Spark Training questions

Please answer the questions below

Exercise 1

Q2. Add print-screen of the stage proving you have 5 tasks

Tasks (5)

Show 20 entries   Search:

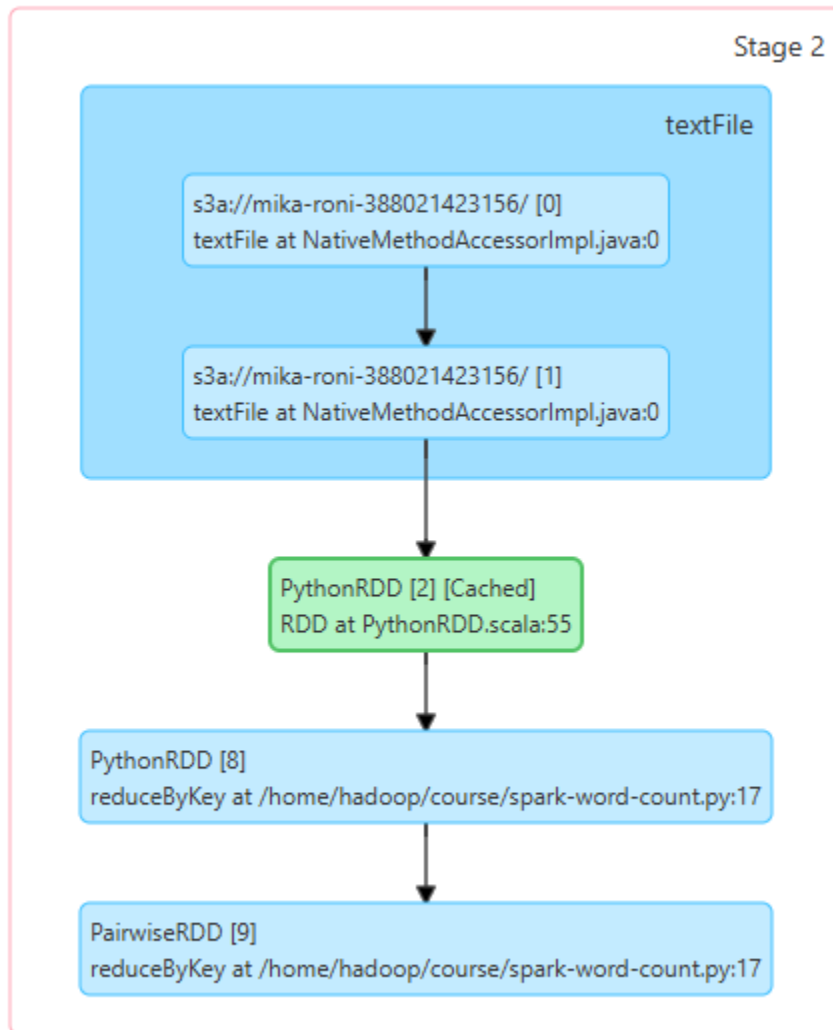| Index	Task ID	Attempt	Status	Locality level	Executor ID	Host	Logs	Launch Time	Duration	GC Time	Shuffle Read Size / Records	Errors
0	6	0	SUCCESS	NODE_LOCAL	1	ip-172-31-61-231.ec2.internal	stderr stdout	2025-12-16 16:19:51	0.2 s		174.8 KiB / 1559	
1	7	0	SUCCESS	NODE_LOCAL	1	ip-172-31-61-231.ec2.internal	stderr stdout	2025-12-16 16:19:51	0.2 s		173.9 KiB / 1559	
2	8	0	SUCCESS	NODE_LOCAL	1	ip-172-31-61-231.ec2.internal	stderr stdout	2025-12-16 16:19:51	0.2 s		173.8 KiB / 1558	
3	9	0	SUCCESS	NODE_LOCAL	1	ip-172-31-61-231.ec2.internal	stderr stdout	2025-12-16 16:19:51	0.2 s		173.8 KiB / 1557	
4	10	0	SUCCESS	RACK_LOCAL	2	ip-172-31-52-86.ec2.internal	stderr stdout	2025-12-16 16:19:51	2 s	0.1 s	174.4 KiB / 1560	

Exercise 2

Q2. Write the number of words found

```
( understand , 400)
------------------------------------------------
Total distinct words (all words): 77928
25/12/16 14:58:00 INFO SparkContext: Invoki
```

Exercise 3

Put a print-scrin with the DAG of the first stage, which shows it reads the files from s3a://<your_bucket_name>

▼ DAG Visualization

Stage 2

textFile

s3a://mika-roni-388021423156/ [0]
textFile at NativeMethodAccessorImpl.java:0

s3a://mika-roni-388021423156/ [1]
textFile at NativeMethodAccessorImpl.java:0

PythonRDD [2] [Cached]
RDD at PythonRDD.scala:55

PythonRDD [8]
reduceByKey at /home/hadoop/course/spark-word-count.py:17

PairwiseRDD [9]
reduceByKey at /home/hadoop/course/spark-word-count.py:17

▶ Show Additional Metrics

Exercise 4

Q2. Put here the printout of the longest word:

```
-------------------------------------------------
Longest word: straightforwardness
Length: 19
-------------------------------------------------
```

Exercise 5

Q2. Put here the printout of the line with the most words:

```
-------------------------------------------------
Line with the most words:
Archimedes, on the centre of gravity [Footnote 9: The works of Archimedes were not printed during Leonardo's life-time.]; anatomy [Footnote 10: Comp
are No. 1494.] Alessandro Benedetto; The Dante of Niccolo della Croce; Inflate the lungs of a pig and observe whether they increase in width and in
length, or in width
Number of words: 51
-------------------------------------------------
```