

MongoDB Spark Lab

CS 6350

In class we worked through certain queries on the restaurants database in MongoDB. In this lab, you will connect MongoDB to Apache Spark, load the restaurants database into a Spark SQL dataframe and run queries 1 – 10.

Following are the requirements:

1. You should host the MongoDB database on Atlas cluster.
2. You can use Databricks for Apache Spark.
3. Connect Databricks to MongoDB. An example can be seen at <https://docs.databricks.com/spark/latest/data-sources/mongodb.html>
4. Solve questions 1 – 10 of the lab using Spark. You are free to use Scala or PySpark.
5. Submit the public URL of your notebook.