

**Introduction to Azure Databricks using R**

**Course Number:** AZDB-142
**Duration:** 2 days

**Overview**

This in-person or online Introduction to Azure Databricks using R training course teaches attendees how to scale R applications for complex analytics and data science operations on the Azure Databricks, Microsoft’s cloud-based Apache Spark platform. This class is hands-on and can be customized to your team's goals and needs.

**Prerequisites**

Prior knowledge of R and SQL are presumed.

**Materials**

All Azure Datatbricks training attendees receive a copy of the instructor’s handout and all code created during the class.

**Software Needed on Each Student PC**

Attendees will write applications using the Databricks service running on the cloud.

**Objectives**

* Understand What Databricks is and its architecture.
* Work in the Databricks environment.
* Learn What Spark SQL is and How to Write Spark Applications using it.
* Understand Concepts of running R on Spark.
* Write Spark Applications Using the SparkR API (library).
* Write Spark Applications Using the sparklyr API (library).

**Outline**

* Databricks Introduction
	+ Getting Things Ready
	+ Tour the Databricks Workspace
	+ Create a Spark Cluster
	+ Create Spark Tables
* Using Databricks Notebooks
	+ Using Spark SQL in a Databricks Notebook
	+ Touring the Databricks Notebook
	+ Managing cells
	+ Managing Notebooks
	+ Finding Sample Notebooks
* Visuals and Dashboards
	+ Creating and Customizing Visuals
	+ Creating Dashboards
* Exploring Spark SQL
	+ Creating Tables Over Flat Files (Schema On Read)
	+ Common SQL Operations
	+ JOINS
	+ UNION
	+ Scalar Functions
	+ Aggregations
	+ Creating Views and Tables
	+ Common Table Expressions (CTE)
	+ Reading and Writing Data
	+ Saving to parquet files
	+ Saving to Delta Tables
	+ Using SQL from R
* Intro to R on Spark
	+ Running R locally on Spark
	+ Importing R Libraries
* Using SparkR
	+ Intro to SparkR
	+ Differences Between R and SparkR
	+ Understanding Apache Arrow
* Performing Exploratory Data Analysis (EDA)
	+ Reading and Writing Data
	+ Writing Custom User Defined Functions
* Intro to Sparklyr
	+ Differences Between SparkR and sparklyr
	+ Using sparkly
* Performing EDA with sparklyr
	+ Reading and Writing Data
* Conclusion