

Hadoop Project

The project aims to display show-to-channel relationship is Many-to-Many. In other words, each show might appear on many channels, and each channel might broadcast many shows.

The description of the data is as below

- TV show titles do not have spaces
- Channels have 3 letters
- TV show titles can appear multiple times, with different counts
- A TV show and channel combination might appear multiple times
- TV shows could appear on multiple channels
- The output should have no commas or punctuation, only 1 space between the TV show title and number.

Problem statement:

- What is the total number of viewers for shows on ABC?
- What is the number of viewers for the BAT channel?
- What is the most viewed show on ABC channel?
- What are the aired shows on ZOO,NOX, ABC channels ?

Lab Environment: You need to have Hadoop setup in order to perform this project.

The above problem has to be solved using either MapReduce or Hive or Pig programming constructs and codes should be shared.

Please find attached files as the input data sets and provide solution to the given questions.

Hadoop Project

Guidelines for Project

Please find the below information

join2_gennum*.txt consist of <TV show, count> (A TV show title and the number of viewers)

Example join2_gennum*.txt:

```
Almost_News, 25
Hourly_Show,30
Hot_Cooking,7
Almost_News, 35
Postmodern_Family,8
Baked_News,15
Dumb_Games,60
...
```

join2_genchan*.txt consists of <TV show title, channel> (A TV show title and the channel it was on)

Example join2_genchan*.txt:

```
Almost_News, ABC
Hourly_Show, COM
Hot_Cooking, FNT
Postmodern_Family, NBC
Baked_News, FNT
Dumb_Games, ABC
...
```

We need to perform join operations to solve the problem.