**Copy the file from local to hadoop:**

scp -i ~/.ssh/hsiangct\_mpcs53014.pem NBA\_player\_stat.csv [hadoop@ec2-3-131-137-149.us-east-2.compute.amazonaws.com](mailto:hadoop@ec2-3-131-137-149.us-east-2.compute.amazonaws.com):~

ssh -i ~/.ssh/hsiangct\_mpcs53014.pem -L 8070:ec2-3-131-137-149.us-east-2.compute.amazonaws.com:8070 hadoop@ec2-3-131-137-149.us-east-2.compute.amazonaws.com

**Create the Hive table and link HBase with it:**

beeline -u jdbc:hive2://localhost:10000/default -n hadoop -d org.apache.hive.jdbc.HiveDriver

hbase shell

create "hsiangct\_final\_NBA\_stats", "cf"

CREATE EXTERNAL TABLE IF NOT EXISTS hsiangct\_temp\_NBA\_stats (

id INT,

player\_name STRING,

team\_abbreviation STRING,

age INT,

player\_height FLOAT,

player\_weight FLOAT,

college STRING,

country STRING,

draft\_year INT,

draft\_round INT,

draft\_number INT,

gp INT,

pts FLOAT,

reb FLOAT,

ast FLOAT,

net\_rating FLOAT,

oreb\_pct FLOAT,

dreb\_pct FLOAT,

usg\_pct FLOAT,

ts\_pct FLOAT,

ast\_pct FLOAT,

season STRING

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ‘,’

LOCATION ‘hdfs:///user/hadoop/input’;

LOAD DATA INPATH '/hsiangct/final/data/NBA\_player\_stat.csv'

INTO TABLE hsiangct\_temp\_NBA\_stats;

INSERT OVERWRITE TABLE hsiangct\_final\_NBA\_stats

SELECT CONCAT(player\_name, season) AS id,

FROM hsiangct\_temp\_NBA\_stats;

CREATE EXTERNAL TABLE IF NOT EXISTS hsiangct\_final\_NBA\_stats (

id STRING,

player\_name STRING,

team\_abbreviation STRING,

age INT,

player\_height FLOAT,

player\_weight FLOAT,

college STRING,

country STRING,

draft\_year INT,

draft\_round INT,

draft\_number INT,

gp INT,

pts FLOAT,

reb FLOAT,

ast FLOAT,

net\_rating FLOAT,

oreb\_pct FLOAT,

dreb\_pct FLOAT,

usg\_pct FLOAT,

ts\_pct FLOAT,

ast\_pct FLOAT,

season STRING

)

STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'

WITH SERDEPROPERTIES (

"hbase.columns.mapping" = ":key,cf:player\_name,cf:team\_abbreviation,cf:age,cf:player\_height,cf:player\_weight,cf:college,cf:country,cf:draft\_year,cf:draft\_round,cf:draft\_number,cf:gp,cf:pts,cf:reb,cf:ast,cf:net\_rating,cf:oreb\_pct,cf:dreb\_pct,cf:usg\_pct,cf:ts\_pct,cf:ast\_pct,cf:season"

)

TBLPROPERTIES (

"hbase.table.name" = "hsiangct\_final\_NBA\_stats"

);

**Create a Kafka topic:**

kafka-topics.sh --create --zookeeper *z-3.mpcs53014kafka.o5ok5i.c4.kafka.us-east-2.amazonaws.com:2181,z-1.mpcs53014kafka.o5ok5i.c4.kafka.us-east-2.amazonaws.com:2181,z-2.mpcs53014kafka.o5ok5i.c4.kafka.us-east-2.amazonaws.com:2181* --replication-factor 1 --partitions 1 --topic hsiangct\_final\_NBA

**Test if the kafka topic is correctly listening:**

./kafka-console-consumer.sh --bootstrap-server b-3.mpcs53014kafka.o5ok5i.c4.kafka.us-east-2.amazonaws.com:9092,b-1.mpcs53014kafka.o5ok5i.c4.kafka.us-east-2.amazonaws.com:9092,b-2.mpcs53014kafka.o5ok5i.c4.kafka.us-east-2.amazonaws.com:9092 --topic hsiangct\_final\_NBA –from-beginning

**Run the webapp:**

ssh -i ~/.ssh/hsiangct\_mpcs53014.pem ec2-user@ec2-3-143-113-170.us-east-2.compute.amazonaws.com

node app.js 3074 ec2-3-131-137-149.us-east-2.compute.amazonaws.com 8070 b-2.mpcs53014kafka.o5ok5i.c4.kafka.us-east-2.amazonaws.com:9092,b-1.mpcs53014kafka.o5ok5i.c4.kafka.us-east-2.amazonaws.com:9092,b-3.mpcs53014kafka.o5ok5i.c4.kafka.us-east-2.amazonaws.com:9092

ec2-3-143-113-170.us-east-2.compute.amazonaws.com:3074/submit-stat.html

**Run the spark job:**

spark-submit --master local[2] --driver-java-options "-Dlog4j.configuration=file:///home/hadoop/ss.log4j.properties" --class StreamNBAStats uber-SpeedLayer-1.0-SNAPSHOT.jar b-2.mpcs53014kafka.o5ok5i.c4.kafka.us-east-2.amazonaws.com:9092,b-1.mpcs53014kafka.o5ok5i.c4.kafka.us-east-2.amazonaws.com:9092,b-3.mpcs53014kafka.o5ok5i.c4.kafka.us-east-2.amazonaws.com:9092