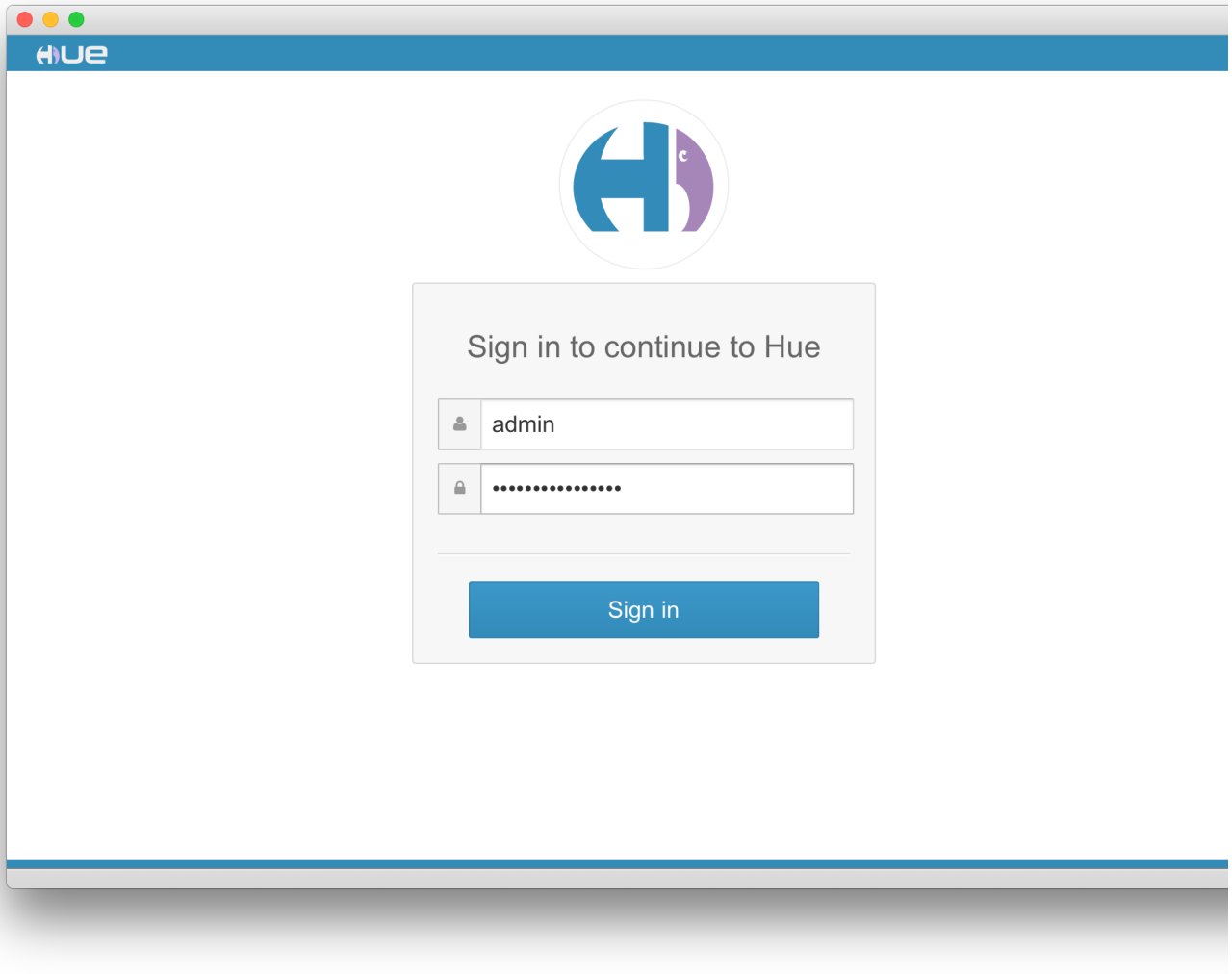
< Tutorial Exercise 1 **Showing Big Data**

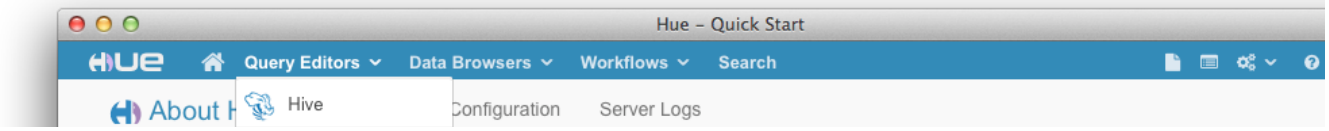
Tutorial Exercise 2

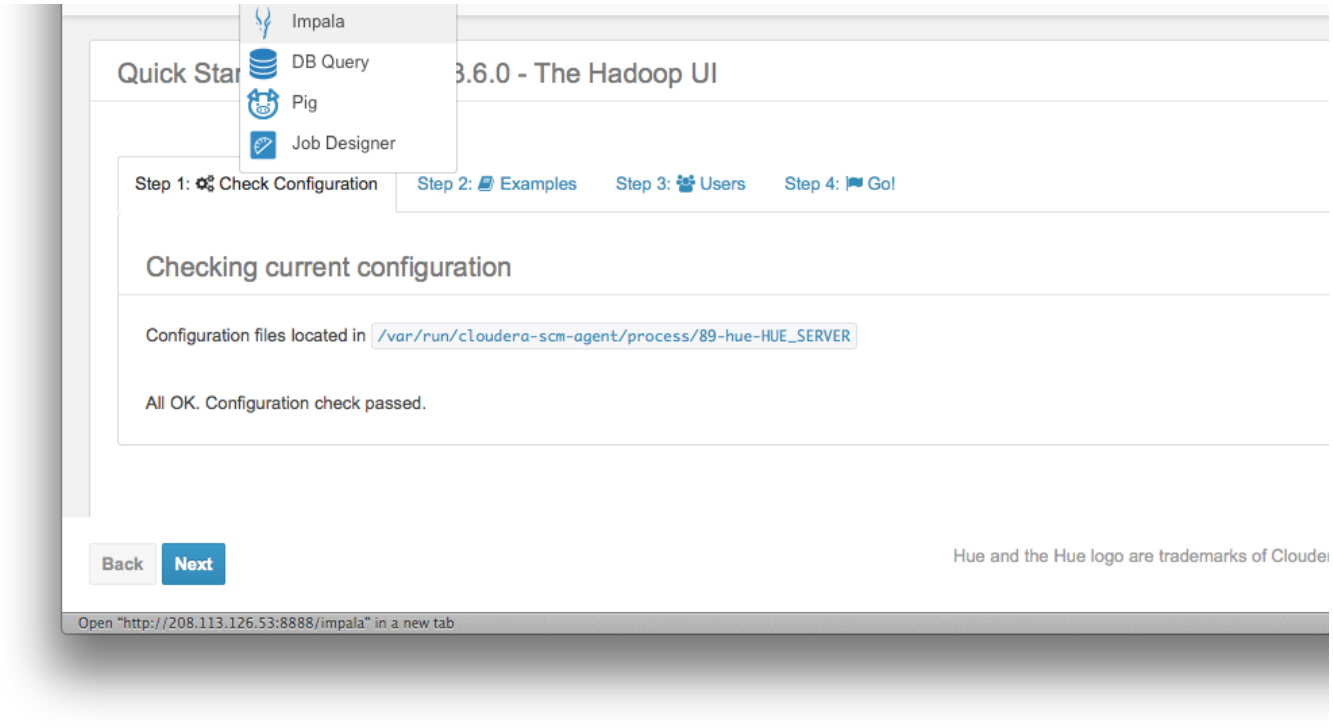
Query Structured Data

We're going to use Hue's Impala app to create the metadata for our tables in Hue, and then query them. Hue provides web-based interface for many of the tools in CDH and can be found on port 8888 of your Master Node (here (http://127.0.0.1:8888)). In the QuickStart VM, the administrator username for Hue is 'cloudera' and the password is 'c



Once you are inside of Hue, click on Query Editors, and open the Impala Query Editor.





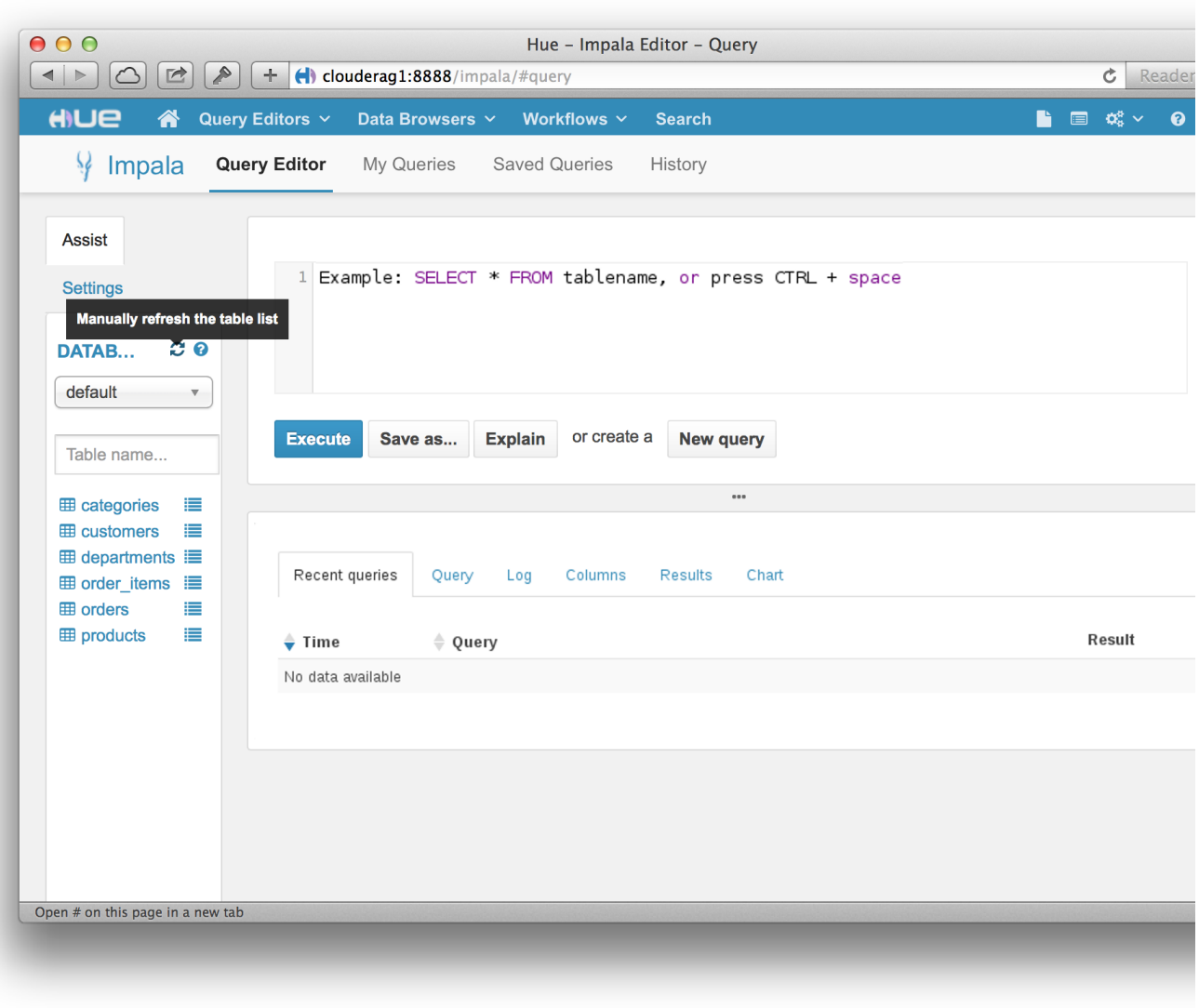
Copy and paste the queries below, and hit enter.

```
CREATE EXTERNAL TABLE categories STORED AS AVRO
LOCATION 'hdfs:///user/hive/warehouse/categories'
TBLPROPERTIES ('avro.schema.url'='hdfs://quickstart/user/examples/sqoop_import_catego

CREATE EXTERNAL TABLE customers STORED AS AVRO
LOCATION 'hdfs:///user/hive/warehouse/customers'
TBLPROPERTIES ('avro.schema.url'='hdfs://quickstart/user/examples/sqoop_import_custom

CREATE EXTERNAL TABLE departments STORED AS AVRO
LOCATION 'hdfs:///user/hive/warehouse/departments'
TBLPROPERTIES ('avro.schema.url'='hdfs://quickstart/user/examples/sqoop_import_depart

CREATE EXTERNAL TABLE orders STORED AS AVRO
LOCATION 'hdfs:///user/hive/warehouse/orders'
TBLPROPERTIES ('avro.schema.url'='hdfs://quickstart/user/examples/sqoop_import_orders

CREATE EXTERNAL TABLE order_items STORED AS AVRO
LOCATION 'hdfs:///user/hive/warehouse/order_items'
TBLPROPERTIES ('avro.schema.url'='hdfs://quickstart/user/examples/sqoop_import_order_

CREATE EXTERNAL TABLE products STORED AS AVRO
LOCATION 'hdfs:///user/hive/warehouse/products'
TBLPROPERTIES ('avro.schema.url'='hdfs://quickstart/user/examples/sqoop_import_produc
```

Delete the queries currently in the editor, and run the following to verify all of the tables were created.

```
show tables;
```

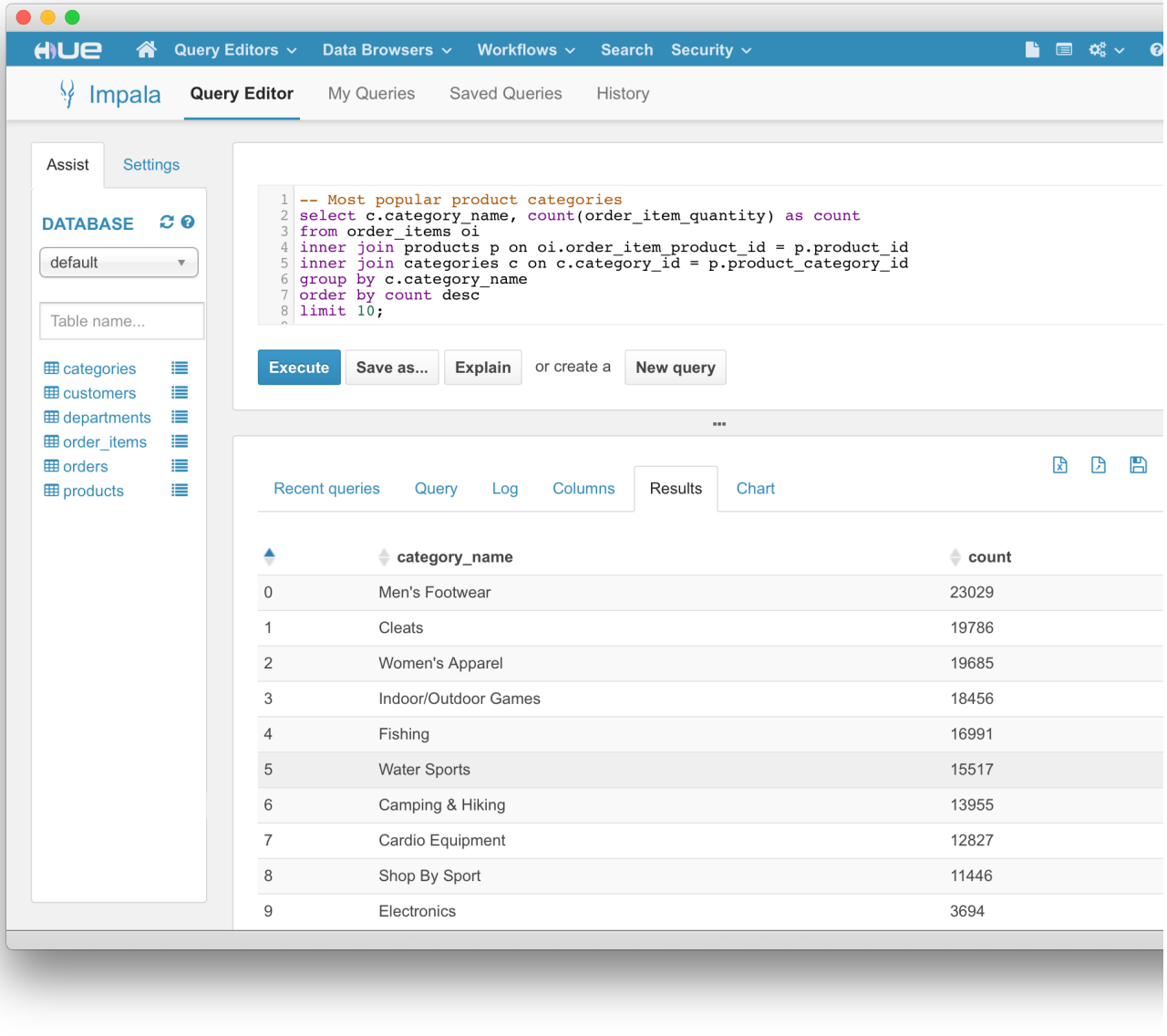
You can also click on the "Refresh Table List" icon on the left to see your new tables.



Now that your transaction data is readily available for structured queries in CDH, it's time to address DataCo's busine
Copy and paste or type in the following standard SQL example queries for calculating total revenue per product and s
top 10 revenue generating products:

```
-- Most popular product categories
select c.category_name, count(order_item_quantity) as count
from order_items oi
inner join products p on oi.order_item_product_id = p.product_id
inner join categories c on c.category_id = p.product_category_id
group by c.category_name
order by count desc
limit 10;
```

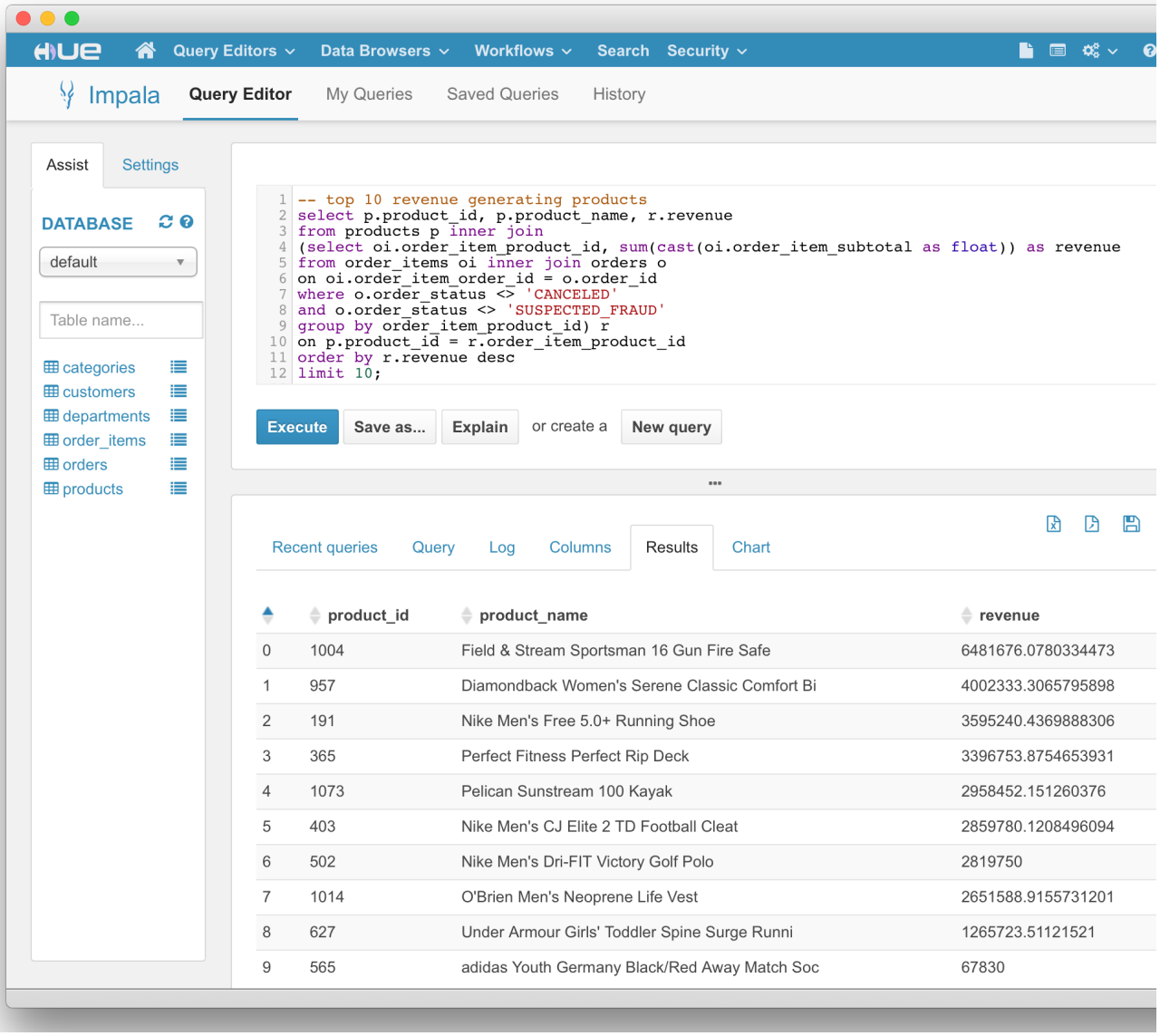
You should see results of the following form:



Clear out the previous query, and replace it with the following:

```
-- top 10 revenue generating products
select p.product_id, p.product_name, r.revenue
from products p inner join
(select oi.order_item_product_id, sum(cast(oi.order_item_subtotal as float)) as reven
from order_items oi inner join orders o
on oi.order_item_order_id = o.order_id
where o.order_status <> 'CANCELED'
and o.order_status <> 'SUSPECTED_FRAUD'
group by order_item_product_id) r
on p.product_id = r.order_item_product_id
order by r.revenue desc
limit 10;
```

You should see results similar to this:



	product_id	product_name	revenue
0	1004	Field & Stream Sportsman 16 Gun Fire Safe	6481676.0780334473
1	957	Diamondback Women's Serene Classic Comfort Bi	4002333.3065795898
2	191	Nike Men's Free 5.0+ Running Shoe	3595240.4369888306
3	365	Perfect Fitness Perfect Rip Deck	3396753.8754653931
4	1073	Pelican Sunstream 100 Kayak	2958452.151260376
5	403	Nike Men's CJ Elite 2 TD Football Cleat	2859780.1208496094
6	502	Nike Men's Dri-FIT Victory Golf Polo	2819750
7	1014	O'Brien Men's Neoprene Life Vest	2651588.9155731201
8	627	Under Armour Girls' Toddler Spine Surge Runni	1265723.51121521
9	565	adidas Youth Germany Black/Red Away Match Soc	67830

If one of these steps fails, please reach out to our Cloudera Live Forum (http://community.cloudera.com/t5/Cloudera-L p/ClouderaLive) and get help. Otherwise continue.

CONCLUSION

Now you have learned how to create and query tables using Impala and that you can use regular interfaces and tools SQL) within a Hadoop environment as well. The idea here being that you can do the same reports you usually do, but architecture of Hadoop vs traditional systems provides much larger scale and flexibility.

< Tutorial Exercise 1 **Showing Big Data**

© 2015 Cloudera (http://www.cloudera.com), Inc. All rights reserved | Terms & Conditions (http://www.cloudera.com/content/cloudera/en/terms-of-service.html) | Privacy Policy (http://www.cloudera.com/content/cloudera/en/privacy-policy.html)
Hadoop and the Hadoop elephant logo are trademarks of the Apache Software Foundation.