# Databricks Foundation Training

**Curriculum 2 Days**

## Day - 1

- **Introduction to Databricks**
  - Overview of Databricks: Introduction to the platform and its components.
  - Databricks Architecture: Understanding clusters, workspaces, and notebooks.
  - Setting up Databricks: Creating and managing clusters.
  - Core Features: Overview of notebooks, version control, jobs, and collaboration.
- **Understanding Apache Spark Basics**
  - Introduction to Apache Spark: Spark's core components, RDDs, DataFrames, and Spark SQL.
  - The Spark DataFrame API: How DataFrames are used for large-scale data processing.
  - Spark SQL: Performing queries on large datasets.
- **Data Engineering with Databricks**
  - Building Data Pipelines: Techniques for transforming and storing data using Spark.
  - Working with Delta Lake: Introduction to Delta Lake for reliable and scalable data storage.
  - Data Transformation: Using Spark's DataFrame API for data cleaning and manipulation.
- **Hands-on**
  - Setting up Databricks Workspace: Create a Databricks workspace, manage clusters, and explore Databricks notebooks.
  - Running Spark Queries: Perform basic data transformations using Spark SQL and DataFrames.
  - Data Pipeline Creation: Build a simple data pipeline using Databricks, including reading and writing data with Delta Lake.

# Day - 2

- **Machine Learning Fundamentals in Databricks**
  - Introduction to MLlib: Using Spark's MLlib for machine learning tasks.
  - Creating Machine Learning Models: Data preparation, feature extraction, model training, and evaluation.
  - MLflow: Managing machine learning experiments, model tracking, and hyperparameter tuning.

- **Collaboration in Databricks**
  - Collaborative Notebooks: Sharing and collaborating on Databricks notebooks.
  - Git Integration: Managing version control for notebooks and data science projects.
  - Team-based Workflow: Using Databricks for collaborative data engineering and machine learning projects.

- **Data Visualization in Databricks**
  - Creating Visualizations: Using built-in tools to visualize data within Databricks notebooks.
  - Building Interactive Dashboards: Building dashboards to explore and present data insights.

- **Hands-on Lab**
  - Building a Machine Learning Model: Use Spark's MLlib to build and evaluate a simple machine learning model.
  - Collaborative Work: Share and collaborate on Databricks notebooks with team members.
  - Visualization: Create visualizations based on the data, analyze results, and build an interactive dashboard.