SPARK STREAMING with KAFA and HBASE

The main purpose of this task is to produce the messages to kafka-topic, and using Spark Streaming consume the messages from kafka-topic, then commit the consumed messages to HBASE.

Prerequisites:

1. install hadoop

2. install hbase: hbase depends on hadoop for storage.

3. install spark: to start spark streaming.

4. install kafka: start kafka without zookeeper, as zookeeper is started along with hbase.

How to Run:

step 1: run hadoop services

start-all.sh: starts Resource Manager, Node Manager, NameNode, DataNode, SecondaryNameNode.

step 2: run hbase services

Change hbase.zookeeper.property.clientPort to 2182 in hbase-site.xml

start-hbase.sh: starts HMaster, HRegionServer, HQuorumPeer.

hbase shell: create 'ultratendency','weather-data'.

step 3: run spark

cd ${SPARK_HOME}

sbin/start-all.sh: starts Master, Worker

step 4: run kafka

cd ${KAFKA_HOME}

Change zookeeper.connect to localhost:2182 in config/server.properties

bin/kafka-server-start.sh config/server.properties : starts kafka

Step 5: Create topics

${KAFKA_HOME}/bin/kafka-topics.sh --create --zookeeper localhost:2182 --replication-factor 1 --partitions 1 --topic topic 1

${KAFKA_HOME}/bin/kafka-topics.sh --create --zookeeper localhost:2182 --replication-factor 1 --partitions 1 --topic topic 1

${KAFKA_HOME}/bin/kafka-topics.sh --create --zookeeper localhost:2182 --replication-factor 1 --partitions 1 --topic topic 1

step 6: run producer.jar

java -jar producer.jar ultratendency.task01.Producer

step 7: run consumer.jar

java -jar consumer.jar ultratendency.task01.Consumer