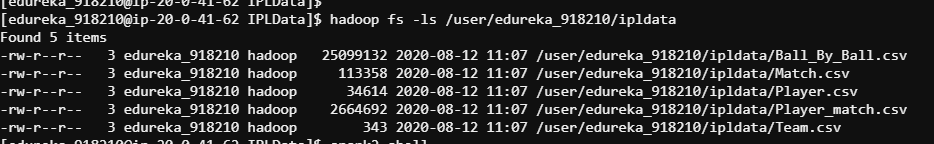
**PLAYING WITH IPL DATA IN SPARK**

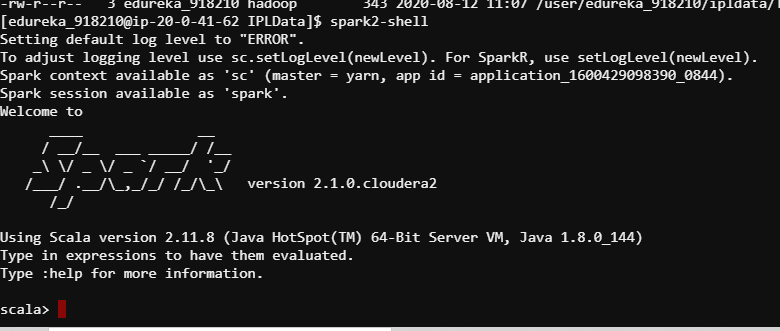
IPL DATA DOWNLOADED :

<https://data.world/raghu543/ipl-data-till-2017#:~:text=Overview,-Description&text=This%20data%20set%20has%20the,R%20Script%20%2CSQL%2CSSIS>.

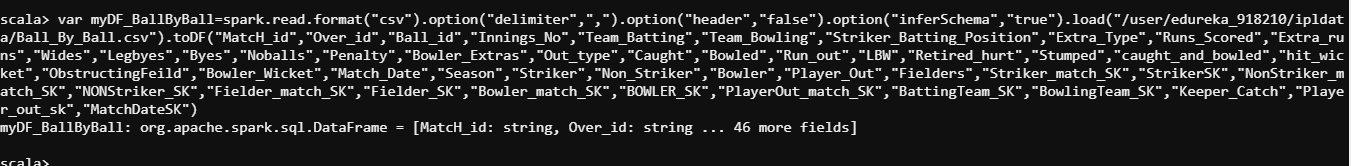
SITE : http : IPL DATASETS

hadoop fs -ls /user/edureka\_918210/ipldata

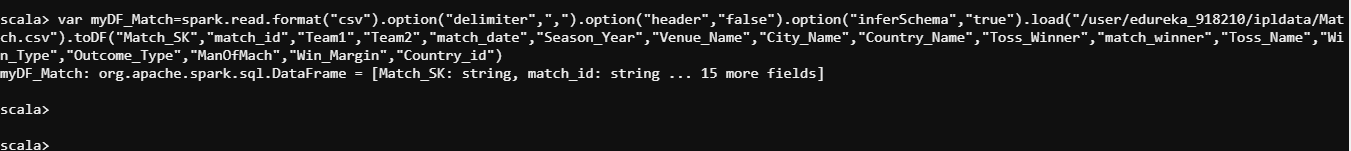




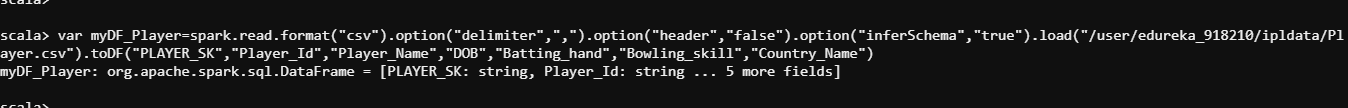
var myDF\_BallByBall=spark.read.format("csv").option("delimiter",",").option("header","false").option("inferSchema","true").load("/user/edureka\_918210/ipldata/Ball\_By\_Ball.csv").toDF("MatcH\_id","Over\_id","Ball\_id","Innings\_No","Team\_Batting","Team\_Bowling","Striker\_Batting\_Position","Extra\_Type","Runs\_Scored","Extra\_runs","Wides","Legbyes","Byes","Noballs","Penalty","Bowler\_Extras","Out\_type","Caught","Bowled","Run\_out","LBW","Retired\_hurt","Stumped","caught\_and\_bowled","hit\_wicket","ObstructingFeild","Bowler\_Wicket","Match\_Date","Season","Striker","Non\_Striker","Bowler","Player\_Out","Fielders","Striker\_match\_SK","StrikerSK","NonStriker\_match\_SK","NONStriker\_SK","Fielder\_match\_SK","Fielder\_SK","Bowler\_match\_SK","BOWLER\_SK","PlayerOut\_match\_SK","BattingTeam\_SK","BowlingTeam\_SK","Keeper\_Catch","Player\_out\_sk","MatchDateSK")



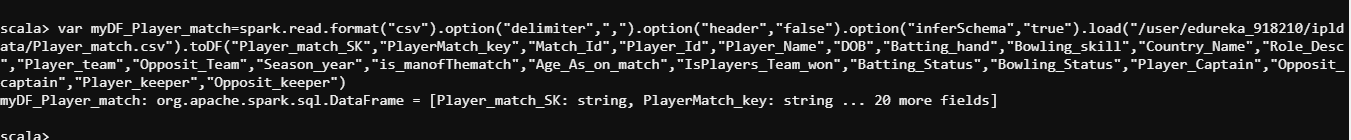
var myDF\_Match=spark.read.format("csv").option("delimiter",",").option("header","false").option("inferSchema","true").load("/user/edureka\_918210/ipldata/Match.csv").toDF("Match\_SK","match\_id","Team1","Team2","match\_date","Season\_Year","Venue\_Name","City\_Name","Country\_Name","Toss\_Winner","match\_winner","Toss\_Name","Win\_Type","Outcome\_Type","ManOfMach","Win\_Margin","Country\_id")



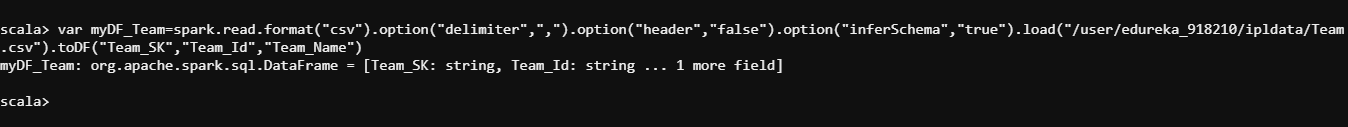
var myDF\_Player=spark.read.format("csv").option("delimiter",",").option("header","false").option("inferSchema","true").load("/user/edureka\_918210/ipldata/Player.csv").toDF("PLAYER\_SK","Player\_Id","Player\_Name","DOB","Batting\_hand","Bowling\_skill","Country\_Name")



var myDF\_Player\_match=spark.read.format("csv").option("delimiter",",").option("header","false").option("inferSchema","true").load("/user/edureka\_918210/ipldata/Player\_match.csv").toDF("Player\_match\_SK","PlayerMatch\_key","Match\_Id","Player\_Id","Player\_Name","DOB","Batting\_hand","Bowling\_skill","Country\_Name","Role\_Desc","Player\_team","Opposit\_Team","Season\_year","is\_manofThematch","Age\_As\_on\_match","IsPlayers\_Team\_won","Batting\_Status","Bowling\_Status","Player\_Captain","Opposit\_captain","Player\_keeper","Opposit\_keeper")



var myDF\_Team=spark.read.format("csv").option("delimiter",",").option("header","false").option("inferSchema","true").load("/user/edureka\_918210/ipldata/Team.csv").toDF("Team\_SK","Team\_Id","Team\_Name")



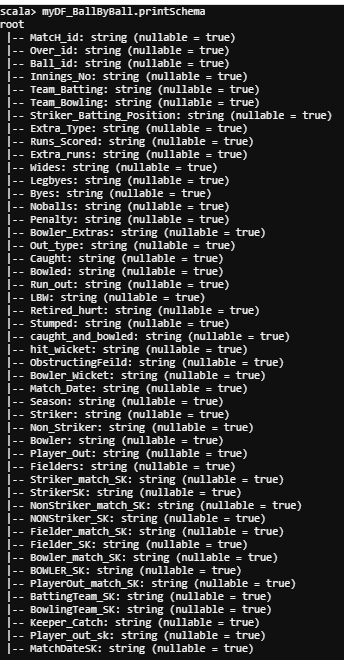
myDF\_BallByBall.printSchema

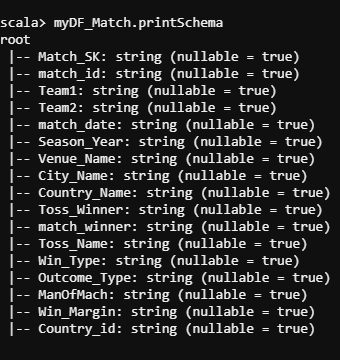
myDF\_Match.printSchema

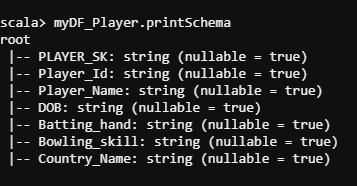
myDF\_Player.printSchema

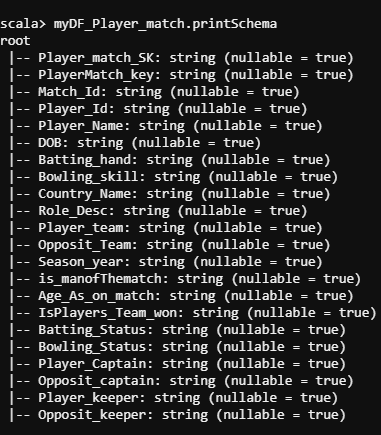
myDF\_Player\_match.printSchema

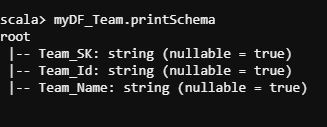
myDF\_Team.printSchema











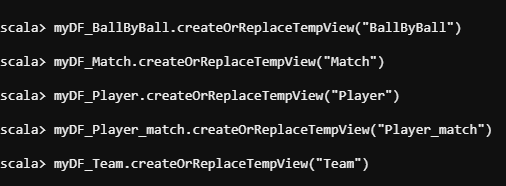
myDF\_BallByBall.createOrReplaceTempView("BallByBall")

myDF\_Match.createOrReplaceTempView("Match")

myDF\_Player.createOrReplaceTempView("Player")

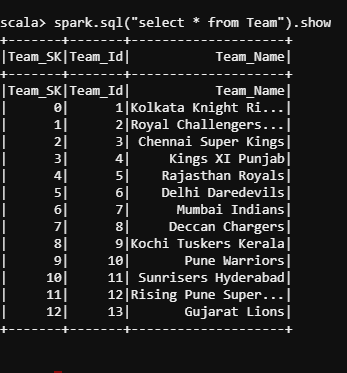
myDF\_Player\_match.createOrReplaceTempView("Player\_match")

myDF\_Team.createOrReplaceTempView("Team")

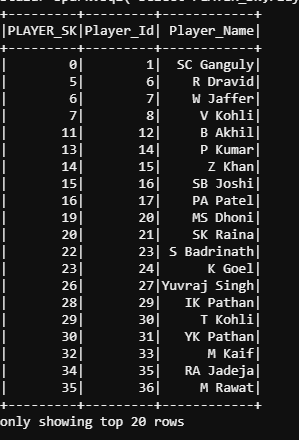


spark.sql("describe Team").show

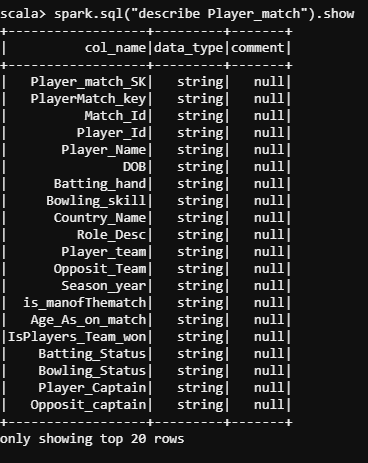
spark.sql("select \* from Team").show



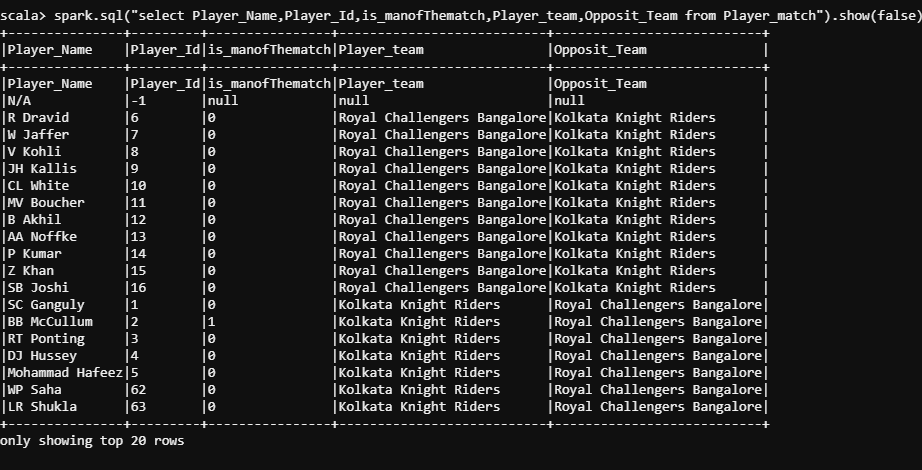
spark.sql("select PLAYER\_SK,Player\_Id,Player\_Name from Player where Country\_Name='India'").show



spark.sql("describe Player\_match").show

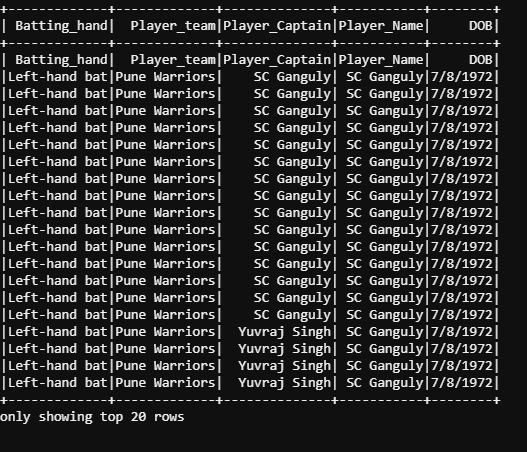


spark.sql("select Player\_Name,Player\_Id,is\_manofThematch,Player\_team,Opposit\_Team from Player\_match").show(false)



**INNER JOIN**

spark.sql("SELECT Player\_match.Batting\_hand,Player\_match.Player\_team,Player\_match.Player\_Captain,Player.Player\_Name,Player.DOB FROM Player INNER JOIN Player\_match ON Player.Player\_Id = Player\_match.Player\_Id").show



spark.sql("SELECT Player\_match.Batting\_hand,Player\_match.Player\_team,Player\_match.Player\_Captain,Player.Player\_Name,Player.DOB FROM Player RIGHT JOIN Player\_match ON Player.Player\_Id = Player\_match.Player\_Id").show

