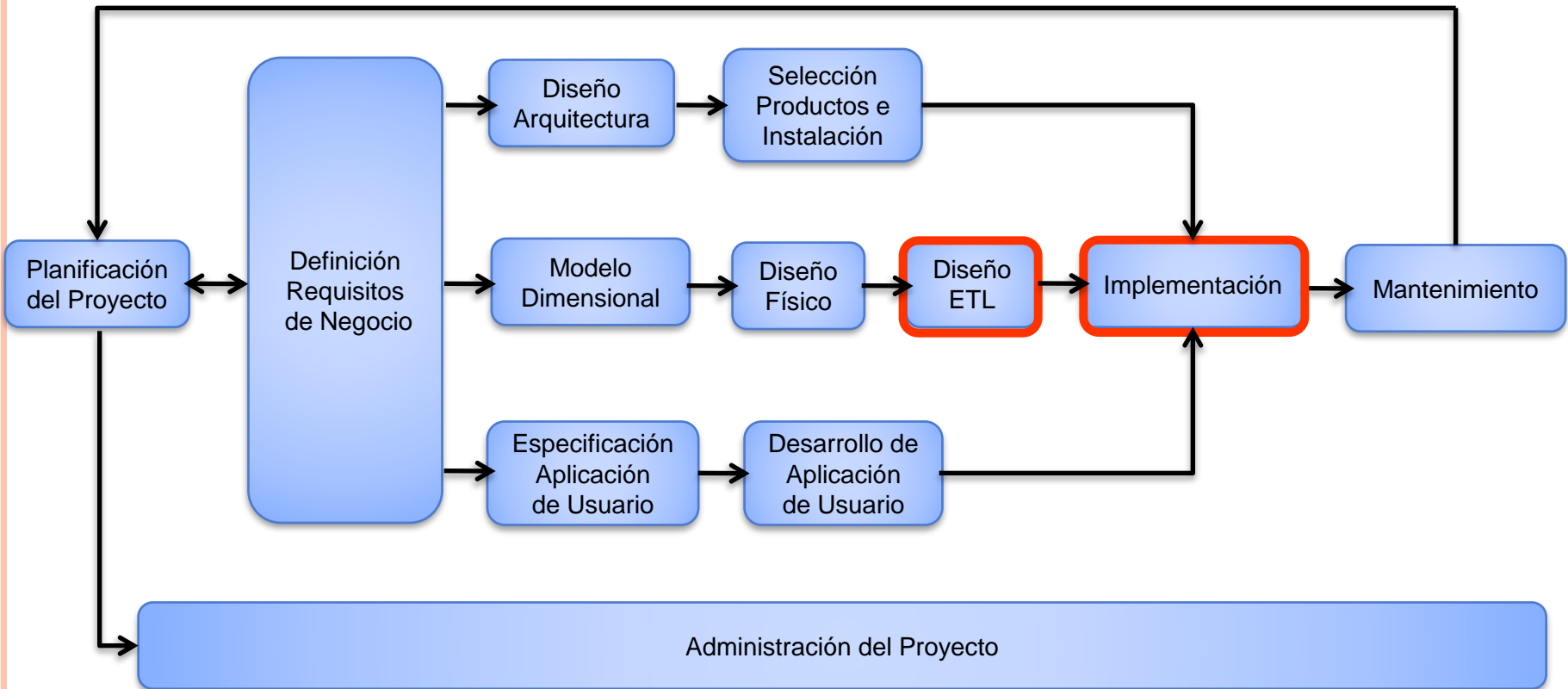
Autores ppt orig.: Lic. HUGO M. CASTRO / MG. DIEGO BASSO



CICLO DE VIDA DE UN PROYECTO DE BI

14/5/2025

IN2025



INTEGRACIÓN Y PROCESAMIENTO DE DATOS EN ARQUITECTURAS



¿Por qué es clave?

La integración de datos permite unificar la información proveniente de distintas fuentes (ERP, CRM, sensores, redes sociales, etc.) para convertirla en conocimiento útil para la toma de decisiones. El procesamiento transforma esos datos crudos en estructuras limpias, accesibles y valiosas.

Enfoques Clásicos: ETL/ELT

○ ETL (Extract – Transform – Load)

- Extrae datos → los transforma → luego los carga a un Data Warehouse.
- Ideal para ambientes controlados, estructuras rígidas y necesidades de reportes estables.

○ ELT (Extract – Load – Transform)

- Extrae datos → los carga sin procesar → luego transforma usando la potencia del motor de base de datos (muy usado en la nube).

INTEGRACIÓN Y PROCESAMIENTO DE DATOS EN ARQUITECTURAS



El Desafío de los Datos Modernos

- Volumen y variedad de datos sin precedentes provenientes de numerosas fuentes.
- Las organizaciones luchan por gestionar y aprovechar estos datos de manera efectiva.
- La creciente prevalencia de silos de datos y aplicaciones dificulta el flujo de información.
- Los métodos tradicionales de gestión de datos son complejos y requieren mucho tiempo.




INTEGRACIÓN Y PROCESAMIENTO DE DATOS EN ARQUITECTURAS



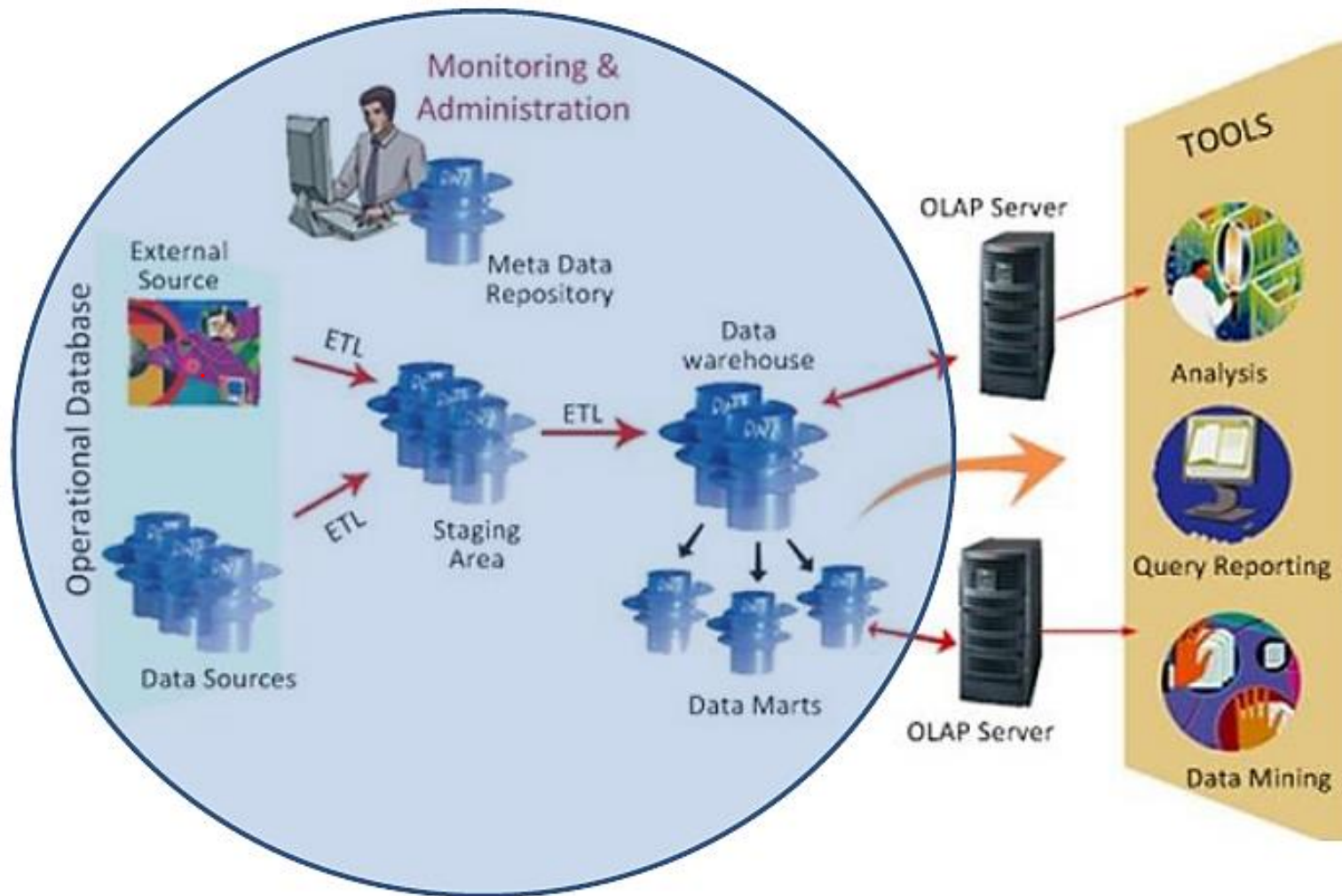
Tendencias Modernas en Integración y Procesamiento

- Data Vault
- Data Mesh
- Data Fabric

Veremos cómo cada uno de estos modelos cambia la manera en que integramos y transformamos los datos.

 *"Del ETL clásico a las arquitecturas distribuidas y automatizadas: comprender los modelos modernos es clave para un BI efectivo y escalable."*

PROCESO ETL - UBICACIÓN





PROCESO ETL

- ¿Qué es ETL?
 - ✓ E - Extract (Extracción)
 - ✓ T - Transform (Transformación)
 - ✓ L - Load (Carga)
- Es el proceso que se define para tomar los datos de los sistemas fuente y cargarlos en el DW.
- Consume el 70-80% del tiempo y esfuerzo de la construcción de un DW.
- El proceso ETL es muy complejo y consume muchos recursos.



PROCESO ETL - FUNCIONES

- Transfiere información de los sistemas transaccionales y fuentes externas al DW.
- Guarda los datos luego de transferirlos al área de preparación o *Staging Area*.
 - Minimizar al máximo nivel los posibles errores o problemas en la fase de carga de los procesos ETL, normalmente se reserva un área de disco para poder recuperar los datos por etapas.
- Normaliza diversas fuentes de datos.
- Corrige datos incorrectos.
- Completa información faltante.
- Detecta problemas de calidad en los datos.



PROCESO ETL - FUNCIONES

- Mantiene las dimensiones del modelo dimensional.
- Agrega filas a las tablas de hechos.
- Genera métricas
 - Tiempos
 - Volúmenes de datos
 - Incidentes
- Permite puntos de control y auditoría.



PROCESO ETL

- Alguno de los subsistemas que tiene que incluir:
 - **Data Quality:** son procesos y tecnologías que permiten asegurar la calidad de los datos a las necesidades de negocio.
 - **Data Profiling:** es el proceso de examinar los datos que existen en las fuentes de origen de una organización y recopilar estadísticas e información sobre los mismos.
 - **Data Cleansing:** es el proceso de detectar o descubrir y corregir datos corruptos, incoherentes o erróneos.



PROCESO ETL

✓ Extracción

- Se extraen los datos desde los distintos sistemas fuentes (internos y externos).
- Datos con diferentes organización y formatos.
- Se programa en horarios en que el impacto sea mínimo.

✓ Transformación

- Aplica reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados.
- Los datos se filtran, limpian, completan, homogenizan y se agrupan.



PROCESO ETL

✓ Ejemplos de Transformación

- Seleccionar sólo ciertas columnas para su carga
 - ❑ No cargar columnas con valores nulos
- Traducir códigos
 - ❑ si la fuente almacena una "H" para Hombre y "M" para Mujer pero el DW tiene que guardar "1" para Hombre y "2" para Mujer.
- Codificar valores libres
 - ❑ Convertir "Hombre" en "H"
- Obtener nuevas medidas calculadas
 - ❑ $\text{Importe_Venta} = \text{cantidad} * \text{precio}$
- Calcular totales de múltiples filas de datos
 - ❑ Ventas totales por cada producto



PROCESO ETL

✓ Carga

- Los datos transformados se ordenan, se consolidan, se verifica la integridad y se incorporan al DW (conjunto de tablas que van a consultar los usuarios).
- En las dimensiones en general se hacen Inserts o Updates.
- Lógica para manejar claves SK y claves operacionales.
- Tabla de Hechos:
 - ❑ Separar inserts de updates: ideal es que sólo hayan inserts.
 - ❑ Carga incremental: distintos períodos de tiempo para la carga de datos.
 - ❑ Recuperación de fallas: Estar preparado ante fallas.
- Es un proceso repetitivo.



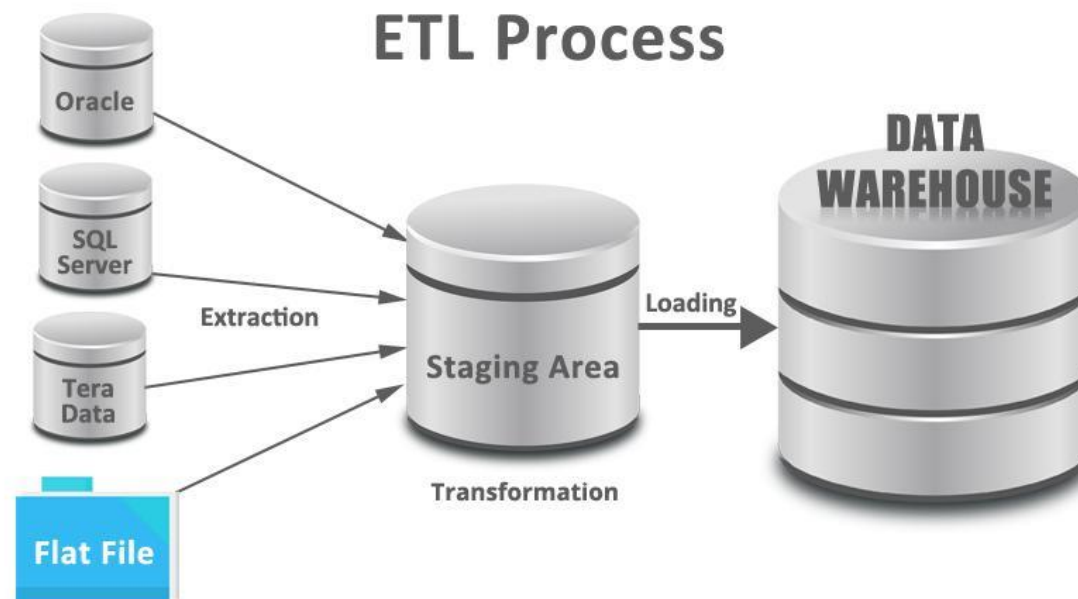
STAGING AREA

- Es un área temporal de almacenamiento de datos utilizada para el procesamiento de los mismos durante los procesos de ETL.
 - Se toman los datos necesarios para las cargas, y se aplica el mínimo de transformaciones a los mismos.
 - Una vez que los datos han sido traspasados, el DW se independiza de los sistemas fuentes hasta la siguiente carga.
 - Se suele añadir algún campo de fecha para que en el caso de falla evita empezar todo desde el principio.



STAGING AREA

- Se construyen y se implementan los procesos de extracción, limpieza, transporte, transformación y carga de los datos.
- Se utiliza una herramienta especializada en el tratamiento de grandes volúmenes de datos.





OTRAS ESTRATEGIAS DE INTEGRACIÓN DE DATOS

- ✓ **ELT (Extract, Load, Transform):** Carga de datos sin transformar en almacenamiento en la nube para su procesamiento posterior.
- ✓ **Data Mesh:** Arquitectura distribuida que descentraliza la gestión de datos y permite su acceso mediante dominios organizacionales.
- ✓ **Data Fabric:** Enfoque unificado que automatiza el movimiento y transformación de datos en entornos híbridos y multi-cloud.



ERRORES TÍPICOS EN LOS DATOS

- Datos Incompletos
 - Registros o campos faltantes
- Datos Incorrectos
 - Códigos erróneos
 - Cálculos o agregaciones incorrectas
 - Registros duplicados
- Datos Incomprensibles
 - Varios campos dentro de un campo
 - Formato extraño en los datos
 - Códigos desconocidos
 - Formatos de archivo extraños



ERRORES TÍPICOS EN LOS DATOS

○ Datos Inconsistentes

- Uso inconsistente de la codificación
- Inconsistencia en el significado de los códigos
- Códigos superpuestos
- Códigos distintos con igual significado
- Inconsistencia en nombres y direcciones
- Reglas de negocio inconsistentes
- Agregaciones inconsistentes
- Inconsistencia en la granularidad del nivel atómico de datos



HERRAMIENTAS ETL

- Adoptar una herramienta o hacerlo nosotros mismos.

Herramientas ETL	Codificación Manual
Específicas	Flexibilidad
Metodologías ETL	Técnicas de programación estándar
Personal capacitado en herramienta	Recursos propios
Mayor costo inicial	Menor costo inicial
Conectores múltiples	Se deben desarrollar conectores
Simplifica el mantenimiento	El mantenimiento se complejiza con el tiempo
Auto documentación (meta data)	Se debe desarrollar
Trazabilidad de los datos	Muy difícil de conseguir trazabilidad



CUADRANTE MÁGICO DE GARTNER

- Proveedores de herramientas de Integración de Datos (Nov 2024)

Figure 1: Magic Quadrant for Data Integration Tools



Gartner

Fuente: [Complimentary 2024 Gartner® Magic Quadrant™ for Data Integration Tools | Informatica](https://www.gartner.com/reviews/market/data-integration-tools)
<https://www.gartner.com/reviews/market/data-integration-tools>



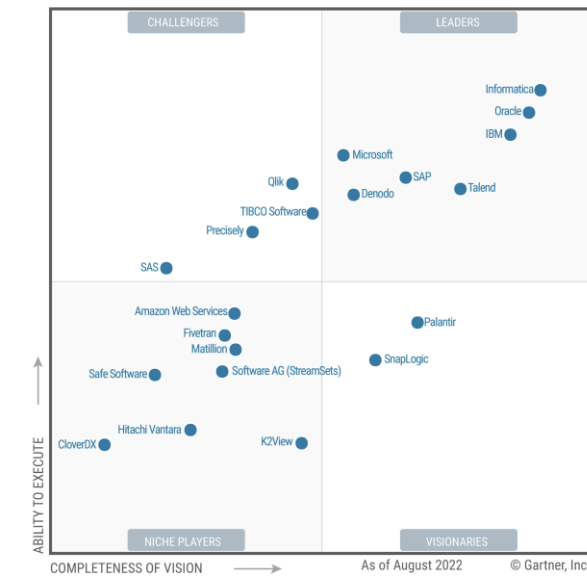
CUADRANTE MÁGICO DE BI

○ 2023-2022-2021

Figure 1: Magic Quadrant for Data Integration Tools



Figure 1: Magic Quadrant for Data Integration Tools



Source: Gartner (August 2022)

Magic Quadrant

Figure 1: Magic Quadrant for Data Integration Tools



Source: Gartner (August 2021)

CUADRANTE MÁGICO DATA AND ANALYTICS GOVERNANCE PLATFORMS



○ Nov 2024





CUADRANTE MÁGICO DE GARTNER

Si queremos conocer cuáles son las mejores herramientas de ETL, una referencia importante es el Cuadrante Mágico de Gartner sobre Herramientas de Integración de Datos, que cada año indica cuáles son los proveedores *leader* del mercado. A Noviembre de 2023 se seleccionaron :

- **Informática:** líder según Gartner. Su suite empresarial de integración de datos incluye la solución PowerCenter, una de las más populares.
- **IBM:** proporciona la suite de soluciones InfoSphere, en la cual destaca su herramienta DataStage.
- **Oracle:** proporciona la herramienta ELT Data Integrator, que permite gestionar procesos de integración de datos en sistemas de inteligencia de negocio.



CUADRANTE MÁGICO DE GARTNER

- **Denodo**: ofrece la virtualización como principal tecnología de integración de datos estructurados y no estructurados. Con Denodo Express apuestan a ofrecer el servicio de virtualización de manera gratuita.
- **Talend (Qlik)**: conocido por su software de integración. Adquirido por Qlik, suspenden el producto de código abierto gratuito Talend Open Studio, a partir del 31 de enero de 2024.
- **SAP**: ofrece la herramienta ETL Data Services como parte de SAP BO (Business Objects).
- **SAS**: proporciona una solución de integración de datos llamada Data Management.



CUADRANTE MÁGICO DE GARTNER

- **Microsoft:** Sus herramientas de integración de datos cuentan con funcionalidades relevantes para migración de datos, flujos de trabajo, administración de metadatos y procesos ETL, requiriendo de un mínimo trabajo de programación. Trabajan en la plataforma de análisis unificada: [Azure Synapse Analytics](#).
- Así el ETL (DWH) se encuentra en el mismo entorno que la administración del Data Lake, el almacenamiento de los datos y los proyectos de aprendizaje automático, optimizando la colaboración, simplificando el cumplimiento normativo y la seguridad, y acelerando la obtención de información.
- Por ende, se puede dedicar menos tiempo a la administración de los datos y más tiempo a detectar información en esos datos, actuando en consecuencia.

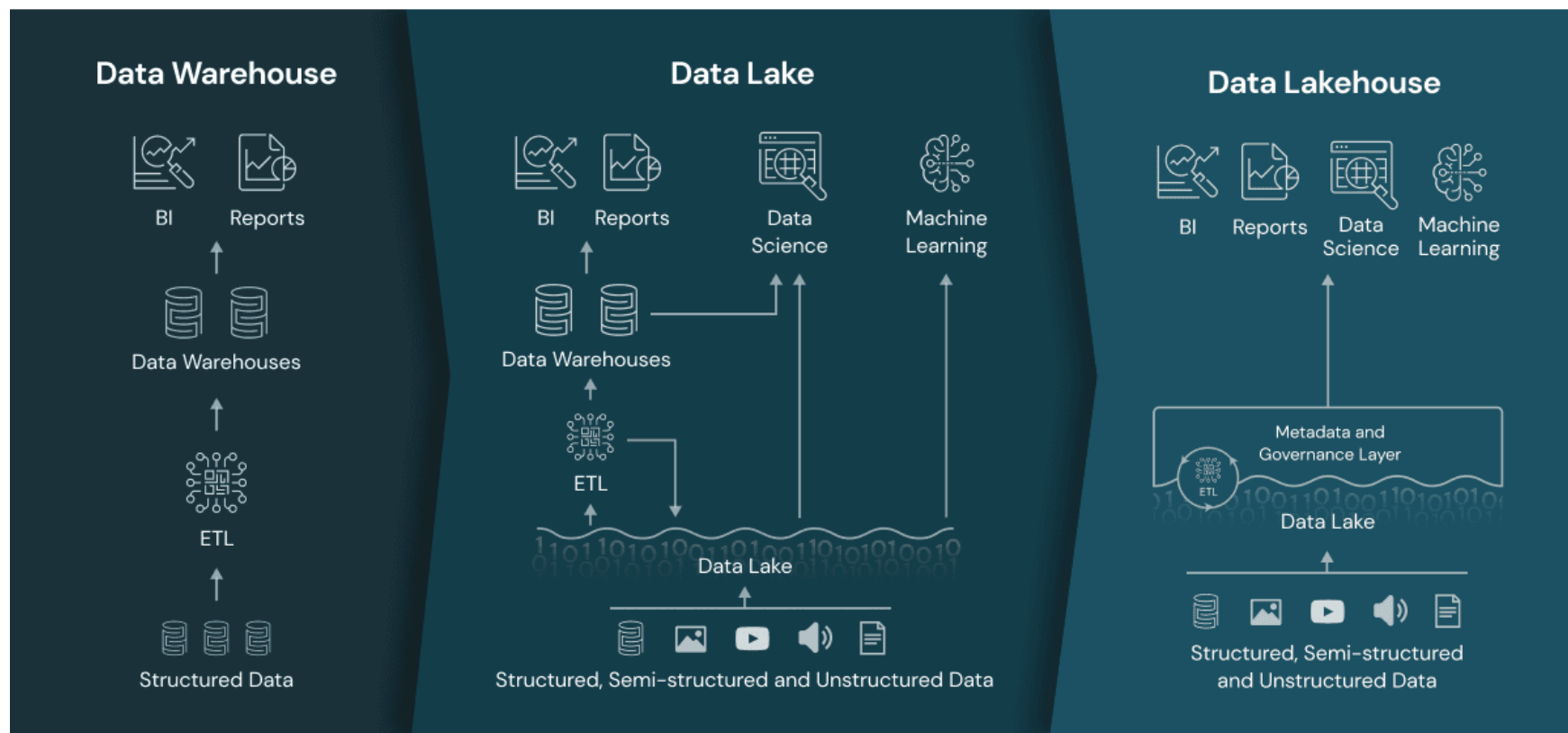


TENDENCIAS ARQUITECTURA BI

Un Data Warehouse es una base de datos estructurados, centralizada y optimizada para el análisis y la generación de informes. Esos datos se extraen, transforman y cargan desde sistemas transaccionales a través del proceso ETL.

Un Data Lake es un repositorio que permite almacenar grandes volúmenes de datos en su formato natural. Es una solución más flexible en comparación con los DWs tradicionales, especialmente para manejar datos no estructurados.

Un Lakehouse es un paradigma emergente que combina las capacidades de un Data Lake y un DW. Proporciona el almacenamiento y la flexibilidad de un Data Lake con las capacidades de gestión y procesamiento de consultas de un DW.



TENDENCIAS ARQUITECTURA BI

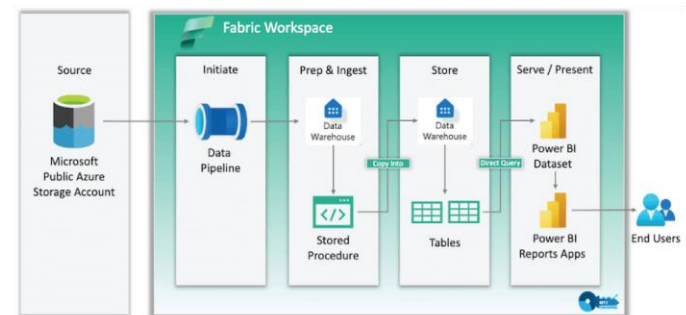
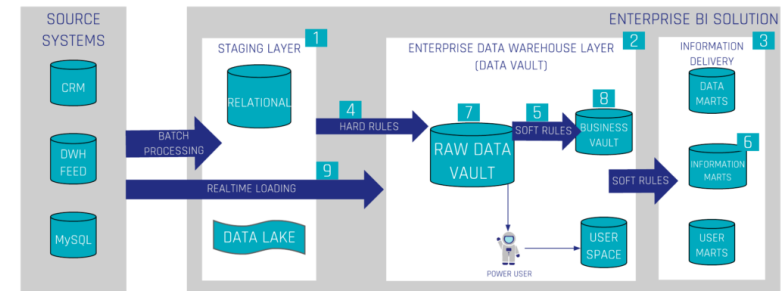
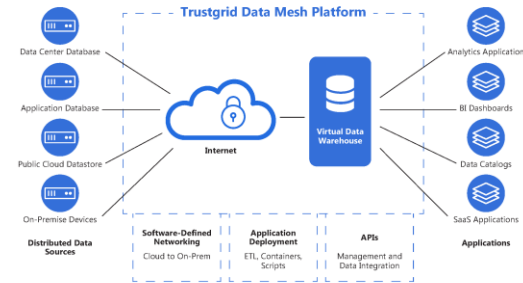
DATA MESH & DATA VAULT



• **Data Mesh** es una arquitectura y una forma de organizar los equipos alrededor de los datos.

• **Data Vault** es una técnica de modelado de datos para construir un data warehouse.

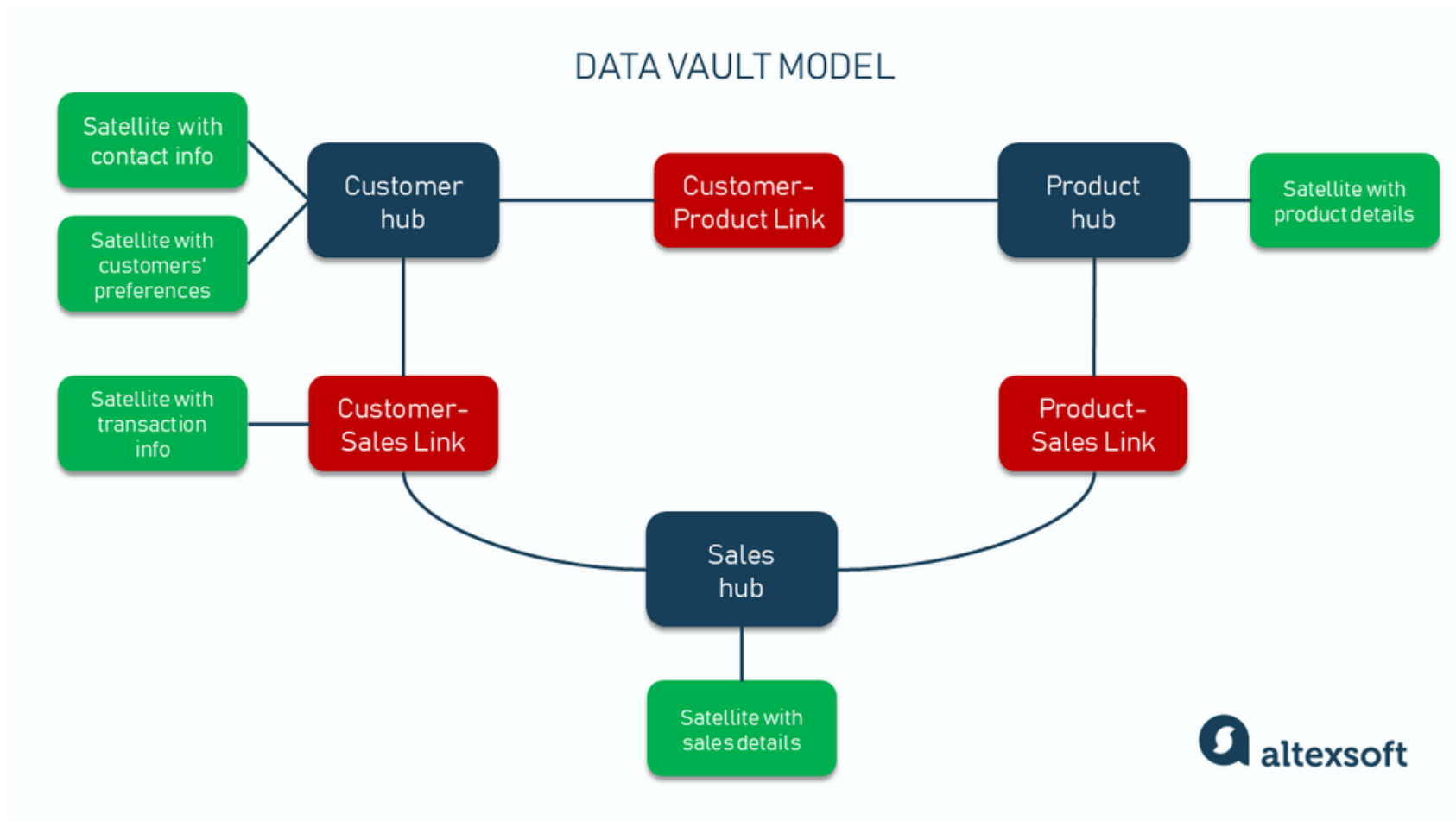
• **Data Fabric** Diversas soluciones heterogéneas que comprenden una plataforma de datos moderna (o stack de datos moderno) unidas por una capa de gestión virtual.



TENDENCIAS ARQUITECTURA BI -DATA VAULT



- **Data Vault:** Un modelo de almacenamiento de datos diseñado para proporcionar trazabilidad, auditoría e integración histórica, utilizando una arquitectura de hubs, satélites y enlaces.



TENDENCIAS ARQUITECTURA BI -DATA VAULT



Imagina que tienes un gran armario donde quieres guardar información importante de tu negocio, como clientes, productos y ventas. Quieres guardar no solo la información actual, sino también cómo ha cambiado con el tiempo y de dónde vino cada pieza de información.

Data Vault es como un sistema de organización muy inteligente para ese armario. En lugar de simplemente apilar todo, lo organiza en tres tipos principales de cajas:

- Cajas de "Quiénes" y "Qué" (Hubs):
- Cajas de "Detalles" (Satélites):
- Cajas de "Relaciones" (Links):

Piensa en Data Vault como un archivista muy meticuloso que organiza la historia de tus datos de una manera clara, rastreable y fácil de entender a largo plazo. No se enfoca en cómo analizar los datos directamente, sino en cómo guardarlos de forma inteligente para que el análisis sea más confiable y completo.

TENDENCIAS ARQUITECTURA BI -DATA VAULT



- **Modelo de Almacenamiento de Datos:** Diseñado para auditoría, trazabilidad e integración histórica de datos.
- **Arquitectura Orientada a Hubs, Satélites y Enlaces:**
 - **Hubs:** Contienen las claves de negocio únicas.
 - **Satélites:** Almacenan atributos descriptivos con fechas de inicio y fin para el historial.
 - **Enlaces (Links):** Modelan las relaciones entre las claves de negocio.
- **Énfasis en la Integración y la Historia:** Captura todos los cambios a lo largo del tiempo sin sobrescribir datos.
- **Diseño Robusto y Flexible:** Adaptable a diferentes fuentes y cambios en los sistemas origen.
- **Ideal para Data Warehousing:** Fundamento para construir un data warehouse empresarial escalable y auditable.
- **Soporte para Auditoría y Trazabilidad:** Permite rastrear el origen y la evolución de los datos.
- **Facilita la Carga Paralela:** Permite cargar datos de forma independiente.

TENDENCIAS ARQUITECTURA BI -DATA VAULT



- **Cajas de "Quiénes" y "Qué" (Hubs):** Estas cajas guardan la **identidad única** de las cosas importantes, como el número de identificación de cada cliente o el código de cada producto. Piensa en ellas como las etiquetas principales de cada cosa. No guardan detalles cambiantes, solo la identidad.
- **Cajas de "Detalles" (Satélites):** Estas cajas guardan toda la **información descriptiva** sobre las cosas que están en las cajas de "Quiénes" y "Qué". Por ejemplo, el nombre del cliente, su dirección o el precio del producto. **Lo importante es que cada vez que esta información cambia, se guarda una nueva versión en una nueva caja de "Detalles", indicando cuándo empezó a ser válida esa información.** Así, siempre tienes el historial de cómo han cambiado los detalles. Además, cada caja de "Detalles" sabe **de dónde vino** esa información (de qué sistema o fuente).
- **Cajas de "Relaciones" (Links):** Estas cajas guardan **cómo se conectan las cosas** entre sí. Por ejemplo, qué cliente compró qué producto, o qué empleado trabaja en qué departamento. Al igual que con los "Detalles", estas cajas también pueden tener información sobre **cuándo comenzó esa relación y de dónde se obtuvo la información de la relación.**

TENDENCIAS ARQUITECTURA BI -DATA VAULT



Data Vault ayuda a:

- Identificar claramente las cosas importantes (Hubs).
- Guardar todos los detalles sobre esas cosas, manteniendo un historial de cada cambio y sabiendo de dónde vino la información (Satélites).
- Registrar cómo se relacionan esas cosas entre sí (Links).

¿Para qué sirve todo esto?

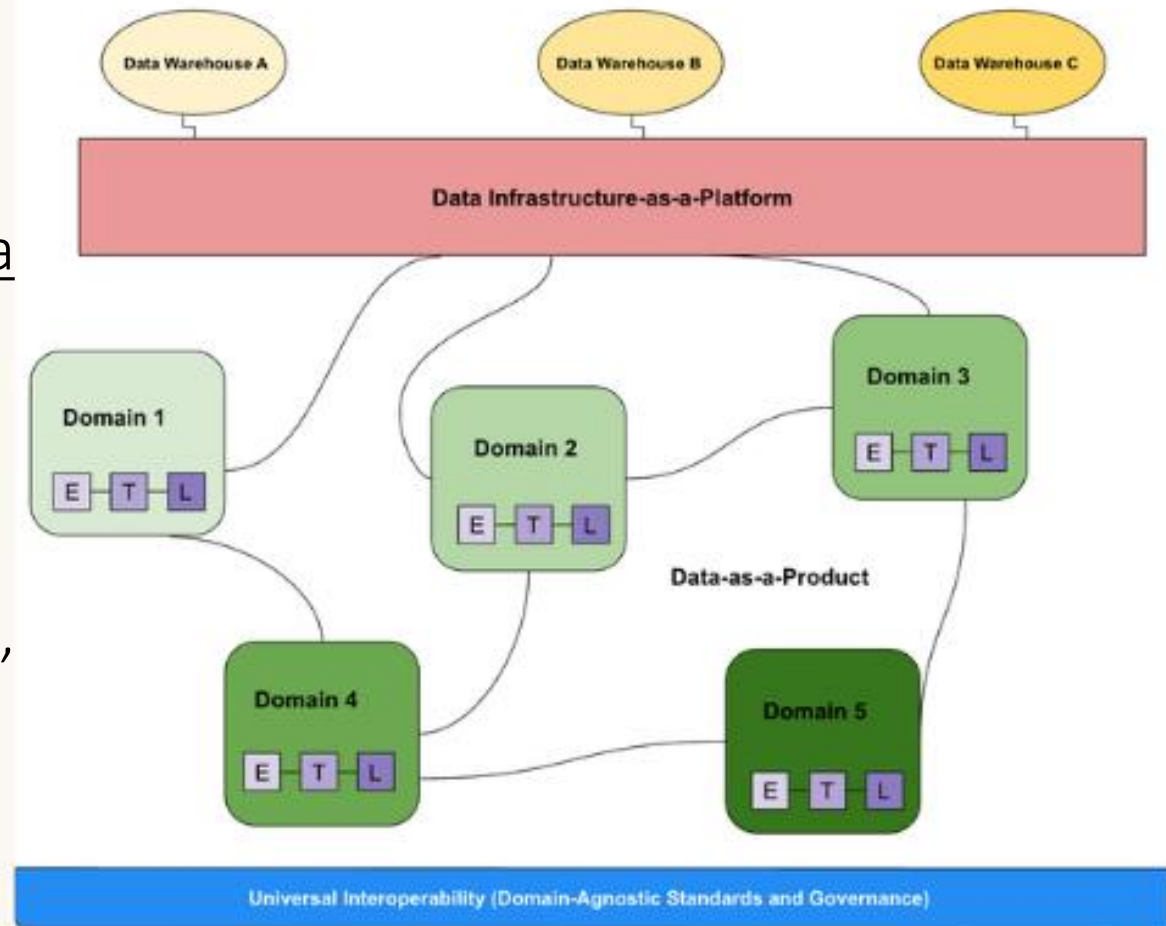
- Tener una visión completa y consistente de tus datos a lo largo del tiempo.
- Saber exactamente cuándo cambió algo y por qué.
- Rastrear de dónde vino cada pieza de información, lo cual es muy útil para la calidad de los datos y la auditoría.
- Construir una base de datos que puede crecer y adaptarse fácilmente a medida que cambian tus sistemas y necesidades.

TENDENCIAS ARQUITECTURA BI -DATA MESH



○ Data Mesh:

Un enfoque arquitectónico descentralizado para la gestión de datos, donde los equipos de dominio son responsables de sus datos como productos, facilitando el acceso y la colaboración en toda la organización



.(Zhamak Dehghani en 2019,)

TENDENCIAS ARQUITECTURA BI - DATA MESH

- **Descentralización de la propiedad de datos:** Los equipos de dominio son dueños de sus datos como productos. (Domain Oriented)

- **Datos como producto:**

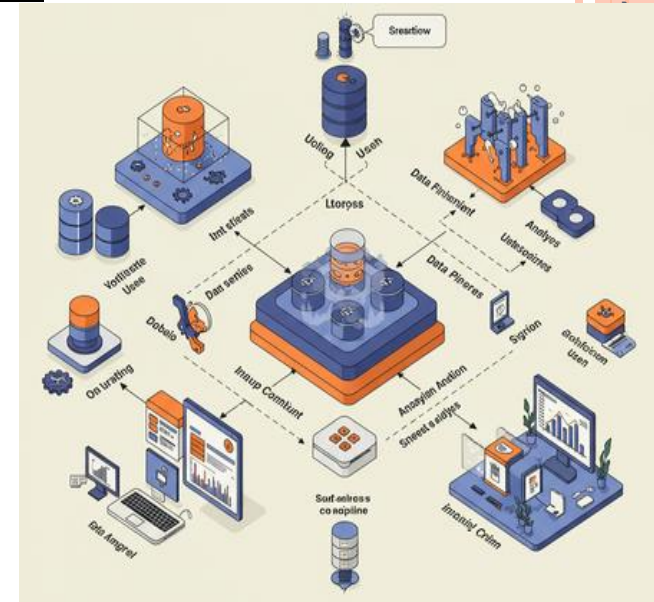
Los equipos de dominio proveen datos descubribles, accesibles, confiables y valiosos para otros. (Data Products)

- **Infraestructura de datos autoservicio:**

Plataforma tecnológica que facilita a los equipos de dominio gestionar y compartir sus datos de forma independiente. (Self Service Architecture)

- **Gobernanza federada:**

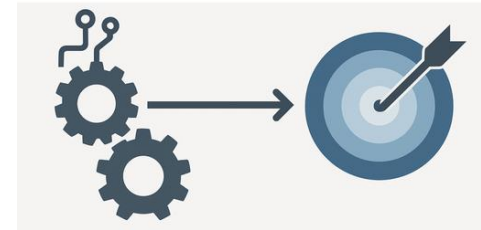
Políticas y estándares globales aplicados de forma descentralizada por los equipos de dominio. (Federated Governance)



TENDENCIAS ARQUITECTURA BI -DATA MESH

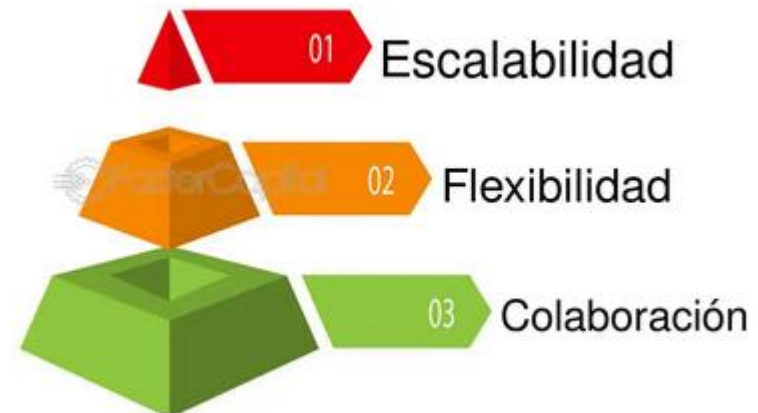


- **Foco en el valor del negocio:** Alineación de la gestión de datos con los objetivos estratégicos de la organización.



- **Agilidad y escalabilidad:** Permite una evolución más rápida y una mejor adaptación al crecimiento de los datos.

- **Rompe silos de datos:** Facilita el acceso y la colaboración entre diferentes áreas de la empresa.





On Premise vs Cloud vs Híbrido

<https://www.cloudcenterandalucia.es/blog/iaas-paas-y-saas-que-son-ejemplos-y-diferencias/>



On Premise vs Cloud vs Híbrido

On-Premise (Local):

- Se refiere a la infraestructura de TI (hardware, software, servidores, etc.) que se encuentra **físicamente dentro de las instalaciones de la empresa**.
- La organización es **dueña y responsable** de la adquisición, instalación, gestión, seguridad y mantenimiento de toda la infraestructura.
- Los datos se almacenan y procesan en servidores **locales**.

Cloud (Nube):

- Es la disponibilidad de recursos informáticos (servidores, almacenamiento, bases de datos, software, etc.) **bajo demanda a través de Internet**.
- Un **proveedor externo** es el propietario y el encargado de gestionar la infraestructura.
- Las empresas **pagan por los servicios** que utilizan, generalmente bajo un modelo de suscripción o pago por uso.
- Los datos y las aplicaciones se almacenan y se acceden a través de **servidores remotos** en los centros de datos del proveedor.



On Premise vs Cloud vs Híbrido

Híbrido:

- Es un **entorno informático que combina una infraestructura on-premise con servicios en la nube.**
- Permite a las organizaciones **ejecutar cargas de trabajo y almacenar datos en ambos entornos**, eligiendo la mejor opción para cada necesidad específica.
- Implica la **interconexión** entre la infraestructura local y la nube para permitir la portabilidad de datos y aplicaciones.
- Puede utilizarse para diversas estrategias, como **extender la capacidad** de la infraestructura local con recursos de la nube, **mantener datos sensibles** on-premise mientras se aprovechan servicios en la nube para otras cargas de trabajo, o para la **recuperación ante desastres**.

TENDENCIAS ARQUITECTURA BI

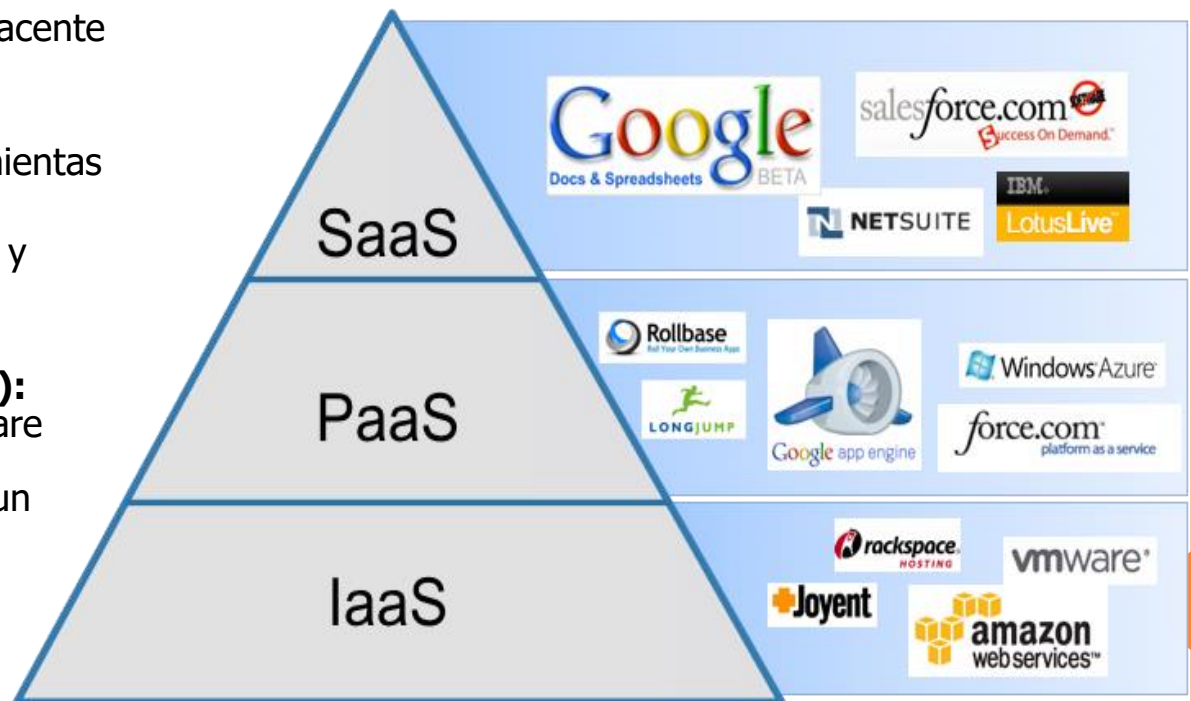


Modelos de Servicio de Cloud Computing

- **IaaS (Infraestructura como Servicio):**
- Proporciona recursos informáticos fundamentales como servidores virtuales, almacenamiento, redes y sistemas operativos a través de la nube.
- El proveedor gestiona la infraestructura subyacente.

- **PaaS (Plataforma como Servicio):**
- Ofrece una plataforma completa para desarrollar, ejecutar y gestionar aplicaciones sin la complejidad de administrar la infraestructura subyacente (hardware, sistemas operativos, almacenamiento, redes).
- Proporciona el entorno, las herramientas de desarrollo, las bibliotecas y los servicios necesarios para construir y desplegar aplicaciones.

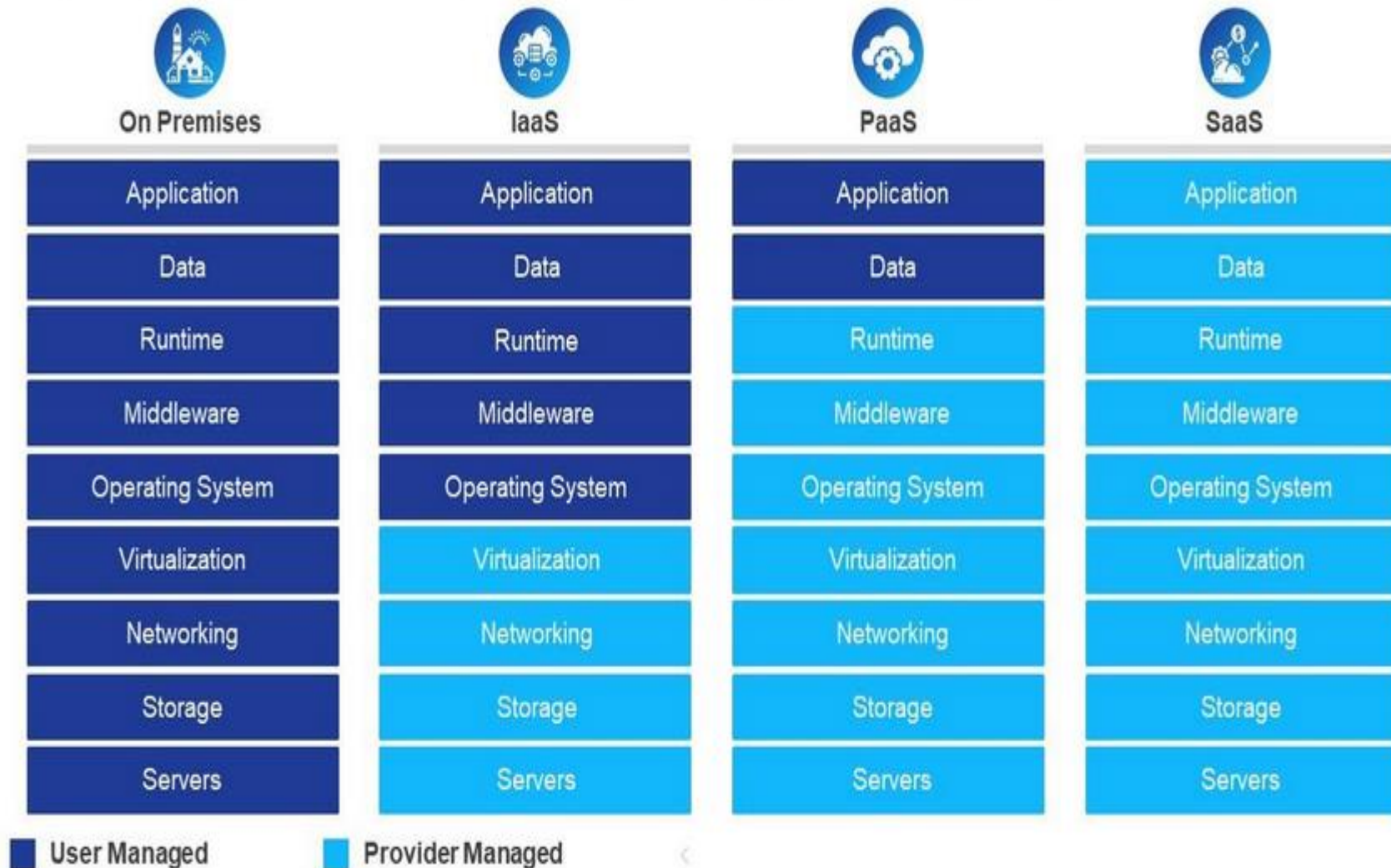
- **SaaS (Software como Servicio):**
- Proporciona aplicaciones de software completas accesibles a través de Internet, generalmente mediante un navegador web.
- El proveedor gestiona toda la infraestructura, el software de la aplicación y los datos.



TENDENCIAS ARQUITECTURA BI



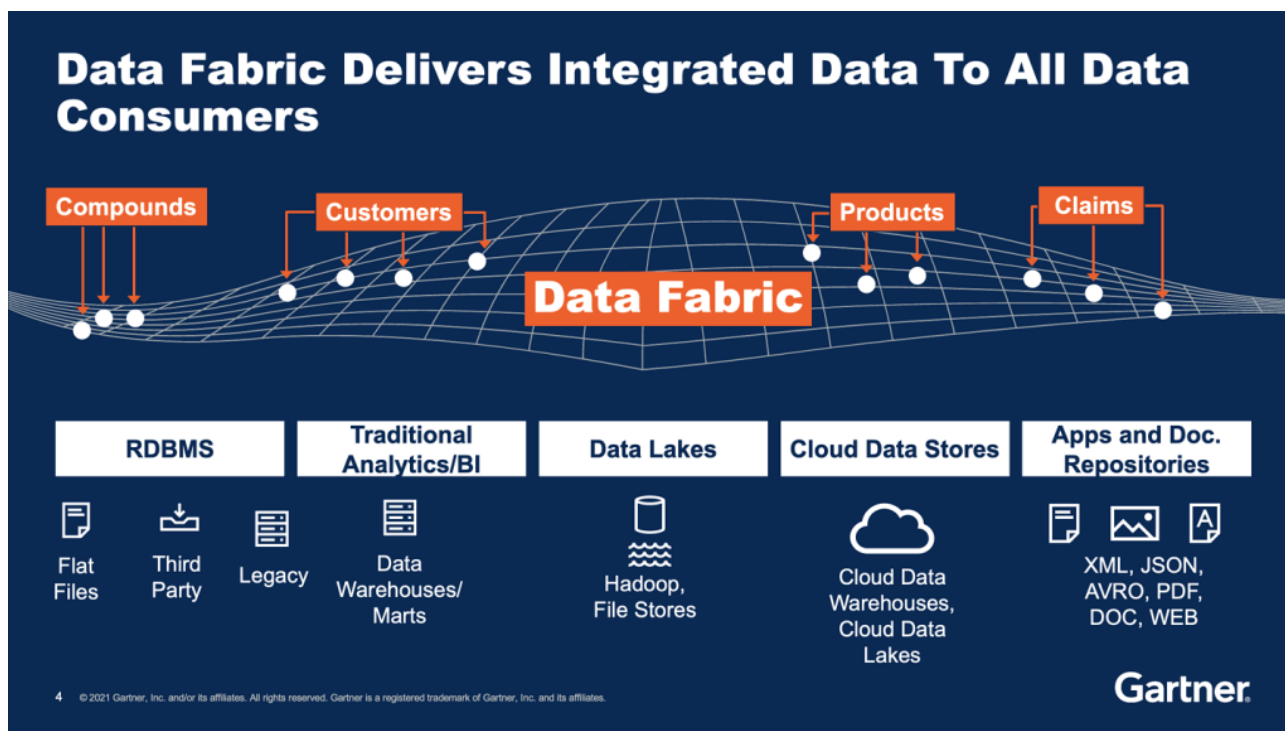
- **Comparación de TI de modelos de servicios en la nube**





TENDENCIAS ARQUITECTURA BI – DATA FABRIC

Un **Data Fabric** es una arquitectura que unifica la integración, acceso y gestión de datos de múltiples fuentes — locales, en la nube, estructurados o no — en una sola plataforma flexible, automatizada y escalable. Es decir, una capa de arquitectura que conecta los datos y los procesos analíticos



Fuente y Video: [Data Fabric: Clave para la gestión de datos • Fusiona](#)
[Data Fabric Explained - YouTube](#)



TENDENCIAS ARQUITECTURA BI – DATA FABRIC

Funciones Específicas del Data Fabric:

- **Unificación y acceso centralizado:** Integra datos dispersos (bases, lakes, APIs, etc.) bajo una sola capa lógica.
- **Integración flexible:** Admite flujos en lote o en tiempo real, con componentes reutilizables.
- **Automatización con IA:** Clasifica, prepara y cataloga datos automáticamente.
- **Gobernanza y seguridad de datos estandarizada:** Aplica políticas para privacidad, calidad y cumplimiento normativo.
- **Escalabilidad:** Se adapta a distintos entornos (on-premise, cloud, híbrido, multicloud).
- **Catálogo inteligente:** Facilita la búsqueda y comprensión de datos por parte de los usuarios.
- **Canalizaciones automáticas:** Optimiza la ingesta, transformación y orquestación de datos.



TENDENCIAS ARQUITECTURA BI – DATA FABRIC

Beneficios del Data Fabric:

- Mejora la eficiencia en gestión e integración de datos.
- Reduce costos y uso de herramientas dispersas.
- Acelera la toma de decisiones basadas en datos confiables.
- Potencia la innovación en productos y servicios.
- Democratiza el acceso a los datos (también para perfiles no técnicos).
- Prepara datos para proyectos de IA generativa y automatización avanzada.

Ejemplos de plataformas Data Fabric

- **IBM Cloud Data Fabric:** Solución para integración y gestión de datos en la nube y entornos locales.
- **SAP Data Fabric:** Plataforma para administración de datos en distintos entornos.
- **Microsoft Fabric (antes OneLake):** Herramienta integral para ingesta, procesamiento y análisis de datos.
- **Informatica Data Fabric:** Solución enfocada en gestión e integración de datos.
- **DataCamp Data Fabric:** Explicación y beneficios del concepto.



RESUMEN TENDENCIAS ARQUITECTURA BI

Propiedad	Data Warehouse	Data Mart	Data Lake	Data Lakehouse
Tipos de datos	Solo datos estructurados	Solo datos estructurados (segmentados por área)	Estructurados, semiestructurados y no estructurados (sin procesar)	Estructurados, semiestructurados y no estructurados (sin procesar)
Costo	\$\$\$ (elevado, requiere infraestructura dedicada)	\$\$ (moderado, depende del DW)	\$ (bajo costo de almacenamiento)	\$ (bajo costo con optimización para análisis)
Formato	Cerrado y propietario	Cerrado o abierto según el DW	Abierto (almacenamiento en bruto)	Abierto (estructura de DW con flexibilidad de un Data Lake)
Escalabilidad	Media-baja (costosa)	Media (limitada por el DW)	Muy alta	Alta
Usuarios	Analistas de datos	Analistas de negocio por área	Científicos de datos	Analistas, científicos de datos, ingenieros de ML
Fiabilidad	Alta calidad y gobernanza	Alta calidad (depende del DW)	Baja calidad (riesgo de data swamp: pantano de datos)	Alta calidad con estructura
Facilidad de uso	Estructura rígida pero fácil de usar para reportes	Alta, pero enfocada en áreas específicas	Difícil, requiere herramientas para organización	Simple, combina estructura de DW con flexibilidad de un Data Lake
Performance	Alto	Alto (para su dominio específico)	Bajo	Alto
Gobernanza	Alta (control en TI, cumplimiento de normativas)	Alta (pero depende del DW principal)	Baja (sin control centralizado)	Media (estructura de DW con gobernanza flexible)
Procesamiento	ETL (Transform antes de cargar)	ETL (diseñado para consultas rápidas)	Almacenamiento masivo sin transformación previa	Integrado con procesamiento en tiempo real
Casos de Uso	Reportes y análisis tradicionales	Reportes operativos y análisis específicos de negocio	Almacenamiento en bruto de grandes volúmenes de datos	Empresas que combinan BI con AI/ML
Ejemplo de Uso	Bancos y aseguradoras	Ventas, Finanzas, Recursos Humanos	Empresas de AI/ML	Retailers y empresas con Data Science



RESUMEN TENDENCIAS ARQUITECTURA BI

Propiedad	Data Vault	Data Mesh	Data Fabric
Tipos de datos	Estructurados con historial	Estructurados y semiestructurados	Todos los tipos de datos
Costo	\$\$ (moderado, depende de la transformación)	\$\$ (moderado, distribuido por dominio)	\$\$ (varía según la integración)
Formato	Híbrido (estructura flexible con hubs y satélites)	Abierto y distribuido	Virtualización de múltiples formatos
Escalabilidad	Alta	Muy alta	Muy alta (arquitectura dinámica)
Usuarios	Analistas de datos, auditores	Científicos de datos, analistas, equipos de negocio	Todos los usuarios con permisos de acceso
Fiabilidad	Alta calidad y trazabilidad	Variable según cada dominio	Alta calidad con automatización y monitoreo
Facilidad de uso	Moderada, requiere procesos ETL/ELT	Compleja, depende de la madurez de cada dominio	Alta, con acceso en tiempo real a múltiples fuentes
Performance	Medio-alto	Medio (depende de la infraestructura)	Alto (procesamiento en tiempo real)
Gobernanza	Media (historial de datos, auditable)	Baja (cada dominio gestiona sus datos)	Alta (automatización y gobierno de datos)
Procesamiento	ELT (Carga antes de transformar)	Enfoque distribuido	Integración en tiempo real con orígenes múltiples
Casos de Uso	Empresas con necesidades de auditoría y evolución rápida	Empresas con Big Data y múltiples fuentes	Análisis de datos en entornos híbridos
Ejemplo de Uso	Hospitales y empresas reguladas	Netflix, Airbnb, Amazon	Empresas con múltiples fuentes de datos interconectadas

Fuente: <https://todobi.com/diccionario-de-arquitecturas-de-datos/>

TENDENCIAS ARQUITECTURA BI - ON-PREMISE

Fuente:
<https://blog.softexpert.com/es/on-premise-vs-cloud/>



Ventajas

Control total de los datos: La organización tiene control directo sobre la ubicación, seguridad y acceso a sus datos.

Seguridad personalizada: Se pueden implementar medidas de seguridad específicas y adaptadas a las necesidades de la empresa

Personalización e integración: Mayor capacidad de personalización del software y su integración con sistemas existentes.

Rendimiento confiable: Al no depender de la conexión a internet, el rendimiento puede ser más predecible.

Potenciales ahorros a largo plazo: Si bien la inversión inicial es alta, a largo plazo podría ser más económico que las suscripciones en la nube.

Menor dependencia de proveedores externos: La continuidad operativa no depende de la estabilidad de un proveedor de la nube.

Desventajas

Altos costos iniciales: Requiere una inversión significativa en hardware, software, licencias e infraestructura.

Mantenimiento y gestión: La empresa es responsable del mantenimiento, actualizaciones, copias de seguridad y seguridad de la infraestructura.

Escalabilidad limitada: La expansión requiere inversión adicional en hardware y tiempo de configuración.

Accesibilidad limitada: El acceso a los datos y las aplicaciones puede estar limitado a la red interna o requerir soluciones de acceso remoto complejas.

Necesidad de personal técnico especializado: Se requiere personal cualificado para la gestión y el mantenimiento de la infraestructura.

Obsolescencia del hardware: El hardware requiere actualizaciones y reemplazos periódicos, generando costos adicionales.

Mayor consumo energético: Los servidores locales implican un mayor consumo de electricidad y costos asociados.

TENDENCIAS ARQUITECTURA BI - CLOUD

Fuente: <https://blog.softexpert.com/es/on-premise-vs-cloud/>



Ventajas	Desventajas
Escalabilidad y flexibilidad: Los recursos se pueden escalar rápidamente a necesidad.	Dependencia de la conexión a internet: Se requiere una conexión a internet estable para acceder a los servicios.
Menores costos iniciales: No se requiere una gran inversión en hardware e infraestructura, se opera bajo un modelo de pago por uso o suscripción.	Preocupaciones por la seguridad y privacidad de los datos: Aunque los proveedores implementan medidas de seguridad, existe la preocupación por el acceso no autorizado o las brechas de seguridad.
Accesibilidad remota: Se puede acceder a los datos y las aplicaciones desde cualquier lugar con conexión a internet.	Menor control sobre los datos y la infraestructura: La organización tiene un control limitado sobre la ubicación y la gestión física de sus datos.
Recuperación ante desastres: Los proveedores suelen ofrecer soluciones robustas de copia de seguridad y recuperación ante fallos.	Costos a largo plazo: Los costos de suscripción pueden acumularse y superar la inversión inicial de una solución on-premise.
Mantenimiento y actualizaciones automáticas: El proveedor de la nube se encarga del mantenimiento, las actualizaciones y la seguridad de la infraestructura.	Problemas de personalización e integración: La personalización de los servicios en la nube puede ser limitada, y la integración con sistemas locales puede ser compleja.
Colaboración mejorada: Facilita el acceso y la colaboración para equipos distribuidos.	Riesgo de dependencia del proveedor: Cambiar de proveedor de nube puede ser complicado y costoso.
Implementación rápida: Los servicios y las aplicaciones se pueden implementar rápidamente.	Posibles problemas de rendimiento: La latencia y el rendimiento pueden verse afectados por la conexión a internet y la infraestructura del proveedor.

TENDENCIAS ARQUITECTURA BI – ONPREMISE & CLOUD



14/5/2025

IN2025

Característica	On-Premise	Cloud
Inversión Inicial	Alta (hardware, software, infraestructura)	Baja (modelo de pago por uso o suscripción)
Costos a Largo Plazo	Potencialmente menor, pero con gastos puntuales	Potencialmente mayor (costos recurrentes)
Control de Datos	Total	Limitado
Seguridad	Personalizable, responsabilidad del cliente	Compartida, proveedor implementa medidas base
Cumplimiento	Mayor facilidad para regulaciones estrictas	Puede ser desafiante según la ubicación del proveedor
Mantenimiento	Responsabilidad total del cliente	A cargo del proveedor
Escalabilidad	Limitada por la inversión en hardware	Alta, elástica y rápida
Accesibilidad	Principalmente local, acceso remoto complejo	Remota desde cualquier lugar con internet
Personalización	Mayor flexibilidad	Limitada por las opciones del proveedor
Integración	Mayor control sobre la integración con sistemas	Puede ser compleja con sistemas locales
Rendimiento	Potencialmente más confiable (sin internet)	Dependiente de la conexión a internet y proveedor
Personal Técnico	Requerido	Menor necesidad interna
Obsolescencia	Riesgo para el hardware del cliente	A cargo del proveedor
Consumo Energético	Mayor	Generalmente menor para el cliente individual
Recuperación ante Desastres	Responsabilidad del cliente	Generalmente robusta por parte del proveedor
Dependencia	Menor de proveedores externos	Mayor de un proveedor específico

50

Fuente: <https://blog.softexpert.com/es/on-premise-vs-cloud/>



LINKS DE INTERÉS

Canal Business Intelligence LATAM

- [Business Intelligence LATAM – YouTube](#)
- [Curso Data Warehouse + Business Intelligence - YouTube](#)[Modern Data Warehouse - C1: Intro DW – YouTube](#)

ETL con Power BI

- [Modificar los datos de origen desde Power BI. – YouTube](#)
- [ETL Presentación | Creación de un Data Warehouse para BI | Webminar Viernes de BI – YouTube](#)
- [Global AI 2022 - Importancia del proceso ETL en Power BI \(youtube.com\)](#)

Casos de Estudio ETL

- [Análisis y desarrollo del proceso de ETL en la BD Northwind - Grupo Javeros \(youtube.com\)](#)

Otros Links de Interés

El Rincon del BI - Descubriendo el Business Intelligence...

- [Todo lo que necesitas saber sobre ETL \(bismart.com\)](#)
- [Diccionario de Arquitecturas de Datos](#)
- [Data Mesh and Data Vault working together – a case study](#)
- [Data Mesh and Data Vault - did you know that they can work well together?](#)



LINKS DE INTERÉS

Otros Links de Interés

- [What is Data Fabric? Uses, Definition & Trends | Gartner](#)
- [Data Fabric Explained](#)
- [<https://blog.softexpert.com/es/on-premise-vs-cloud/>](#)



TRABAJO PRÁCTICO DE APLICACIÓN (PARTE3)

ENTREGA PRÓXIMA CLASE Ó 1+ (*)

Parte 3 (Power BI Desktop)

- Construya el Modelo de Inteligencia de Negocios del caso propuesto. Se proveerán las fuentes de datos correspondientes.
- Construya diferentes Tableros, Reportes, Gráficos, Filtros que permitan dar respuesta a los interrogantes de información planteados por la empresa.
- Cree una página de Sugeridos con visualizaciones que mejor respondan a las nuevas preguntas detectadas en la Parte 1.

Links de Interés:

- [Obtener Power BI Desktop - Power BI | Microsoft Docs](#)
- [Documentación de Power BI - Power BI | Microsoft Docs](#)
- [Power BI en Microsoft Learn | Microsoft Docs](#)
- [Documentación de introducción a Power BI - Power BI | Microsoft Docs](#)
- [Blog de Power BI: actualizaciones y novedades | Microsoft Power BI](#)
- [Galleries - Microsoft Power BI Community](#)

() Cada trabajo práctico tiene una Fecha Límite de Entrega que deberá ser cumplida sin excepción. (Ver Condiciones de Cursada en MIEl – Sección: "Plazos y condiciones de Entrega Trabajos Prácticos y Casos de Estudio")*

RESUMEN TEMA 5

