



31 DE OCTUBRE DE 2023

CONCEPTOS DE INTEGRACIÓN DE DATOS



Contenido

1. Integración de datos	2
a. Activo de datos	2
b. Ingeniería de datos	2
c. Limpieza de datos	2
2. Calidad de los datos	3
a. Enriquecimiento de datos	3
b. Protección de datos	3
c. Validación de datos	4
d. Arquitectura de datos	4
3. Gobernanza de datos	5
a. Catálogo de datos	5
b. Linaje de datos	5
c. Política de datos y flujo de trabajo	5
d. Política de Datos	5
e. Flujo de Trabajo de Datos	6
4. Estado de los datos	7
a. Resultados comerciales	7
b. Preparación de datos y API de datos	7
c. Alfabetización de datos	7
5. Otros conceptos	8
a. Perfil de datos	8
b. Protección de datos	8
c. Curación de datos	8
d. Enriquecimiento de datos	8
e. Validación de datos	8
f. Observabilidad de datos	9
6. Anexo	10

1. Integración de datos

La integración de datos es un proceso que consiste en reunir datos de diferentes fuentes para obtener una vista unificada y más valiosa de ellos, de modo que las empresas puedan tomar mejores decisiones y con mayor rapidez.

a. Activo de datos

El término "activos de datos" se refiere a los conjuntos de datos, información o recursos digitales que una organización considera valiosos y críticos para sus operaciones o sus objetivos estratégicos. Estos activos de datos pueden incluir una amplia variedad de tipos de datos, como datos de clientes, datos financieros, datos de inventario, registros de transacciones, información de empleados y cualquier otro tipo de información que sea esencial para el funcionamiento y la toma de decisiones de una empresa u organización.

b. Ingeniería de datos

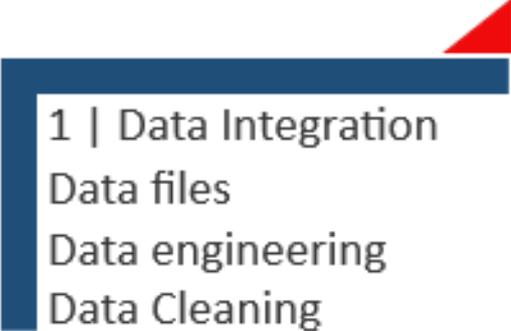
La ingeniería de datos es una disciplina que basa su enfoque en diseñar, construir y mantener sistemas de procesamientos de datos para el almacenamiento y procesamiento de grandes cantidades de datos tanto estructurados como no estructurados.

c. Limpieza de datos

La limpieza de datos, también conocida como depuración de datos, es el proceso de identificar y corregir errores, incoherencias y problemas en conjuntos de datos. Este proceso es esencial para garantizar la calidad de los datos y la confiabilidad de la información que se encuentra en una base de datos, sistema de información o conjunto de datos en general. La limpieza de datos implica una serie de tareas, que pueden incluir:

- Detección y corrección de errores tipográficos y ortográficos.
- Eliminación de duplicados.
- Estandarización de datos.
- Validación de datos.
- Manejo de valores faltantes.
- Verificación de integridad referencial.

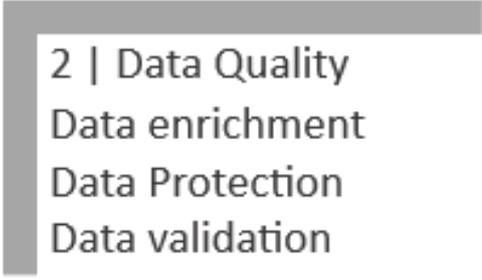
La limpieza de datos es un paso crítico en el proceso de gestión de datos, ya que datos inexactos o sucios pueden llevar a decisiones erróneas y problemas en el análisis.



1 | Data Integration
Data files
Data engineering
Data Cleaning

2. Calidad de los datos

La calidad de datos se refiere a la medida en que los datos son precisos, confiables, coherentes y adecuados para su propósito previsto en una organización. Es fundamental para asegurar que los datos utilizados para la toma de decisiones, análisis, operaciones y otros procesos sean de alta calidad y reflejen la realidad de manera precisa. La calidad de datos involucra varios aspectos clave, que incluyen:



2 | Data Quality
Data enrichment
Data Protection
Data validation

- Exactitud.
- Integridad.
- Coherencia.
- Relevancia.
- Actualidad.
- Confiabilidad

La mejora de la calidad de datos es esencial para que una organización tome decisiones informadas y obtenga resultados precisos de análisis y procesos. La gestión de calidad de datos implica la implementación de políticas, procesos y tecnologías para mantener y mejorar continuamente la calidad de los datos a lo largo del tiempo.

a. Enriquecimiento de datos

El enriquecimiento de datos es un proceso mediante el cual se agregan o mejoran datos existentes con información adicional, más detallada o relevante. El objetivo principal del enriquecimiento de datos es mejorar la calidad y la utilidad de los datos, lo que puede ayudar a las organizaciones a tomar decisiones más informadas, a comprender mejor a sus clientes y a mejorar la precisión de sus análisis y modelos.

b. Protección de datos

La protección de datos se refiere a las medidas y prácticas diseñadas para garantizar la seguridad, la privacidad y la integridad de la información personal o sensible. Esto es fundamental para proteger la información confidencial de individuos y organizaciones de posibles amenazas y abusos.

Algunos aspectos clave de la protección de datos incluyen:

- Privacidad.
- Seguridad de la información.
- Cumplimiento legal.
- Gestión de consentimiento.
- Retención y eliminación de datos.
- Monitoreo y auditoría.
- Respuesta a incidentes.

c. Validación de datos

La validación de datos es un proceso que implica verificar la precisión y la integridad de los datos ingresados o almacenados en un sistema o en una base de datos. El objetivo principal de la validación de datos es garantizar que los datos sean coherentes, confiables y cumplan con ciertos criterios o reglas predefinidas. Este proceso es fundamental para mantener la calidad de los datos y prevenir errores que puedan afectar las operaciones y la toma de decisiones.

Aquí hay algunas técnicas y enfoques comunes en la validación de datos:

- Verificación de formato.
- Validación numérica.
- Validación de longitud.
- Validación de patrones.
- Validación de reglas de negocio.

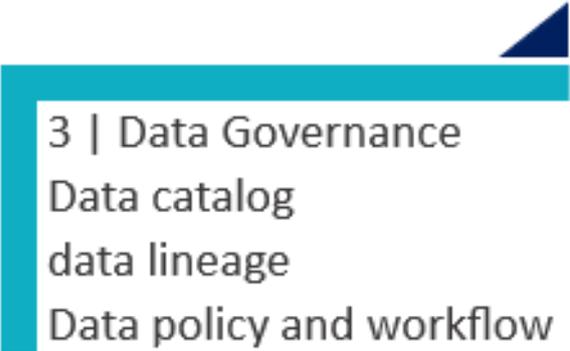
La validación de datos es esencial para garantizar la calidad de los datos y evitar problemas como datos incorrectos o inconsistentes que pueden afectar la precisión de los informes, la toma de decisiones y la eficiencia de los procesos empresariales.

d. Arquitectura de datos

La arquitectura de datos se refiere a la estructura y diseño de cómo una organización almacena, organiza, procesa y gestiona sus datos. Es un componente esencial de la gestión de datos en una empresa o entidad, y su objetivo principal es asegurarse de que los datos estén disponibles, sean accesibles, sean confiables y cumplan con los requisitos comerciales y tecnológicos de la organización.

3. Gobernanza de datos

La gobernanza de datos es un conjunto de procesos, políticas, estándares y prácticas que se implementan en una organización para garantizar la gestión efectiva, la calidad, la seguridad y el cumplimiento de los datos a lo largo de toda la empresa. El objetivo principal de la gobernanza de datos es establecer un marco sólido que permita a una organización aprovechar al máximo sus datos mientras se minimizan los riesgos y se asegura la integridad y la confidencialidad de la información.



3 | Data Governance
Data catalog
data lineage
Data policy and workflow

a. Catálogo de datos

Un catálogo de datos es una herramienta o sistema que actúa como un repositorio centralizado de información sobre los datos dentro de una organización. Su propósito principal es proporcionar una visión organizada y detallada de los activos de datos disponibles, lo que facilita su descubrimiento, acceso y gestión.

El catálogo de datos desempeña un papel crucial en la gestión de datos y la gobernanza de datos al proporcionar visibilidad y control sobre los activos de datos de una organización.

b. Linaje de datos

El linaje de datos es un concepto que se refiere al rastreo y la documentación de la procedencia y los cambios que ha experimentado un conjunto de datos a lo largo de su ciclo de vida. En otras palabras, el linaje de datos muestra la historia completa de un dato, desde su origen hasta su estado actual, incluyendo todas las transformaciones y procesos a los que ha sido sometido.

c. Política de datos y flujo de trabajo

Las políticas de datos y los flujos de trabajo de datos son dos componentes esenciales de la gestión de datos en una organización. Juntos, ayudan a definir cómo se manejan, almacenan, protegen y utilizan los datos de manera coherente y eficiente.

d. Política de Datos

Una política de datos es un conjunto de directrices, reglas y principios que establecen cómo se deben administrar y utilizar los datos en una organización. Estas políticas se crean para garantizar la calidad de los datos, la privacidad, la seguridad, la conformidad con las regulaciones y la toma de decisiones basada en datos confiables.

e. Flujo de Trabajo de Datos

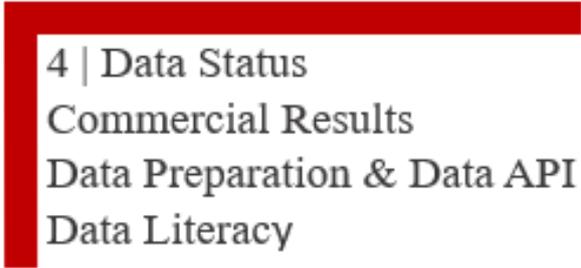
Un flujo de trabajo de datos, también conocido como proceso de datos, describe la secuencia de pasos y tareas que se siguen para mover, transformar y utilizar los datos en una organización. Estos flujos de trabajo son esenciales para garantizar que los datos se procesen de manera eficiente y efectiva desde su origen hasta su destino final. Algunos elementos clave de un flujo de trabajo de datos incluyen:

- Extracción.
- Transformación.
- Carga.
- Programación y automatización.
- Monitoreo y gestión.

Tanto las políticas de datos como los flujos de trabajo de datos son esenciales para la gestión efectiva de datos en una organización. Las políticas establecen el marco de cómo se deben tratar los datos, mientras que los flujos de trabajo permiten la implementación práctica de esas políticas en la vida cotidiana de la organización.

4. Estado de los datos

El "estado de los datos" se refiere a la condición actual de los datos dentro de una organización o sistema en un momento específico. Describe si los datos son precisos, actualizados, completos, coherentes y disponibles para su uso previsto. El estado de los datos es un indicador crítico de la calidad y la utilidad de la información que una organización utiliza para tomar decisiones, realizar análisis y llevar a cabo sus operaciones.



4 | Data Status
Commercial Results
Data Preparation & Data API
Data Literacy

a. Resultados comerciales

Los "resultados comerciales" se refieren a los logros, métricas y datos que una organización obtiene en el curso de sus operaciones comerciales. Estos resultados comerciales pueden variar según la industria, el tipo de empresa y los objetivos específicos de la organización, pero en general, se utilizan para evaluar el rendimiento y el éxito de la empresa en términos financieros y operativos.

b. Preparación de datos y API de datos

La "preparación de datos" y las "API de datos" son dos aspectos importantes en la gestión y el uso efectivo de los datos en una organización. Aquí se describen ambos conceptos:

Preparación de Datos: La preparación de datos es el proceso de limpiar, transformar y organizar los datos para que estén en un formato adecuado y sean utilizables para análisis, informes u otras aplicaciones.

API de Datos: Una API de datos, o interfaz de programación de aplicaciones de datos, es un conjunto de reglas y protocolos que permiten que las aplicaciones y sistemas informáticos se comuniquen entre sí y compartan datos de manera estructurada.

c. Alfabetización de datos

La "alfabetización de datos" se refiere a la capacidad de una persona para comprender, analizar y utilizar datos de manera efectiva. Implica la habilidad de leer, interpretar y comunicar información basada en datos de manera crítica y precisa. En un mundo donde los datos desempeñan un papel cada vez más importante en la toma de decisiones, la alfabetización de datos se ha convertido en una habilidad fundamental tanto en el ámbito personal como profesional.

5. Otros conceptos

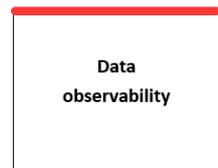
a. Perfil de datos

Un perfil de datos se refiere a una descripción detallada y completa de los datos en una base de datos, conjunto de datos o sistema de información. El objetivo de un perfil de datos es proporcionar una visión general de los datos, incluyendo información sobre su estructura, calidad, contenido y otros aspectos relevantes.



b. Protección de datos

La protección de datos se refiere a las prácticas y medidas implementadas para garantizar la seguridad y la privacidad de la información personal y sensible. Esto es fundamental en un mundo cada vez más digital, donde la información se comparte y almacena en una variedad de plataformas y sistemas.



c. Curación de datos

La curación de datos es un proceso integral y continuo que implica la organización, limpieza, mejora y mantenimiento de datos para garantizar su calidad, confiabilidad y utilidad a lo largo del tiempo. Es esencial para gestionar grandes conjuntos de datos, ya que ayuda a garantizar que la información sea precisa, coherente y esté lista para su uso en análisis, investigación y toma de decisiones.

d. Enriquecimiento de datos

El enriquecimiento de datos es una práctica esencial que se emplea para mejorar y ampliar la calidad y utilidad de los datos. Este proceso implica la incorporación de información adicional, más detallada o relevante a conjuntos de datos existentes. Su finalidad es enriquecer la información disponible, permitiendo a las organizaciones obtener una comprensión más profunda de sus datos, clientes y operaciones.

e. Validación de datos

La validación de datos es un procedimiento esencial que consiste en verificar la exactitud y la integridad de los datos introducidos o almacenados en un sistema o base de datos. Su propósito principal es asegurar la coherencia, confiabilidad y conformidad de los datos con criterios y reglas preestablecidos. Este proceso se erige como un pilar fundamental para preservar la calidad de los datos y evitar posibles errores que pudieran impactar las operaciones y la toma de decisiones.

f. Observabilidad de datos

El objetivo de la observabilidad de datos es proporcionar una visión completa y transparente de lo que sucede con los datos, lo que facilita la detección temprana de posibles fallos o cuellos de botella en los sistemas. Esto, a su vez, permite tomar medidas proactivas para garantizar la calidad y el rendimiento de los datos, así como mejorar la toma de decisiones basada en datos en una organización.

6. Anexo

Journey to Data Health

Turning raw data assets into trusted business results

