

Exercise 2:

Learning goal:

- Field indexing for csv file
- Case sensitivity
- Prepare a production ready script
- Pig script automation
- Passing configuration parameters to pig script
- Read json file

Load csv file without schema:

```
data = load 'books.csv' using PigStorage(',');
```

Extract first field:

```
b = foreach data generate $0;  
dump b;
```

Relation is case sensitive:

```
Dump B;
```

Running a pig script:

Step 1: Create a file say myPig.script with following content within that file

```
data = load 'books.csv' using PigStorage(',');  
b = foreach data generate $0;  
dump b;
```

Step 2: Execute script as
pig myPig.script

Passing a parameter file to pig script:

Step 1: Create parameter file myParams.keyval as following
dilim=",,"

Step 2: Modify script as following:

```
data = load 'books.csv' using PigStorage('$dilim');  
b = foreach data generate $0;  
dump b;
```

Step 3: Pass param file while running script
pig -param_file myParams.keyval myPig.script

Load Json file:

```
data = load 'pig_bookreviews.json' using JsonLoader('author:chararray, title:chararray, year:int,  
reviews:{review:(name:chararray,stars:int)}');  
dump data;
```

Reference: courses.cognitiveclass.ai