

## 2 System Requirements

### 2.1 Functional Requirements – Operational Database

#### 2.1.1 Transaction Details Module

##### 2.1.1 Transaction Details

Credit Card System Req-2.1.1	Customer Details
Functional Requirements	<ol style="list-style-type: none"><li>1) To display the transactions made by customers living in a given zipcode for a given month and year. Order by day in descending order.</li><li>2) To display the number and total values of transactions for a given type.</li><li>3) To display the number and total values of transactions for branches in a given state</li></ol>

#### 2.1.2 Customer Details Module

##### 2.1.2 Customer Details

Credit Card System Req-2.1.2	Customer Details
Functional Requirements	<ol style="list-style-type: none"><li>1) To check the existing account details of a customer.</li><li>2) To modify the existing account details of a customer</li><li>3) To generate monthly bill for a credit card number for a given month and year.</li><li>4) To display the transactions made by a customer between two dates. Order by year, month, and day in descending order.</li></ol>

## 2.2 Functional Requirements – ETL of Data

### 2.2.1 Data Extraction and Transportation Module

#### 2.2.1 Data Extraction and Transportation with Sqoop

Credit Card System Req-2.2.1	Data Extraction and Transportation with Sqoop
Functional Requirements	<p>Utilize Sqoop to extract the following data <b>according to the specifications found in the mapping document:</b></p> <ol style="list-style-type: none"><li>1. Branch data into CDW_SAPP_BRANCH</li><li>2. Credit Card Data into CDW_SAPP_CREDITCARD</li><li>3. Time data into CDW_SAPP_TIME</li><li>4. Customer Data into CDW_SAPP_CUSTOMER</li></ol> <p>Notes:</p> <ul style="list-style-type: none"><li>• <b>Data Engineers will be required to transform the data based on requirements found in the Mapping Document prior to loading the data into Hadoop.</b></li><li>• TIMEID is a field that the Data Engineers should create based on the DAY, MONTH, and TIME fields located in the CUSTOMER table. Format should be YYYYMMDD. For instance, January 4<sup>th</sup>, 2017 would become 20170104</li><li>• Data Engineers should extract the above data to the /Credit_Card_System/ folder in the Hadoop Filesystem</li></ul>

### 2.2.2 Data Loading Module

#### 2.2.2 Data Loading with Hive

Credit Card System Req-2.2.2	Data Loading with Hive
Functional Requirements	<p>Utilize Hive to create tables in the Hadoop Filesystem and then load the data extracted via Sqoop into those tables.</p> <p><b>Data Engineers will be map to transform the data based on requirements found in the Mapping Document.</b></p>

## 2.2.3 Process Automation Module

### 2.2.3 Automating the Process with Oozie

Credit Card System Req-2.2.3	Automating the Process with Oozie
Functional Requirements	<ol style="list-style-type: none"><li>1) Create an Oozie Workflow that will automate the processes of steps 2.2.1 and 2.2.2.<ul style="list-style-type: none"><li>• Each of the files in step 2.2.1 should be deleted before the workflow is executed in order to prevent storage of redundant data</li><li>• The tables created in step 2.2.2 should be dropped before executing the hive workflow in order to prevent redundancy.</li></ul></li><li>2) Incorporate that workflow into an Oozie Coordinator that will execute with the following conditions:<ul style="list-style-type: none"><li>• Every weekday between 08:00 and 18:00 EST</li><li>• Executes once every 20 minutes</li><li>• Starts on April 2<sup>nd</sup> 2018 at 08:00 EST</li><li>• Ends on March 29<sup>th</sup> 2025 at 18:00 EST</li></ul></li></ol>

## 2.2.4 Process Optimization Module

### 2.2.4 Optimizing the Process

Credit Card System Req-2.2.4	Optimizing the Process
Functional Requirements	<ol style="list-style-type: none"><li>1) Create a new Oozie workflow similar to the process of 2.2.3. This time, however, Sqoop should <b>only import new data</b>. Hive should then import <b>only that new data</b>. Original data should not be deleted in this process.</li><li>2) Modify the Oozie Coordinator to use this workflow rather than the original, unoptimized, workflow.</li></ol>

## 2.2.5 Data Visualization

### 2.2.5 Visualization of Dataset

Credit Card System Req-2.2.5	Data exploration and visualization
Functional Requirements	<p><b>Use Hive Query and Hive Visualization tool.</b></p> <ol style="list-style-type: none"><li>1) Find the top 20 zip codes(hint: branch_zip) by total transaction value</li><li>2) Find total transaction value for each transaction type by Quarter in 2018</li></ol> <p>Hint: Find quarter from 'creditcard' table using month or use 'time' table if you already added transaction_id column there.</p>