**Explore roles and responsibilities in the world of data**

1. Over the last decade, the amount of data that systems and devices generate has increased significantly.
2. Because of this increase, new technologies, roles, and approaches to working with data are affecting data professionals.
3. Data professionals typically fulfill different roles when managing, using, and controlling data.

# Explore job roles in the world of data

There's a wide variety of roles involved in managing, controlling, and using data.

Some roles are business-oriented, some involve more engineering, some focus on research, and some are hybrid roles that combine different aspects of data management.

## **What are the roles in the world of data?**

* **Database Administrators** manage databases, assigning permissions to users, storing backup copies of data and restore data in case of any failures.
* **Data Engineers** are vital in working with data, applying data cleaning routines, identifying business rules, and turning data into useful information.
* **Data Analysts** explore and analyze data to create visualizations and charts to enable organizations to make informed decisions.

## **Azure Database Administrator role**



1. An Azure database administrator is responsible for the **design, implementation, maintenance, and operational aspects of on-premises and cloud-based database solutions built on Azure data services and SQL Server**.
2. They're responsible for the **overall availability and consistent performance and optimizations** of the database solutions.
3. They work with **stakeholders to implement policies, tools, and processes for backup and recovery plans to recover following a natural disaster or human-made error.**
4. The database administrator is also responsible for **managing the security** of the data in the database, **granting privileges** over the data, granting or **denying access** to users as appropriate.

## **Data Engineer role**



1. A data engineer **collaborates with stakeholders** to **design and implement data-related assets** that include **data ingestion pipelines, cleansing and transformation activities, and data stores for analytical workloads.**
2. They use a wide range of data platform technologies, including relational and nonrelational databases, file stores, and data streams.
3. They're also responsible for **ensuring that the privacy** of data is maintained within the cloud and spanning from on-premises to the cloud data stores.
4. They also own the **management and monitoring of data stores and data pipelines to ensure that data loads perform as expected.**

## **Data Analyst role**



1. A data analyst **enables businesses to maximize the value of their data assets**.
2. They're responsible for **designing and building scalable models, cleaning and transforming data, and enabling advanced analytics capabilities through reports and visualizations.**
3. A data analyst processes raw data into relevant insights based on identified business requirements to deliver relevant insights.

# Review tasks and tools for database administration

# Database Administrators are tasked with managing and organizing databases.

# A database administrator's primary job is to ensure that

# data is available,

# protected from loss ,

# corruption,

# theft,

# easily accessible as needed.

## **Database Administrator tasks and responsibilities**

* Installing and upgrading the database server and application tools.
* Allocating system storage and planning storage requirements for the database system.
* Modifying the database structure, as necessary, from information given by application developers.
* Enrolling users and maintaining system security.
* Ensuring compliance with database vendor license agreement.
* Controlling and monitoring user access to the database.
* Monitoring and optimizing the performance of the database.
* Planning for backup and recovery of database information.
* Maintaining archived data.
* Backing up and restoring databases.
* Contacting database vendor for technical support.
* Generating various reports by querying from database as per need.
* Managing and monitoring data replication.

## **Common database administrator tools**

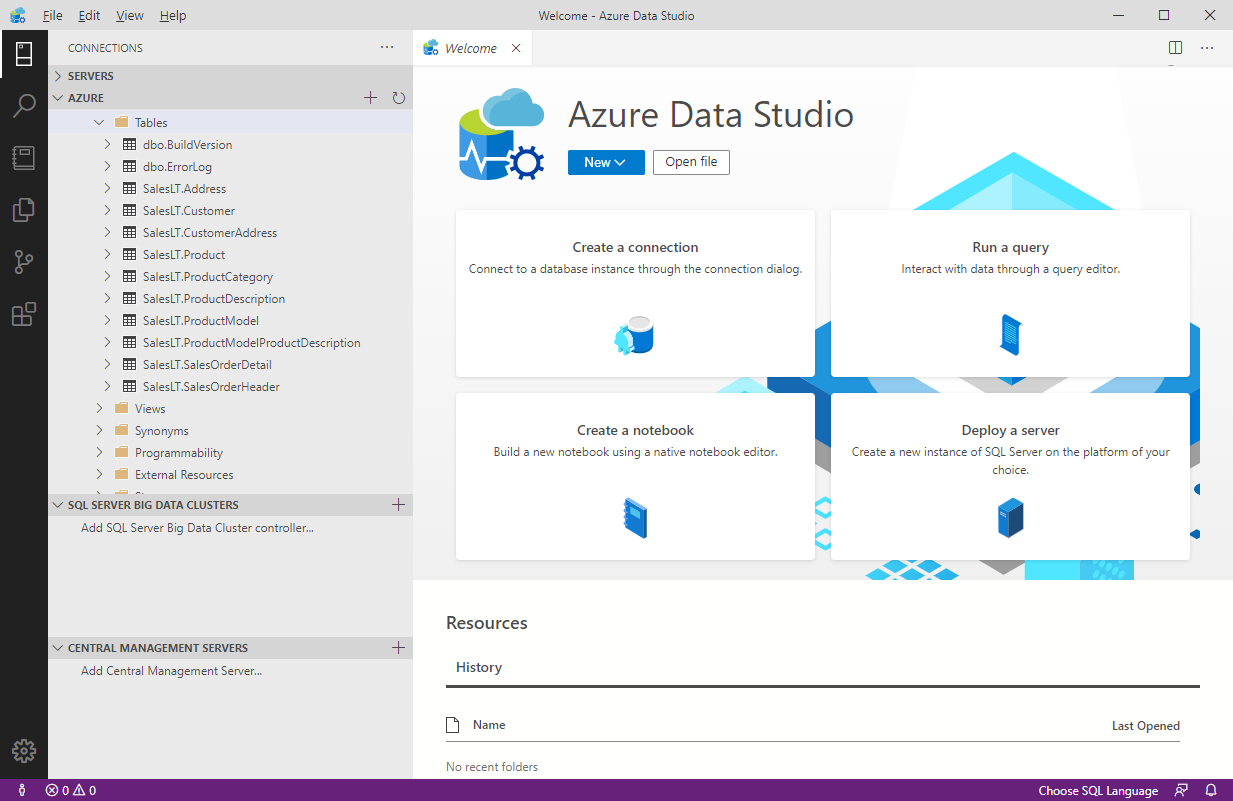
* 1. Most database management systems provide their own set of tools to assist with database administration.
  2. For example, **SQL Server Database Administrators use SQL Server Management Studio** for most of their day-to-day database maintenance activities.
  3. Other systems have their own database-specific interfaces, such **as pgAdmin for PostgreSQL systems**, **or MySQL Workbench for MySQL**.
  4. There are also a number of **cross-platform database** administration tools available. One example is **Azure Data Studio.**

### **What is Azure Data Studio?**

1. Azure Data Studio provides a **graphical user interface for managing many different database systems**.
2. It currently provides connections to

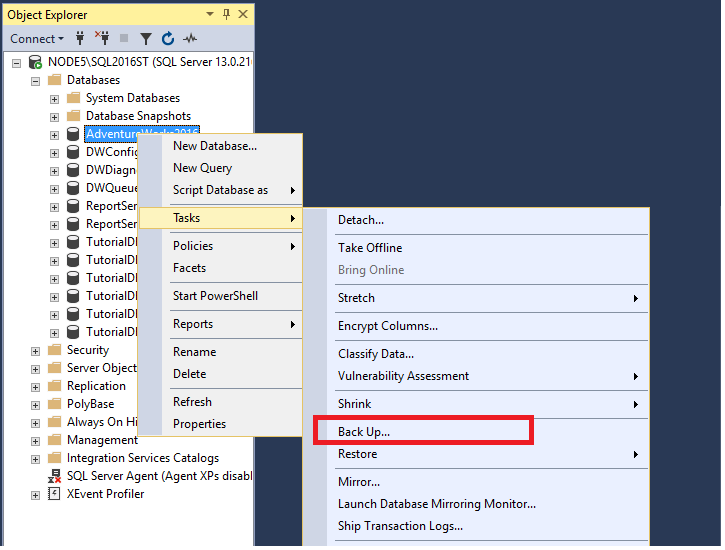
* **on-premises SQL Server databases,**
* **Azure SQL Database,**
* **PostgreSQL,**
* **Azure SQL Data Warehouse,**
* **SQL Server Big Data Clusters,**
* **amongst others**.

1. It's an extensible tool, and you can download and install extensions from third-party developers that connect to other systems, or provide wizards that help to automate many administrative tasks.



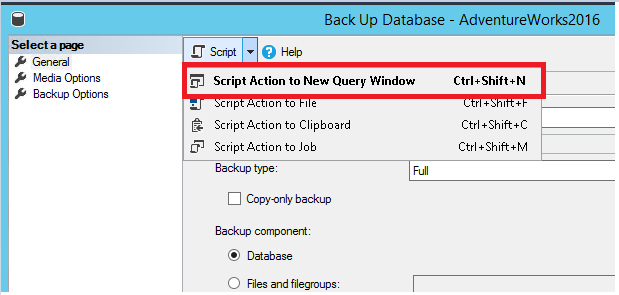
### **What is SQL Server Management Studio?**

1. SQL Server Management Studio {SSMS} provides a graphical interface, enabling you to **query data, perform general database administration tasks, and generate scripts for automating database maintenance and support operations.**
2. The example below shows SQL Server Management Studio being used to back up a database.



A useful feature of SQL Server Management Studio is the ability to generate Transact-SQL scripts for almost all of the functionality that SSMS provides.

This gives the DBA the ability to schedule and automate many common tasks.



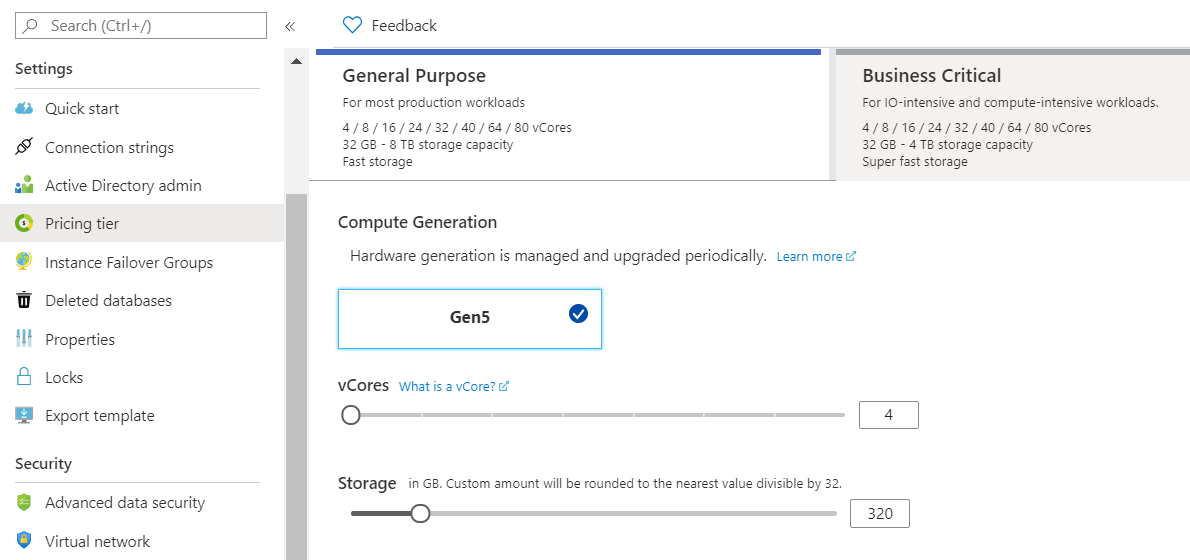
**Note**

**Transact-SQL is a set of programming extensions from Microsoft that adds several features to the Structured Query Language (SQL), including transaction control, exception and error handling, row processing, and declared variables.**

### **Use the Azure portal to manage Azure SQL Database**

Azure SQL database provides database services in Azure. It's similar to SQL Server, except that it runs in the cloud. You can manage Azure SQL database using [Azure portal](https://portal.azure.com/#home).

Typical configuration tasks such as increasing the database size, creating a new database, and deleting an existing database are done using the Azure portal.



You can use the Azure portal to dynamically manage and adjust resources such as the data storage size and the number of cores available for the database processing. These tasks would require the support of a system administrator if you were running the database on-premises.

# Review tasks and tools for data engineering

1. Data engineers are tasked with managing and organizing data, while also monitoring for trends or inconsistencies that will impact business goals.
2. It’s a **highly technical position,** requiring experience and skills in areas like programming, mathematics, and computer science.
3. 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.

1. 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.
2. 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.**
3. In some cases, you may need to interact with a database from the **command line**.
4. 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.
5. As a SQL Server professional, your primary data manipulation tool might be Transact-SQL.
6. **As a data engineer you might use additional technologies, such as**[**Azure Databricks**](https://docs.microsoft.com/en-us/azure/azure-databricks/what-is-azure-databricks)**, and**[**Azure HDInsight**](https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-overview)**to generate and test predictive models. If you're working in the non-relational field, you might use**[**Azure Cosmos DB**](https://docs.microsoft.com/en-us/azure/cosmos-db/introduction)**as your primary data store.**
7. To manipulate and query the data, you might use languages such as **HiveQL, R, or Python.**

# Review tasks and tools for data visualization and reporting

**1.Data analysts** are responsible for understanding what data actually means.

A skilled data analyst will explore the data and use it to determine trends, issues, and gain other insights that might be of benefit to the company.

1. A large part of the data analyst role is concerned with **communication** and **visualization**.
2. Data visualization is key to presenting large amounts of information in ways that are universally understandable or easy to interpret and spot **patterns, trends, and correlations**.
3. These representations include charts, graphs, infographics, and other pictorial diagrams.
4. Data visualization analysts use visualization tools and software to communicate information in these ways, for clients or for their own company.
5. A good data analyst requires experience and skills in reporting tools such as **Microsoft Power BI and SQL Server Reporting Services.**

## **Data Analyst tasks and responsibilities**

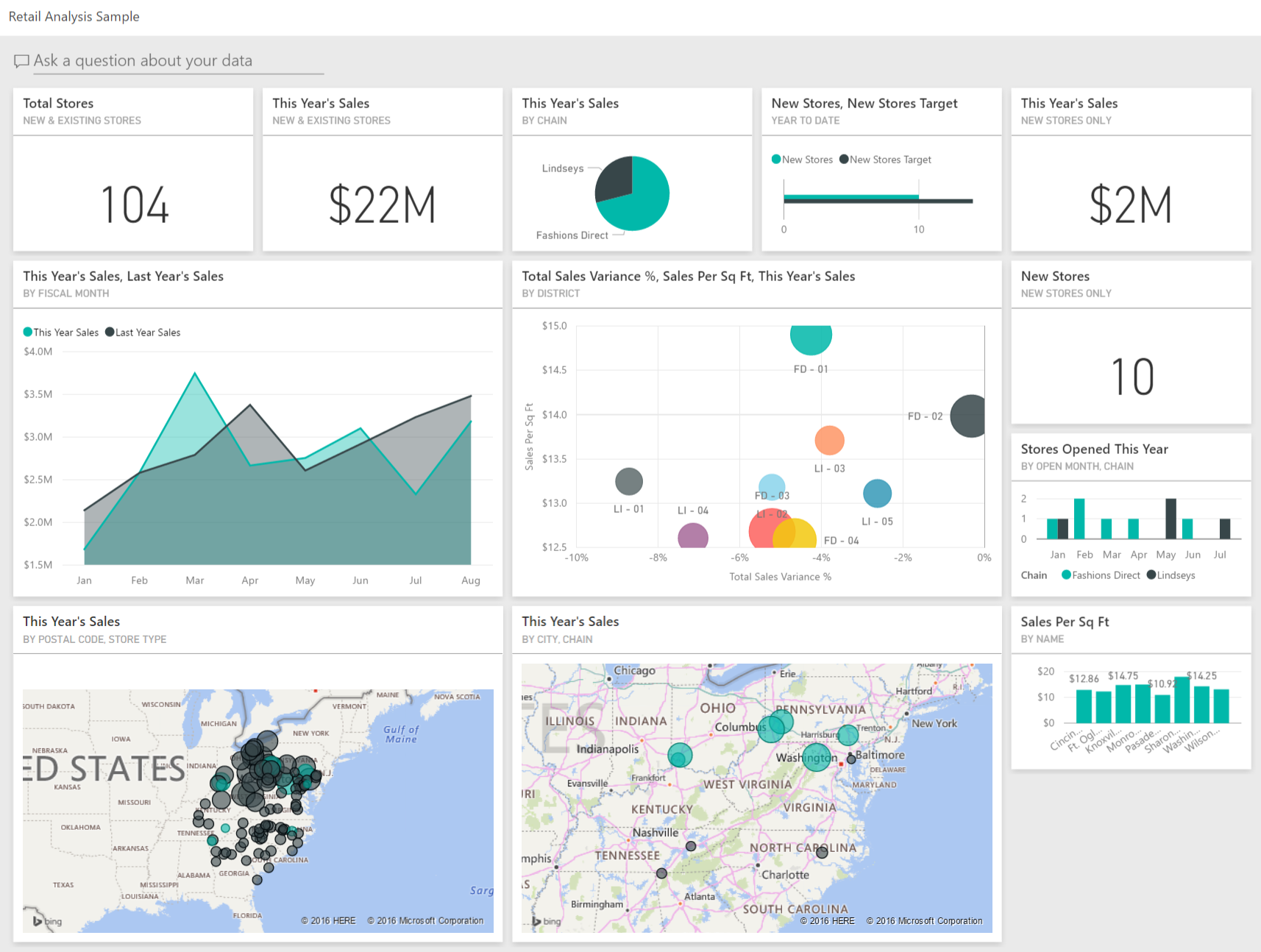
The primary functions of a data analyst usually include the following:

* Making large or complex data more accessible, understandable, and usable.
* Creating charts and graphs, histograms, geographical maps, and other visual models that help to explain the meaning of large volumes of data, and isolate areas of interest.
* Transforming, improving, and integrating data from many sources, depending on the business requirements.
* Combining the data result sets across multiple sources. For example, combining sales data and weather data provides a useful insight into how weather influenced sales of certain products such as ice creams.
* Finding hidden patterns using data.
* Delivering information in a useful and appealing way to users by creating rich graphical dashboards and reports.

## **Common data visualization tools**

* 1. Traditionally, many data analysts used Microsoft Office Apps such as **Microsoft Excel** for creating rich visual reports. Many analysts now use **Microsoft Power BI**, a powerful visualization platform, to create rich, graphical dashboards and reports over data that can vary dynamically.
  2. **Power BI** is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights. Your data might be held somewhere local such as an Excel spreadsheet, or in a collection of cloud-based and on-premises databases, or some other set of data sources. Power BI lets you easily connect to your data sources, discover what's important in that data, and share your findings with others in the organization.

The image below shows an example of a dashboard created using Power BI. In this example, the analyst is using Power BI to examine retail sales data for items sold across multiple stores and districts. The metrics compare this year's performance to last year's for sales, units, gross margin, and variance, as well as new-store analysis.



* [Overview of Azure Databricks](https://docs.microsoft.com/en-us/azure/azure-databricks/what-is-azure-databricks)
* [Overview of Azure HDInsight](https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-overview)
* [Introduction to Azure Cosmos DB](https://docs.microsoft.com/en-us/azure/cosmos-db/introduction)
* [Overview of Power BI](https://docs.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview)
* [SQL Server Technical Documentation](https://docs.microsoft.com/en-us/sql/sql-server/?view=sql-server-ver15)
* [Introduction to Azure Data Factory](https://docs.microsoft.com/en-us/azure/data-factory/introduction)