

### Module 1 – Data Engineering Concepts

- Data Engineering Defined
  - What is engineering?
  - What is software engineering?
  - What is database engineering
  - What is data engineering?
- The Data Engineer Role
  - Data engineering deliverables
  - Data engineer vs. software engineer
  - Data engineer vs. data scientist
  - Qualities of a good data engineer
- Why Data Engineering?
  - The data quake
  - The data in data science
  - Big data complexities
  - Data pipeline complexities
- Data Engineers and Teams
  - The data engineer as a collaborator
  - Kinds of data engineering teams

### Module 2 – Data Engineering Knowledge and Skills

- Database Technologies
  - Relational databases
  - NoSQL databases
  - Delimited files
  - Tagged data files
- Data Modeling
  - Relational modeling
  - Dimensional modeling
  - Big data modeling
- Microservices
  - Microservices architecture
  - Applied microservices
- Distributed Systems
  - Distributed systems architecture
  - Distributed systems challenges

### Module 3 – Data Engineering Technologies

- Hadoop

- Hadoop ecosystem
- Hadoop distribution
- Hadoop data ingestion
- Hadoop data processing
- Hadoop management
- SQL on Hadoop
- Hadoop distributions
- Hadoop alternatives
- Data Lake Management
- Data Stream Processing
- Data Integration
  - iPaaS
  - Data warehouse automation
  - ETL
- Data Virtualization
- Data Preparation
- Programming Languages

#### Module 4 – Data Engineering Processes and Projects

- Database Processes
  - Designing databases
  - Building databases
  - Database refactoring
  - Cloud migration
- Data Pipeline Processes
- Data Services Processes
- Kinds of Data Engineering Projects
  - Variety of goals
  - One-time vs. enduring
  - Variety of lifecycles
  - Waterfall
  - Iterative
  - Agile
  - DevOps

#### Module 5 – Becoming a Data Engineer

- Skills Inventory
- Gap Analysis
- Next Steps