

DISTRIBUTED SYSTEMS ASSIGNMENT REPORT



Assignment ID: 3

Student Name: Peiqi YIN

Student ID: 11810305

DESIGN

Describe the design of your system by providing the following information:

- *Module design: structure block - flow chart and description, if any*
 - *Class design: UML class diagram and description, if any*
- I use pyspark in python to finish this job.

RUNNING RESULT

Provide an illustration of your running system by providing screenshots of all the different steps and dialogues and any other relevant information you consider important.

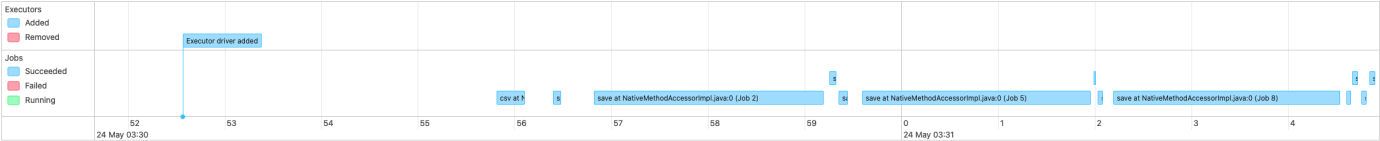
- *Screenshot (at least 3)*
 - *Explanation of each screenshot*
- I use pyspark 3.2.1 to finish this task.

The running result csv are stored in the files.

Spark Jobs ^(?)

User: ubuntu
Total Uptime: 24 min
Scheduling Mode: FIFO
Completed Jobs: 13

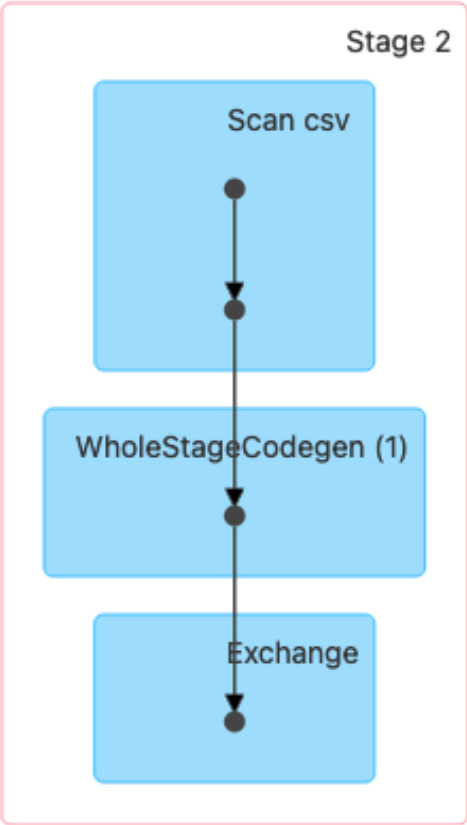
Event Timeline
Enable zooming



Details for Job 2

Status: SUCCEEDED
Submitted: 2022/05/24 03:30:56
Duration: 2 s
Associated SQL Query: 2
Completed Stages: 1

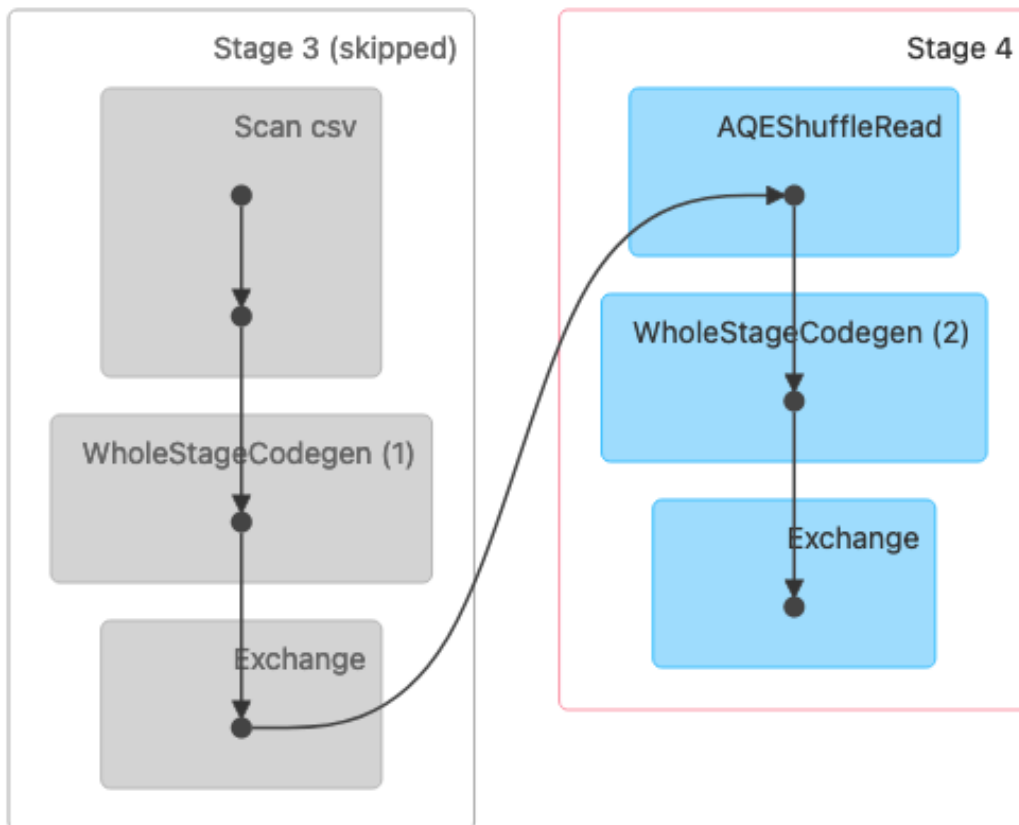
- ▶ Event Timeline
- ▼ DAG Visualization



Details for Job 3

Status: SUCCEEDED
Submitted: 2022/05/24 03:30:59
Duration: 66 ms
Associated SQL Query: 2
Completed Stages: 1
Skipped Stages: 1

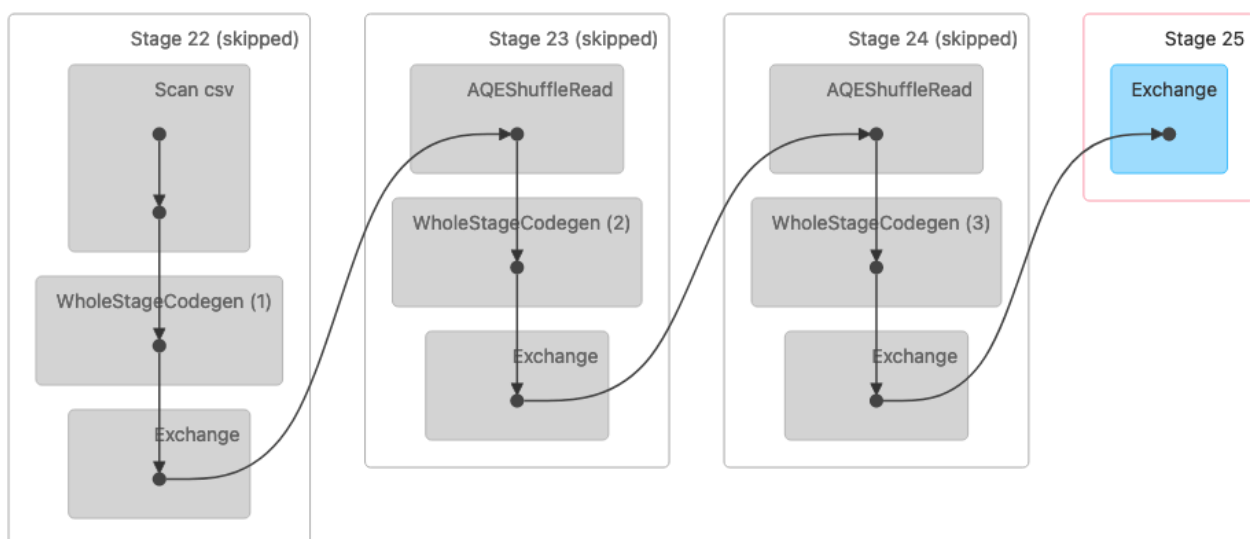
- ▶ Event Timeline
- ▼ DAG Visualization



Details for Job 12

Status: SUCCEEDED
Submitted: 2022/05/24 03:31:04
Duration: 58 ms
Associated SQL Query: 4
Completed Stages: 1
Skipped Stages: 3

- ▶ Event Timeline
- ▼ DAG Visualization



PROBLEMS

Describe major design decisions you had to make and the problems you encountered when implementing the system together with the solutions you devised.

- *Problems and solutions*

The question 4 is hard and I got no idea how to solve it.
Maybe I will try on it later.