Querying Big Data with PySpark

Overview

For today's lesson, you'll learn how to group, aggregate, and order data, and learn how to parse, format a date to a timestamp, and plot time series data. Then, you'll learn how to create temporary views to run Spark SQL queries faster with large datasets. Finally, you'll combine PySpark and Spark SQL to run SQL queries.

What You'll Learn

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

- Apply grouping and aggregation functions to a dataset by using Spark.

- Parse and format date timestamps by using Spark.

- Use temporary tables to prepare data for SQL.

- Combine PySpark and SQL to run queries.

22.2 Activity Files

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

22.2 Activity Files (https://static.bc-edx.com/data/dl-1-2/m22/lms/activities/Class_2_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.

© 2024 edX Boot Camps LLC