

Gestión de Datos

Gestión y Arquitectura de Datos



La gestión de datos dentro de una organización es el proceso de:



Organizar



Proteger

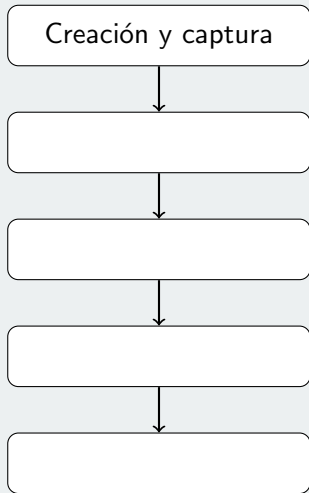


Recuperar

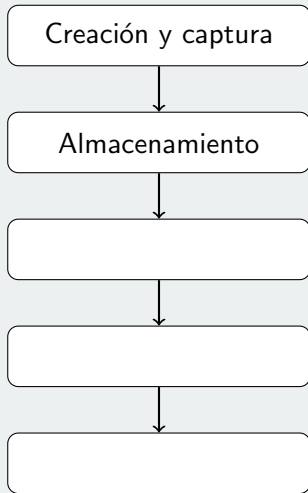


Mantener

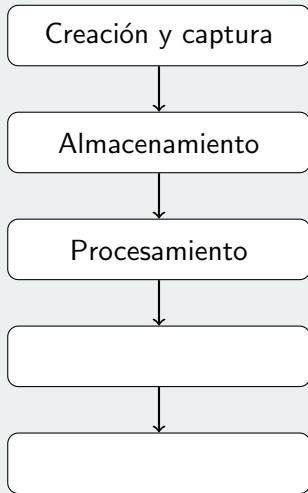
Ciclo de vida de los datos



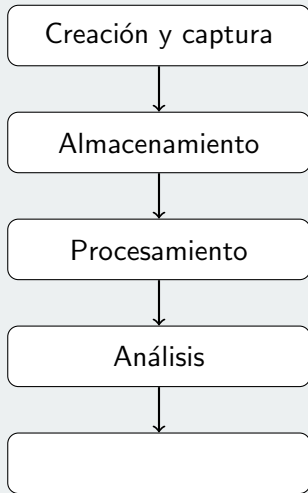
Ciclo de vida de los datos



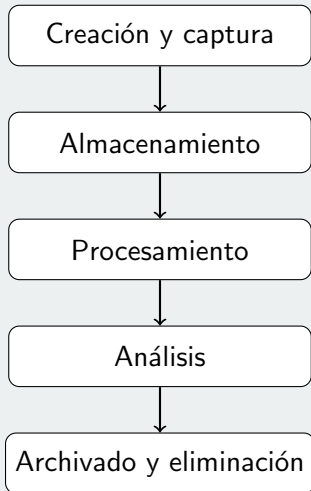
Ciclo de vida de los datos



Ciclo de vida de los datos




Ciclo de vida de los datos









Ingreso manual de datos

Creación y captura de datos





 Ingreso manual de datos

 Captura automática

Creación y captura de datos


-  Ingreso manual de datos
-  Captura automática
-  Integración de sistemas


Creación y captura de datos


-  Ingreso manual de datos
-  Captura automática
-  Integración de sistemas
-  Sensores y dispositivos IoT



Bases de datos relacionales


 Bases de datos relacionales

 Bases de datos NoSQL

 Bases de datos relacionales


 Bases de datos NoSQL

 Data Lakes

 Bases de datos relacionales

 Bases de datos NoSQL

 Data Lakes

 Sistemas de archivos distribuidos

⚙️ ETL (Extract, Transform, Load)

⚙️ ETL (Extract, Transform, Load)

🧹 Limpieza de datos

⚙️ ETL (Extract, Transform, Load)

🧹 Limpieza de datos


✓ Validación


⚙️ ETL (Extract, Transform, Load)


🧹 Limpieza de datos


✓ Validación


✚ Enriquecimiento


 Business Intelligence


 Business Intelligence


 Analytics


 Business Intelligence


 Analytics

 Machine Learning

 Business Intelligence




 Analytics





 Machine Learning

 Visualización

Políticas de retención

- 📁 Políticas de retención
- 🔑 Cumplimiento regulatorio

-  Políticas de retención
-  Cumplimiento regulatorio
-  Backup y recuperación

-  Políticas de retención
-  Cumplimiento regulatorio
-  Backup y recuperación
-  Eliminación segura

Dimensiones de la calidad de los datos (I)

- Precisión: grado en que los datos son exactos y libres de errores.
- Completitud: grado en que los datos cubren todas las partes relevantes de la información.
- Consistencia: grado en que los datos son coherentes y no contienen contradicciones.
- Actualidad: grado en que los datos reflejan la situación actual.

Dimensiones de la calidad de los datos (II)

- Unicidad: grado en que los datos son únicos y no duplicados.
- Accesibilidad: grado en que los datos son fáciles de encontrar y usar.
- Seguridad: grado en que los datos están protegidos contra accesos no autorizados.

- ⚠ Tasa de error: proporción de datos que contienen errores.
- 📄 Tasa de duplicación: proporción de datos que son duplicados.
- 🔄 Tasa de actualización: proporción de datos que se actualizan regularmente.

Gobierno de datos: roles y responsabilidades



CDO

Dirección y gestión
de todos los datos
de la organización



Data Owner

Responsable de un
conjunto de datos
específico



Data Steward

Responsable
operativo, mantiene
datos actualizados

Gobierno de datos: políticas y procedimientos



Seguridad

Protección de datos



Privacidad

Control de acceso



Retención

Tiempo de vida



Uso

Normas



Acceso

Permisos y roles



Fuentes de datos:

- Sistemas operacionales (ERP, CRM)
- APIs externas
- Datos no estructurados (documentos, imagenes, videos)



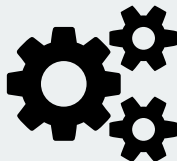
Almacenamiento:

- Data Warehouse: datos estructurados y análisis histórico
- Data Lake: datos en formato original
- Data Mart: subconjuntos especializados



Procesamiento:

- Procesamiento por lotes (Batch)
- Procesamiento en streaming
- Arquitectura Lambda (combinación de ambos)



Consumo:

- Reportes estructurados
- Dashboards interactivos
- APIs para integración



¿Dudas?
¿Consultas?



Universidad de
SanAndrés