2 System Requirements

2.1 Functional Requirements – Operational Database

2.1.1 Transaction Module

2.1.1 Transaction Incidents

Credit Card System Req-2.1.1	Customer Details
Functional Requirements	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. 2) To display the number and total values of transactions for a given type. *3)* To display the number and total values of transactions for branches in a given state

2.1.2 Customer Module

2.1.2 Customer module Incidents

Credit Card System Req-2.1.2	Customer Details
Functional Requirements	1) To check the existing account details of a customer. 2) To modify the existing account details of a customer 3) To generate monthly bill for a credit card number for a given month and year. *4)* To display the transactions made by a customer between two dates. Order by year, month, and day in descending order.

2.2 Functional Requirements – ETL of Data

2.2.1 Data Extraction and Transformation 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	Utilize Sqoop to extract the following data **according to the specifications found in the mapping document**: 1. Branch data into CDW_SAPP_BRANCH.txt 2. Credit Card Data into CDW_SAPP_CREDITCARD.txt 3. Time data into CDW_SAPP_TIME.txt 4. Customer Data into CDW_SAPP_CUSTOMER.txt Notes: • **Data Engineers will be required to transform the data based on requirements found in the Mapping Document prior to loading the data into Hadoop.** • TIMEID is a field that the Data Engineers should create based on the DAY, MONTH, and TIME fields located in the "Credit Card" table. Format should be YYYYMMDD. For instance, January 4th, 2017 would become 20170104 • Data Engineers should extract the above data to the /Credit_Card_System/ folder in the Hadoop Filesystem

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	Utilize Hive to create tables in the Hadoop Filesystem and then load the data extracted via Sqoop into those tables. **Data Engineers will be required to transform/map the data based on requirements found in the Mapping Document.**

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	1) Create an Oozie Workflow that will automate the processes of steps 2.2.1 and 2.2.2 • 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 • The tables created in step 2.2.2 should be dropped before executing the hive workflow in order to prevent redundancy. • Be sure to implement Partitioning in Hive

2.2.4 Process Optimization Module

2.2.4 Optimizing the Process

Credit Card System Req-2.2.4	Optimizing the Process
Functional Requirements	1) Create a new Oozie workflow similar to the process of 2.2.3. This time, however, Sqoop should **only import new data**. Hive should then import **only that new data**. Original data should not be deleted in this sqoop/hive process. 2) Incorporate that workflow into an Oozie Coordinator that will execute with the following conditions: • Every weekday between 08:00 and 18:00 EST • Executes once every 20 minutes • Starts today Date for example start on 15th Jan 2019 at 08:00 EST 2) Ends on March 29th, 2025 at 18:00 EST

2.2.5 Process Filter and Spark

2.2.5 Optimizing the Process

Credit Card System Req-2.2.4	Optimizing the Process
Functional Requirements	1) Create "credit_card_number" directory inside the HDFS and store below information in it. 2) Find which state has the highest number of branches 3) Find which customers have the greatest number of transactions on their credit cards 4) Find which states have the greatest number of credit card transactions

2.2.6 Data Visualization

2.2.6 Visualization of Dataset

Credit Card System Req-2.2.5	Data exploration and visualization
Functional Requirements	**Use Hive Query and Hive Visualization tool.** 1. Find the top 20 zip codes(hint: branch_zip) by total transaction value 2. Find total transaction value for each transaction type by Quarter in 2018 Hint: Find quarter from 'creditcard' table using month or use 'time' table if you already added transaction_id column there..