

¿QUÉ ES LA TRANSFORMACIÓN DE DATOS?

La transformación de datos es una parte crítica del proceso de integración de datos, en el que los datos sin procesar se convierten en un formato o estructura unificado. La transformación de datos garantiza la compatibilidad con los sistemas de destino y mejora la calidad y la usabilidad de los datos. Es un aspecto esencial de las prácticas de gestión de datos, incluida la contención de datos, el análisis de datos y el almacenamiento de datos.

Si bien los especialistas pueden lograr la transformación de datos manualmente, las grandes franjas de datos necesarias para impulsar las aplicaciones empresariales modernas suelen requerir cierto nivel de **automatización**. Las herramientas y tecnologías desplegadas a través del proceso de conversión de datos pueden ser simples o complejas.

Por ejemplo, una transformación de datos puede ser tan sencilla como convertir un campo de fecha (por ejemplo: MM/DD/AA) en otro, o dividir una sola columna de Excel en dos. Pero las transformaciones de datos complejas, que limpian y estandarizan datos de múltiples fuentes dispares y constan de múltiples flujos de trabajo, pueden requerir habilidades avanzadas de **ciencia de datos**.

Estas funciones avanzadas de **ingeniería de datos** incluyen la normalización de datos, que define las relaciones entre los puntos de datos, y enriquecimiento de datos, que complementa la información existente con conjuntos de datos de terceros.

En la economía global actual, digital-first, las transformaciones de datos ayudan a las organizaciones a aprovechar grandes volúmenes de datos de diferentes fuentes para mejorar el servicio, capacitar modelos de machine learning y desplegar analytics de big data.

Casos de uso de transformación de datos

Al estandarizar los conjuntos de datos y prepararlos para su posterior procesamiento, la transformación de datos hace posibles varias prácticas cruciales de datos empresariales.

¿QUÉ ES LA TRANSFORMACIÓN DE DATOS?

Las razones comunes para la transformación de datos en el mundo empresarial incluyen:

- **Inteligencia empresarial.**

Las organizaciones transforman los datos para su uso en aplicaciones de business intelligence, como paneles en tiempo real e informes de pronósticos, lo que permite una toma de decisiones basada en datos que tiene en cuenta grandes cantidades de información.

- **Almacenamiento de datos.**

La transformación de datos prepara los datos para su almacenamiento y gestión en un almacén de datos o lago de datos, lo que facilita la realización de consultas y análisis eficientes.

- **Aprendizaje automático.**

Los modelos de machine learning requieren datos limpios y organizados. Garantizar que los datos sean confiables y estén en el formato correcto permite a las organizaciones usarlos para capacitar y ajustar herramientas de inteligencia artificial (IA).

- **Analytics de big data.**

Antes de poder analizar los big data para business intelligence, investigación de mercado u otras aplicaciones, deben cotejarse y formatearse adecuadamente.

- **Migración de datos.**

Mover datos de sistemas on-premises más antiguos a plataformas modernas, como un almacén de datos en la nube o un lago de datos a menudo implica transformaciones de datos complejas.

Proceso de transformación de datos

Las transformaciones de datos suelen seguir un proceso estructurado para producir datos utilizables y valiosos a partir de su forma sin procesar. Los pasos comunes en un proceso de transformación de datos incluyen:

¿QUÉ ES LA TRANSFORMACIÓN DE DATOS?

1. Descubrimiento de datos.

Durante el proceso de descubrimiento, se recopilan los datos de origen. Este proceso puede incluir el raspado de datos sin procesar de API, una SQL database o archivos internos en formatos dispares. Al identificar y extraer esta información, los profesionales de datos se cercioran de que la información recopilada sea completa y relevante para su eventual aplicación. Durante el descubrimiento, los ingenieros también comienzan a comprender las características y la estructura de los datos en un proceso conocido como perfilado de datos.

2. Limpieza de datos.

La preparación y limpieza de datos requiere identificar y corregir errores, inconsistencias e imprecisiones en los datos brutos. Este paso garantiza la calidad y confiabilidad de los datos mediante la eliminación de duplicados y valores atípicos o el manejo de missing values.

3. Mapeo de datos.

El mapeo de datos implica la creación de un esquema o proceso de mapeo para guiar el proceso de transformación. Durante este proceso, los ingenieros de datos definen cómo los elementos del sistema de origen corresponden a elementos específicos en el formato de destino.

4. Generación de código.

Ya sea empleando una herramienta de terceros o generando código internamente, durante este paso una organización crea el código que transformará los datos.

5. Ejecución y validación de código.

Durante esta fase, la transformación real tiene lugar a medida que se aplica el código a los datos sin procesar. Los datos transformados se cargan en su sistema de destino para su posterior análisis o procesamiento. A continuación, los datos transformados y el modelo de datos se validan para garantizar la coherencia y la corrección.

¿QUÉ ES LA TRANSFORMACIÓN DE DATOS?

6. Revisión.

Durante el proceso de revisión, analistas de datos, ingenieros o usuarios finales revisan los datos de salida, confirmando que cumplen con los requerimientos.

Referencia:

Hayes, Molly; Downie, Amanda. (2024) ¿Qué es la transformación de datos? IBM.

Recuperado de: <https://www.ibm.com/mx-es/think/topics/data-transformation>