Azure Data Bricks Course Syllabus

Section 1: Introduction to Azure Databricks

- Overview of Azure Databricks and its features
- Introduction to Apache Spark
- Understanding the benefits of using Azure Databricks with Spark
- Setting up an Azure Databricks Workspace

Section 2: Cluster Management

- Understanding the concept of Clusters in Azure Databricks
- Creating, configuring, and managing Clusters
- Cluster Auto-scaling and Performance Tuning
- Cluster Monitoring and Logs

Section 3: Data Processing with Spark

- Understanding Spark RDDs and DataFrames
- Creating and transforming RDDs and DataFrames
- Spark SQL: Querying structured data
- DataFrame Operations
- Spark Data Sources (CSV, Parquet, JSON, Delta Lake)
- Caching and Persisting Data

Section 4: Databricks Delta Lake

- Overview of Delta Lake and its features
- Delta Lake vs Traditional Data Lakes
- Working with Delta tables and Delta Lake format
- Upserts, Deletes, and Merges with Delta Lake

Section 5: Spark MLlib for Machine Learning

- Introduction to Spark MLlib and its components
- Loading data into Spark MLlib
- Feature Engineering and Data Preprocessing
- Building and Evaluating Machine Learning Models
- Using MLlib for Regression, Classification, and Clustering

Section 6: Azure Integration

- Integrating Azure Blob Storage, Data Lake, and SQL Databases
- Accessing Azure Key Vault from Databricks
- Using Azure Databricks with Azure Synapse Analytics
- Working with Azure Databricks Workspace APIs

Section 7: Collaboration and Sharing

- Sharing Notebooks and Results with Teams
- Version Control and Git Integration in Databricks
- Working with Collaboration features in Databricks

Section 8: Security and Access Control

- Authentication and Authorization in Azure Databricks
- Configuring User Roles and Permissions
- Securing Data with Encryption
- Managing Secrets and Credentials

Section 9: Advanced Topics

- Advanced Spark Optimizations and Tuning
- Streaming Data with Spark Structured Streaming
- Real-time Data Processing with Databricks
- Integrating Databricks with Azure Machine Learning Service
- Implementing Custom Libraries and UDFs in Databricks

Section 10: Real-World Use Cases and Best Practices

- Best practices for cluster management and resource allocation
- Using Databricks for big data processing and real-time analytics
- Ensuring scalability, security, and performance optimization