Databricks Certified Associate Developer for Apache Spark – Syllabus Overview

# Exam Overview

The Databricks Certified Associate Developer for Apache Spark exam evaluates your understanding of PySpark in the Databricks environment.

Topic Weightage:  
- Understand Spark architecture & components – 20%  
- Using Spark SQL – 20%  
- Developing Spark DataFrame/DataSet applications – 30%  
- Troubleshooting & tuning Spark DataFrame apps – 10%  
- Structured Streaming – 10%  
- Spark Connect – 5%  
- Pandas API on Spark – 5%

# Breakdown of Topics

## 1. Apache Spark Architecture & Components (20%)

Understand executor-driver model, deployment modes, spark execution hierarchy, lazy evaluation, shuffles, and fault tolerance.

## 2. Spark SQL (20%)

Write SQL queries on DataFrames. Use views, UDFs, built-in functions, and work with schemas.

## 3. DataFrame/DataSet API (30%)

Perform selection, renaming, adding/dropping columns. Filter, dedupe, group, aggregate, sort data.  
Handle missing/null values, read/write data, partitions, schemas. Work with nested data, arrays, JSON, explode, pivot operations.

## 4. Performance Tuning & Debugging (10%)

Identify and resolve common performance issues like skew, shuffles, partitioning. Use Spark UI to analyze jobs, stages, and optimize query execution.

## 5. Structured Streaming (10%)

Build streaming pipelines with concepts like triggers, watermarks, increments. Understand micro-batching and continuous streaming modes.

## 6. Spark Connect (5%)

Use Spark Connect to separate application client and execution engine.

## 7. Pandas API on Spark (5%)

Use pandas-style syntax through PySpark API for compatibility and easy migration.

# How to Prepare

- Official Exam Guide: Provided by Databricks –   
 https://www.databricks.com/learn/certification/apache-spark-developer-associate  
- Databricks Academy Course – Apache Spark Programming with Databricks –   
 https://www.databricks.com/learn/certification/apache-spark-developer-associate  
- Practice Questions and Certification Guide –   
 https://www.certfun.com/databricks/databricks-certified-data-engineer-associate-exam-syllabus

# Follow the following

- Focus especially on DataFrame API usage (core 30%).  
- Don’t neglect architecture, streaming, and Spark Connect components.  
- Practice on Databricks Community Edition to mimic the exam environment.  
- Try small hands-on notebooks with PySpark and Spark SQL to solidify your understanding.