

Week-5

Two running modes $\begin{cases} \text{MapReduce (Pig)} \text{ [Default]} \\ \text{Local (Pig -x local)} \end{cases}$

```
grunt> fd = LOAD '/home/hadoop/data.csv' using PigStorage();  
grunt> dump fd;
```

→ loading the input file using pig storage.
→ To view the content of the file.

```
grunt> fad = LOAD '/home/hadoop/data.csv' using PigStorage(',')  
AS (model, hp);
```

→ To select the specified columns (no. of columns)

```
grunt> dump fad;
```

```
grunt> fs = FOREACH fad generate model;
```

→ To select the specified column from dataset

```
grunt> data = FILTER fs by model == 'volant';  
grunt> dump data;
```

→ To filter out data by a specific value

```
grunt> ord = ORDER fs BY hp desc;
```

→ To order the dataset by a specified column's value
[only works for first 3 columns]