Data Engineer Training

Arun Bonam

Prerequisites : Basic knowledge in Python and Sql

Module 1 - Overview - 4 hrs
==========

1. Data Engineering Role - Introduction

2. Python in Realtime

3. SQL for Interviews

5. SQL and NOSQL Paradigms

6. ETL Introduction

7. Data Formats

Module 2 - Bigdata Fundamentals - 9 hrs
==========

1. Introduction to Clusters - Distributed storage and processing systems

2. Hadoop core components and architecture

4. Advantages and Limitations of Hadoop

5. Onprem vs Cloud

6. Popular data storage technologies - Onprem

7. Popular distributive data processing technologies - Onprem

8. Batch vs Realtime data processing

9. Data pipeline Orchestration

Module 3 - Apache Spark deep-dive and introduction to other equivalents - 6 hrs
==========

1. Apache Spark Fundamentals
2. Apache Spark UseCases
3. Developing Apache Spark Jobs
4. Optimising Spark Jobs
5. Deploying Spark Jobs Onprem and Cloud(GCP)
6. Introduction to other bigdata processing systems.

Module 4 - Apache Kafka and introduction to other equivalents - 6 hrs
==========

1. Realtime messaging systems -Introduction
2. Kafka Cluster Components
3. Kafka Topic creation
4. Kafka Producer Consumer Mechanism
5. Kafka integration with other services.
6. Introduction to kafka Equivalents -onprem and cloud

Module 5- Cloud Data Engineering - 7 hrs
==========

1. Introduction to different clouds
2. Data Storage services offered by cloud providers
3. Data Processing services offered by cloud providers
4. Migrating workloads from onprem to Cloud
5. Introduction to cloud services in Google cloud and AWS for Data Engineers
6. Data pipeline Orchestration tools in Cloud
7. Building data-pipelines in cloud - Realtime case studies

Module 6 - Datawarehousing and Data Modelling - 7 hrs
==========

1. OLTP VS OLAP systems

2. Datawarehouse vs Datalake vs Database

3. Fact tables vs Dimension tables

4. Slowly Changing Dimensions (SCD)

5. Data Modelling techniques for Data warehouses

6. Datawarehousing in Cloud.

7. Case studies and realtime examples in GCP Bigquery .

==========
==========

Module 7 - DataEnigeers in Collaborative Environment - 6 hrs
==========

1. Day to Day activities of Data Engineer in organisations.

2. Realtime issues and how Data Engineers Solve them.

3. How Data Engineers collaborate with each other with other departments.

4. Github and Code reviews

5. CI/CD pipeline building

6 .Infrastructure tools that Data Engineer should have hands-on .

Module 8 - Lets Build and Achieve the Goal
==========

1. Project 1 : Building Batch Data Pipelines in GCP

2. Project 2 : Building Realtime Data Pipelines in GCP

3. Project3 : Building Datawarehouse in GCP

4 .Interview Prep and Resume Building

==========
==========
==========