Telecom Use case

**Data Elements Explanation**

**calls.csv and messages.csv** user - phone number of user other - phone number of the other end of the call/message direction - incoming or outgoing duration - call duration in seconds timestamp - timestamp is the one recorded by the phone when the call/message was originally recorded. length - number of characters in the message

**presence.csv** user - bluetooth/wifi id other - bluetooth/wifi id name - human readable name of device class - bluetooth/wifi class timestamp - as explained before

Use Case Problems :

1. Load data in hdfs, give self explanatory names and not just the ones given in csv files. The 'presence' data should be created as partitioned(think for yourself what column(s) will be best suited for partitioning). Also upload some sample data (1-2 lines) for the sake of clarity on exactly what kind of data you worked on.
2. Identify which numbers made calls as well as made sms messages, where number of calls made should be more than 10 and number of messages should be more than 5.
3. Find which users sent smses but did not make calls (can you use NOT IN, how to replace NOT IN by JOINS)
4. Find which users sent at least 10 more messages in the month of Jan than in the month of Feb.
5. What is the nth longest SMS conversation length.