

Azure Data Engineering Curriculum

TOPICS	Hours
Introduction to Cloud Computing - Azure	1
Types of cloud, Cloud services	
Azure Portal Explore – Various Services	
Azure Storage Services	
Azure Accounts – Subscription, Resources, Management	
Blob Storage , Azure SQL server – Introduction, Evolution of Storages, and creation in Portal	
Azure Data Lake vs Blob	
Data Processing – Data Engineer	2
Stages of Data Processing	
Azure Data Factory – Introduction, Why ADF?	
ADF Architecture	
Data Factory Components	
Working of ADF	
ADF Creation	
Exploring ADF	3
Integration RunTime	
DataSets, LinkedServices	
Triggers	
ARM Templates	
Linked Services - Authorization	
Creation of Linked Services & Datasets	
Brief introduction to CI/CD	
Pipelines & Activities	4
Parameters & Variables	
Data Movement Activities	
Control Flow Activities	
Creation of Pipelines	
Scenario – Copy /Move Data by building pipelines [4 Possible Scenarios]	
Move Files from ADL's to Azure SQL Server	
Data Flows - Introduction	3
Mapping Data Flows	
Azure Functions	
Working with Azure Functions (different pattern)	
Invoke Azure Function through ADF Pipeline	

Data Bricks – Introduction	1
ADF Mapping Data Flows VS Azure Data Bricks	
Connecting ADF & ADB	
Why Azure Data Bricks?	
Understanding Spark – Brief Introduction	

Spark Architecture	2
How Spark Works?	
Exploring ADB Workspace	
Clusters, Cluster Pool, Data Bricks Runtime	
Cluster Creation	
Data Bricks File System (DBFS)	
Note book Creation & Manipulation	
Mounting ADL's in ADB	
Data Bricks utilities	1
RDD, DataFrame, Datasets	
Lazy Evaluation Spark	
What is a DataFrame? Creating 1 st DataFrame	
Introduction to Pyspark	
Pyspark API Functions – Transformation	
Delta Lake- Introduction	2
Slowly Changing Dimensions , Types	
Fact & Dimensions – Brief Introduction	
Understanding Various File Formats	
Parquet & Avro Format	
Shuffle Operations, Partitioning	1
Narrow & Wide Transformation	
Transformation with frequently used Data Frame API Methods	
Difference between Pyspark & Pandas DF	
Pyspark to Pandas & Pandas to Pyspark	
User Defined Functions	
Performance Optimization	1
Analytic Functions in Pyspark	
Slowly Changing Dimensions	
Scheduling of Notebooks	

Azure Devops - Introduction	2
Exploring GIT	
Creating Feature Branch	
Merging with Master	
CI/CD with ADF & ADB	
Azure Synapse	1
Work Flows From END to END	
Layers of Data Handling	
Project [End to End] – Supply Chain	2
	26 Hours