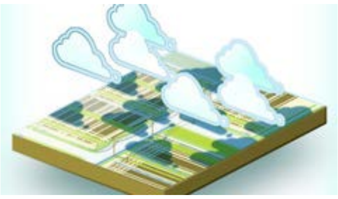
coursera Courses Gleb Filimonov

ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Cloud Computing Capstone
...

Course Certificate



Quick Questions 15

Announcements

Syllabus

Orientation (Mandatory; Start Here)

AWS Credit Opt-in Consent Form

Responses to Commonly Asked Questions

TASKS

Project Overview ←

Task 1

Task 1 Queries

Task 1 Example Solutions

Task 2

Task 2 Queries

Task 2 Example Solutions

COURSE COMMUNITY

All Forums

Know Your Classmates

Help Center

Task 2: Stream Processing with Spark Streaming

Help Center

Instructions

The goal of this task is to process streaming data. You cannot use batch processing for this task. As such, you must answer two questions from Group 1, three questions from Group 2, and Question 3.2, using Spark Streaming. You will then store the results for questions from Group 2 and Question 3.2 in Cassandra or DynamoDB. Please note that you may answer the same questions you chose for Task 1, but the sample queries for Groups 2 and 3 will be different for this task.

More specifically, determine a way to feed the cleaned aviation data into Spark Streaming. Although pulling directly from S3 is supported, we recommend using a system such as Kafka (for more information, consult the tutorial "Storm and Kafka Together"). Then, use Spark Streaming to answer your chosen questions and store the results of questions from Group 2 and Question 3.2 in Cassandra.

Note: You can use the Spark Shell for interactive prototyping and debugging. Explore the API Docs (e.g., for Scala or Python) if you are stuck.

Submission

PDF Report

The submission for Task 2 is due Wednesday, February 17. You must submit your report in PDF format. Your report should be no longer than **5 pages, 11 point font**. Your report should include the following:

1. Give a brief overview of how you integrated each system.
2. What approaches and algorithms did you use to answer each question in each system?
3. What are the results of each question? Use only the provided subset for questions from Group 2 and Question 3.2.
4. What system-level or application-level optimizations (if any) did you employ?
5. Your opinion about whether the results make sense and are useful in any way.
6. How did the different stacks (Hadoop and Spark) from Task 1 and Task 2 compare? Which stack did you find the easiest to use? The fastest?

Video Demonstration Link

In your report, you will also need to submit a video demonstration of your approach. Your video should be **no more than 5 minutes long**. Your video should include the following:

1. Ingesting and analyzing data for each question.
2. Displaying/querying the results for each question.

Following is a list of suggested websites to upload your video.

- YouTube
- Vimeo
- Youku
- Vidme

Sendvid

Submit Task 2

Evaluation

Your peers will evaluate your submission based on the **Task 2 Rubric**. This assignment is worth 50 points. The evaluation period will begin immediately after the submission deadline. You must evaluate **five** of your peers' submissions or your own submission score will be penalized by 20%.

Evaluate Task 2

Deadlines

See the Syllabus for detailed information about deadlines for this task.

Getting Help

You can discuss Task 2 with your peers in the Task 2 Discussion forum.

Created Wed 2 Sep 2015 7:07 PM MSK

Last Modified Tue 9 Feb 2016 8:54 PM MSK