100 XP

Review tasks and tools for data engineering

3 minutes

Data engineers are tasked with managing and organizing data, while also monitoring for trends or inconsistencies that will impact business goals. It's a highly technical position, requiring experience and skills in areas like programming, mathematics, and computer science. But data engineers also need soft skills to communicate data trends to others in the organization and to help the business make use of the data it collects.

Data Engineer tasks and responsibilities

Some of the most common roles and responsibilities of a data engineer include:

- Developing, constructing, testing, and maintaining databases and data structures.
- Aligning the data architecture with business requirements.
- Data acquisition.
- Developing processes for creating and retrieving information from data sets.
- Using programming languages and tools to examine the data.
- Identifying ways to improve data reliability, efficiency, and quality.
- Conducting research for industry and business questions.
- Deploying sophisticated analytics programs, machine learning, and statistical methods.
- Preparing data for predictive and prescriptive modeling.
- Using data to discover tasks that can be automated.

Common data engineering tools

To master data engineering, you'll need to be familiar with a range of tools that enable you to create well-designed databases, optimized for the business processes that will be run. You must have a thorough understanding of the architecture of the database management system, the platform on which the system runs, and the business requirements for the data being stored in the database.

If you're using a relational database management system, you need to be fluent in SQL. You must be able to use SQL to create databases, tables, indexes, views, and the other objects required by the database. Many database management systems provide tools that enable you to create and run SQL scripts. For example, *SQL Server Management Studio* (described in the

previous unit), lets you create and query tables visually, but you can also create your own SQL scripts manually.

In some cases, you may need to interact with a database from the command line. Many database management systems provide a command-line interface that supports these operations. For example, you can use the *sqlcmd* utility to connect to Microsoft SQL Server and Azure SQL Database, and run ad-hoc queries and commands.

As a SQL Server professional, your primary data manipulation tool might be Transact-SQL. As a data engineer you might use additional technologies, such as Azure Databricks, and Azure HDInsight to generate and test predictive models. If you're working in the non-relational field, you might use Azure Cosmos DB as your primary data store. To manipulate and query the data, you might use languages such as HiveQL, R, or Python.

Next unit: Review tasks and tools for data visualization and reporting

Continue >