**Analysis of Book Data using Hadoop-Hive and PIG**

1. **creating table book:**

create table book (isbn string,title string,author string, year string,publisher string,urls string,urlm string,urll string) row format SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde' with SERDEPROPERTIES ( "separatorChar" = "\;", "quoteChar" = "\"") stored as TEXTFILE;

create table rating (id string, isbn string, rating string) row format SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde' with SERDEPROPERTIES ( "separatorChar" = "\;", "quoteChar" = "\"") stored as TEXTFILE;

2. **loading data:**

load data inpath 'data/BX-Books.csv' into table book;

load data inpath 'data/BX-Book-Ratings.csv' into table rating;

**Problem solution**:

1. **Find out the frequency of books published each year. (Hint: Use Boooks.csv file for this).**

**HIVE Solution:**

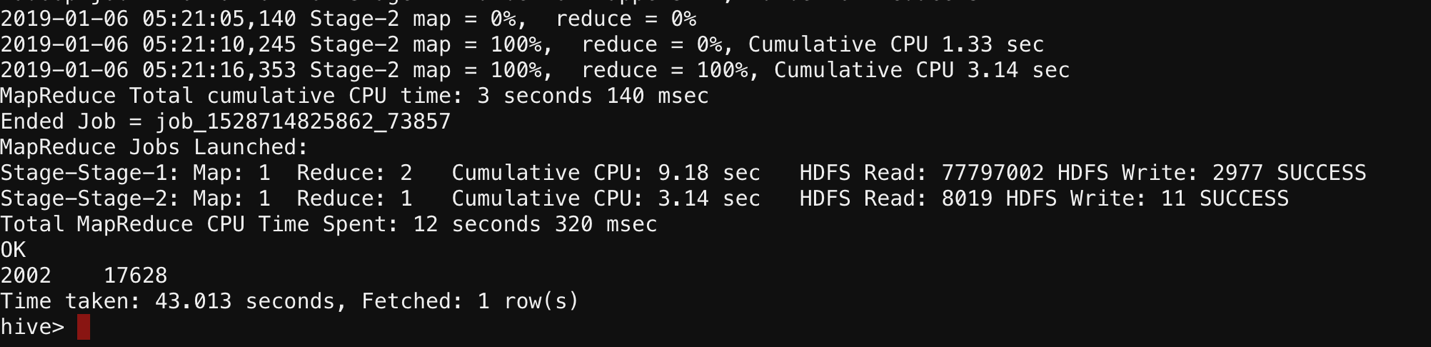
select year, count(\*) as cnt from book group by year order by cntb desc;



1. **Find out in which year maximum number of books were published .**

HIVE Solution:

select year, count(\*) as cnt from book group by year order by cnt desc limit 1;



1. **Find out how many books were published based on ranking in the year 2002.**

select b.rating, count(a.isbn) from book a join rating b on a.isbn = b.isbn where a.year = '2002' group by b.rating order by b.rating;

