****

Technical Exercise

**Project Overview**

**Background**

Upon analysis, you find out that company collects real time data for all the calls done by customer. The data team informs you that the following files are present.

1. Call\_logs- date,hour,signal,status,description,net,speed, activity ,postal\_code, satellites, precission
2. location\_dim – zip,city,state\_id,state\_name,population,density

Note: - All files are comma separated. Sample data for 2 months call logs for USA has been provided to you for analysis.

Assume that location\_dim has data for all the location of USA. Metadata of each file is attached below-



**Request 1**: Design a data platform for both Batch and realtime Data with below considerations-

* Build a platform to ingest and process real time data for call logs with minimum latency. This data needs to be integrated with company’s datawarehouse having customer and location data (Batch).
* Platform should be capable of detecting Network Failure and send notification to Maintenance and Customer care.
* Need to build real time network coverage visualization with self service capability.
* System should also be able to predict any outage based on historical data using AI/ML.

**Request 2**: Provide current state analytics to the CIO. Use the above data files to calculate the following

1. What is the peak hour of the day when the call traffic is maximum?
2. List top 5 states where speed is lower than average of the country.
3. Which state has maximum Emergency state of network (STATE\_EMERGENCY\_ONLY)?
4. What is the relation between user activity with the Speed (Calculate avg speed for each activity like IN\_VEHICLE)
5. Which state has maximum call volume density (number of calls/area)?

**Note-** There could be some bad data in column speed of file call\_logs, do validation and filter all such bad records.

**Your Task Today**

**Duration: 150 minutes**

1. **Solution & Design –** Use a paper/whiteboard and clearly explain your design & architecture for **Request 1**. Take a snap and send it over email to hiring team.
   1. Create a solution architecture to address the following
      1. Logical architecture with tools (based on your choice of cloud)
      2. What are the preferred databases?
      3. Elaborate the solution flows for both Real time and batch data/processing
      4. Physical architecture/Hosting
      5. Which Reporting Tool would be most suitable?
      6. How would you design Recommendation Engine for new infrastructure?
   2. Define the data storage model for the platform. List down all the entities and reference tables.
   3. Please list down any assumptions.
2. **Coding –** Use the AWS Workspace to code for **Request 2**.
   1. AWS Workspace provided to you has the following software preinstalled
      1. Java 8
      2. Python 3.7
      3. Scala 2.11
      4. Spark 2.4
      5. Anaconda
      6. IDEs like Spyder, PyCharm, IntelliJ, Jupyter and ScalaIDE
      7. Mozilla Developer Web Browser

In case you want to launch pyspark shell, launch it from Anaconda prompt. Spark-shell can be launched from CMD. Run ScalaIDE by doing a right click on its icon and select ‘run as administrator’.

It is recommended to perform coding using the above tools. But additional plugins or libraries can be installed by you on the provided windows desktop.

* 1. Please use below location for your coding exercise
     1. Data Files Directory – C:/Data/telecom\_data
     2. Workspace or Source Code Directory- D:/SourceCode
     3. Output Directory – D:/Results

1. **Enhancements –** Use a paper/whiteboard and clearly explain your design & architecture. Take a snap and send it over email to hiring team.
2. How will the architecture/solution change to expand globally and process data at very high scale?
3. What would you recommend for DevOps Implementation such as CI/CD, QA Automation etc?
4. Please suggest design of common reusable framework for Data Acquisition, Ingestion, Transformations, Loading, Error Mgmt, Auditing, Recovery/Restart, Reconciliation, Scheduling and Monitoring Frameworks.