databricks

Databricks Training Content Outline

1. Introduction to Databricks

- What is Databricks?

- Key Features and Benefits

- Use Cases of Databricks

2. Setting Up Databricks

- Creating a Databricks Account

- Navigating the User Interface

- Important Components (Workspaces, Clusters, Notebooks)

3. Basics of Apache Spark

- Introduction to Apache Spark

- Spark Architecture

- Spark Components (RDDs, DataFrames, Datasets)

4. Working with Notebooks

- Creating and Managing Notebooks

- Markdown and Code Cells

- Visualizing Data in Notebooks

5. Data Management

- Connecting to Data Sources (Databases, Data Lakes)

- Reading and Writing Data (CSV, Parquet, Delta Lake)

- Data Ingestion Techniques

6. Data Analysis and Manipulation

- Basic DataFrame Operations

- Advanced DataFrame Functions

- Data Cleaning and Transformation

7. Machine Learning with Databricks

- Introduction to MLlib

- Building and Training Models

- Evaluating Model Performance

8. Jobs in Databricks

- Introduction to Databricks Jobs

- Creating and Managing Jobs

- Scheduling and Monitoring Jobs

- Integrating Jobs with Notebooks and Workflows

9. Advanced Spark Concepts

- Spark SQL

- Performance Tuning

- Managing Partitions and Caching

10. Collaboration and Sharing

- Sharing Notebooks

- Using Libraries and APIs

- Collaboration Best Practices

11. Real-World Applications and Projects

- Case Studies

- Hands-On Projects

- Best Practices and Tips

12. Conclusion

- Recap of Key Learnings

- Next Steps and Resources