

# Azure Data Bricks

• Real-Time Working • Detailed Subject • Interview Suitable

-  Mr.Pavan [10+ Years of IT Experience]
-  30 hrs (Normal Track) , Daily 1 hour
-  Normal / Fast Track / 1 on 1 / Customized
-  Rs. 15,000
-  Multiple timings available



## INSTITUTE PROVIDES

- \* Two **Near Real Time** Projects
- \* In-Depth **Theory & Practical** Material
- \* Complete **Practical Oriented** Training
- \* Mock Interviews
- \* Case Studies / User Stories / Sprints
- \* Daily **Live Class Videos**
- \* **FAQ Tests** in Whatsapp
- \* **Resume Preparation** Sessions

## COMPONENTS

|                                           |                                                              |
|-------------------------------------------|--------------------------------------------------------------|
| Introduction to Big Data                  | Databricks Integration                                       |
| Introduction to Hadoop                    | Databricks – Streaming API                                   |
| Introduction to Spark                     | Databricks – Lakehouse (Delta Lake)                          |
| Introduction To Databricks                | Workflows in Databricks                                      |
| Working with Databricks Filesystem - DBFS | Azure DevOps – Repos                                         |
| Databricks -Spark Core                    | SDLC and Agile methodology                                   |
| Databricks - Spark-SQL- Data Frames       | End to End Data Migration Project from On Premises to Cloud. |
| Databricks- Handle multiple file formats  |                                                              |
| Databricks utilities                      | Interview Questions                                          |
| Databricks Cluster Management             | Mock Interviews                                              |
| Databricks – Batch Processing             |                                                              |
| Introduction to Azure                     |                                                              |

## INTRODUCTION TO BIG DATA

- ▶ What is Data?
- ▶ What is Database?
- ▶ What is Big Data?
- ▶ What are the challenges of Big Data?
- ▶ Why Traditional Databases Doesn't handle Bigdata

## INTRODUCTION TO HADOOP

- ▶ What is Hadoop?
- ▶ How Hadoop overcomes bigdata challenges
- ▶ Hadoop Architecture
- ▶ Hadoop Daemons
- ▶ HDFS
- ▶ YARN
- ▶ MapReduce

## INTRODUCTION TO SPARK

- ▶ Spark Architecture
- ▶ Spark internals
- ▶ Spark RDD
- ▶ Spark Data Frame
- ▶ Spark Streaming

## INTRODUCTION TO DATABRICKS

- ▶ What is Databricks?
- ▶ Databricks Architecture
- ▶ Working in Databricks workspace
- ▶ Working with Databricks notebook

## WORKING WITH DEATRICKS FILESYSTEM - DBFS

- ▶ What is DBFS?
- ▶ DBFS commands - mkdirs , cp , mv , head, put, rm , rmdir
- ▶ How to handle multiple files in DBFS
- ▶ How to process the files in DBFS
- ▶ How to archive the files in DBFS

## DATABRICKS -SPARCK CORE

- ▶ RDD Programming
- ▶ Operations on RDD
- ▶ Transformations- Narrow

->

- ▶ Transformations -Wide
- ▶ Actions
- ▶ Loading Data and Saving Data
- ▶ Key Value Pair RDD
- ▶ Broadcast variables

#### DATABRICKS - SPARK-SQL- DATA FRAMES

- ▶ Creating Data Frames
- ▶ Data Frames internal execution
- ▶ Transformations using Data Frame API
- ▶ Actions using Data Frame API
- ▶ User-defined functions in Spark SQL

#### DATABRICKS- HANDLE MULTIPLE FILE FORMATS

- ▶ CSV Data
- ▶ JSON Data
- ▶ parquet files
- ▶ Excel files
- ▶ ORC file format

#### DATABRICKS UTILITIES

- ▶ credentials utility
- ▶ Filesystem utility
- ▶ Notebook utility
- ▶ secrets utility
- ▶ widgets utility

#### DATABRICKS CLUSTER MANAGEMENT

- ▶ Creating and configuring clusters
- ▶ Managing Clusters
- ▶ Displaying clusters
- ▶ Starting a cluster
- ▶ Terminating a cluster
- ▶ Delete a cluster
- ▶ Cluster Information
- ▶ Types of Clusters
- ▶ All-purpose clusters
- ▶ Job cluster
- ▶ Clusters Mode
- ▶ Standard
- ▶ High Concurrency
- ▶ Autoscaling
- ▶ Databricks runtime versions



### DATABRICKS – BATCH PROCESSING

- ▶ Historical Data load
- ▶ Incremental Data load
- ▶ Date Transformations
- ▶ Aggregations
- ▶ Join Operations
- ▶ Window Functions
- ▶ Union Operations

### INTRODUCTION TO AZURE

- ▶ Azure Portal Walkthrough
- ▶ What is Subscription?
- ▶ What is a Resource Group?
- ▶ What is a Resource?
- ▶ Overview of Azure Resources / Services
- ▶ Azure Data bricks
- ▶ BLOB Storage, Data Lake Storage Gen2
- ▶ Azure SQL Server, SQL Database
- ▶ Key Vault

### DATABRICKS INTEGRATION WITH

- ▶ Blob storage
- ▶ Azure Data Lake storage gen2
- ▶ Azure SQL Database
- ▶ Synapse
- ▶ Azure Key vault

### DATABRICKS – STREAMING API

- ▶ What is streaming?
- ▶ Process streaming using Pyspark API
- ▶ Handling bad records
- ▶ Stream data into Gen2lake
- ▶ Load the data into Tables

### DATABRICKS – LAKEHOUSE (DELTA LAKE)

- ▶ Difference between Data Lake and Delta Lake
- ▶ Introduction to Delta Lake
- ▶ Features of Delta Lake
- ▶ How to create delta table
- ▶ How to DML operations in Delta Table
- ▶ Merge statements
- ▶ Handling SCD Type1 and Type2

->

- ▶ Handling Data Deduplication in delta tables
- ▶ Handling streaming Data in Delta Lake
- ▶ **Delta Lake: Medallion Architecture**
- ▶ Implement the Bronze Layer (Raw Data)
- ▶ Implement the Silver Layer (Cleansed & Transformed Data)
- ▶ Implement the Gold Layer (Curated, Business-Ready Data)

## WORKFLOWS IN DATABRICKS

- ▶ Introduction to workflows
- ▶ Create, run and manage Databricks' jobs
- ▶ Schedule Databricks jobs
- ▶ Monitor Databricks Jobs

## AZURE DEVOPS – REPOS

- ▶ What are DevOps Repos
- ▶ Integrate data bricks notebooks with Repos
- ▶ Commit, Sync notebooks to and from Repos

## SDLC AND AGILE METHODOLOGY

## END TO END DATA MIGRATION PROJECT FROM ON PREMISES TO CLOUD

## INTERVIEW QUESTIONS

## MOCK INTERVIEWS

# A Big Thank you

## COURSES WE OFFER

POWER BI

DATA ANALYST/ ANALYTICS

MICROSOFT FABRIC

SQL SERVER

ADE / AZURE BI

ADF: AZURE DATA FACTORY

ADB: AZURE DATA BRICKS

MSBI

POWER APPS

POWER AUTOMATE

INFORMATICA

EXCEL

PYTHON

### Power BI

**FIRST DEMO**

Taught by

MR. VINAY

1:02:49

Get the Best Power BI Training in  
Hyderabad | Free Demo #1

To watch  
our Latest Demos

**SUBSCRIBE TO OUR YOUTUBE**

**@ Vinay Tech House**

**For Regular Updates on Demos (Free)**

**Follow us on Instagram / Facebook**

**@VINAYTECHHOUSE**

**For more information, Call us on: +91 9859 831 831**

**504, Nilgiri Block, Adithya Enclave, Beside Ameerpet Metro, Hyderabad**