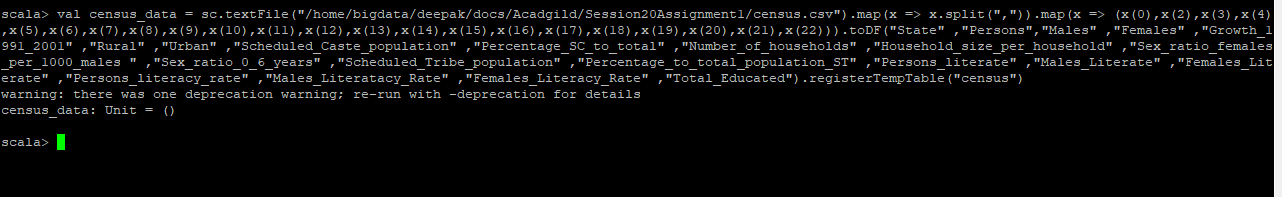
# Session 20 Assignment 1

**First, read the data from the csv file**

val census\_data = sc.textFile("/home/bigdata/deepak/docs/Acadgild/Session20Assignment1/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")



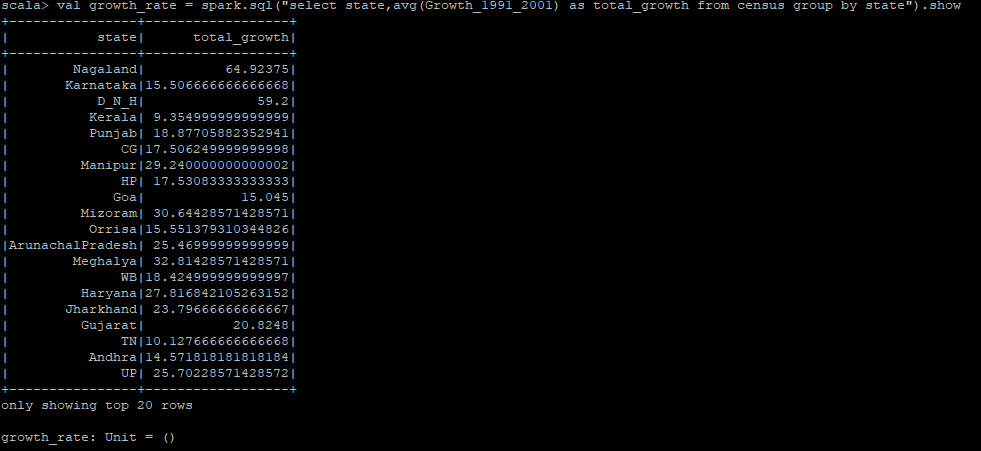
1. **Find out the state wise population and order by state**

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



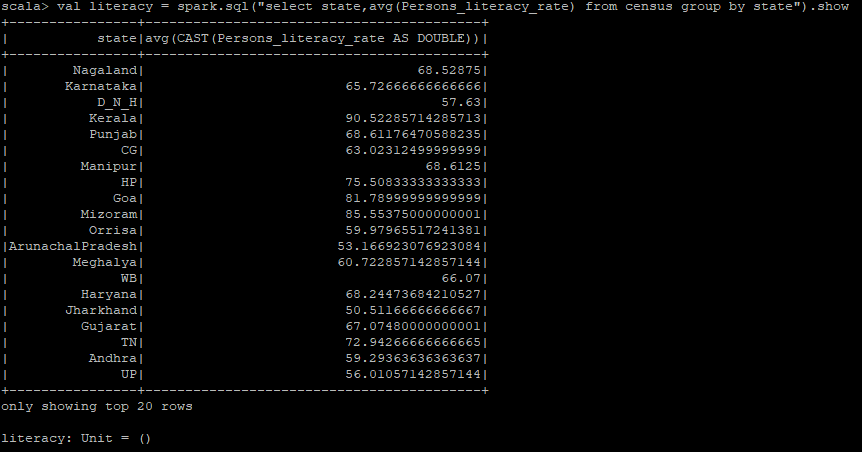
1. **Find out the Growth Rate of Each State Between 1991-2001**

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



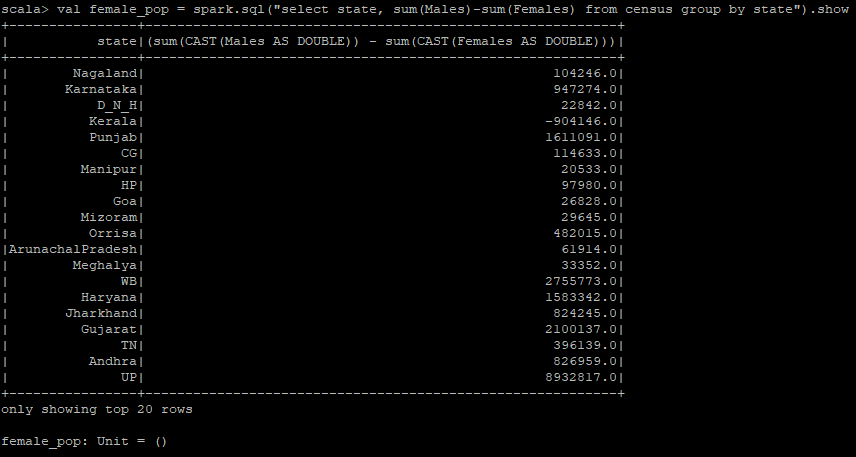
1. **Find the literacy rate of each state**

val literacy = spark.sql("select state,avg(Persons\_literacy\_rate) from census group by state").show



1. **Find out the States with More Female Population**

val female\_pop = spark.sql("select state, sum(Males)-sum(Females) from census group by state").show



1. **Find out the Percentage of Population in Every State**

val percenet\_pop = spark.sql("select state, (sum(persons) \* 100.0) / SUM(sum(persons)) over() as percent\_pop\_by\_state from census group by state").show

