

1. If you were to use the following SQL statement in the Quickstart VM via the **Hue interface** - utilizing the Hive query editor:

```
SELECT product_name, product_price
FROM products
WHERE ( product_price > 10)
ORDER BY product_price DESC
LIMIT 1000
```

What would be the correct answer for the most expensive product?

- a) SOLE E25 Elliptical
- b) SOLE E35 Elliptical
- c) SOLE F85 Treadmill
- d) Spalding Beast 60" Glass portable Basketball

2. In the sample\_08 data set - who had the highest salary in 2008?

(hint: look at the example SQL query from the hands-on exercise)

- a) Chief Executives
- b) Anesthesiologists
- c) Surgeons
- d) Lawyers

3. In the sample\_08 data set - who had the lowest salary above \$50000 in 2008? (hint: look at the example SQL query from the hands-on exercise -ASC vs. DESC)

- a) Food service managers
- b) Millwrights
- c) Postal service clerks
- d) Interior designers

4. If you only had \$10 - what is the most expensive product you could afford to buy from the products table?

- a) Clicgear Rovic Shoe Brush
- b) \$10 gift card
- c) Toronto FC Team Color Soccer Bracelet
- d) adidas Brazuca 2014 Mini Soccer Ball
- e) LIJA Women's golf Beanie

5. In this assignment we will use the Bay Area bike share data set. First, open a browser inside the VM download both the year 1 and year 2 data sets from:

<http://www.bayareabikeshare.com/datachallenge>

You can save it on your desktop or at the location of your preference within the Quickstart VM. (additional info can be found in the reading Titled "Uploading Bike Share data into Hive")

Within the Hue window click on "Data Browsers" and then on the "Metastore Tables". Here, under the Database click on "Create a new table from a file". Choose the 201402\_trip\_data.csv from the folder that you

have downloaded from the website above. Use the default values on the data import, except for Duration change from **Tinyint to an Int**.

Now that you have a new database in the Hadoop - you can query the data. (hint: refresh the database list to show the newly updated database in the Hive query view)

Which startstation has the longest trip duration?

- a) Davis at Jacskon
- b) University and Emerson
- c) Park at Olive
- d) Powell Street BART

6- Utilizing the Bay Area Bike Share database (both Year 1 & 2, Aug. 2013- Aug. 2015)- what is the most popular start station based on trip data?

- a) Embarcadero at Sansome
- b) Market at 4th
- c) San Francisco Caltrain
- d) Townsend at 7th

7. Utilizing the Bay Area Bike Share database (Year 1 only, Aug. 2013- Feb 2014) - Which is the least popular(least used) start station in the Bike share trips data?

(Hint: Use the count of start station, group and order in ascending order)

- a) Townsend at 7th
- b) Mezes Park
- c) Market at 4th
- d) Embarcadero at Sansome

8. Utilizing the Bay Area Bike Share database (for Year 1 only, Aug. 2013 - Aug. 2014 only) - what is the SECOND MOST popular end station based on trip data? (Hint: Use the count of end station, group and order in descending order)

- a) Steuart at Market
- b) Market at Sansome
- c) Embarcadero at Sansome
- d) Harry Bridges Plaza(Ferry Building)

### What to check with your TA?

	The queries you've used to answer the above questions	The computed result or selected option from a to d
1	sample query: SELECT product_name, product_price FROM products WHERE ( product_price > 10) ....	sample answer : b) SOLE E35 Elliptical
...	...	
8		