

3 dni po 8h

I. Databricks Lakehouse Fundamentals

1. What is the Databricks Lakehouse Platform
 - a. Describe what the Databricks Lakehouse Platform is.
 - b. Explain the origin of the Lakehouse data management paradigm.
 - c. Outline fundamental challenges related to managing and using data.
 - d. Describe security features of the Databricks Lakehouse Platform.
 - e. Give examples of organizations that have benefited from using the Databricks Lakehouse Platform

2. What is Databricks SQL
 - a. Summarize fundamental concepts for using Databricks SQL effectively
 - b. Identify tools and features in Databricks SQL for querying data and sharing insights
 - c. Explain how Databricks SQL supports data analysis workflows that allow users to extract and share business insights

3. What is Databricks Machine Learning
 - a. Describe the basic overview of Databricks Machine Learning.
 - b. Identify how using Databricks Machine Learning benefits data science and machine learning teams
 - c. Summarize the fundamental components and functionalities of Databricks Machine Learning
 - d. Exemplify successful use cases of Databricks Machine Learning by real Databricks customers

4. What is Databricks Data Science and Data Engineering Workspace
 - a. Describe the basic overview of Databricks Data Science and Engineering Workspace
 - b. Identify assets provided by the workspace
 - c. Describe a simple development workflow that queries and aggregates data

II. Data Engineering with Databricks

1. Databricks Workspaces And Services
 - a. Databricks Architecture and Services
 - b. Data Science and Engineering Workspace
 - c. Create and Manage Interactive Clusters
 - d. Notebook Basics

- e. Git Versioning with Databricks Repos
 - f. Using Databricks Repos
 - g. Getting Started with the Databricks Platform
2. Delta Lakehouse
 - a. What is Delta Lake
 - b. Managing Delta Tables
 - c. Manipulating Tables with Delta Lake
 - d. Advanced Delta
 3. Relational entities on Databricks
 - a. Databases and Views
 - b. Views and CTEs
 4. ETL with Spark Sql
 - a. Query Files Directly
 - b. Providing Options
 - c. Creating Delta Tables
 - Writing to Tables
 - Cleaning Data
 - Advanced SQL Transformations
 - UDF

III. Data Analysis with Databricks Sql

1. Getting Started with Databricks Sql
 - a. Getting Started with Databricks Sql
 - b. Navigating Databricks SQL
 - c. Unity Catalog on Databricks Sql
 - d. Schemas, Tables and Views on Databricks SQL
2. Basic SQL on Databricks Sql
 - a. Ingesting Data for Databricks Sql
 - b. Ingesting Data
 - c. Joins
 - d. Delta Commands in Databricks Sql
3. Presenting Data Visually
 - a. Data Visualization
 - b. Data Visualizations on Databricks Sql
 - c. Dashboards on Databricks Sql
 - d. Notifying Stakeholders

IV. Apache Spark Programming with Databricks

1. Apache Spark Programming - Dataframes

- a. Databricks platform
 - i. Databricks Ecosystem
 - b. Spark Sql
 - i. Spark Sql
 - ii. Dataframes
 - iii. SparkSession
 - c. Reader and writer
 - i. Data Sources
 - d. Dataframe and column
 - i. Column and Expression
 - ii. Transformation Actions and Rows
2. Apache Spark Programming - Transformations
- a. Aggregation
 - i. Aggregation Functions
 - b. Datetimes
 - i. Dates and Timestamps
 - c. Complex Types
 - i. Complex Types
 - d. Additional Functions
 - i. Additional Functions
 - e. UDFS
 - i. UDFs Vectorized Functions
3. Apache Spark Programming - Spark Internals
- a. Spark Architecture
 - i. Spark Cluster, Spark Execution
 - ii. Shuffling and Caching
 - b. Query Optimization
 - i. Query Optimization
 - c. Partitioning
 - i. Partitioning
4. Apache Spark Programming - Structured Streaming
- a. Apache Spark Programming
 - i. Streaming Concepts