# Project: Runs scored by Players

**Data Set:** This is data for runs scored by players in different countries in different years. Let's assume some external process is writing data into a directory in CSV format where columns of the data are as shown below:

```
Player_id, Year, Country, Opposition_Team, Runs_Scored, Balls_Played

9,2000,India,NewZealand,1619,1684

9,2000,Pakistan,Bangaldesh,1574,1210
```

**Problem Statement:**

Assume data is copied periodically into "/home/cloudera/runs" directory. Write a flume configuration to copy this data to HDFS using flume and then write a PIG script to process data using PIG to find out sum of run scored and balls played by each player.