



Converting mini3

Brandon Haynes authored 17 hours ago

c25aa86a

 **mini1.md** 1.63 KB

Mini-Homework #1

Due date: December 10, 2019

Objectives: GraphX on Spark

Assignment tools:

- [Spark and GraphX](#)
- [Amazon Web Services](#)

What to Turn In

Turn in command and results of a query using GraphX on Spark. Submit everything as a single markdown file, Jupyter notebook, or PDF.

How to submit the assignment

In your GitLab repository, you should see a directory called `Homeworks/mini-hw1`. Put your report in that directory. Remember to `git add`, `git commit`, and `git push`. You can add your report early and keep updating it and pushing it as you do more work. We will collect the final version after the deadline passes. If you need extra time on an assignment, let us know. This is a graduate course, so we are reasonably flexible with deadlines but please do not overuse this flexibility. Use extra time only when you truly need it.

Assignment Details

In this Assignment you will be required to deploy a EMR cluster with Spark and ingest the flights dataset as specified in the [GraphX section \(PDF\)](#).

- List the number of vertices and edges in the graph. Recall the graph has airports as the vertices and flights as edges. We ran this query in the section, so this task is just to check that you successfully created the graph (0 points)
- Which airport has the most number of flights (total flights incoming and outgoing)? (10 points)
- List the top-30 airports in Page Rank order (10 points)

That's it!