

Case Study on

Most Occured Start Letter in a text file

- ✓ Use a large text file for this purpose.
- ✓ Create the relevant script file and write all pig commands in it.

Suggested steps to implement this case study.

1. Load text into a bag
2. Tokenize the text in the 'lines' bag
3. Retain first letter of each token
4. Group by letter
5. Count the number of occurrences in each group
6. Order by descending order
7. Grab the first element => Most occurring letter
8. Persist result on a file system