**1.What is Data Engineering?**

**46 min** of videos left

**20 min** of readings left

**2** graded assessments left

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

**Modern Data Ecosystem**

To quote a Forbes 2020 report on data in the coming decade,

 “The constant increase in data processing speeds and bandwidth, the nonstop invention of new tools for creating, sharing, and consuming data, and the steady addition of new data creators and consumers around the world, ensure that data growth continues unabated. Data begets more data in a constant virtuous cycle.”

* Modern data ecosystems are independent, inter connected and continually evolving.
* Data tegrated from disparate sources

**Technical Skills**

OS : Unix |. Windows |. Windows Administrative Tools | System Utilities & Commands

Infrastructure Components : VM | Networking | Application Services | Cloud based Services

RDBMS : IBM DB2 | MySQL | Oracle Database | PostgreSQL

NoSQL : Redis | MongoDB | Cassandra | Neo4J

Databases & Data Warehouses : Oracle Exadata | IBM DB2 Warehouse on Cloud | IBM Netezza Performance Server | AWS Redshift

Data Pipelines : Apache Beam, AirFlow, DataFlow

ETL Tools : IBM Infosphere, AWS, improvado

Languages: SQL,Python,R,Java,Shell Scripting

BigData Processing tools : Hadoop, Hive, Apache Spark