This assignment will help you to consolidate the concepts learnt in the session.

2. Problem Statement

Go through below blog and reiterate the same at your end.

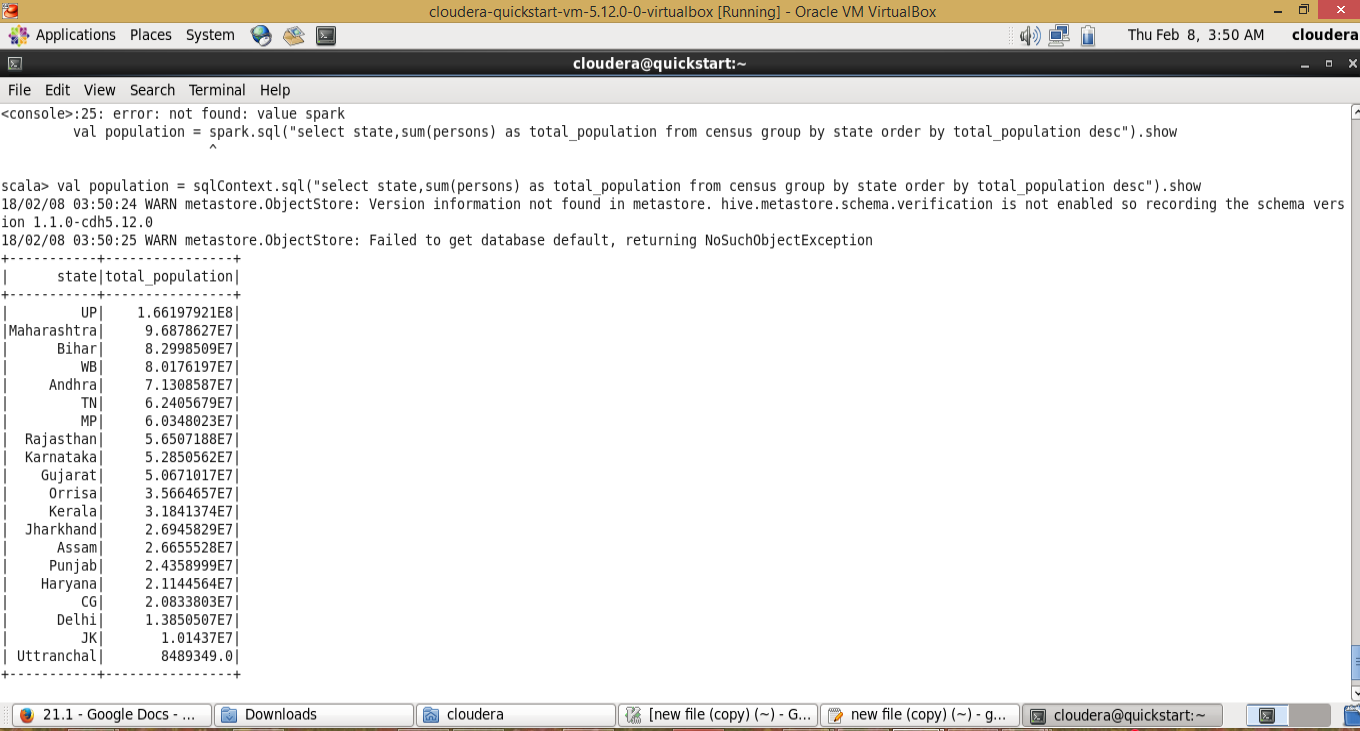
https://docs.google.com/document/d/1csLBlMiEXs\_hXWV2Z8VpBlrj\_R6RoDQLlZUnA0uBT

Ck/edit

1. Find out the state wise population and order by state

*val census\_data = sc.textFile("file:///home/cloudera/census.csv").map(x => x.split(",")).map(x => (x(0),x(2),x(3),x(4),x(5),x(6),x(7),x(8),x(9),x(10),x(11),x(12),x(13),x(14),x(15),x(16),x(17),x(18),x(19),x(20),x(21),x(22))).toDF("State" ,"Persons","Males" ,"Females" ,"Growth\_1991\_2001" ,"Rural" ,"Urban" ,"Scheduled\_Caste\_population" ,"Percentage\_SC\_to\_total" ,"Number\_of\_households" ,"Household\_size\_per\_household" ,"Sex\_ratio\_females\_per\_1000\_males " ,"Sex\_ratio\_0\_6\_years" ,"Scheduled\_Tribe\_population" ,"Percentage\_to\_total\_population\_ST" ,"Persons\_literate" ,"Males\_Literate" ,"Females\_Literate" ,"Persons\_literacy\_rate" ,"Males\_Literatacy\_Rate" ,"Females\_Literacy\_Rate" ,"Total\_Educated").registerTempTable("census")*

*val population = sqlContext.sql("select state,sum(persons) as total\_population from census group by state order by total\_population desc").show*



2. Find out the Growth Rate of Each State Between 1991-2001

*val census\_data = sc.textFile("file:///home/cloudera/census.csv").map(x => x.split(",")).map(x => (x(0),x(2),x(3),x(4),x(5),x(6),x(7),x(8),x(9),x(10),x(11),x(12),x(13),x(14),x(15),x(16),x(17),x(18),x(19),x(20),x(21),x(22))).toDF("State" ,"Persons","Males" ,"Females" ,"Growth\_1991\_2001" ,"Rural" ,"Urban" ,"Scheduled\_Caste\_population" ,"Percentage\_SC\_to\_total" ,"Number\_of\_households" ,"Household\_size\_per\_household" ,"Sex\_ratio\_females\_per\_1000\_males " ,"Sex\_ratio\_0\_6\_years" ,"Scheduled\_Tribe\_population" ,"Percentage\_to\_total\_population\_ST" ,"Persons\_literate" ,"Males\_Literate" ,"Females\_Literate" ,"Persons\_literacy\_rate" ,"Males\_Literatacy\_Rate" ,"Females\_Literacy\_Rate" ,"Total\_Educated").registerTempTable("census")*

*val growth\_rate = sqlContext.sql("select state,avg(Growth\_1991\_2001) as total\_growth from census group by state").show*

