IBM Data Analyst Professional Certificate

# Course 1: Introduction to Data Analytics

## Module 1

## **Modern Data Ecosystem and the Role of Data Analytics**

### **Modern Data Ecosystem**

The video discusses the modern data ecosystem and its components. It starts by highlighting the constant growth of data due to increasing processing speeds, new tools, and the addition of data creators and consumers worldwide. The data ecosystem consists of interconnected entities, including various data sources such as structured and unstructured datasets from text, images, videos, click streams, social media platforms, IoT devices, and more.

+ The first step in working with data is acquiring it from the original sources and pulling it into a data repository. Challenges at this stage include working with different data formats, sources, and interfaces, as well as ensuring data reliability, security, and integrity.

+ Once the raw data is in a common place, it needs to be organized, cleaned up, and optimized for access by end users. This involves conforming to data compliances and standards enforced in the organization, such as guidelines for personal data storage and use.

+ Business stakeholders, including applications, programmers, analysts, and data scientists, then pull the data from the enterprise data repository. Challenges at this stage involve providing interfaces, APIs, and applications that meet the specific needs of different users. For example, data analysts may require access to raw data, while business stakeholders may need reports and dashboards.

The influence of emerging technologies such as cloud computing, machine learning, and big data in shaping the data ecosystem. Cloud technologies provide enterprises with limitless storage, high-performance computing, open-source technologies, and machine learning capabilities. Data scientists can create predictive models by training machine learning algorithms on large and diverse datasets, leading to new tools, techniques, knowledge, and insights.

### **Key Players in the Data Ecosystem**

A black and white text on a black background

Description automatically generated

We learn about the roles of data professionals in helping organizations harness the power of data and turn it into actionable insights. Here is a summary of the key roles:

+ Data Engineer: Data engineers develop and maintain data architectures, making data available for business operations and analysis. They extract, integrate, and organize data from various sources, clean and transform it, and design data repositories. Data engineers need programming skills, knowledge of systems and technology architectures, and an understanding of relational and non-relational databases.

+ Data Analyst: Data analysts translate data and numbers into plain language for decision-making. They inspect and clean data, identify correlations and patterns, and apply statistical methods to analyze and mine data. Data analysts require knowledge of spreadsheets, writing queries, and using statistical tools to create charts and dashboards. They also need programming skills and strong analytical and storytelling abilities.

+ Data Scientist: Data scientists analyze data for actionable insights and build predictive models using machine learning or deep learning techniques. They answer questions like how many new social media followers to expect next month or what percentage of customers may be lost to competition. Data scientists need knowledge of mathematics, statistics, programming languages, databases, and domain expertise.

+ Business Analyst and BI Analyst: Business analysts and business intelligence analysts leverage the work of data analysts and data scientists to drive decisions that benefit their business. They organize and monitor data on different business functions, extract insights, and improve business performance. Business analysts focus on internal implications, while BI analysts consider market forces and external influences.

### **Defining Data Analysis**

Data analysis is the process of gathering, cleaning, analyzing, and mining data to find patterns and correlations.

There are four primary types of data analysis:

+ Descriptive Analytics: Summarizes past data to provide insights into past events.

+ Diagnostic Analytics: Digs deeper to find the cause of outcomes.

+ Predictive Analytics: Uses historical data and trends to predict future outcomes.

+ Prescriptive Analytics: Recommends a course of action based on past decisions and events.

Key steps in the data analysis process include:

+ Understanding the problem and desired result: Before starting the analysis, it's important to clearly define the problem that needs to be solved and the desired outcome. This helps set the direction for the analysis.

+ Setting a clear metric: In this step, you decide what will be measured and how it will be measured. For example, you might measure the number of products sold in a specific region during a particular period.

+ Gathering data: Once you know what you're going to measure and how, you identify the data you need and the sources from which you can obtain that data. This could involve collecting data from databases, surveys, or other sources.

+ Cleaning data: After gathering the data, it's important to clean it to ensure accuracy. This involves fixing any quality issues in the data that could affect the analysis. For example, you might remove missing or incomplete values and handle outliers.

+ Analyzing and mining data: Once the data is clean, you can start analyzing and mining it. This involves extracting insights from the data by looking at it from different perspectives, manipulating it to identify trends, correlations, patterns, and variations.

+ Interpreting results: After analyzing the data, it's time to interpret the results. This may involve conducting further research or iterating on the analysis. It's important to evaluate the defendability of the analysis against objections and consider any limitations or circumstances under which the analysis may not hold true.

+ Presenting findings: The ultimate goal of data analysis is to impact decision-making. To achieve this, it's important to effectively communicate and present the findings. This can be done through reports, dashboards, charts, graphs, maps, case studies, or other visualizations.

### **Viewpoints: What is Data Analytics?**

Data professionals define data analytics as the process of collecting and analyzing information to confirm various hypotheses. They also emphasize that data analytics involves storytelling with data, using it to convey the state of the world concisely. Data analysis is not an abstract concept; it is something we do naturally every day. The process of data analytics starts with defining a problem and creating a hypothesis. Then, data is collected, cleaned, analyzed, and presented to key stakeholders. Data analytics involves reviewing any sets of data to understand what is going on. It can be used to predict the future of a company by analyzing financial data. Data analytics is the process of taking information gathered from a relevant population and using it to make decisions about products, services, or content.

### **Summary and Highlights**

In this lesson, you have learned the following information:

A modern data ecosystem includes a network of interconnected and continually evolving entities that include:

+ Data that is available in a host of different formats, structure, and sources.

+ Enterprise Data Environment in which raw data is staged so it can be organized, cleaned, and optimized for use by end-users.

+ End-users such as business stakeholders, analysts, and programmers who consume data for various purposes.

Emerging technologies such as Cloud Computing, Machine Learning, and Big Data, are continually reshaping the data ecosystem and the possibilities it offers. Data Engineers, Data Analysts, Data Scientists, Business Analysts, and Business Intelligence Analysts, all play a vital role in the ecosystem for deriving insights and business results from data.

Based on the goals and outcomes that need to be achieved, there are four primary types of Data Analysis:

+ Descriptive Analytics, that helps decode “What happened.”

+ Diagnostic Analytics, that helps us understand “Why it happened.”

+ Predictive Analytics, that analyzes historical data and trends to suggest “What will happen next.”

+ Prescriptive Analytics, that prescribes “What should be done next.”

The Data Analysis process involves:

+ Developing an understanding of the problem and the desired outcome.

+ Setting a clear metric for evaluating outcomes.

+ Gathering, cleaning, analyzing, and mining data to interpret results.

+ Communicating the findings in ways that impact decision-making.

## **The Data Analyst Role**

## **Responsibilities of a Data Analyst**

Responsibilities of a Data Analyst:

+ Acquiring data from primary and secondary sources

+ Creating queries to extract data from databases

+ Filtering, cleaning, and reorganizing data

+ Using statistical tools to interpret data sets

+ Identifying patterns and correlations in data

+ Analyzing patterns in complex data sets

+ Preparing reports and charts to communicate trends and patterns

+ Creating documentation to define and demonstrate the data analysis process

Technical Skills for a Data Analyst:

+ Expertise in using spreadsheets like Excel or Google Sheets

+ Proficiency in statistical analysis and visualization tools like IBM Cognos, Tableau, etc.

+ Knowledge of programming languages like R, Python, SQL, etc.

+ Familiarity with databases and data repositories

+ Understanding of Big Data processing tools like Hadoop, Hive, Spark, etc.

Functional Skills for a Data Analyst:

+ Proficiency in Statistics

+ Analytical skills for research and interpretation

+ Problem-solving skills

+ Probing skills for understanding problem statements

+ Data visualization skills

+ Project management skills

Soft Skills for a Data Analyst:

+ Collaboration with business and cross-functional teams

+ Effective communication and presentation skills

+ Curiosity and the ability to challenge assumptions

+ Intuition based on pattern recognition and past experiences

### **Viewpoints: Qualities and Skills to be a Data Analyst**

Qualities of a data analyst:

+ Curiosity: A data analyst should be naturally curious and willing to find answers even when there isn't a specific question. They should enjoy researching and exploring new areas.

+ Attention to detail: A data analyst should be detail-oriented, love numbers and information, and be willing to dive deeper into data. They should not take things at face value and should compare and analyze information to identify any discrepancies or patterns.

+ Adaptability and Continuous learning: Technology in the field of data analytics is constantly evolving. Data analysts need to be able to learn new skills and software as the industry changes.

+ End-to-end skill set: Data analysts need to know how to identify the problem they are trying to solve, extract and structure data using SQL, clean and manipulate data, mine insights from it, and present those insights using visualizations and dashboards.

### **Summary and Highlights**

In this lesson, you have learned the following information:

The role of a Data Analyst spans across:

+ Acquiring data that best serves the use case.

+ Preparing and analyzing data to understand what it represents.

+ Interpreting and effectively communicating the message to stakeholders who need to act on the findings.

+ Ensuring that the process is documented for future reference and repeatability.

In order to play this role successfully, Data Analysts need a mix of technical, functional, and soft skills.

+ Technical Skills include varying levels of proficiency in using spreadsheets, statistical tools, visualization tools, programming and querying languages, and the ability to work with different types of data repositories and big data platforms.

+ An understanding of Statistics, Analytical techniques, problem-solving, the ability to probe a situation from multiple perspectives, data visualization, and project management skills – all of which come under Functional Skills a Data Analyst needs in order to play an effective role.

+ Soft Skills include the ability to work collaboratively, communicate effectively, tell a compelling story with data, and garner support and buy-in from stakeholders. Curiosity to explore different pathways and intuition that helps to give a sense of the future based on past experiences are also essential skills for being a good Data Analyst.