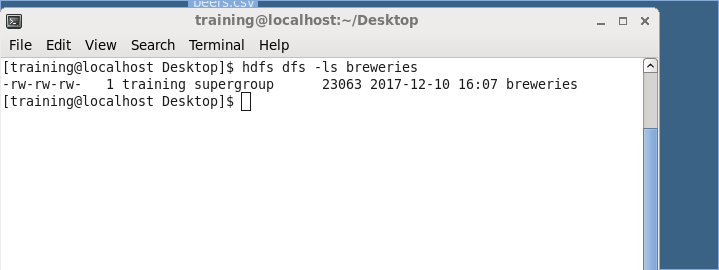
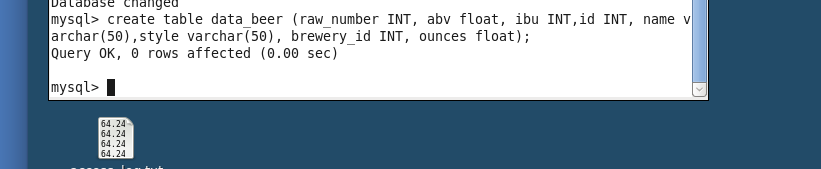
1.)

hdfs dfs -put Desktop/breweries.csv breweries

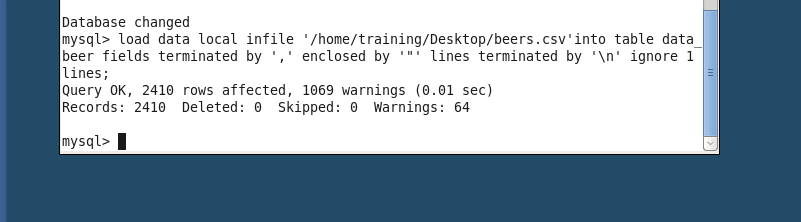


2.)

create table data\_beer (raw\_number INT, abv float, ibu INT,id INT, name varchar(50),style varchar(50), brewery\_id INT, ounces float);

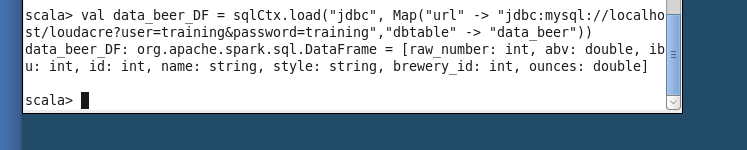


load data local infile '/home/training/Desktop/beers.csv' into table data\_beer fields terminated by ',' enclosed by '"' lines terminated by '\n' ignore 1 lines;

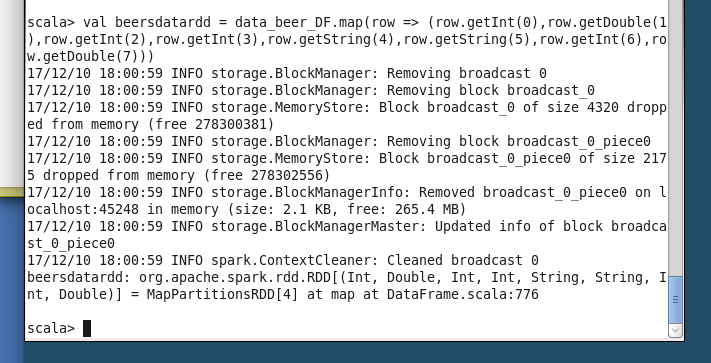


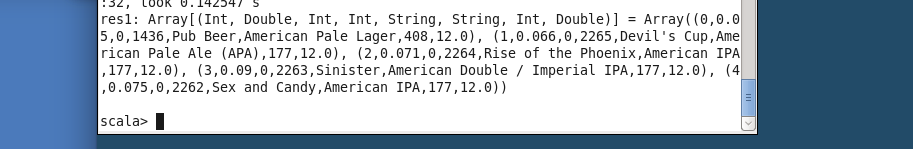


val data\_beer\_DF = sqlCtx.load("jdbc", Map("url" -> "jdbc:mysql://localhost/loudacre?user=training&password=training","dbtable" -> "data\_beer"))



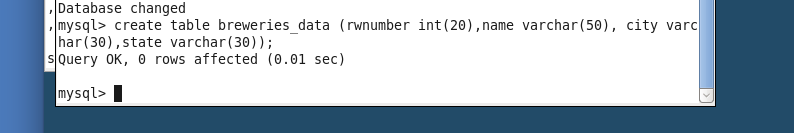
val beersdatardd = data\_beer\_DF.map(row => (row.getInt(0),row.getDouble(1),row.getInt(2),row.getInt(3),row.getString(4),row.getString(5),row.getInt(6),row.getDouble(7)))



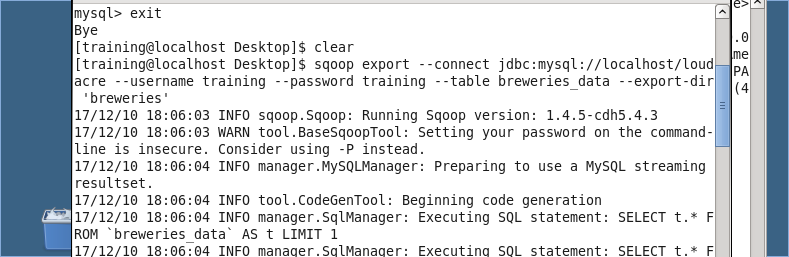


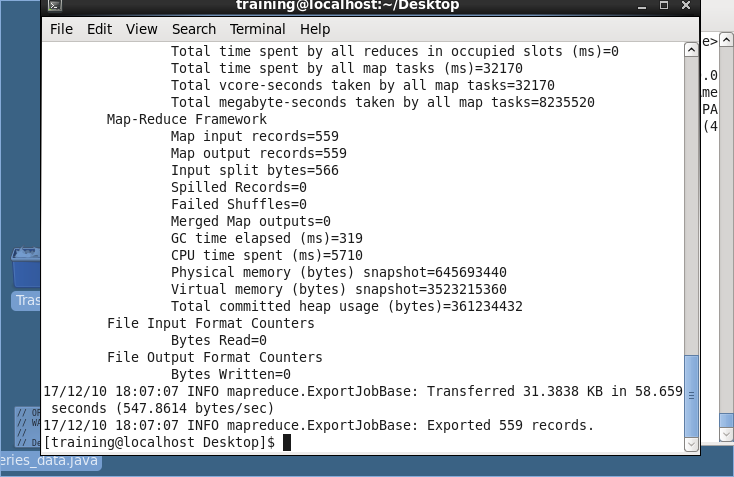
3.)

create table breweries\_data (rwnumber int(20),name varchar(50), city varchar(30),state varchar(30));

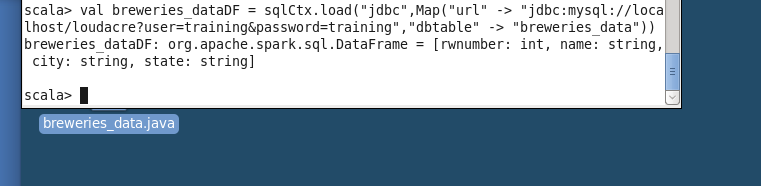


sqoop export --connect jdbc:mysql://localhost/loudacre --username training --password training --table breweries\_data --export-dir 'breweries'





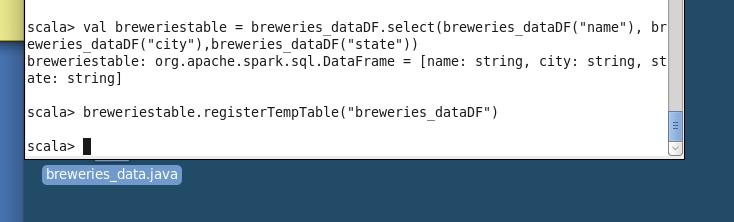
val breweries\_dataDF = sqlCtx.load("jdbc",Map("url"-> "jdbc:mysql://localhost/loudacre?user=training&password=training","dbtable" -> "breweries\_data"))



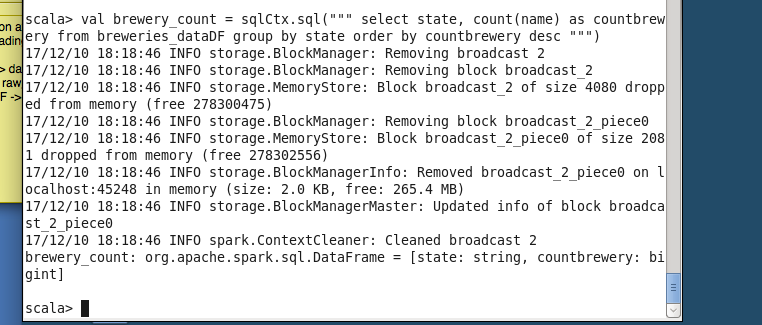
4.a) No. of breweries in each state

val breweriestable = breweries\_dataDF.select(breweries\_dataDF("name"),breweries\_dataDF("city"),breweries\_dataDF("state"))

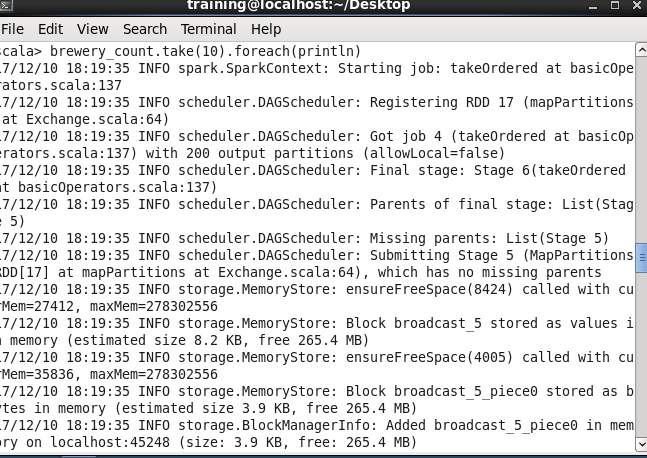
breweriestable.registerTempTable("breweries\_dataDF")

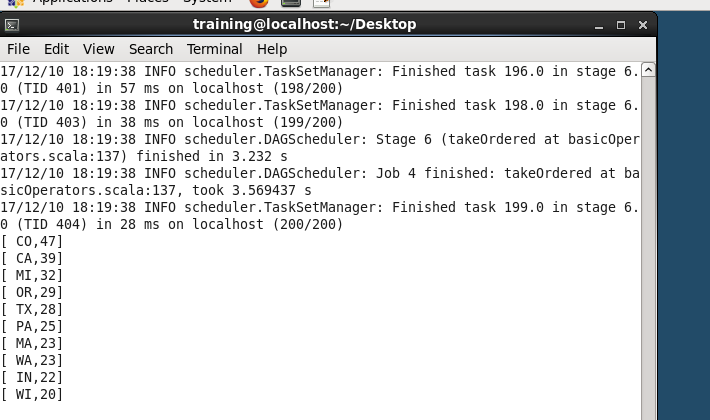


val brewery\_count = sqlCtx.sql("""select state, count(name) as countbrewery from breweries\_dataDF group by state order by countbrewery desc""")



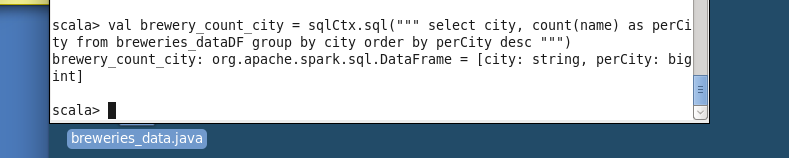
countbrewery.take(10).foreach(println)

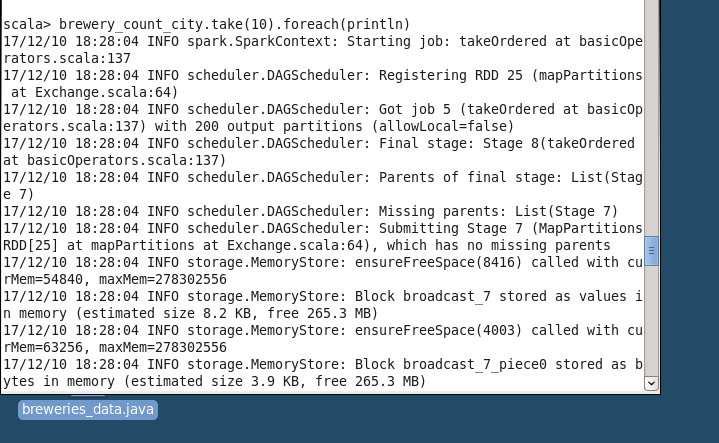


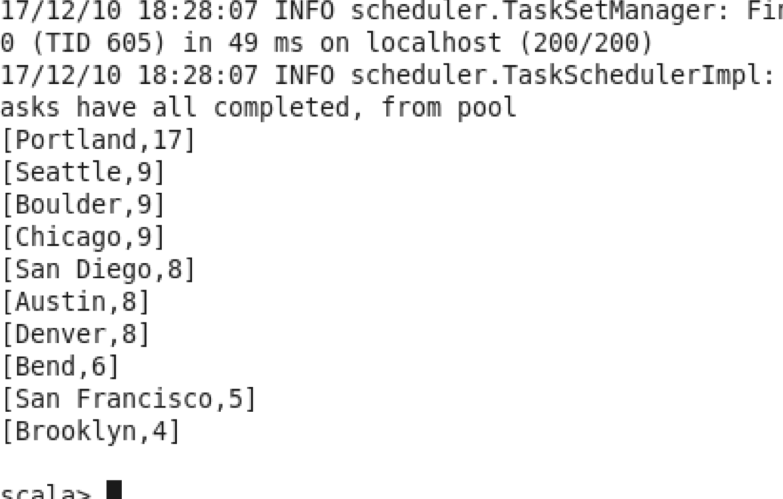


4.b) Portland with 17 breweries

val brewery\_count\_city = sqlCtx.sql(""" select city, count(name) as perCity from breweries\_dataDF group by city order by perCity desc """)



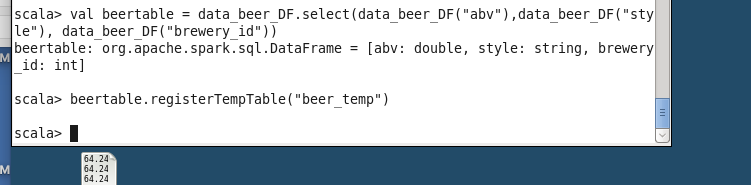




c.) Average alcohol % by Volume

val beertable = data\_beer\_DF.select(data\_beer\_DF(“abv”),data\_beer\_DF(“style”),data\_beer\_DF(“brewery\_id”))

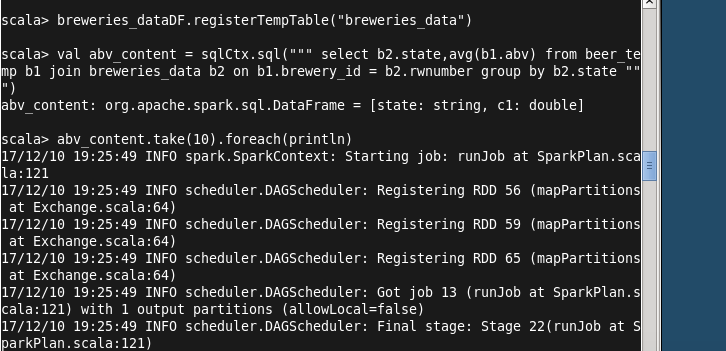
beertable.registerTempTable(“beer\_temp”)

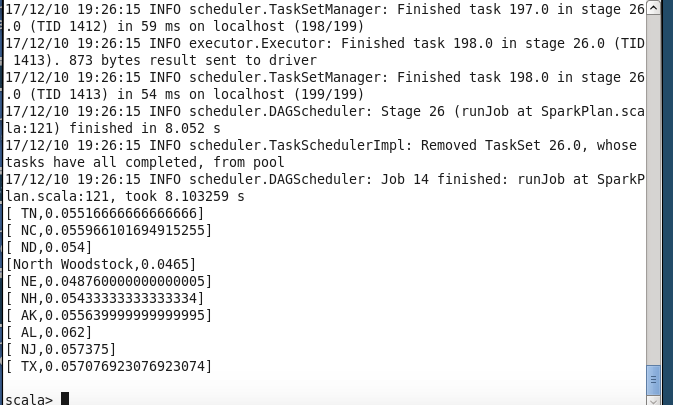


Breweries\_dataDF.registerTempTable(“breweries\_data”)

Val abv\_content = sqlCtx.sql(“”” select b2.state,avg(b1.abv) from beer\_temp b1 join breweries\_data b2 on b1.brewery\_id = b2.rwnumber group by b2.state “””)

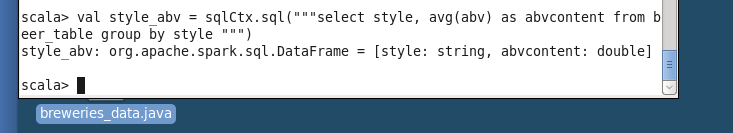
abv\_content.take(10).foreach(println)



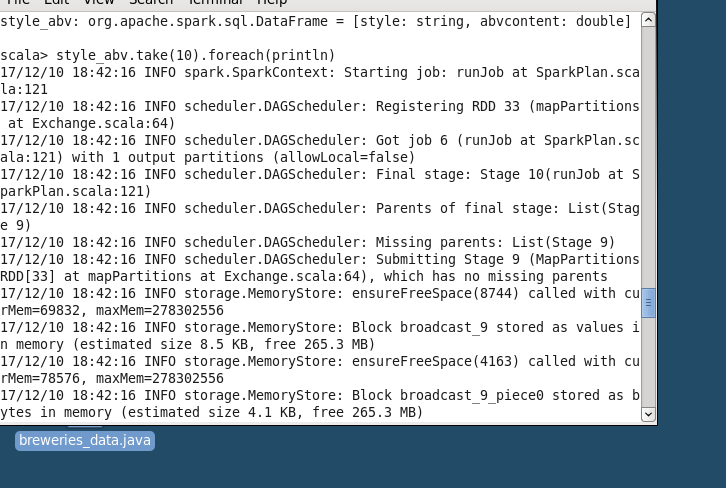


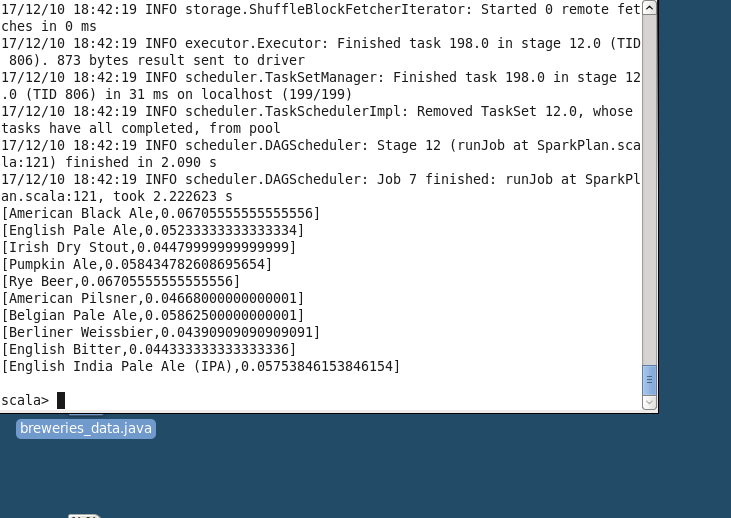
d.) American Black with 6.7% abv

val style\_abv = sqlCtx.sql("""select style, avg(abv) as abvcontent from beer\_table group by style """)



style\_abv.take(10).foreach(println)





4.e) American IPA with 424 count of brews

val count\_style=sqlCtx.sql("""select style, count(style) as countstyle from beer\_table group by style order by countstyle desc""")



count\_style.take(10).foreach(println)



