

Week-6

- Execute the questions using pig script on cluster
Steps

- 1: Create Input File w.r.t problem
- 2: Transfer the dataset to the cluster.
- 3: Enter into pig using pig command (pig -x mapreduce)
↳ Pig MapReduce has been picked
- 4: Create a file (.pig) (locally)
↳ Dump the operators based on problem statement
- 5: Run the pig file using run file.pig
- 6: Go to the port (9870) check SUCCESS part-0000 files.

Pig File

```
1 movies = LOAD '/batch_3/200968048/movs.csv' USING PigStorage(',')  
AS (movie_id: int, title: chararray, genres: chararray);  
  
2 genre_tags = FOREACH movies GENERATE movie_id, FLATTEN (TOKENIZE  
genres; ' ')) AS genre;  
  
3 movs_by_genre = GROUP genre_tags BY genre;  
  
4 movie_count = FOREACH movs_by_genre GENERATE group AS  
genre, COUNT(genre_tags) AS count;  
  
5 STORE movie_counts INTO '/batch_3/200968048/movs2' USING  
PigStorage(',');
```

Out put

"Action, 2702

"Comedy, 3503

"Horror, 591