**Case Study ReadMe**

NOTE: It is assumed that the user has knowledge of the Hadoop ecosystem, Sqoop, and Java. The platform should have VMWare (Hortonworks Sandbox), Eclipse, WinSCP, and VNCViewer. All the login information is “root” for username and “password” for password except for Ambari (“maria\_dev”, “maria\_dev”). “IP” below is equivalent to the host address e.g. 192.168.109.131 (which can be seen on the activated VMWare screen).

**RDBMS/MySQL**

Instructions: Open MySQL Workbench in VNCViewer (for 2.1.1, in the local system) and create a default schema using the SQL file provided.

**2.1.1 Transaction Details & 2.1.2 Customer Details (Java)**

Description: Connects to the database and interacts with a user to output the desired result in Eclipse in DAO format

Instructions: Import the zip file into Eclipse as an existing project and run the MainRunner class

**2.2.1 Data Extraction and Transportation with Sqoop**

Instructions:

1. Open Ambari (IP:8080 in Google Chrome) and make sure there is no directory called Credit\_Card\_System in HDFS.
2. Copy and paste the Sqoop query from each text file into the Sandbox Terminal (IP:4200 in Google Chrome) and press Enter.

**2.2.2 Data Loading with Hive**

Instructions: In Ambari, go to Hive View. Copy and paste the Hive queries from the given files into the Query Editor and execute them, one by one.

**2.2.3 Automating the Process with Oozie**

Instructions:

1. Create a new file named “case\_study\_oozie” in /user/maria\_dev/.
2. Unzip, then upload all the Hive, workflow XML, and coordinator files into Ambari’s path **/user/maria\_dev/case\_study\_oozie**.
3. Use WinSCP to copy job.properties and jobc.properties files to the *Documents* section in the virtual machine.
4. Type (NOT copy and paste) the following command into the Sandbox Terminal:

a) Without coordinator: oozie job -oozie [http://localhost:11000/oozie -config ./Documents/job.properties](http://localhost:11000/oozie%20-config%20./Documents/job.properties) -run

b) With coordinator: oozie job -oozie [http://localhost:11000/oozie -config ./Documents/jobc.properties](http://localhost:11000/oozie%20-config%20./Documents/jobc.properties) -run

Track the oozie job progress by typing in the following address on Google Chrome: IP:11000/oozie/

**2.2.4 Optimizing the Process**

Instructions:

1. Open a new Terminal tab, log in, and enter “sqoop-metastore”.
2. Come back to the original Terminal tab and do not touch the newly opened one.
3. Unzip, then in the original Terminal tab, copy and paste each of the Sqoop job queries in the unzipped file and press Enter, effectively saving the Sqoop jobs in the metastore.
4. In Ambari, create a new file named “case\_study\_oozie\_optimized” in /user/maria\_dev/.
5. Upload all the optimized Hive, workflow, and coordinator files in the unzipped file into Ambari’s path **/user/maria\_dev/case\_study\_oozie\_optimized**.
6. Use WinSCP to copy joboptimized.properties and joboptimizedc.properties files to the *Documents* section in the virtual machine
7. Go to /usr/hdp/current/sqoop-client/lib, /usr/hdp/<version>/sqoop/lib, and **/user/oozie/oozie\_scripts/lib/** and upload java-json file if it does not exist in the directories (ls first two directories in the Terminal to check, use Ambari HDFS for the third)
8. Type the following command into the Sansbox Terminal:

a) Without coordinator: oozie job -oozie [http://localhost:11000/oozie -config ./Documents/joboptimized.properties](http://localhost:11000/oozie%20-config%20./Documents/joboptimized.properties) -run

b) With coordinator: oozie job -oozie [http://localhost:11000/oozie -config ./Documents/joboptimizedc.properties](http://localhost:11000/oozie%20-config%20./Documents/joboptimizedc.properties) -run

**2.2.5 Data Visualization using Hive**

Description: Hive queries and the corresponding graphs are inside the files