**2.1.0\_RDBMS/MySQL data transformation**

* ***updated\_cdw\_data*** – updated database
* ***sql.txt*** - SQL server database, updated with LAST\_UPDATED TIMESTAMP column added to cdw\_sapp\_customer and cdw\_sapp\_brunch tables, so it can have updated timestamps every time data is modified or added to these tables.

**2.1.1 – 2.1.2 - Java**

2.1.1 - 2.1.2- Java/

* ***Case\_Study.zip*** -

**2.2.1 Data Extraction and Transportation with Sqoop**

2.2.1\_Sqoop/

* ***2.2.1\_Data\_Import\_SQOOP.txt*** – Sqoop statements for Importing data from SQL server into HDFS

**2.2.2 Data Loading with Hive**

2.2.2\_Hive/

* ***222\_Populate\_External\_Tables\_with\_HIVE.txt*** - script with Hive statements to create External tables and link to the data imported into HDFS
* ***222\_Populate\_Internal\_Partitioned\_Tables\_with\_HIVE.txt*** – script with Hive statements to create partitioned Internal tables and overwrite with data from respective External tables

**2.2.3 Automating Sqoop and Hive processes from steps 2.2.1 and 2.2.2 with Oozie Workflow.**

2.2.3\_Oozie/

* ***223\_workflow.xml*** – oozie xml script for extracting and modifying data based on requirements with Sqoop, creating and populating tables with Hive using scripts.

* ***223\_Populate\_External\_Tables\_with\_HIVE.txt*** – script with hive statements to drop, create and populate External tables with extracted data.
* ***223\_Populate\_Internal\_Partitioned\_Tables\_with\_HIVE.txt*** - script with hive statements to create Internal Tables and populate with data from respective External ones.
* ***223\_job.properties*** - configuration file, containing all parameters needed to run Oozie workflow.
* ***223\_coordinator.xml*** – Oozie workflow scheduler.

**2.2.4 Process Optimization Module Oozie (Sqoop and Hive optimized)**

2.2.4\_Oozie/

* ***224\_workflow.xml*** – updated Oozie workflow script for extracting only *new data* using Sqoop Job and updating original tables with new data using Hive.
* ***224\_sqoop\_jobs.txt*** – script with Sqoop statements that needs to be run from terminal before running Oozie Workflow.
* ***224\_Populate\_External\_Tables\_with\_HIVE.txt*** - script with hive statements to drop if exists and create Incremental/External 224\_branch\_incremental, 224\_customer\_incremental, 224\_creditcard\_incremental tables.
* ***224\_Populate\_Internal\_Tables\_with\_HIVE.txt***- script with hive statements to create and populate Base/Internal tables for 224\_branch\_base, 224\_customer\_base, 224\_creditcard\_base with data from incremental tables.
* ***224\_merge\_new\_BRANCH\_data.txt***- script with hive statements to update originals 224\_branch\_base table with new data.
* ***224\_merge\_new\_CUSTOMER\_data.txt***- script with hive statements to update originals 224\_customer\_base table with new data.
* ***224\_merge\_new\_CC\_TIME\_data.txt***- script with hive statements to update originals 224\_creditcard\_base and 224\_time\_base tables with new data.
* ***224\_job.properties*** - configuration file, containing all parameters needed to run Oozie workflow.
* ***224\_coordinator.xml*** – Oozie workflow scheduler.

**2.2.5 Data Visualization**

***225\_Visualization\_n1.txt*** and***225\_Visualization\_n2.txt***- hive queries requested.

***225\_Visualization\_n1.png*** and***225\_Visualization\_n2.png***- the graph snapshots of hive queries requested.