**Customer Mobile Usage Analysis - Use case**

**XYZ is a Telecom based company.** With the explosive growth of smart phones, communications service providers (CSPs) are seeing huge expansion in the volume of data travelling across their networks. **CSP need new storage and analysis solutions so they can cost-effectively store and derive insights from this growing and invaluable volume of customer data.**

**XYZ company provided you Call and SMS logs details of customer along with Presence details which provides insight into mobile behavior.**

**Dataset:**

1. calls.csv

* Customer phone number, Receiver phone number, Direction of call, Duration of call, Date & Time

1. messages.csv

* Customer phone number, Receiver phone number, Direction of message, Message Length, Date & Time

1. presence.csv

* Customer device, Receiver device, Device name, Device class, Date & Time

Presence provides insight into mobile behaviour across locations

1. MySQL Table – customer

* Customer ID, City, Phone number, Device and Age

**Objective:**

**To analyze customer mobile usage behavior based on the source files providing using Hadoop components (Hive/Sqoop/HBase).**

**Problem Statement:**

1. Load these csv files into HDFS directory
2. Create hive tables based on calls & messages data by providing meaningful field names.
3. Load respective data into those tables
4. Create partitioned table on presence data. Choose the field which is best suited for partition. Also store it as ORC file with Snappy compression. This table will be used in future by other application.
5. Load data into partition table
6. Analyze and write query to find answers for the below
7. Identify which numbers made calls as well as made messages, where number of calls made should be more than 5 and number of messages should be more than 5. Display with header column.
8. Find which number sent messages but did not make calls (can you use “not in”? how to replace “not in” by joins?)
9. What is the frequency of calls by each day of week and by hour of day?
10. Load customer data from MySQL database to hive table
11. Display customer id, city, duration of call, receiver of call and timestamp when call is made for all the customers.
12. Generate the above query output from hive to local environment in any file format. Check the contents of the file.
13. Load customer data from MySQL database to HBase table. Think and choose appropriate rowkey to load.
14. Find out the city details for each customer id in HBase