

Spark Azure Databricks

Databricks Spark Certification and beyond

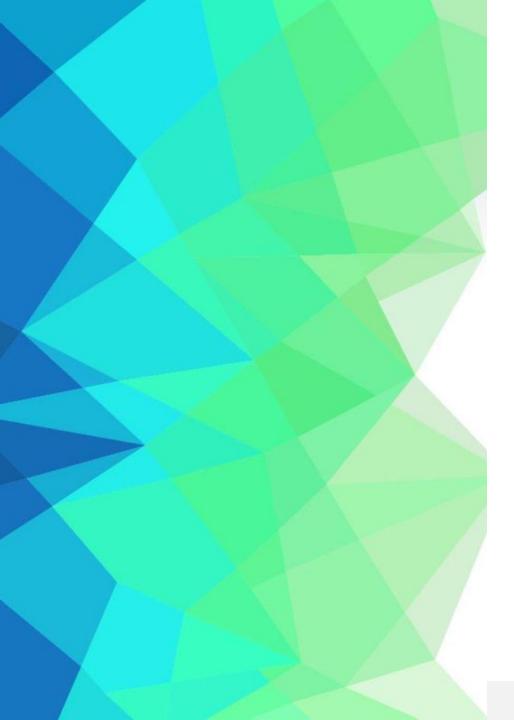
Instructor: Prashant Kumar Pandey





Absolute Beginner to Specialization in Apache Spark and Azure Databricks





Assignment - 1

Load the apache-logs.txt file into the Spark data frame and do the following.

- 1. Extract Date and Referrer fields from the log file
- 2. Filter the records for 17th May 2015
- 3. Find out the number of references by the website
- 4. Eliminate bad Referrers
- 5. Show top 3 Referrers



Assignment - 2

Import the Assignment Notebook (SCA05-Module Assignment-2.ipynb).

Rewrite all 10 Data frame transformations using column APIs and functions. You cannot use SQL like string expressions and expr() function.

