¿Qué es la metadata?

La metadata es datos que proporcionan información sobre otros datos.

Esta es una definición muy amplia. Aquí consideraremos el concepto de metadata dentro del contexto de bases de datos, almacenamiento de datos, sistemas de inteligencia empresarial y todo tipo de repositorios y plataformas de datos.

Consideraremos los siguientes tres tipos principales de metadata:

- Metadata técnica
- Metadata de proceso, y
- Metadata empresarial

Metadatos técnicos

Los metadatos técnicos son metadatos que definen las estructuras de datos en repositorios o plataformas de datos, principalmente desde una perspectiva técnica.

Por ejemplo, los metadatos técnicos en un almacén de datos incluyen activos como:

- Tablas que registran información sobre las tablas almacenadas en una base de datos, como:
 - o el nombre de cada tabla
 - o el número de columnas y filas que tiene cada tabla
- Un catálogo de datos, que es un inventario de tablas que contienen información, como:
 - o el nombre de cada base de datos en el almacén de datos de la empresa
 - o el nombre de cada columna presente en cada base de datos
 - o los nombres de cada tabla en la que se encuentra cada columna
 - o el tipo de datos que contiene cada columna

Los metadatos técnicos para bases de datos relacionales se almacenan típicamente en tablas especializadas en la base de datos llamadas Catálogo del Sistema.

Metadatos del proceso

Los metadatos del proceso describen los procesos que operan detrás de sistemas empresariales como almacenes de datos, sistemas contables o herramientas de gestión de relaciones con clientes.

Muchos sistemas empresariales importantes son responsables de recopilar y procesar datos de diversas fuentes. Tales sistemas críticos necesitan ser monitoreados para detectar fallos y cualquier anomalía de rendimiento que surja. Los metadatos del proceso para tales sistemas incluyen el seguimiento de cosas como:

- tiempos de inicio y fin del proceso
- uso del disco
- de dónde se movieron los datos y a dónde, y
- cuántos usuarios acceden al sistema en un momento dado

Este tipo de datos es invaluable para la solución de problemas y la optimización de flujos de trabajo y consultas ad hoc.

Metadatos empresariales

Los usuarios que desean explorar y analizar datos dentro y fuera de la empresa suelen estar interesados en *el descubrimiento de datos*. Necesitan poder encontrar datos que sean significativos y valiosos para ellos y saber de dónde se puede acceder a esos datos. Por lo tanto, estos usuarios con mentalidad empresarial están interesados en los metadatos empresariales, que son información sobre los datos descrita de maneras fácilmente interpretables, como:

- cómo se adquieren los datos
- qué están midiendo o describiendo los datos
- la conexión entre los datos y otras fuentes de datos

Los metadatos empresariales también sirven como documentación para todo el sistema de almacenamiento de datos.

Gestión de metadatos

La gestión de metadatos incluye el desarrollo y la administración de políticas y procesos para garantizar que la información pueda ser accedida e integrada desde diversas fuentes y compartida adecuadamente en toda la empresa.

La creación de un catálogo de datos confiable y fácil de usar es un objetivo principal de un modelo de gestión de metadatos. El catálogo de datos es un componente clave

de un sistema moderno de gestión de metadatos, sirviendo como el principal activo alrededor del cual se administra la gestión de metadatos. Sirve como la base sobre la cual las empresas pueden inventariar y organizar de manera eficiente sus sistemas de datos. Un modelo moderno de gestión de metadatos incluirá una interfaz de usuario basada en la web que permite a los ingenieros y usuarios de negocios buscar y encontrar fácilmente información sobre atributos clave como CustomerName o ProductType. Este tipo de modelo es central para cualquier iniciativa de Gobernanza de Datos.

¿Por qué es importante la gestión de metadatos?

Una buena gestión de metadatos tiene muchos beneficios valiosos. Tener acceso a un catálogo de datos bien implementado mejora enormemente el descubrimiento de datos, la repetibilidad, la gobernanza y también puede facilitar el acceso a los datos.

Los metadatos bien gestionados te ayudan a comprender tanto el contexto empresarial asociado con los datos de la empresa como la línea de datos, lo que ayuda a mejorar la gobernanza de datos. La línea de datos proporciona información sobre el origen de los datos y cómo se transforman y trasladan, y así facilita el rastreo de errores de datos hasta su causa raíz. La gobernanza de datos es un concepto de gestión de datos que se refiere a la capacidad que permite a una organización garantizar que exista alta calidad de datos a lo largo de todo el ciclo de vida de los datos, y se implementan controles de datos que apoyan los objetivos empresariales.

Las áreas clave de enfoque de la gobernanza de datos incluyen disponibilidad, usabilidad, consistencia, integridad de datos y seguridad de datos, e incluye el establecimiento de procesos para asegurar una gestión efectiva de datos en toda la empresa, como la responsabilidad por los efectos adversos de la mala calidad de los datos y asegurando que los datos que tiene una empresa puedan ser utilizados por toda la organización.

Herramientas populares para la gestión de metadatos

Las herramientas populares para la gestión de metadatos incluyen:

- IBM InfoSphere Information Server
- CA Erwin Data Modeler
- Oracle Warehouse Builder
- SAS Data Integration Server

- Talend Data Fabric
- Alation Data Catalog
- SAP Information Steward
- Microsoft Azure Data Catalog
- IBM Watson Knowledge Catalog
- Oracle Enterprise Metadata Management (OEMM)
- Adaptive Metadata Manager
- Unifi Data Catalog
- data.world
- Informatica Enterprise Data Catalog