Big Data Aplicado

Introducción

Contenido

Módulo profesional Big Data Aplicado

(Real Decreto 279/2021, de 20 de abril)

Resultados de aprendizaje y criterios de evaluación.

1. Solucionar problemas propuestos mediante Big Data
    a. Diseño y Construcción de soluciones Big Data
    b. Ingestión de datos
    c. Almacenamiento
    d. Procesado de datos
    e. Presentación de resultados
2. Gestionar sistemas de almacenamiento y plataformas de Big Data
    a. Volumen, Variedad, Velocidad
    b. Computación distribuida, tolerancia a fallos
3. Establecer mecanismos de integridad en sistemas distribuidos
4. Monitorizar la solución, asegurando fiabilidad y estabilidad
    a. Herramientas de monitorización, métricas, alertas
    b. Fiabilidad de los datos, estabilidad de servicios
5. Validar la solución de Big Data para facilitar la toma de decisiones
    a. Selección y transformación de datos
    b. Análisis, extracción de valor e interpretación de la información
    c. Modelo de inteligencia de negocios BI

Contenidos básicos

1. Gestión de soluciones con sistemas de almacenamiento y herramientas del centro de datos para la resolución de problemas
    a. Almacenamiento de datos masivo.
    b. Procesamiento de datos.
    c. Analítica de Big Data en los ecosistemas de almacenamiento.
    d. Big Data y Cloud.
2. Gestión de sistemas de almacenamiento y ecosistemas big data
    a. Computación distribuida. Computación paralela,
    b. Sistemas de almacenamiento distribuidos. Tolerancia a fallos.
    c. Herramientas:
        i. Map Reduce.
        ii. Pig, Hive, Flume.
        iii. Sqoop, Oozie.
        iv. Automatización de Jobs.
        v. Consultas Pig y Hive.

vi. Otras herramientas.

3. Generación de mecanismos de integridad de los datos. Comprobación de mantenimiento de sistemas de ficheros
    a. Calidad de los datos.
    b. Comprobación de la integridad de datos de los sistemas de ficheros distribuidos. Sumas de verificación.
    c. Movimiento de datos entre clusters. Actualización y migración. Metadatos.
4. Monitorización, optimización y solución de problemas
    a. Herramientas de monitorización: Interfaz web del Jobtracker y Namenode, entre otras.
    b. Análisis de los históricos.
    c. Monitorización del clúster: Ganglia, entre otros.
5. Validación de técnicas big data en la toma de decisiones en inteligencia de negocios BI
    a. Modelos de Inteligencia de negocios.
    b. Proceso del modelo KDD (Knowledge Discovery in Databases).
    c. Etapas: Selección, limpieza, transformación de datos, minería de datos, interpretación y evaluación de datos.
    d. Implantación de modelos de inteligencia de negocios BI.
    e. Técnicas de validación de modelos BI.

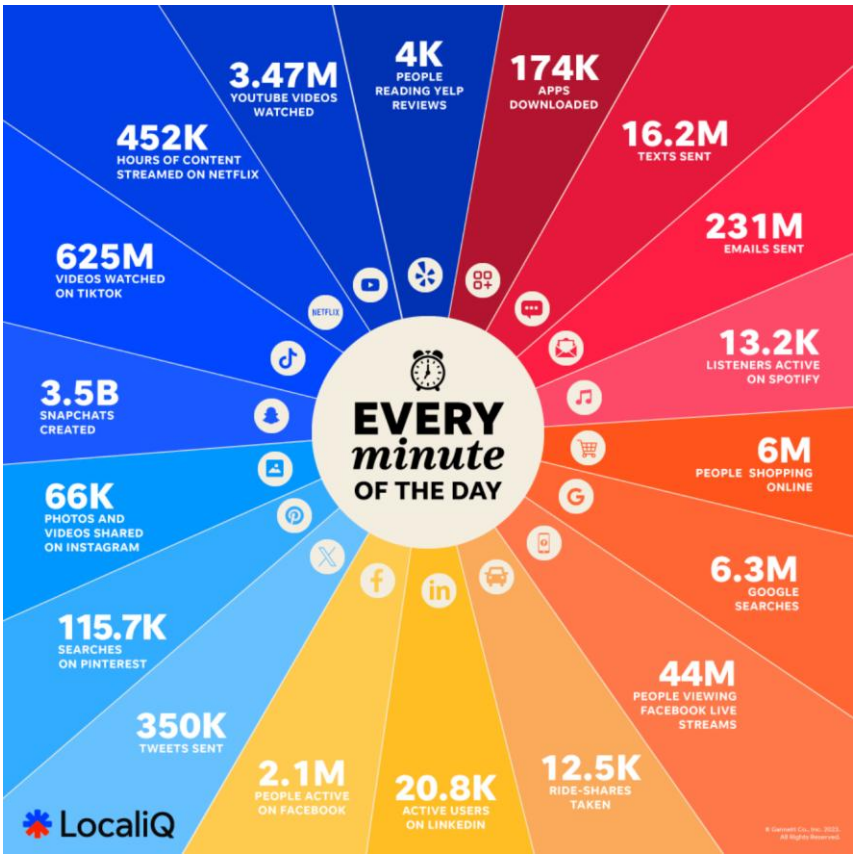
¿Qué es el Big Data?

El término “Big Data”, o “datos masivos”, se refiere a los métodos, herramientas y aplicaciones utilizados para recopilar, procesar y obtener información a partir de conjuntos de datos variados, de gran volumen y alta velocidad.

Estos conjuntos de datos pueden proceder de diversas fuentes, como la web, el móvil, el correo electrónico, las redes sociales y los dispositivos inteligentes conectados en red (IoT).

Son datos generados a gran velocidad y de formas variadas, desde estructurados (tablas de bases de datos, hojas de Excel) a semiestructurados (archivos XML, páginas web) o no estructurados (imágenes, archivos de audio).

Las formas tradicionales de software de análisis de datos no están equipadas para soportar este nivel de complejidad y escala, que es donde entran en juego los sistemas, herramientas y aplicaciones diseñados específicamente para el análisis de big data.



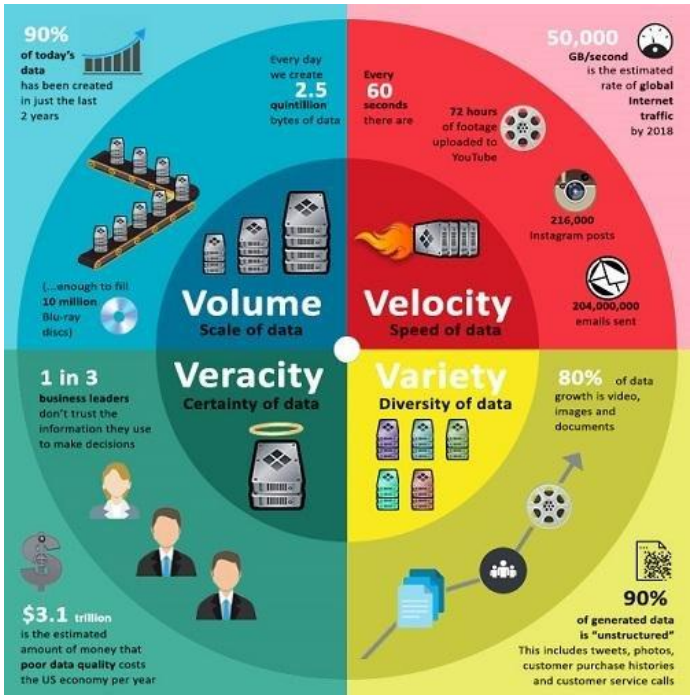
V's de Big Data

- **Volumen**: se refiere a la cantidad masiva de datos generados y almacenados. El volumen de datos ha crecido exponencialmente en las últimas décadas, superando la capacidad de las herramientas tradicionales de gestión de datos.

- **Velocidad**: se relaciona con la velocidad a la que se generan, recopilan y procesan los datos. La velocidad del flujo de datos en el Big Data requiere sistemas capaces de procesar información en tiempo real o casi real.

- **Variedad**: describe la diversidad de los datos, que pueden ser estructurados, semiestructurados o no estructurados. Esta variedad de formatos y estructuras dificulta la integración y el análisis de los datos.

- **Veracidad**: la calidad, la fiabilidad y la precisión de los datos es variable. La veracidad del Big Data es crucial para garantizar que las decisiones tomadas con base en estos datos sean acertadas.



¿Por qué es importante el análisis de big data?

Los datos están en todos los ámbitos de nuestras vidas. Con la digitalización de los procesos en todo tipo de empresas, el uso de móviles y redes sociales, y las tecnologías inteligentes asociadas al Internet de las cosas (IoT), ahora transmitimos más datos que nunca, y a gran velocidad.

El análisis de big data es importante porque permite obtener conocimientos valiosos a partir de grandes volúmenes de datos complejos, identificar tendencias, patrones y correlaciones ocultas, y mejorar la toma de decisiones.

Algunos ejemplos específicos de la importancia del análisis de big data:

- **Mejora de la experiencia del cliente**: al analizar los datos de las redes sociales, las interacciones con el sitio web y otras fuentes, las empresas pueden obtener una comprensión más profunda de las necesidades y preferencias de sus clientes, lo que les permite personalizar sus productos y servicios para una mejor experiencia.
- **Prevención del fraude**: al analizar patrones en tiempo real, las instituciones financieras y otras organizaciones pueden detectar y prevenir actividades fraudulentas de manera más efectiva.
- **Optimización de la cadena de suministro**: al analizar datos de sensores, GPS y otras fuentes, las empresas pueden optimizar sus cadenas de suministro para mejorar la eficiencia, reducir costos y garantizar la entrega oportuna de productos.
- **Avances en la atención médica**: al analizar datos de pacientes, registros médicos e investigaciones, los profesionales de la salud pueden mejorar los diagnósticos, personalizar los tratamientos y desarrollar nuevos medicamentos y terapias.
- **Desarrollo de productos**: análisis de necesidades de clientes a través de grandes volúmenes de datos de análisis de negocio, dirigiendo el desarrollo de características y la estrategia de la hoja de ruta.
- **Fijación de precios**: los datos de ventas y transacciones pueden analizarse para crear modelos de precios optimizados, lo que ayuda a las empresas a tomar decisiones de precios que maximicen los ingresos.
- **Operaciones**: análisis de datos financieros ayuda a las organizaciones a detectar y reducir los costes operativos ocultos, lo que a su vez permite ahorrar dinero y aumentar la productividad.
- **Adquisición y retención de clientes**: historial de pedidos, los datos de búsqueda, las reseñas en línea y otras fuentes de datos para predecir el comportamiento de los clientes, que pueden utilizar para mejorar la retención.

¿Cómo funciona la analítica de big data?

Las soluciones analíticas obtienen información y predicen resultados mediante el análisis de conjuntos de datos. Para que los datos se puedan analizar, primero deben almacenarse, organizarse y limpiarse mediante una serie de aplicaciones en un proceso de preparación integrado paso a paso:

- **Recopilar**: los datos se recopilan de múltiples fuentes a través de la web, el móvil y la nube, y se almacenan en un repositorio -un lago de datos o un almacén de datos.
- **Procesamiento**: los datos almacenados se verifican, clasifican y filtran, lo que los prepara para su uso posterior y mejora el rendimiento de las consultas.
- **Depuración**: los datos se corrigen y se limpian los conflictos, las redundancias, los campos no válidos o incompletos y los errores de formato.

- **Análisis**: los datos se analizan mediante herramientas y tecnologías como la minería de datos, la IA, el análisis predictivo, el aprendizaje automático y el análisis estadístico, que ayudan a definir y predecir patrones y comportamientos en los datos.

Evolución histórica

Tipos de Datos

1. Datos Estructurados: tienen una longitud y un formato predefinidos, lo que facilita su almacenamiento y análisis mediante bases de datos relacionales tradicionales. Ejemplos: números, fechas, cadenas de texto con formato específico (nombres, direcciones, etc.)
2. Datos No Estructurados: no se ajustan a un formato predefinido y pueden variar ampliamente en su estructura y contenido. Ejemplos: texto libre en documentos, correos electrónicos y redes sociales, imágenes, audio, vídeo, etc.
3. Datos semiestructurados: no presentan un esquema rígido, pero contienen etiquetas o marcadores que proporcionan organización. Ejemplos: archivos JSON o XML.

Sistemas de Almacenamiento

La evolución de los dispositivos y sistemas de almacenamiento viene marcada por la creciente digitalización de los procesos y la necesidad de gestionar cantidades cada vez mayores de información. Se busca un equilibrio óptimo entre capacidad, rendimiento, fiabilidad, seguridad y coste.

Dispositivos más usados

- **Discos**: datos magnéticos en discos giratorios. Bajo coste, rendimiento limitado.
- Unidades de estado sólido (**SSD**): chips de memoria flash. Acceso más rápido y menor consumo.
- **Cintas** magnéticas: copias de seguridad de grandes volúmenes de datos a bajo coste.

Métodos Avanzados de Almacenamiento

Inicialmente su ventaja clave era la habilidad de combinar varios dispositivos de bajo coste y tecnología más vieja en un conjunto que ofrecía mayor capacidad, integridad, tolerancia a fallos y/o tasa de transferencia.

- **RAID 0** (Striping): distribuye datos en varios discos para aumentar la velocidad, no ofrece redundancia.
- **RAID 1** (Mirroring): duplica datos en dos discos para garantizar la redundancia.
- **RAID 5**: redundancia y mejora del rendimiento al distribuir datos y paridad en tres o más discos.

Sistemas de Almacenamiento en Red

- **NAS** (Network Attached Storage): Dispositivos conectados a la red que permiten a múltiples usuarios acceder y compartir archivos.
- **SAN** (Storage Area Network): Red de almacenamiento dedicada de alto rendimiento que conecta servidores a dispositivos de almacenamiento.
    - Componentes principales: servidores conectados por fibra o iSCSI, conmutadores de red, cabinas de discos, librerías de cintas, software de gestión especializado.
    - Beneficios: alto rendimiento, alta disponibilidad, escalabilidad, mejora de la eficiencia, gestión centralizada.

Arquitectura de datos

La evolución de las arquitecturas de datos se puede trazar a través de distintos patrones que han surgido para satisfacer necesidades concretas de la gestión y análisis de datos a lo largo del tiempo.

Generación 0 (1970): Sistemas Transaccionales

En la década de 1960, con la aparición de los primeros sistemas de procesamiento de transacciones en línea (OLTP), los datos se gestionaban sin una arquitectura definida. Estos sistemas se centraban en registrar las transacciones del negocio, como pedidos, pagos, inventarios, etc., y se basaban en el modelo relacional y el lenguaje SQL.

- Las bases de datos transaccionales debían mantener la integridad y la consistencia de los datos, cumpliendo con las propiedades **ACID** (Atomicidad, Consistencia, Aislamiento y Durabilidad).
- Si bien estos sistemas eran muy eficientes para las operaciones diarias, presentaban limitaciones para el análisis de datos, ya que no estaban optimizados para consultas complejas.
- Además, cada aplicación era responsable de sus propios datos, lo que dificultaba la integración y la obtención de una visión global del negocio.
- El acceso a los datos era también complejo y lento, ya que implicaba sentencias SQL que podían afectar al rendimiento de las operaciones transaccionales.

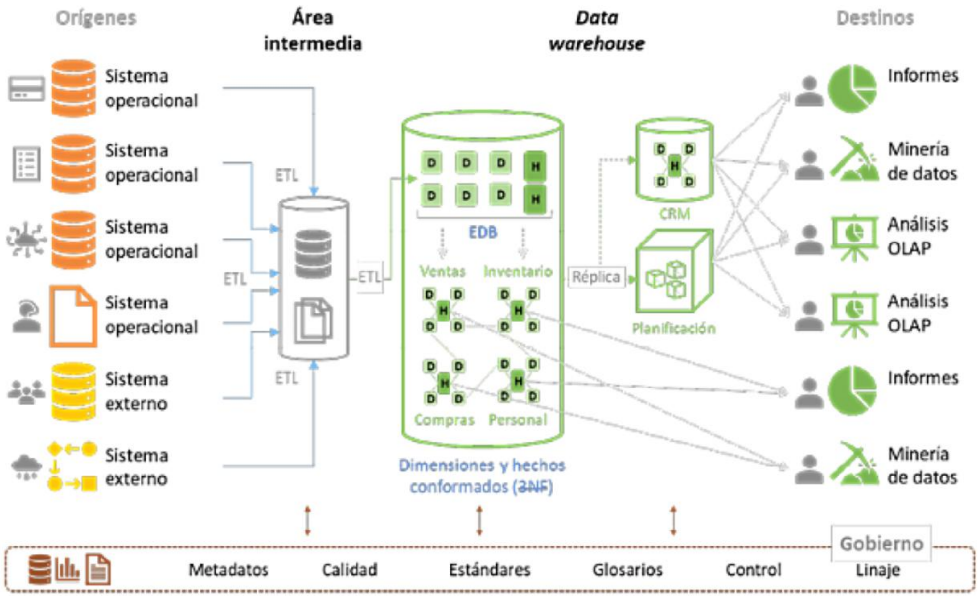
Generación 1 (1980): Data Warehouse

Para superar las limitaciones de los sistemas transaccionales para el análisis, en la década de 1980 surge el concepto de **data warehouse**, un repositorio central de datos optimizado para la consulta y el análisis.

- El data warehouse se alimenta de los datos de los sistemas transaccionales y otras fuentes, integrándolos y transformándolos en un formato adecuado para el análisis.
- Esta separación de los entornos transaccionales y analíticos supuso un gran avance, permitiendo a los usuarios de negocio acceder a la información histórica sin afectar al rendimiento de los sistemas operacionales.
- Surgieron dos modelos principales de data warehouse: **Inmon** (top-down) y **Kimball** (bottom-up). Inmon se basa en un modelo de datos corporativo, mientras

que Kimball se centra en los procesos de negocio y las necesidades de los usuarios.

- La creciente necesidad de herramientas y procesos para la creación de informes y la inteligencia empresarial (BI) impulsó la aparición de roles como ingeniero de BI, desarrollador ETL e ingeniero de almacenamiento de datos.



Data Warehouse – Modelo de Kimball

Generación 2 (1990): Almacenes de Datos Operacionales

Con la necesidad de acceder a información más actualizada, surge el concepto de **Operational Data Store (ODS)**, un almacén de datos que integra datos de distintos sistemas operacionales en tiempo real.

- El ODS permite una **visión integrada y actualizada** de la información, útil para la toma de decisiones operativas y la gestión de procesos en tiempo real.
- Sin embargo, la frecuencia de actualización del ODS implicaba la imposibilidad de implementar grandes transformaciones como en el data warehouse.
- La llegada de Internet a mediados de la década de 1990 y el auge de las empresas basadas en la web (Amazon, Yahoo, AOL) plantearon nuevos desafíos de escalabilidad y gestión de datos.

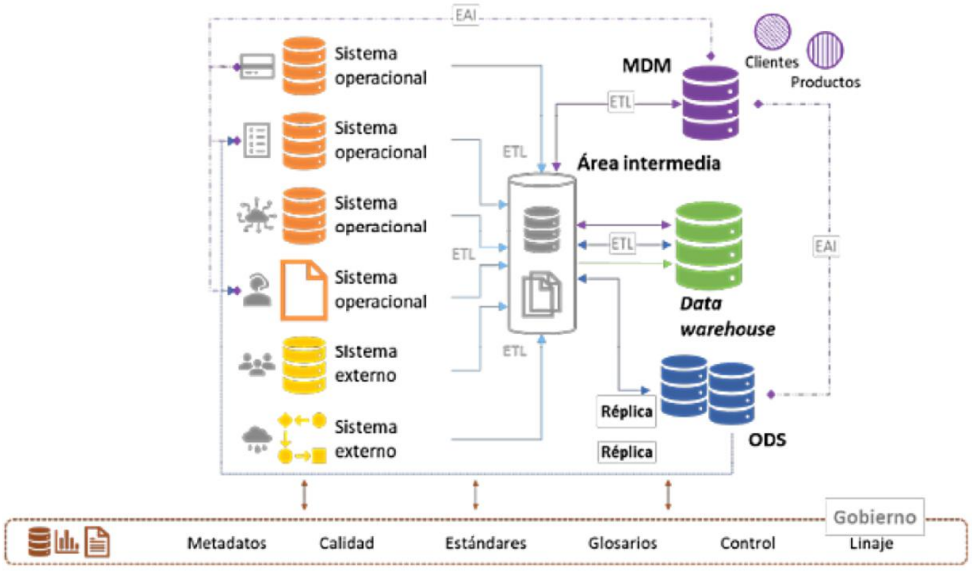
Generación 3 (2000): Gestión de Datos Maestros y Comienzo del Big Data

La proliferación de sistemas operacionales con información duplicada y a veces inconsistente llevó a la aparición de la gestión de **datos maestros (MDM)**.

- El MDM se centra en **unificar y centralizar** la gestión de los datos maestros de la empresa, como clientes, productos, proveedores, etc., asegurando su consistencia y calidad.
- Esta información maestra se utiliza para alimentar los demás sistemas de la arquitectura, garantizando la coherencia de la información en toda la organización.

El colapso de la burbuja puntocom a principios de la década de 2000 dejó un pequeño grupo de empresas supervivientes (Google, Amazon, Yahoo) que se convertirían en gigantes tecnológicos.

- Estas empresas se enfrentaron a una explosión de datos generados por sus plataformas web, lo que llevó a la necesidad de sistemas de datos más rentables, escalables, disponibles y fiables.
- La convergencia de la disminución del costo del hardware, la computación distribuida y el almacenamiento en clústeres masivos marcó el comienzo de la era del Big Data.
- Google publicó documentos sobre el Sistema de Archivos de Google (GFS) en 2003 y MapReduce en 2004, sentando las bases para el procesamiento de Big Data a gran escala.
- Inspirados en los trabajos de Google, los ingenieros de Yahoo desarrollaron y lanzaron Apache Hadoop en 2006, un framework de código abierto para el procesamiento distribuido de Big Data.
- Amazon creó una serie de servicios de computación en la nube, incluyendo Amazon EC2, Amazon S3 y Amazon DynamoDB, que ofrecían escalabilidad y flexibilidad sin precedentes.



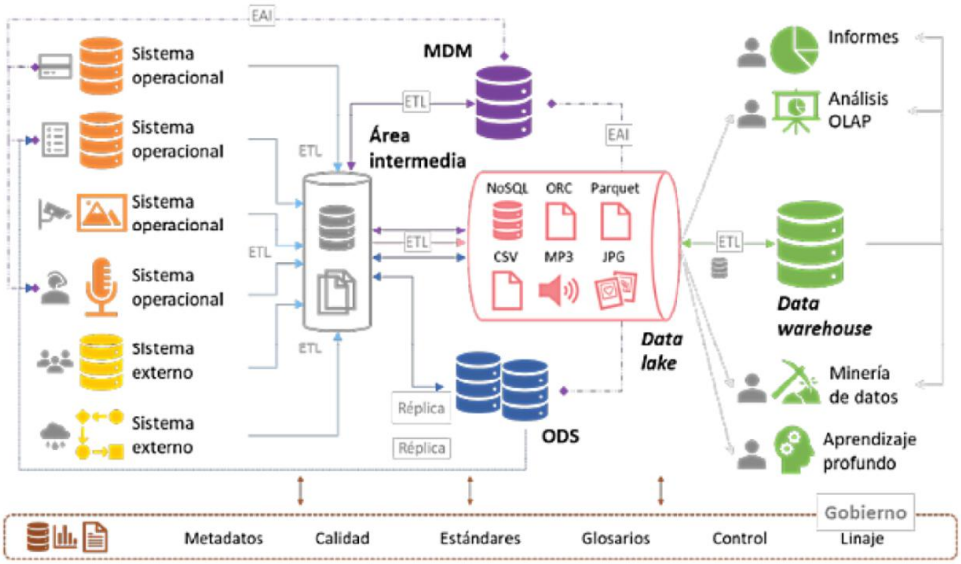
Gestión de Datos Maestros (MDM)

Generación 4 (2010): La era del Big Data. Data Lake

La explosión de datos no estructurados provenientes de redes sociales, dispositivos móviles y sensores, entre otros, junto con la necesidad de un almacenamiento más flexible y económico, dio lugar a la aparición del **data lake**.

- A diferencia del data warehouse, que se basa en un esquema de datos definido de antemano, el data lake permite almacenar datos en su formato original, sin ningún tipo de procesamiento ni esquema predefinido.

- Este enfoque "schema-on-read" ofrece flexibilidad y agilidad a la hora de analizar los datos, pero también implica una mayor complejidad en la gestión y el gobierno de los datos.
- El data lake también habilita el análisis de datos hasta el momento complicados, como por ejemplo audio e imágenes.
- Las herramientas de código abierto de Big Data en el ecosistema Hadoop se popularizaron rápidamente, brindando a las empresas de todos los tamaños acceso a tecnologías de vanguardia.
- La transición de la computación por lotes al procesamiento de flujo de eventos permitió el análisis de datos en tiempo real.
- Surgieron nuevas tecnologías como Apache Pig, Apache Hive, Apache HBase, Apache Cassandra y Apache Spark, lo que llevó a la especialización de los ingenieros de datos en el campo del Big Data.
- La creciente complejidad de las herramientas y la infraestructura de Big Data llevó a la búsqueda de soluciones más abstractas, simplificadas y gestionadas.



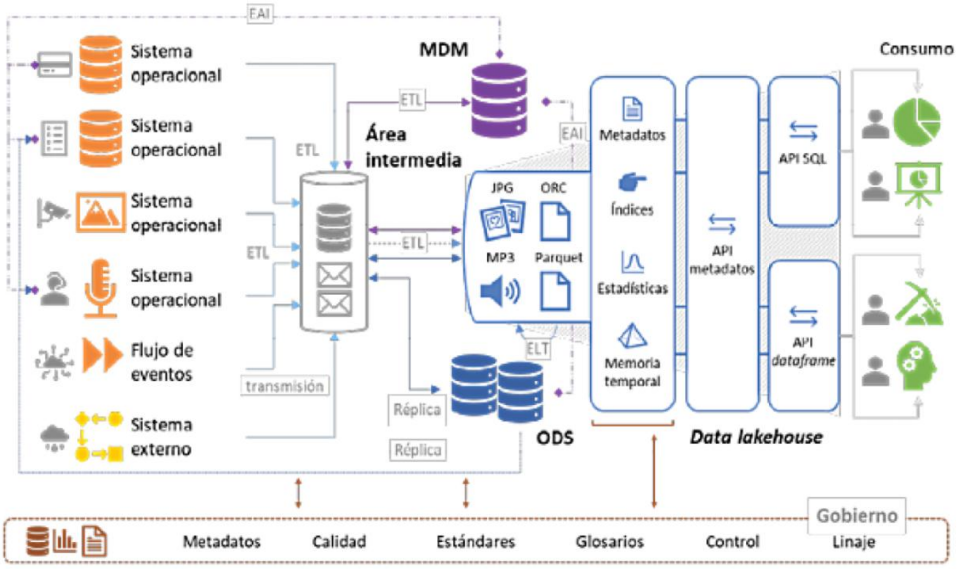
Data Lake

Generación 5 (2020): Data Lakehouse

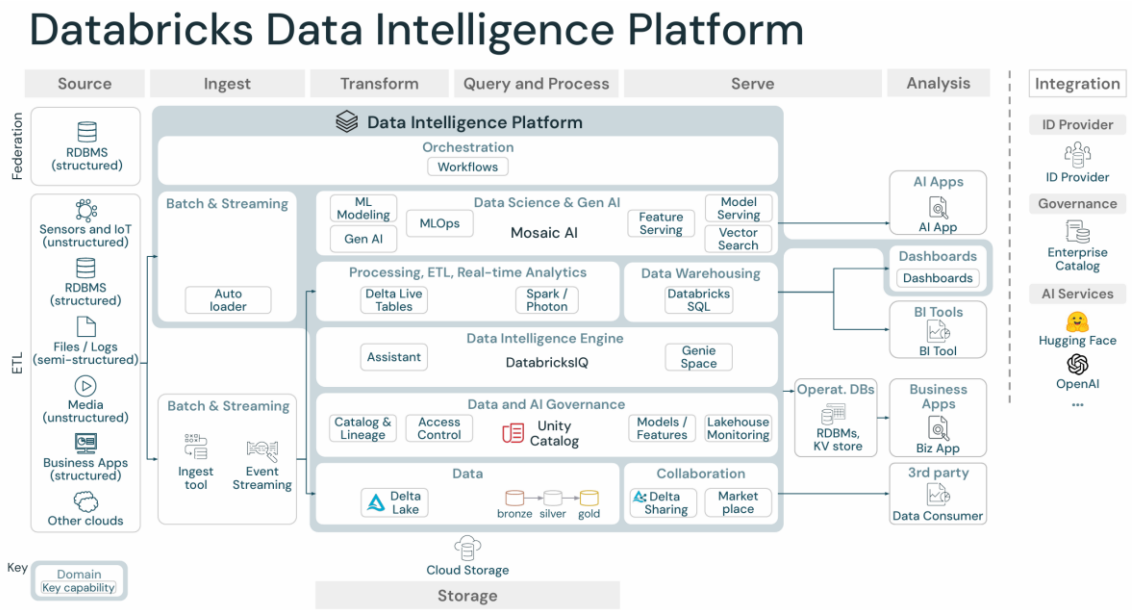
El **data lakehouse** es un nuevo paradigma que busca combinar las ventajas del data warehouse y el data lake en una única plataforma unificada.

- El objetivo es eliminar las duplicidades que se dan al tener que mantener ambos sistemas en paralelo.
- El data lakehouse se basa en estándares abiertos para formatos de archivo, lo que facilita la interoperabilidad entre los distintos componentes de la arquitectura.
- Además, incorpora una capa de metadatos que permite una gestión y gobierno de los datos más eficiente.
- La ingeniería de datos está evolucionando hacia una disciplina centrada en la **gestión del ciclo de vida completo de los datos**, desde la generación hasta el consumo.

- El enfoque se ha desplazado de las tareas de bajo nivel a aspectos de mayor valor añadido como la seguridad, la privacidad de los datos, la gobernanza de datos, DataOps, la arquitectura de datos y la orquestación.



Data Lakehouse



La evolución de las arquitecturas de datos ha sido impulsada por la necesidad de gestionar y analizar volúmenes de datos cada vez mayores y más complejos. Desde los sistemas transaccionales hasta el data lakehouse, cada nuevo paradigma ha buscado superar las limitaciones del anterior, ofreciendo una mayor flexibilidad, escalabilidad y agilidad. La tendencia actual se dirige hacia arquitecturas descentralizadas, como el data mesh, que permiten una gestión de los datos más cercana a los usuarios de negocio y a las necesidades del negocio.

Evolución de los Roles en el Área de Datos

Años 90 y principios de los 2000

- Roles Principales:
  - **Data Analyst** (Analista de Datos): Se encargaba de analizar datos históricos para generar informes y apoyar la toma de decisiones.
  - **Database Administrator** (Administrador de Bases de Datos): Responsable de la gestión y mantenimiento de bases de datos.
- Tecnologías:
  - SQL: Lenguaje de consulta estructurado para gestionar bases de datos.
  - Excel: Herramienta principal para el análisis de datos.
- Tareas:
  - Extracción y limpieza de datos.
  - Generación de informes y gráficos.

Mediados de los 2000

- Roles Principales:
  - **Business Intelligence (BI) Developer**: Desarrolla soluciones de inteligencia empresarial para transformar datos en información útil.
  - **Data Warehouse Specialist**: Especialista en la creación y mantenimiento de almacenes de datos.
- Tecnologías:
  - ETL Tools (Extract, Transform, Load): Informatica, Talend.
  - BI Tools: Microstrategy, Cognos, Business Objects.
- Tareas:
  - Integración de datos de múltiples fuentes.
  - Creación de dashboards y reportes interactivos.

2010s

- Roles Principales:
  - **Data Scientist** (Científico de Datos): Analiza grandes volúmenes de datos para descubrir patrones y tendencias.
  - **Data Engineer** (Ingeniero de Datos): Diseña y mantiene la infraestructura de datos.
- Tecnologías:
  - Big Data Technologies: Hadoop, Spark.
  - Machine Learning Libraries: Scikit-learn, TensorFlow.
- Tareas:
  - Modelado predictivo y análisis avanzado.
  - Construcción de pipelines de datos.

2020s

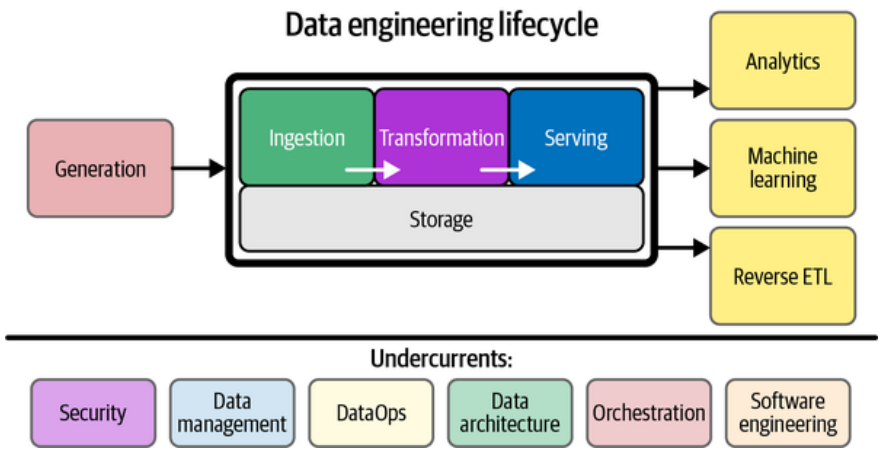
- Roles Principales:
  - **Chief Data Officer** (CDO): Lidera la estrategia de datos de la empresa.
  - **Machine Learning Engineer** (Ingeniero de Aprendizaje Automático): Desarrolla y optimiza modelos de aprendizaje automático.

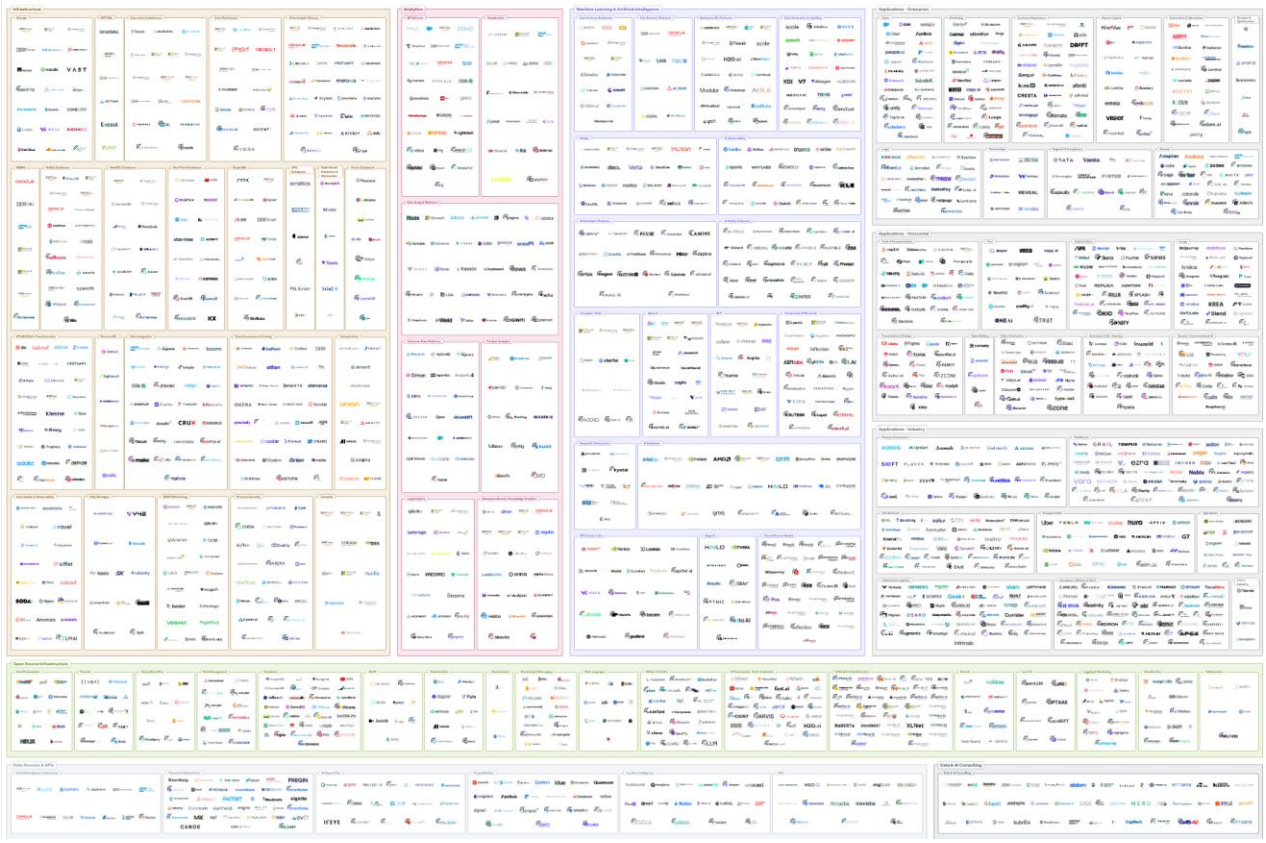
- Tecnologías:
    - Cloud Platforms: AWS, Azure, Google Cloud.
    - AI and ML Tools: PyTorch, Keras.
- Tareas:
    - Implementación de soluciones de inteligencia artificial.
    - Gestión de la gobernanza y calidad de los datos.

Habilidades y Actividades del Ingeniero de Datos

Los ingenieros de datos deben poseer una combinación de habilidades técnicas y empresariales.

- Principios de **arquitectura de datos**, **gestión de datos** y mejores prácticas de **ingeniería de software**.
- Capacidad de **evaluar y seleccionar las tecnologías adecuadas** en función de los requisitos específicos del proyecto.
- Deben colaborar estrechamente con otros roles dentro de los equipos de datos, incluidos científicos de datos, analistas de datos e ingenieros de ML.
- Los ingenieros de datos deben ser competentes en una variedad de tecnologías:
    - Lenguaje de consulta (**SQL**), para la transformación de datos y analítica.
    - **Bases de datos relacionales y NoSQL**, modelado.
    - Lenguajes de programación (**Python**, Java, Scala).
    - Frameworks de procesamiento distribuido (Hadoop, **Spark**) y herramientas de orquestación (**Airflow**).
    - Linux y bash para la realización de operaciones en el sistema operativo.
    - SSH y conceptos de redes para acceder a máquinas remotas.
    - Uso de APIs REST para incorporar datos de fuentes externas.
    - Git, GitHub y GitHub Actions, para conocer todo el ciclo de **CI/CD**.
    - **Docker**, como herramienta de gestión de contenedores para lanzar las diferentes herramientas.
    - Cuadernos **Jupyter** sobre el que desarrollar procesos de extracción, transformación y carga de datos.





The 2024 MAD (ML, AI & Data) Landscape (https://mad.firstmark.com/)

Ejemplos Reales de Uso de Big Data

Sector Comercial

- **Walmart**: Esta multinacional minorista recopila más de **2.5 petabytes de datos por hora** de las transacciones de sus clientes. El análisis de estos datos permite a Walmart optimizar sus operaciones, como la gestión de inventario, la previsión de la demanda y la personalización de ofertas para sus clientes.
- **Facebook**: Esta red social recopila una cantidad asombrosa de datos, incluyendo **300 millones de fotos** y **2.7 millones de "me gusta" por día**. Estos datos, caracterizados por su **volumen**, **velocidad** y **variedad**, se analizan para comprender el comportamiento de los usuarios, personalizar el contenido y los anuncios, e identificar tendencias emergentes.
- **Amazon**: Con **278 millones de clientes activos**, Amazon recopila datos sobre búsquedas en línea, compras y comportamiento de navegación. Estos datos permiten a la empresa recomendar productos, personalizar la experiencia del usuario y optimizar sus estrategias de marketing y ventas.

Ciencia e Ingeniería

- **Proyecto Sloan Digital Sky Survey**: Este proyecto de astronomía, uno de los primeros en utilizar el término "Big Data," tiene como objetivo identificar y

documentar objetos en el espacio. La gran cantidad de datos astronómicos recopilados requiere soluciones de Big Data para su almacenamiento, procesamiento y análisis.

- **Física de Altas Energías**: Los experimentos en el Gran Colisionador de Hadrones (LHC) generan **petabytes de datos por segundo**. El análisis de Big Data es crucial para procesar estos datos y buscar nuevas partículas subatómicas.
- **Genoma Humano**: El proyecto Genoma Humano tenía como objetivo encontrar, secuenciar y elaborar mapas genéticos y físicos de gran resolución del ADN humano, del orden de 100 Gigabytes.

Otros Sectores

- **Nest Labs**: Esta empresa (ahora propiedad de Google) fabrica termostatos inteligentes que aprenden de los patrones de uso de los usuarios para optimizar la eficiencia energética. Estos dispositivos generan grandes cantidades de datos que se analizan para proporcionar recomendaciones personalizadas a los usuarios y ayudar a las empresas de energía a predecir la demanda.
- **Vodafone**: Esta empresa de telecomunicaciones comercializa datos agregados, anónimos y geolocalizados de sus usuarios a empresas como TomTom para la optimización del tráfico en base a datos históricos y en tiempo real.
- **Siemens**: La industria 4.0 se basa en gran medida en utilizar las grandes cantidades de datos provenientes de sensores IoT (Internet of Things). Siemens utiliza Industrial Edge para procesar datos de fabricación de alta frecuencia cerca de la máquina, lo que permite una baja latencia y el uso directo de los datos en el proceso de producción.

Ejercicios

1. Contesta a las siguientes preguntas justificando tus respuestas:
    a. ¿Qué herramienta / servicio / tecnología basada en *Big Data* utilizas más a menudo?
    b. Desde tu punto de vista, ¿qué uso del Big Data podría aportar mayor valor a la sociedad en un futuro cercano?
2. Entra en Linkedin (www.linkedin.com) con tu cuenta y busca empleos de "Ingeniero de datos" en el área geográfica de tu interés.
    a. Revisa de 3 a 5 ofertas de empleo que sean de tu interés, y anota las principales tecnologías que solicitan.
    b. Búsca 5 tecnologías / herramientas en el mapa de herramientas de *Matt Truck* (https://mad.firstmark.com/) y anota a qué categoría pertenecen y el nombre de otra herramienta de su/s misma/s categoría.