# Data Engineering

## Introduction to Data Engineering

### Week 1: What is Data Engineering?

In this module, you will learn about the different entities that come together to form a modern data ecosystem and the role Data Engineers, Data Scientists, Data Analysts, Business Analysts, and Business Intelligence Analysts play in this ecosystem. You will learn what data engineering is and the key tasks in a data engineering lifecycle. You will also gain an understanding of the responsibilities of a data engineer, the skillsets they need in order to be successful, and what a typical day in the life of a data engineer looks like. At the end of the module, you will be guided to create a Lite account on IBM Cloud

#### Learning Objectives

* Recall the different entities that form a modern data ecosystem.
* Describe and differentiate between the role and responsibilities of Data Engineers, Data Scientists, Data Analysts, Business Analysts, and Business Intelligence Analysts.
* Explain what Data Engineering is.
* List the tasks that need to be performed in a typical data engineering lifecycle.
* Recall the essential skills and qualities for data engineering as identified by data professionals.
* Summarize how Data Engineering has evolved over the past few decades.
* Discuss the responsibilities and skillsets of a Data Engineer.
* Recall the various ways that data professionals define data engineering and differentiate it from data analysis and data science.
* Describe what a day in the life of a Data Engineer looks like.
* Create an IBM Cloud account.

### Modern Data Ecosystem and role of Data Engineering

#### The value we derive from data depends on:

* Accuracy of data
* Accessibility of data when we need it

#### Data Professional excited about engineering

* Statistician
* Data Scientist
* Data Analyst
* BI Analyst

#### Working professional with a technical role

* IT support
* Software Tester
* Programmer

This course will introduce you to Core Concepts, Ecosystem, and Life Cycle of Data Engineering.

#### What you will learn

* Data
* Data Repositories
* Data Pipelines
* Data Integration Platforms
* Big Data
* Data Platforms
* Data Stores
* ETL Process
* ELT Process
* Data Security
* Data Privacy
* Governance and Compliance

Raw data needs to get organized, cleaned up, optimized for access, and conform to compliances and standards enforced in the organization

#### Steps in Data

1. **Pull a copy of data from the original sources into a data repository**

Challenges:

* Reliability
* Security
* Integrity

1. **Cleaning**

Challenges:

* Data management
* Repositories that provide high availability, flexibility, accessibility, and security

#### Emerging technologies shaping the modern data ecosystem

* Cloud Technologies
* Every enterprise today has access to limitless storage, high-performance computing, open source technologies, machine learning technologies, and the latest tools and libraries
* Machine Learning
* Data Scientists are creating predictive models by training machine learning algorithms on past data
* Big Data
* Today we’re dealing with datasets that are so massive and so varied that traditional tools and analysis methods are no longer adequate, paving the way for new tools and techniques and also new knowledge and insights

#### Overview

Organizations that are using data to uncover opportunities and are applying that knowledge to differentiate themselves from their competition

* Identifying patterns in financial transactions to detect fraud
* Using recommendation engines to drive conversion
* Mining social media posts for customer voice
* Analyzing customer behavior for personalizing offers

#### Data Professionals

* **Data Engineers**
  + Responsibilities
* Extract, integrate, and organized data from disparate sources
* Clean, transform and prepare data
* Design, store, and manage data in data repositories
  + Skills
* Good knowledge of programming
* Sound knowledge of systems and technology architectures
* In-depth understanding of relational databases and non-relational data stores
* **Data Analyst**

Data Analyst translates data and numbers into plain language, so organizations can make decisions

* + Responsibilities
* Inspect, and clean data for deriving insights
* Identify correlations, find patterns, and apply statistical methods to analyze and mine data
* Visualize data to interpret and present the findings of data analysis.
  + Skills
* Good knowledge of spreadsheets, writing queries and using statistical tools to create charts and dashboards
* Programming skills
* Strong analytical and story-telling skills
  + Answer questions
* "Are the users' search experience generally good or bad with the search functionality on our site?"
* "What is the popular perception of people regarding our rebranding initiatives?"
* "Is there a correlation between sales of one product and another"
* **Data Scientists**
  + Responsibilities
* Analyze data for actionable insights
* Create predictive models using Machine Learning and Deep Learning
  + Skills
* Knowledge of Mathematics and Statistics
* Understanding of programming languages, databases, and building data models
* They also need to have domain knowledge.
  + Answer questions
* “How many new social media followers am I likely to get next month?”
* “What percentage of my customers am I likely to lose to competition in the next quarter”
* “Is this financial transaction unusual for this customer?”
* **Business Analysts and BI Analysts**
  + Business Analysts
* Leverage the work of Data Analysts and Data Scientists to look at possible implications for their business and the actions they need to take or recommend
  + BI Analysts
* Focus is on the market forces and external influences that shape their business
* Organizing and monitoring data on different business functions
* Exploring data to extract insights and actionable that improve business performance

#### Summarize

* Data engineering converts raw data into usable data
* Data Analytics use this data to generate insights
* Data Scientists use Data Analytics and Data Engineering to predict the future using data from the past
* Business Analysts and Business Intelligence Analyst use these insights and predictions to drive decisions that benefit and grow their business