Spark Tutorial Answer Sheet

## Running a notebook follow up questions

1. **Record used and remaining capacity hours**

**Total account capacity unit hours used and remaining account capacity unit hours after loading the data. Zero used. 50 remaining.**

**Total account capacity unit hours used and remaining account capacity unit hours after Graded exercise 6. Zero used. 50 remaining.**

1. Why did the values change or not change?

**It should have decreased. I did have the bug/issue where the tutorial I used was from last semester, which made me chose a different environment (Spark 2.4 & Python 3.6). Created a Salesforce ticket for this problem.**

1. What is the difference between “Computing Service” and “Environment”?

**Computing Service provides computing power to the client like CPU and GPU. An Environment provides that along with the application that the work is being done in.**

1. What is the difference between Apache Spark Service and the IBM Analytic Engine?

**Apache Spark Service is the software that stores the data. The IBM Analytical Engine is the cloud-based environment that utilizes Apache Spark.**

1. What is the difference between Watson Studio and IBM Cloud?

**Watson studio is the IBM service that assists in developing AI models. IBM Cloud supports Watson Studio by hosting the cloud-computing used by Watson Studio.**

## Spark Quiz

Put your answers to Spark Quiz questions below.

* For multiple choice questions, enter the letter of your answer into the second column.
* For multiselect questions, enter the letters of your answer choices into the second column.
* For short answer questions, enter your answer into the second column.

|  |  |
| --- | --- |
| Question number | Your Answer |
| 1 | D |
| 2 | C |
| 3 | D |
| 4 | B |
| 5 | E |
| 6 | D |
| 7 | D |
| 8 | E |
| 9 | C |
| 10 | D |
| 11 | D |
| 12 | Transformations and Actions. Transformations create new RDDs with a new dataset transformed dataset from and old dataset. Actions return results to the driver program and initiate the transformations. |
| 13 | A |
| 14 | E |
| 15 | B |
| 16 | E |
| 17 | B |
| 18 | C |
| 19 | B,C |
| 20 | A,D |
| 21 | E |
| 22 | B |
| 23 | C |
| 24 | A |
| 25 | C |
| 26 | A |
| 27 | A |
| 28 | D |
| 29 | B |
| 30 | B |
| 31 | Spark can be much faster than MapReduce because it can work with RAM while MapReduce has to work from disk. Spark can support Java, Python, Scala and R languages while MapReduce is only Java. Spark also comes with libraries that can help speed up work that is commonly done with data. These libraries help with SQL queries, streaming, graphing and machine learning. MapReduce has none of those. |
| 32 | The hardest part was the SQL part, mostly because I had to remember how to achieve the desired results. The Graded Exercise #5 was harder than the rest just because I wasn’t used to the syntax of aggregating data that way. |