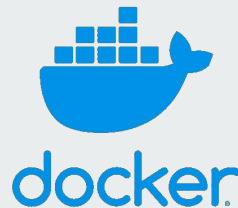




Cloud Data Engineering

Technologies:





Note:

- All the topics will be covered in depth as relevant to real time job.
- From Interview point of view, the interview questions will be shared during the training itself. There will not be a separate session for sharing the interview questions.
- During the completion of each module, Industry standard relevant books will be shared in the PDF format for your reference as a compliment.



Modules

The Core topics will be distributed into various modules. Learning in this order will be efficient

Module 1

Python, SQL, Hadoop Ecosystem

Module 2

Apache Spark, Databricks

Module 3

Delta Lake, Databricks
Optimization Techniques, Unit
Testing

Module 4

Apache Hive Theory Concepts, Kafka
Basics

Module 5

Apache Kafka Integration With
Apache Spark and Hands on Practice
to learn Structured Streaming

Module 6

Data Engineering in AWS Cloud



Program Description

Duration : 3 Months Training (Daily 2 Hours)

Interview Preparation : 15 Days

Project : A Data Pipeline will be built on AWS Cloud

Software Requirements : Minimum Intel Core i3 with 8GB Ram which supports Virtualization (We might be using Docker Environment to avoid installing softwares in local) recommended approach as per latest industry standard.

If Docker is not supported, we will use the Databricks Community Edition

Cloud Environment Requirements: Help will be provided in setting up the cloud environment on your own cost, which may need your Card Payment support

Cloud Environments: Databricks Community Edition for Initial Phase of learning and the premium trial edition for advanced level of topics like performance tuning, delta file handling etc..

AWS Cloud environment and Databricks cost should be bared by the candidate, getting hands on experience highly helps in tackling the interview and during real time project works