

Universidad de Costa Rica

Facultad de Ingeniería

Escuela de Ciencias de la Computación e Informática

Bases de Datos Avanzadas

Resumen de conceptos vistos en el módulo “Introducción a la ingeniería de
datos en Azure”

2025



Autor

Brandon Trigueros Lara C17899

El principal trabajo de un ingeniero de datos es captar, transformar y consolidar datos de diferentes fuentes estructuradas y no estructuradas en una estructura que sea útil para responder preguntas de negocio. Sin embargo, también es el encargado de asegurar la confiabilidad, rendimiento y robustez de las canalizaciones de y los almacenes de datos.

En ingeniería de datos se trabajan con varios tipos de datos distintos. Entre ellos están:

Datos estructurados, son datos con forma de tabla relacional, como base de datos transaccional SQL, archivo CSV, entre otros. Los cuales se encuadran en un esquema de filas y columnas y todos los datos se acoplan en este esquema.

Datos semi-estructurados, son datos que no cuentan con una forma tan rígida como los estructurados pero que sí presentan metadatos o marcadores que permiten encontrar relaciones y propiedades entre los datos.

Datos no estructurados, son aquellos que no siguen ningún tipo de formato que me permite asociar variables con valores, son más libres y cercanos al lenguaje natural, como PDF, TXT, imágenes, entre otros.

Como se mencionó anteriormente, entre las acciones que se realizan en ingeniería de datos podemos encontrar:

Integración, consiste en vincular las fuentes de datos iniciales con los mecanismos analíticos, asegurando que sea correcto y seguro.

Transformación, básicamente consistente en transformar los datos de entrada en una estructura y formato válido para el análisis.

Consolidación, finalmente la consolidación es el proceso en el que se combinan datos provenientes de distintas fuentes, y que han sido previamente transformados, en un poderoso objeto de análisis, como un data warehouse.

El ingeniero de datos se encarga de captar, transformar y consolidar información proveniente de diversas fuentes—ya sean estructuradas, semiestructuradas o no estructuradas—para que esta sea útil en la toma de decisiones de negocio. Su labor abarca la integración de datos operativos (transaccionales) y analíticos (optimizados para el análisis), utilizando canalizaciones ETL que aseguran la transformación y carga de datos en repositorios especializados.


Además, se gestionan flujos de datos en tiempo real (streaming) provenientes de dispositivos IoT y redes sociales, y se emplean herramientas como Apache Spark para procesar grandes volúmenes de información de manera eficiente.

Microsoft Azure ofrece una amplia gama de servicios que facilitan la implementación y administración de cargas de trabajo de ingeniería de datos. En la solución típica de análisis de datos empresariales, los datos operativos generados por aplicaciones y dispositivos se almacenan en servicios como Azure SQL Database, Azure Cosmos DB y Microsoft Dataverse, mientras que los datos de streaming se capturan mediante Azure Event Hubs.

Posteriormente, estos datos se capturan, ingieren y consolidan en almacenes analíticos que permiten modelar y visualizar información a través de herramientas como Power BI. Entre las tecnologías clave para estos procesos se destacan Azure Synapse Analytics, Azure Data Lake Storage Gen2, Azure Stream Analytics, Azure Data Factory y Azure Databricks.

Evaluación de módulos

Evaluación del módulo • 3 minutos



Evaluación del módulo superada

Buen trabajo. Ha superado la evaluación del módulo.

Puntuación: 100%

Elija la mejor respuesta para cada una de las preguntas.

200 XP

Referencias

[1] Microsoft, “¿En qué consiste la ingeniería de datos?,” *Microsoft Learn*, 2024. [Módulo de aprendizaje en línea]. Disponible: <https://learn.microsoft.com/es-es/training/modules/introduction-to-data-engineering-azure/2-what-data-engineering>.