Project Brief

1. Objective

A Home Subscriber switch (HSS) generates logging information of LTE-enabled mobile subscribers who are connected to one or more cellular towers within a given geographical location. The analysis team would like to get a sense of the penetration of LTE subscribers in the area along with other insights, but the team does not have access to the data provided by the HSS. You are tasked to create a data pipeline to connect to the switch and extract data on mobile subscriber registration and connectivity status data from the HSS and load it to Snowflake.

2. Tools
The tools used are:
- ETL tools – Pentaho Data Integration, SQL, Python
- Data store – Snowflake data warehouse

3. Development Process
The project was developed on a test system where all tools required will be installed or configured to be accessible from.
Steps:
1. Create staging and destination tables on Snowflake
2. Develop a script a Python script to do the following:
   o Extract the data from the switch to a text file
   o Parse the data in the text file and output the result to a CSV file
   o Load the data from the CSV file to staging table on Snowflake

3. Create an ETL job in Pentaho to transform data in staging table and load to destination table

4. Testing
Testing consisted of the following steps:

Test item	Output	Verification	Corrective action
Extract data from switch to text file and parse text file and output parsed data to CSV file	CSV file	Verify content of CSV file meets data requirement	If error occur, take any corrective action/s
Load data in CSV file to staging table in data warehouse	Table in data warehouse	Verify data loaded to staging table in desired format	If error occur, take any corrective action/s
Transform data in staging table and load data to destination table	Data in destination table	Verify data in destination table in desired format	