

Chapter 3 – Reading and Writing Data in Pig

Exercise

1. In a terminal, inspect the file called u.user:

```
$ less /home/training/ml-data/u.user
```
2. Log into the Grunt shell and LOAD the u.user file. The columns are: user id, age, gender, occupation and zip code. This file should already be in HDFS, but if not, use copyFromLocal.

```
grunt> users = LOAD 'u.user' USING PigStorage('|') AS  
(userid:int, age:int, gender:chararray, occupation:chararray,  
zipcode:int);
```
3. DESCRIBE the relation.

```
grunt> DESCRIBE users;
```
4. Use DUMP to print the records you loaded.

```
grunt> DUMP users;
```
5. Use STORE to save the records in a file. The resulting file should be comma-delimited.

```
grunt> STORE users INTO 'users' USING PigStorage(',');
```
6. Using a terminal or the Grunt shell, inspect the results.

```
$ hadoop fs -ls users  
$ hadoop fs -cat users/part-r-00000 | less
```

Or:

```
grunt> cat users/part-r-00000
```