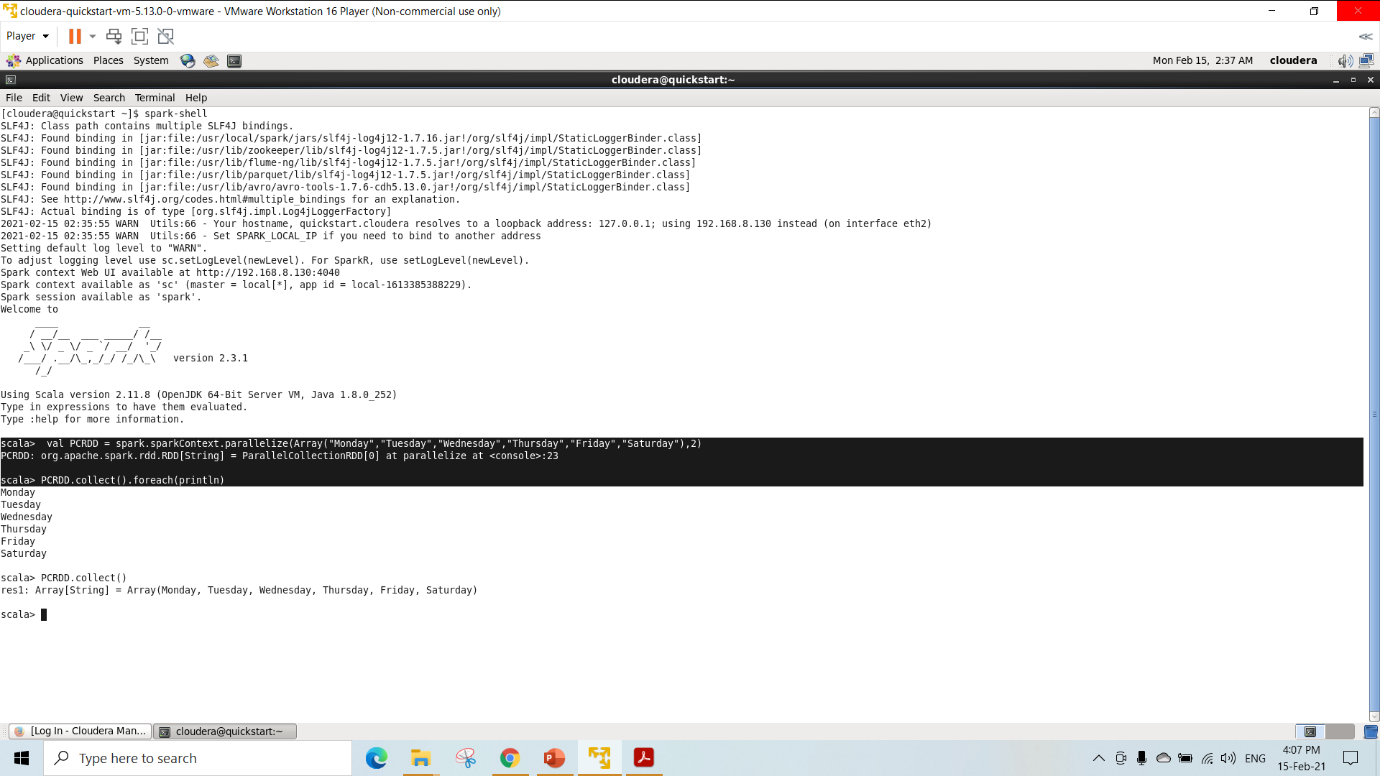
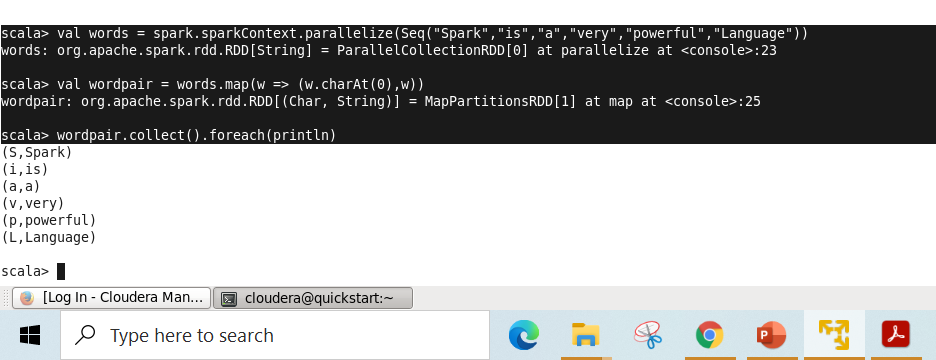
Independent Programming using Spark

Able to analyze data using RDDS and Data Frames

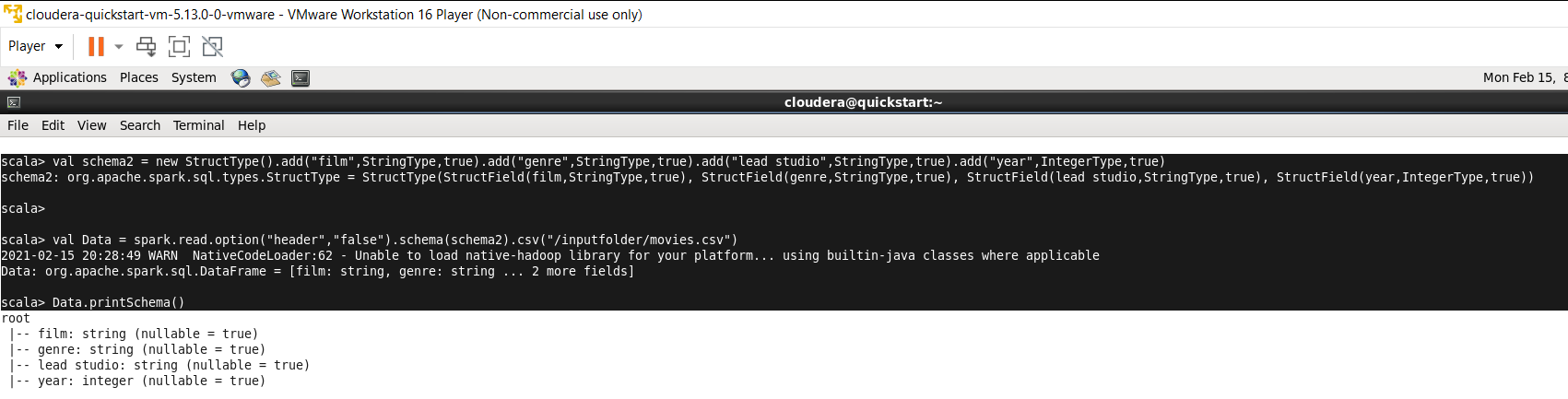
Creating RDD with parallelize



Create RDD from another RDD

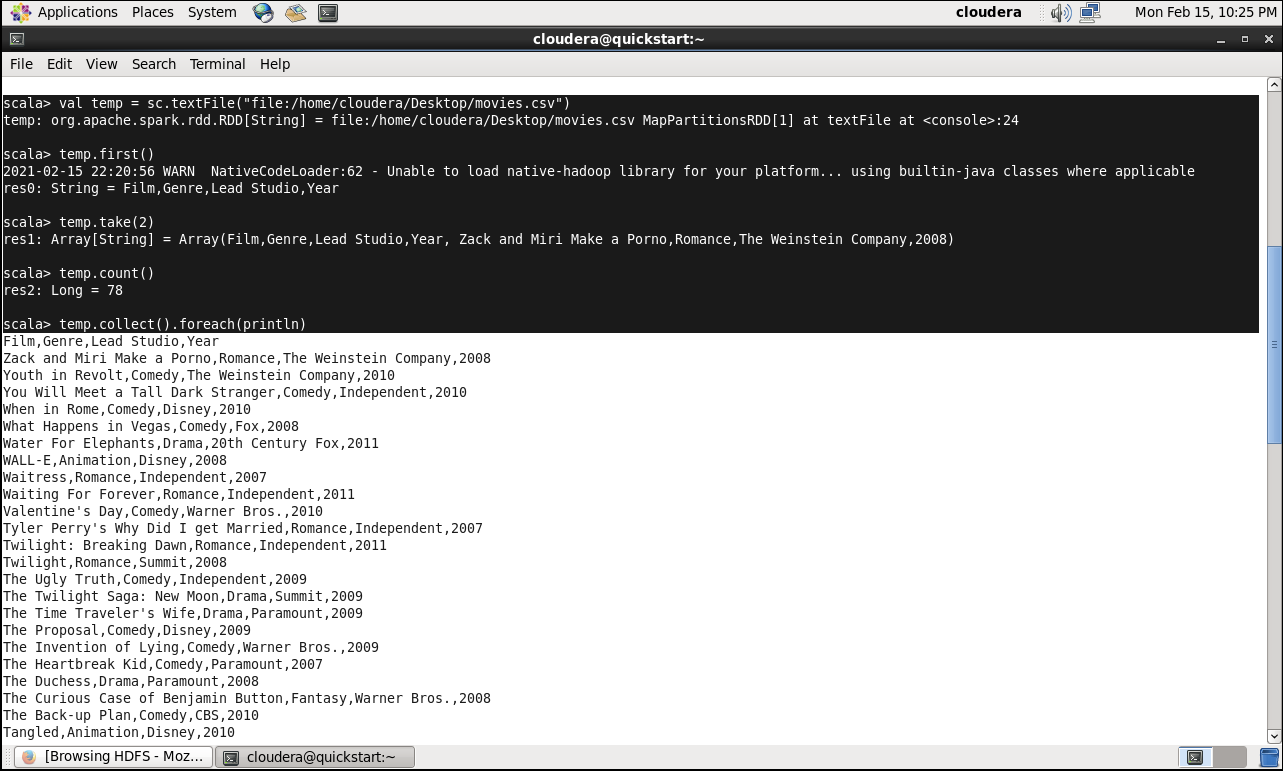


Creating reading and printing schema

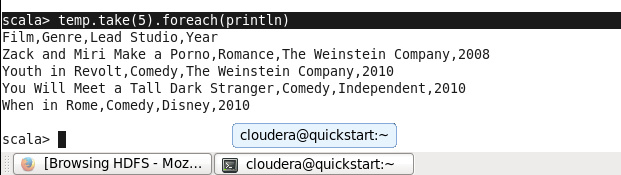


Operations performed on RDD

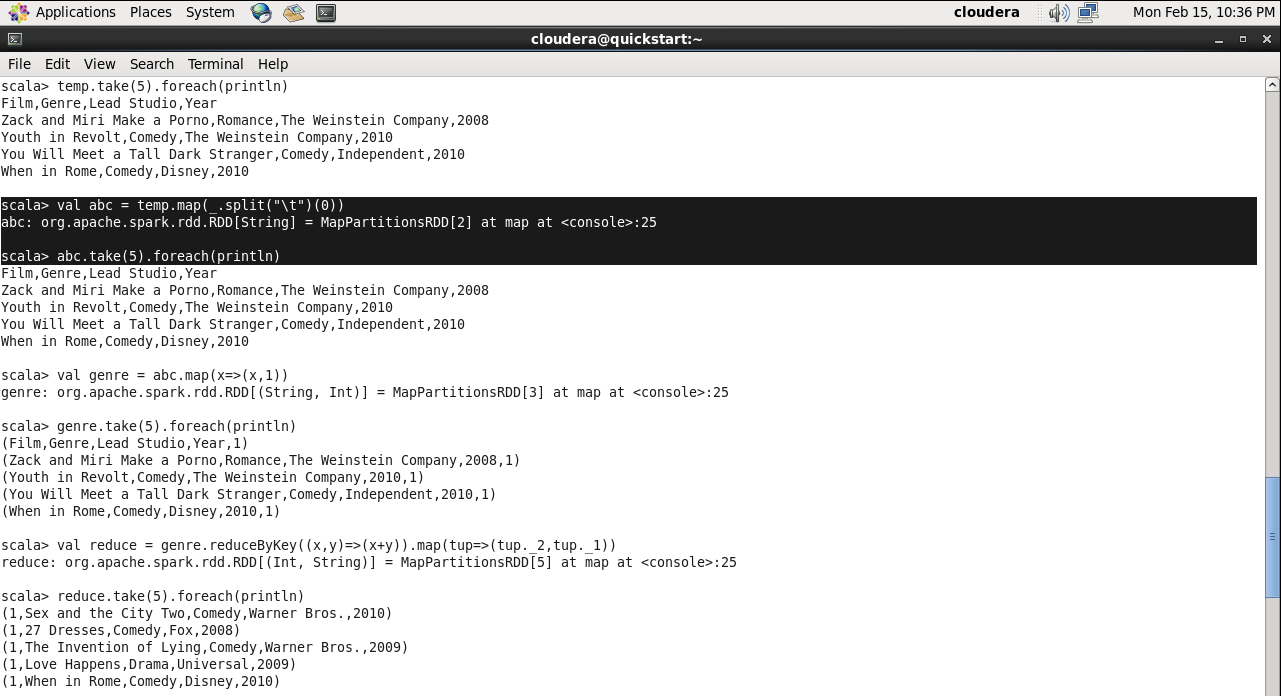
Actions



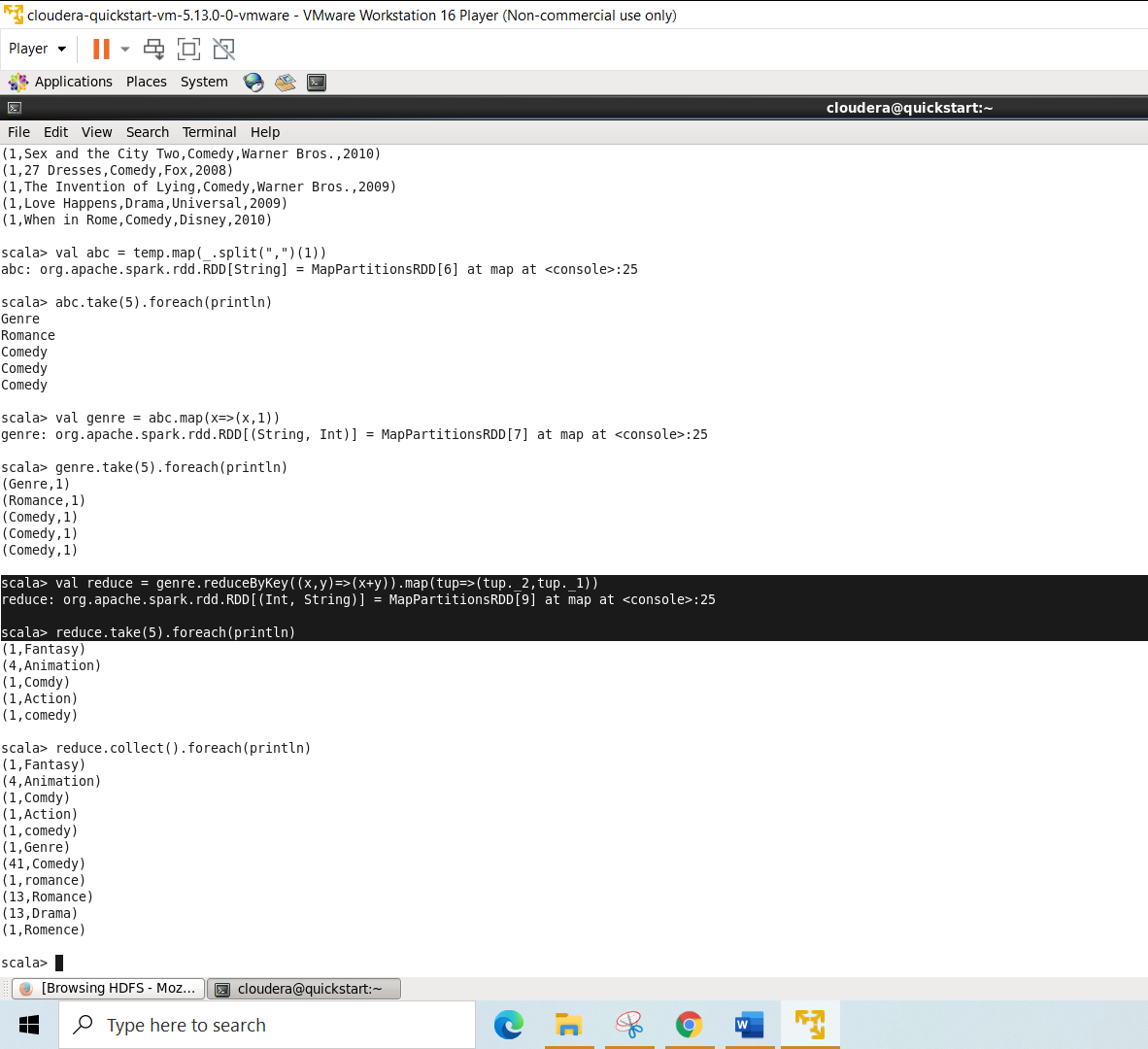
Printing first 5 records

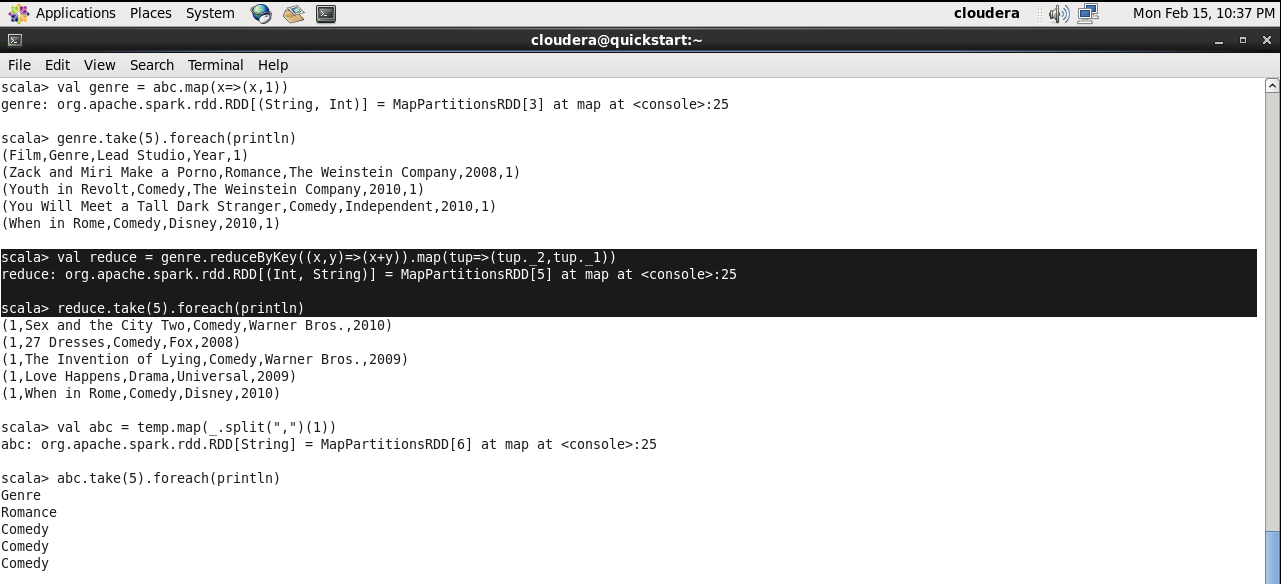


Transformations

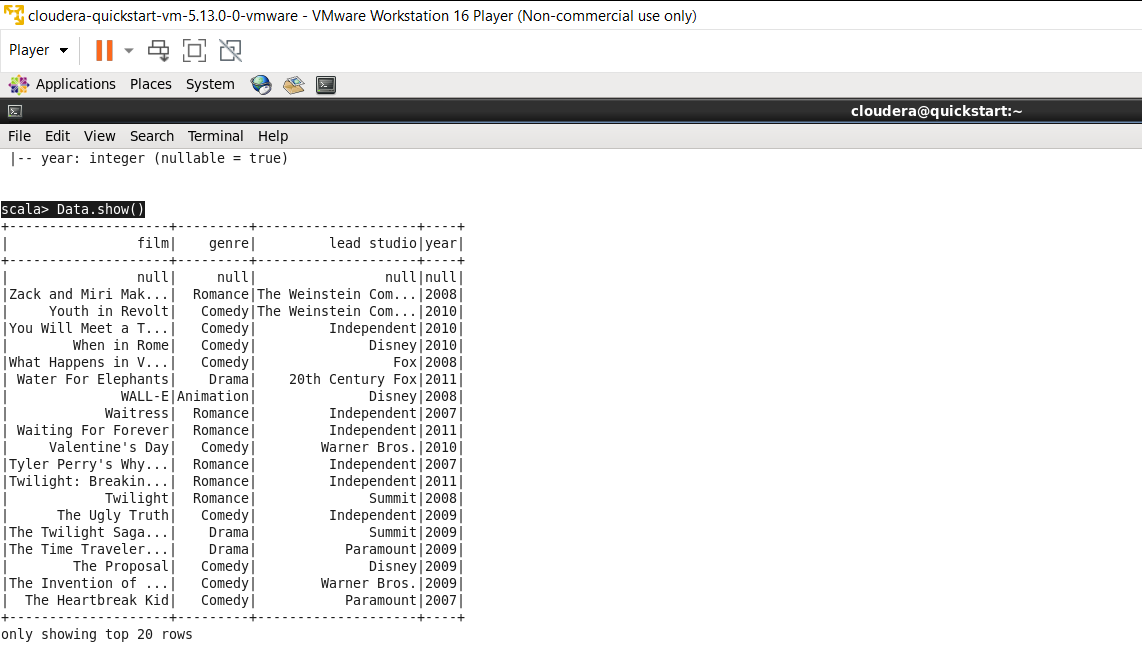


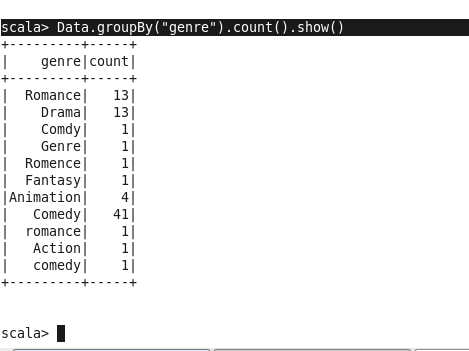
Reduce by key foreach



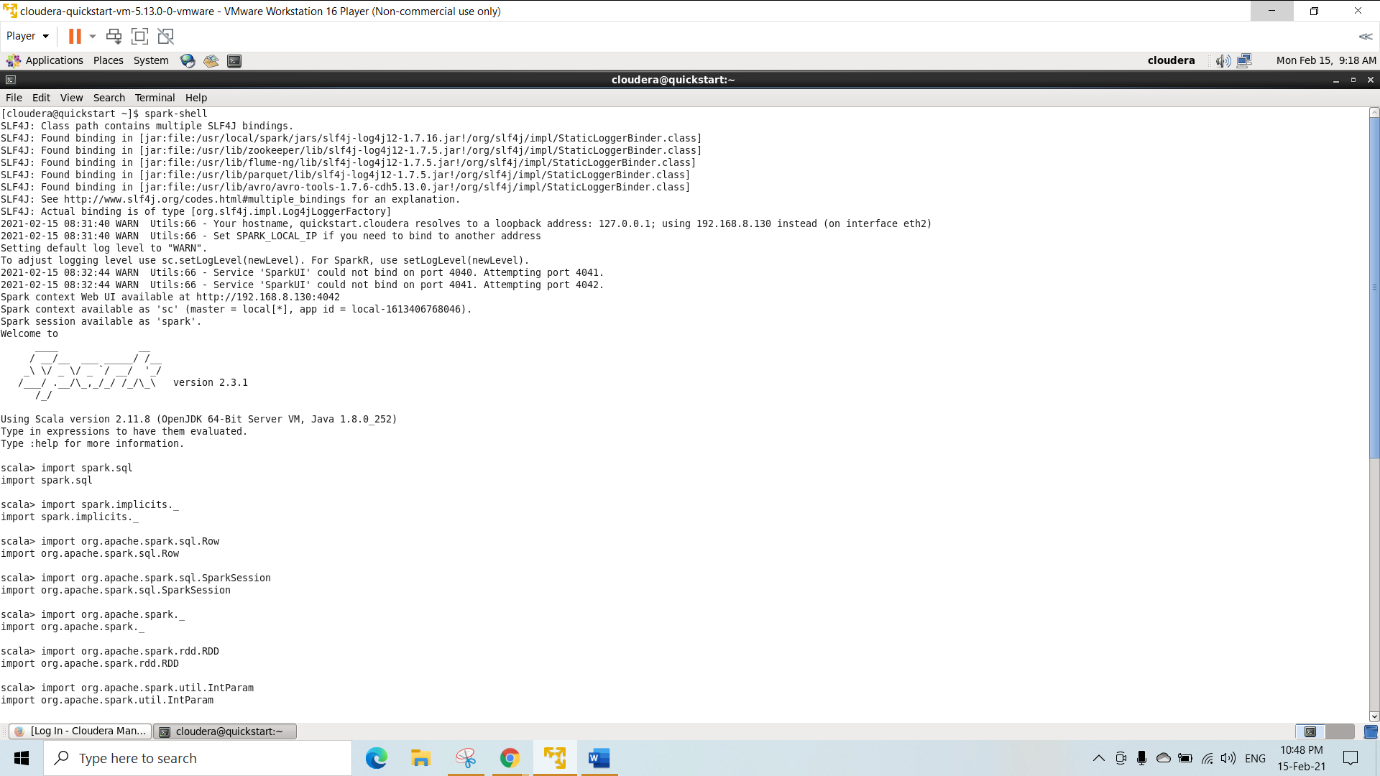
reduceByKey with map

Showing data



Groupthe dataframe according to the genre and display its count

spark-shell command and Importing all packages to run the spark



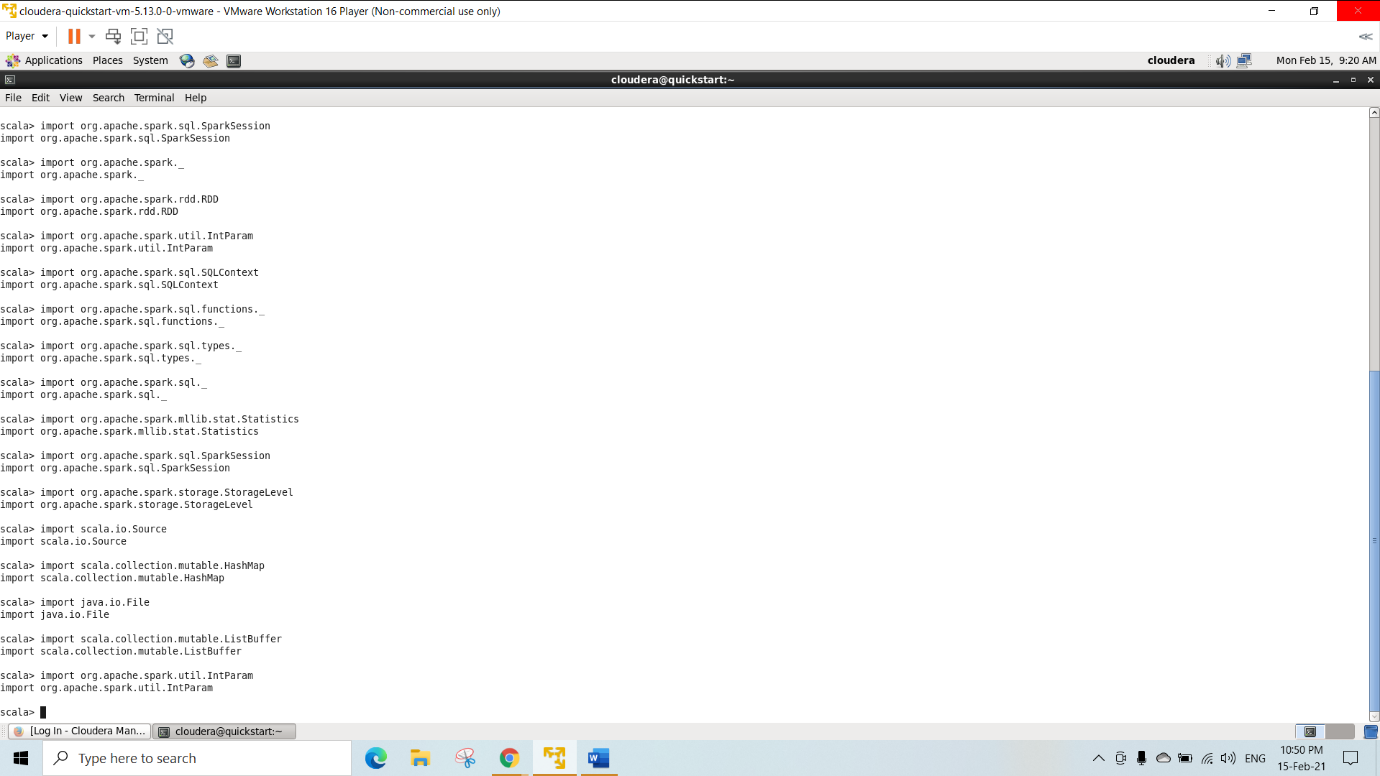
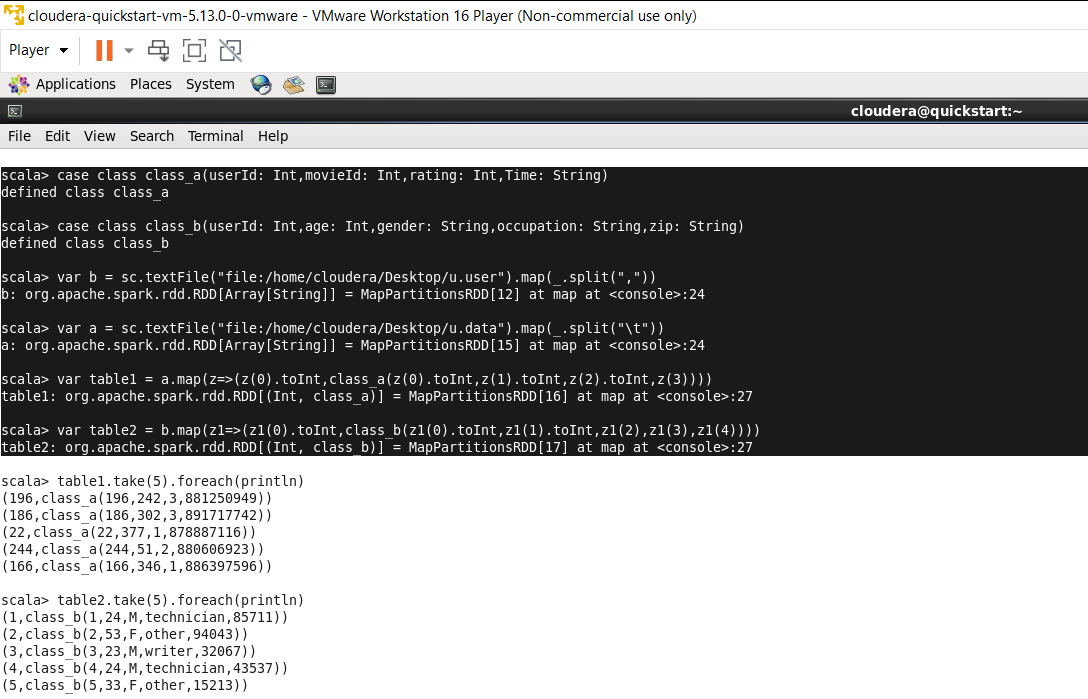
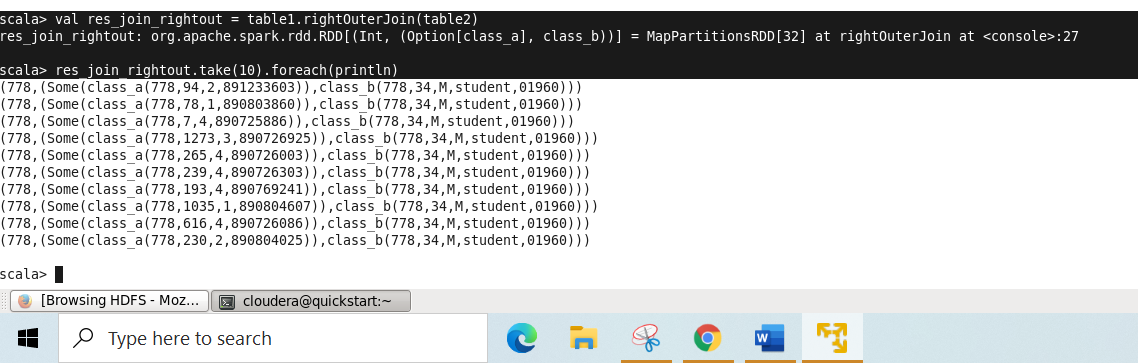


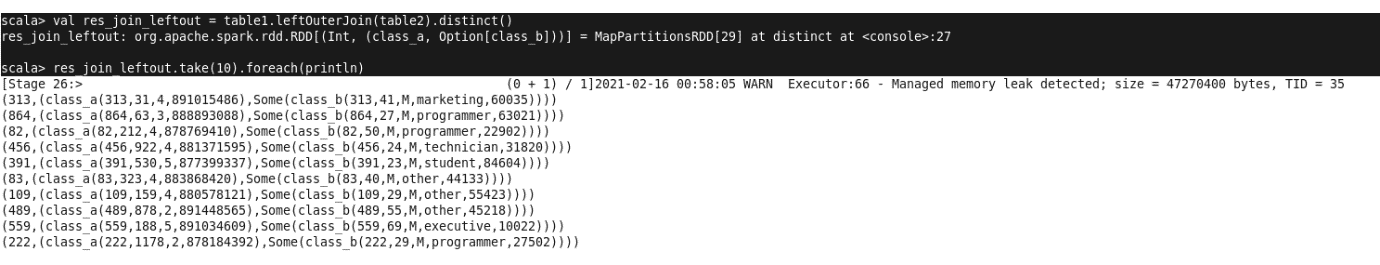
Table creation for joins



right outer join with output



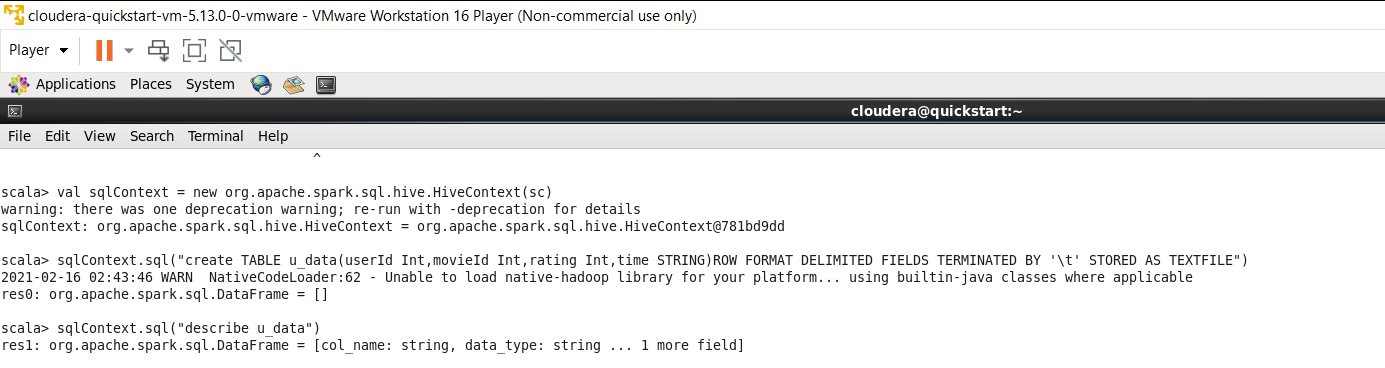
Leftouterjoin with output

Inner or Full join with output

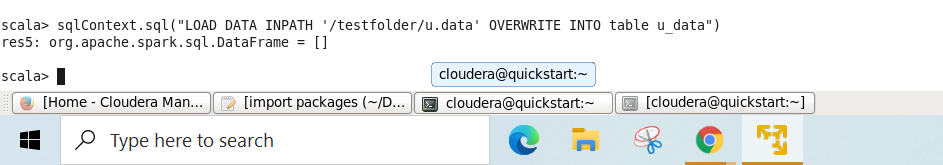


Able to Query data using Spark SQL

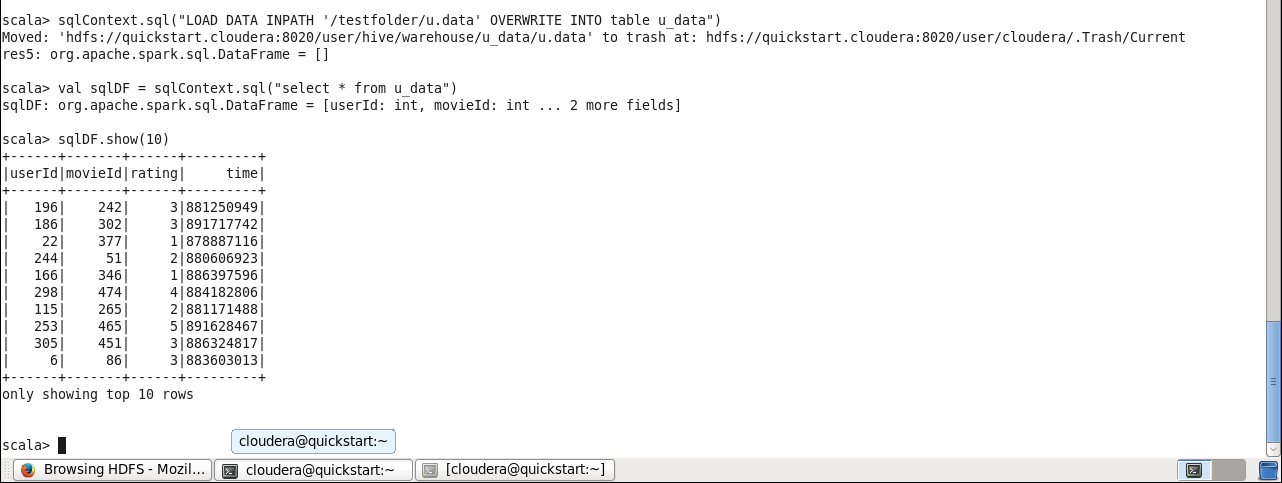
Apache sql and creating table(u\_data)



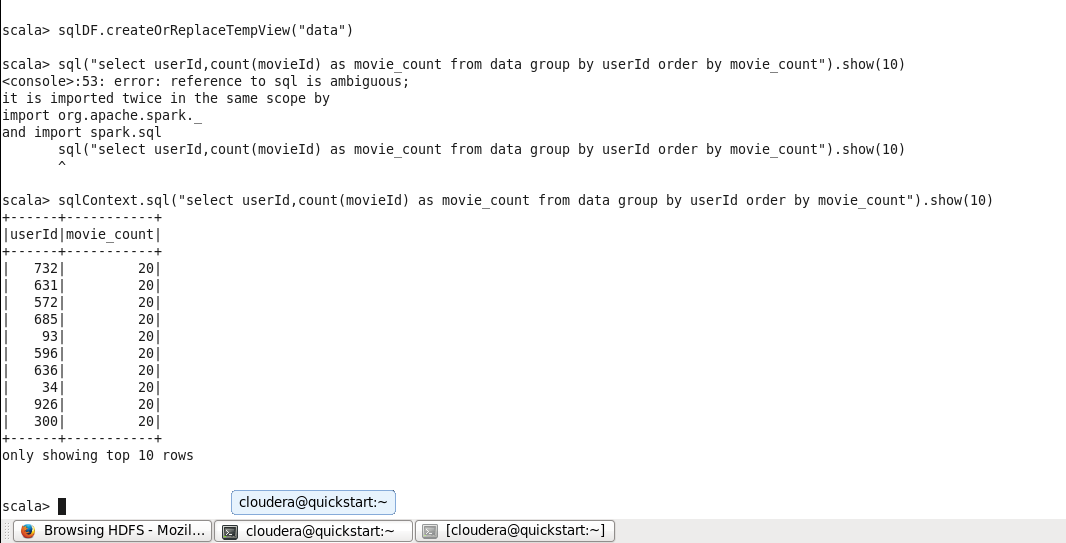
Loading data to the table(put table u.data into testfolder in Hadoop using other terminal)



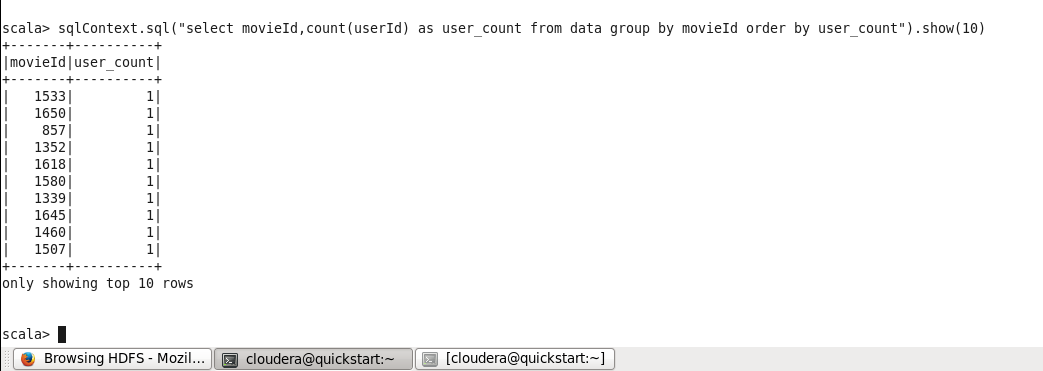
Displaying first 10 rows



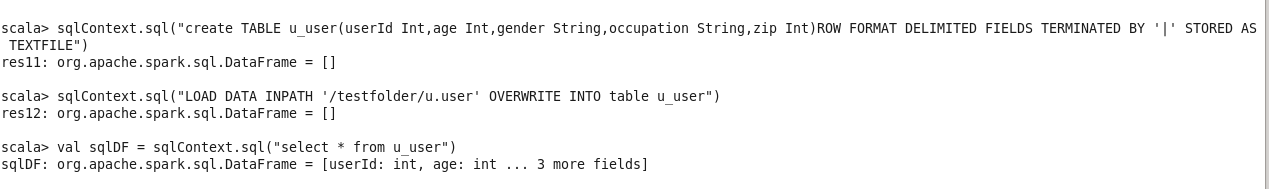
creating temporary view and displaying count of movies watched by each user



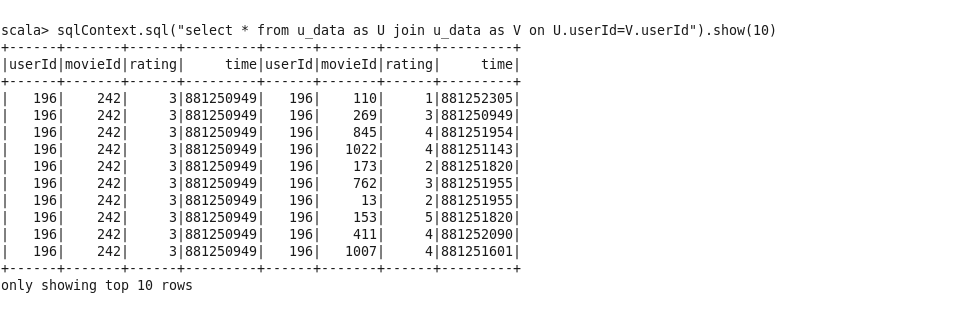
Query to get the count of users who watched particular movie



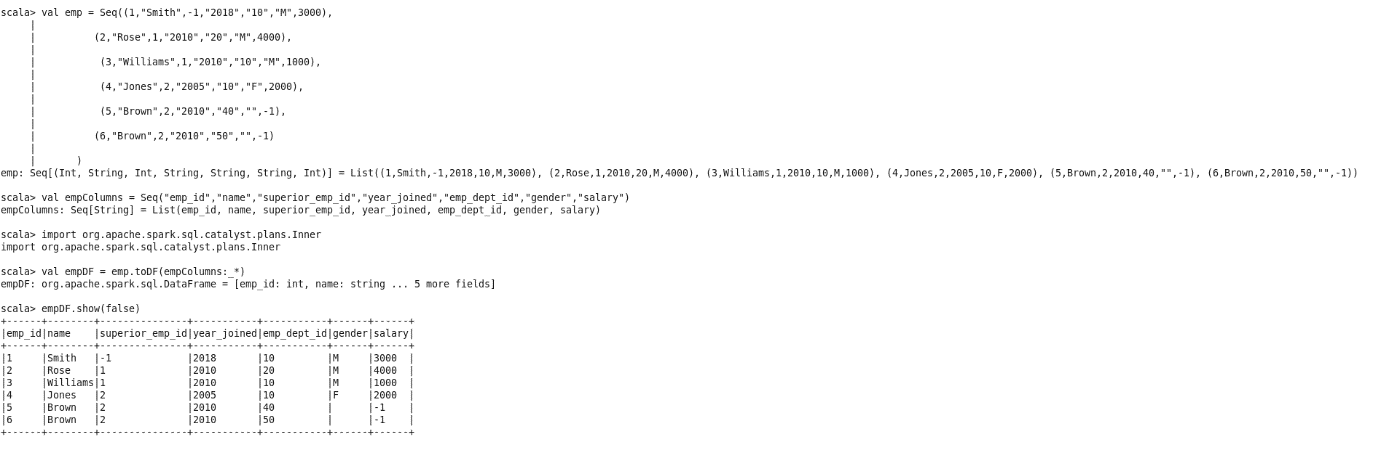
Created table u\_user and get the view

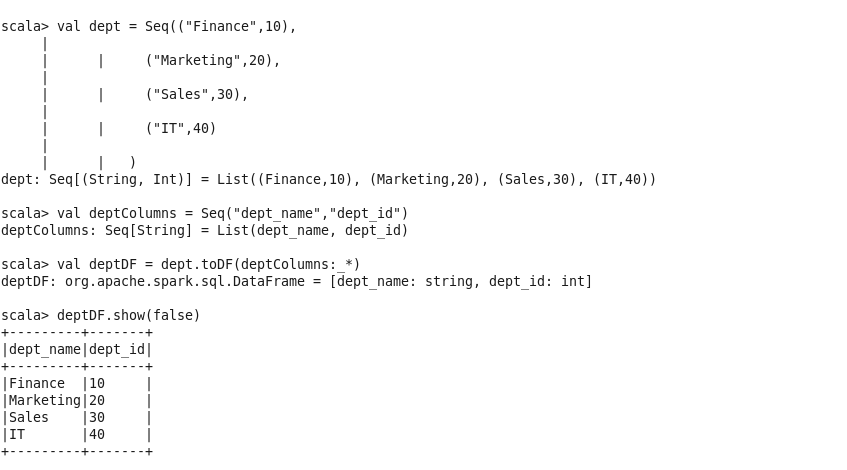


Self Join for u\_data table

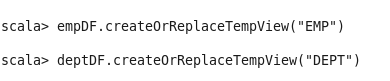


Creating employee sequence and creating schema

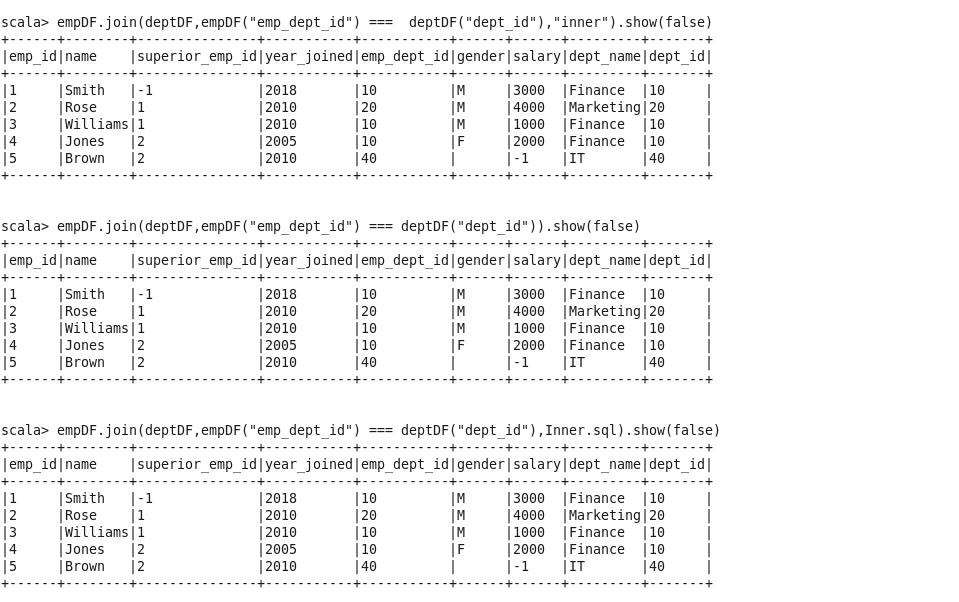


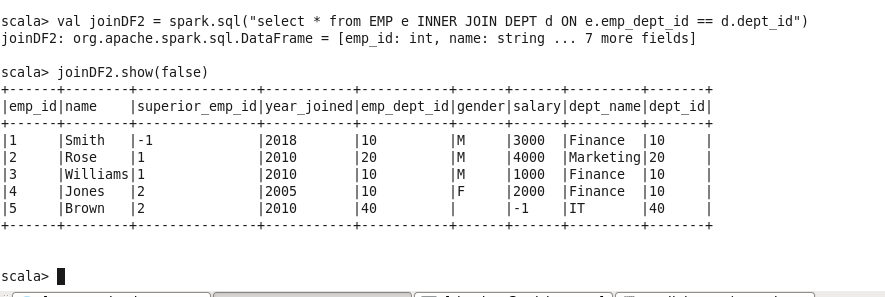
Creating department sequence and creating schema

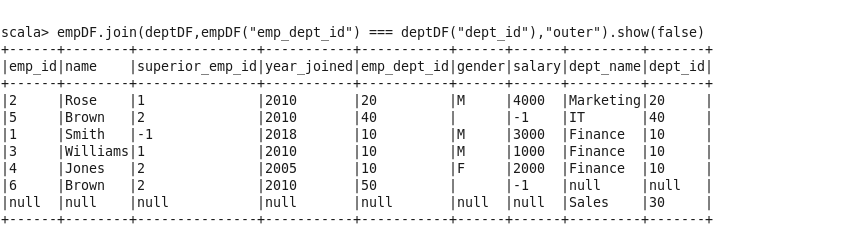
create views



inner joins



outer join



cross join

