**2.1 [Customer and Transaction Details Module]**

**Using the database and respective tables for the case study**

1. If not exists, create the MySQL connection with MySQL workbench with the following properties:
   * User: root
   * Password: password
   * Port: 3306 (Local Instance)

This connection is used throughout the requirements.

PATH: /*unzip path*/Lalita\_Casestudy/Database/cdw\_sapp.sql

1. Open the existing .sql file named cdw\_sapp.sql with MySQL Workbench found under the path above.
2. Execute the file to automatically create and populate the database cdw\_sapp with records used during requirements.

Executing with Eclipse.

PATH: /Lalita\_Casestudy/Lalita\_Java\_Casestudy.zip/

1. Unzip Lalita\_Java\_Casestudy found under the path above.
2. Set the Eclipse’s default workspace to the path above. This is where the project contents is located.
3. On first opening the project, you may experience an error, due to a missing JDBC jar. Follow these steps to add the jar:
   * Right on Referenced Libraries in the Eclipse file explorer and navigate to Build Path > Configure Build Path.
   * Navigate to Add External Jar.

PATH: /*unzip\_path*/Lalita\_Java\_Casestudy/CDW\_SAPP/lib/mysql-connector-java-8.0.11

* + Use the provided file explorer to add the JDBC jar with the above path.

1. Execute the code by clicking on the Run button and available options for customer and transactions

**2.2.1 [Data Extraction and Transportation with Sqoop]**

1. Open Hadoop’s local terminal on any web browser with the URL: *your\_vm\_ip*:4200.

PATH: /Lalita\_Casestudy/Sqoop/sqoop\_casestudy

1. Copy and paste each sqoop job command found in the path above into terminal and execute to create the sqoop jobs.
2. Copy and paste each sqoop run command found in the path above into terminal and execute to run the sqoop jobs.

**2.2.2 [Data Loading Module]**

1. Open Ambari by opening any web browser with the URL: *your\_vm\_ip*:8888.
2. Enter login credentials for user maria\_dev and navigate to Hive View.

PATH: /Lalita\_Casestudy/Hive/hive\_casestudy.sql

1. Copy and paste the contents of the file in the above path into hive and execute.

NOTES:

* User can run the queries in Hive view and check the creation of relevant tables for Branch, Credit card, Time and customer

**Prerequisite for Requirements 2.2.3 and 2.2.4**

Adding the java-json.jar.

PATH: /Lalita\_Casestudy/java-json.jar

1. Place the above jar on HDFS path /user/oozie/share/lib/lib\_20161025075203/sqoop/

NOTES:

* The lib\_20161025075203 directory may be different, if this is the case, the workflows for requirements 2.2.2 and 2.2.3 need to be changed on the <archive> tag.

**2.2.3 [Automation Process with Oozie]**

PATH: /Lalita\_Casestudy/Automation\_oozie\_workflow/job-nonincremental.properties.

1. Place the file located in the above path into package /root/Desktop/initilize on the linux virtual machine.

PATH: /Lalita\_Casestudy/Automation\_oozie\_workflow/coordinator-nonincremental

PATH: /Lalita\_Casestudy/Automation\_oozie/workflow\_init-nonincremental

1. Place the two files located in above paths into package /user/oozie\_workflows/initialize/ on HDFS
2. Make sure to test the hive query specific to the non incremental oozie workflow from the path /Lalita\_Casestudy/Automation\_oozie/oozie\_hive-nonincremental.sql

NOTES:

* The target folder for extracted records (/user/Credit\_Card\_System/) is deleted in the “start-metasore” node.
* Database does not need to be dropped due to the “insert overwrite” clause.
* The <global> tag available in version 0.4 and up allows the nameNode, jobTracker, etcetera to be declared only once.

**2.2.4 [Process Optimization Module]**

PATH: /Lalita\_Casestudy/Automation\_oozie\_Optimized/job.properties.

1. Place the file located in the above path into package /root/Desktop/initilize on the linux virtual machine.

PATH: /Lalita\_Casestudy/Automation\_oozie\_Optimized/coordinator

PATH: /Lalita\_Casestudy/Automation\_oozie\_Optimized/workflow\_init

1. Place the two files located in above paths into package /user/oozie\_workflows/initialize/ HDFS.
2. PATH: /Lalita\_Casestudy/Automation\_oozie\_optimized/sqoop\_casestudy

Execute the sqoop jobs and include them in the workflow for the oozie automation

1. Execute the oozie workflow with necessary changes the hive query and workflow.xml as per the incremental workflow requirements

**2.2.5 [Data Visualization]**

PATH: /Lalita\_Casestudy/Data Visualization/Query\_Datavisualization