

## Apache Pig

### **Task 1: Find the oldest five-star rating movies**

```
ratings = LOAD '/user/maria_dev/u.data' AS (userID:int, movieID:int, rating:int,
ratingTime:int);
```

```
metadata = LOAD '/user/maria_dev/u.item' USING PigStorage('|')
```

```
AS (movieID:int, movieTitle:chararray, releaseDate:chararray,
videoRealese:chararray, imdblink:chararray);
```

```
nameLookup = FOREACH metadata GENERATE movieID, movieTitle,
```

```
ToUnixTime(ToDate(releaseDate, 'dd-MMM-yyyy')) AS releaseTime;
```

```
ratingsByMovie = GROUP ratings BY movieID;
```

```
avgRatings = FOREACH ratingsByMovie GENERATE group as movieID,
AVG(ratings.rating) as avgRating;
```

```
fiveStarMovies = FILTER avgRatings BY avgRating > 4.0;
```

```
fiveStarsWithData = JOIN fiveStarMovies BY movieID, nameLookup BY movieID;
```

```
oldestFiveStarMovies = ORDER fiveStarsWithData BY nameLookup::releaseTime;
```

```
DUMP oldestFiveStarMovies;
```