

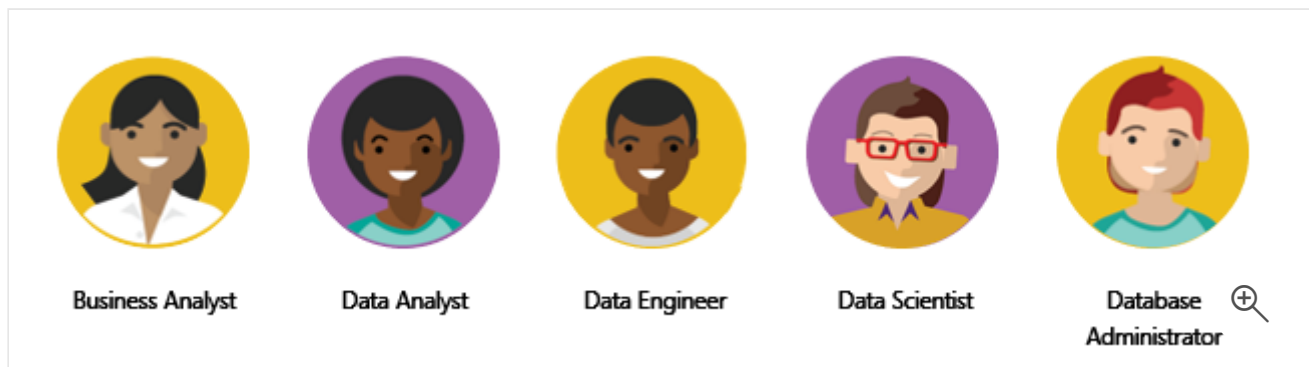
Roles in data

8 minutes

Telling a story with the data is a journey that usually doesn't start with you. The data must come from somewhere. Getting that data into a place that is usable by you takes effort that is likely out of your scope, especially in consideration of the enterprise.

Today's applications and projects can be large and intricate, often involving the use of skills and knowledge from numerous individuals. Each person brings a unique talent and expertise, sharing in the effort of working together and coordinating tasks and responsibilities to see a project through from concept to production.

In the recent past, roles such as business analysts and business intelligence developers were the standard for data processing and understanding. However, excessive expansion of the size and different types of data has caused these roles to evolve into more specialized sets of skills that modernize and streamline the processes of data engineering and analysis.



The following sections highlight these different roles in data and the specific responsibility in the overall spectrum of data discovery and understanding:

- Business analyst
- Data analyst
- Data engineer
- Data scientist
- Database administrator

Business analyst

While some similarities exist between a data analyst and business analyst, the key differentiator between the two roles is what they do with data. A business analyst is closer to the business and is a specialist in interpreting the data that comes from the visualization. Often, the roles of data analyst and business analyst could be the responsibility of a single person.

Data analyst

A data analyst enables businesses to maximize the value of their data assets through visualization and reporting tools such as Microsoft Power BI. Data analysts are responsible for profiling, cleaning, and transforming data. Their responsibilities also include designing and building scalable and effective data models, and enabling and implementing the advanced analytics capabilities into reports for analysis. A data analyst works with the pertinent stakeholders to identify appropriate and necessary data and reporting requirements, and then they are tasked with turning raw data into relevant and meaningful insights.

A data analyst is also responsible for the management of Power BI assets, including reports, dashboards, workspaces, and the underlying datasets that are used in the reports. They are tasked with implementing and configuring proper security procedures, in conjunction with stakeholder requirements, to ensure the safekeeping of all Power BI assets and their data.

Data analysts work with data engineers to determine and locate appropriate data sources that meet stakeholder requirements. Additionally, data analysts work with the data engineer and database administrator to ensure that the analyst has proper access to the needed data sources. The data analyst also works with the data engineer to identify new processes or improve existing processes for collecting data for analysis.

Data engineer

Data engineers provision and set up data platform technologies that are on-premises and in the cloud. They manage and secure the flow of structured and unstructured data from multiple sources. The data platforms that they use can include relational databases, nonrelational databases, data streams, and file stores. Data engineers also ensure that data services securely and seamlessly integrate across data services.

Primary responsibilities of data engineers include the use of on-premises and cloud data services and tools to ingest, egress, and transform data from multiple sources. Data engineers collaborate with business stakeholders to identify and meet data requirements. They design and implement solutions.

While some alignment might exist in the tasks and responsibilities of a data engineer and a database administrator, a data engineer's scope of work goes well beyond looking after a

database and the server where it's hosted and likely doesn't include the overall operational data management.

A data engineer adds tremendous value to business intelligence and data science projects. When the data engineer brings data together, often described as data wrangling, projects move faster because data scientists can focus on their own areas of work.

As a data analyst, you would work closely with a data engineer in making sure that you can access the variety of structured and unstructured data sources because they will support you in optimizing data models, which are typically served from a modern data warehouse or data lake.

Both database administrators and business intelligence professionals can transition to a data engineer role; they need to learn the tools and technology that are used to process large amounts of data.

Data scientist

Data scientists perform advanced analytics to extract value from data. Their work can vary from descriptive analytics to predictive analytics. Descriptive analytics evaluate data through a process known as exploratory data analysis (EDA). Predictive analytics are used in machine learning to apply modeling techniques that can detect anomalies or patterns. These analytics are important parts of forecast models.

Descriptive and predictive analytics are only partial aspects of data scientists' work. Some data scientists might work in the realm of deep learning, performing iterative experiments to solve a complex data problem by using customized algorithms.

Anecdotal evidence suggests that most of the work in a data science project is spent on data wrangling and feature engineering. Data scientists can speed up the experimentation process when data engineers use their skills to successfully wrangle data.

On the surface, it might seem that a data scientist and data analyst are far apart in the work that they do, but this conjecture is untrue. A data scientist looks at data to determine the questions that need answers and will often devise a hypothesis or an experiment and then turn to the data analyst to assist with the data visualization and reporting.

Database administrator

A database administrator implements and manages the operational aspects of cloud-native and hybrid data platform solutions that are built on Microsoft Azure data services and Microsoft SQL Server. A database administrator is responsible for the overall availability and consistent performance and optimizations of the database solutions. They work with

stakeholders to identify and implement the policies, tools, and processes for data backup and recovery plans.

The role of a database administrator is different from the role of a data engineer. A database administrator monitors and manages the overall health of a database and the hardware that it resides on, whereas a data engineer is involved in the process of data wrangling, in other words, ingesting, transforming, validating, and cleaning data to meet business needs and requirements.

The database administrator is also responsible for managing the overall security of the data, granting and restricting user access and privileges to the data as determined by business needs and requirements.

Next unit: Tasks of a data analyst

Continue >