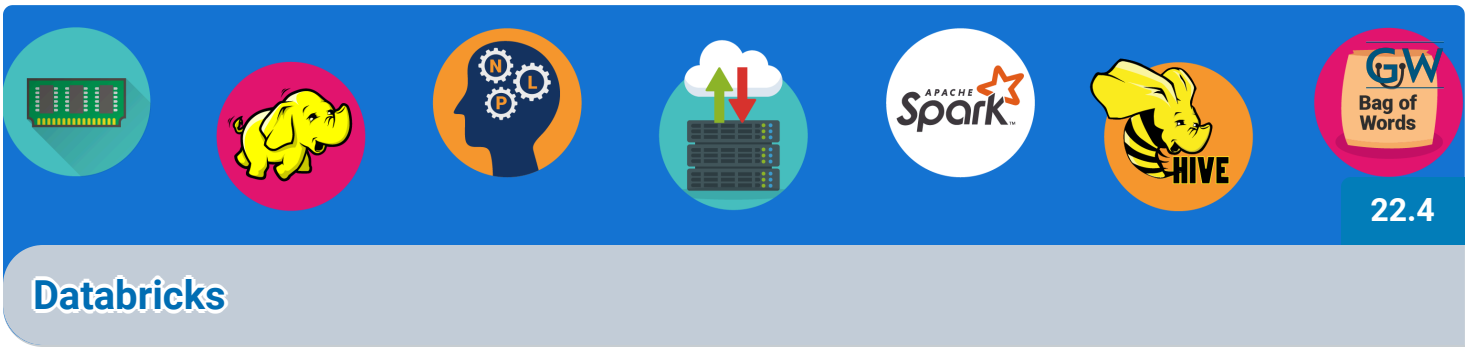


## Overview

In today's class, you will use Spark on Databricks to perform data analysis in the cloud. Through a series of exercises, you will gain hands-on experience with the Python and SQL interfaces of Databricks, with an emphasis on using the SQL interface for increasingly complex queries. The class will conclude with a group activity in which you will query a database in SQL, create a brief report with recommendations, and report your findings to the class.

## What You'll Learn

By the end of this lesson, you will be able to:

- Explain the purpose, key features, and applications of Databricks.
- Set up a Databricks environment.
- Identify the key components of a Databricks environment.
- Navigate the Databricks workspace using `dbutils`.
- Import data into a new notebook by using Parquet files, CSV files, and S3.
- Explain the advantage of Parquet as a big data storage format.
- Perform complex data analysis, including joins, using the Python and SQL interfaces.
- Describe two advantages of using Databricks over PySpark for data analysis.

## 22.4 Activity Files

Download the following files to prepare for today's class:

**22.4 Activity Files (https://static.bc-edx.com/data/dl-1-2/m22/lms/activities/Class_4_Activities.zip)**

You don't need to complete any of the activities before class. But, feel free to review the material ahead of time to preview what the lesson will cover.