1. Read the log file [Raw Data] and push data from TCP Socket into Spark Streaming.
   1. Make sure no transaction log is lost
   2. Find out the number of partitions needed theoretically
   3. Read data from HDFS using Spark Streaming [Stats]
   4. Save aggregate data in HDFS.

Text

Description automatically generated

Graphical user interface, text

Description automatically generated