How to run elasticsearch in local computer

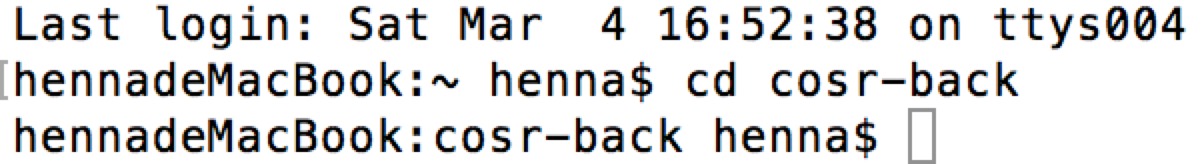
Install cosr-back, follow this link: <https://github.com/commonsearch/cosr-back/blob/master/INSTALL.md>

Install cosr-ops, follow this link: <https://github.com/commonsearch/cosr-ops/blob/master/INSTALL.md>

After installation, you will have Docker, git, cosr-back, cosr-ops in your local computer, through which we can use Spark, Common Crawl and Elasticsearch.

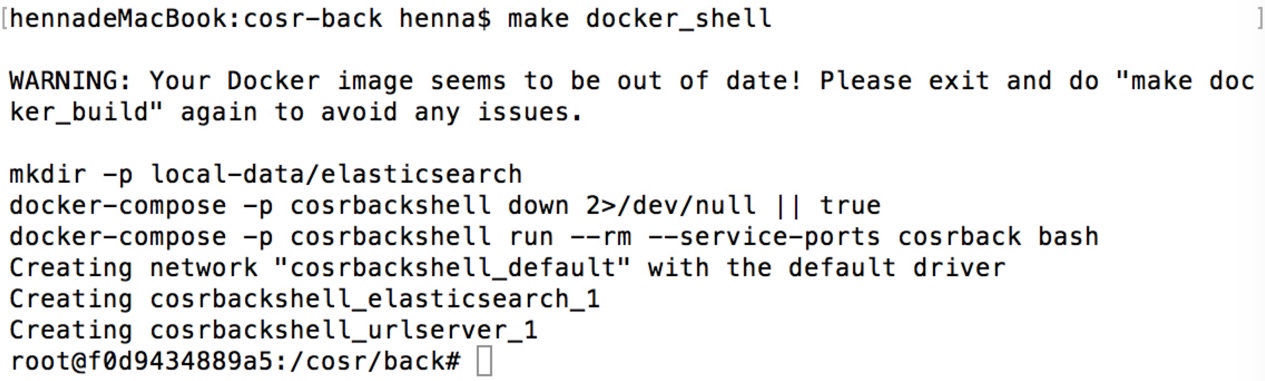
STEPS:

1. Run Docker in your local machine.
2. Open a console, locate into cosr-back



1. Create a Docker repository in console

Code: make docker\_shell



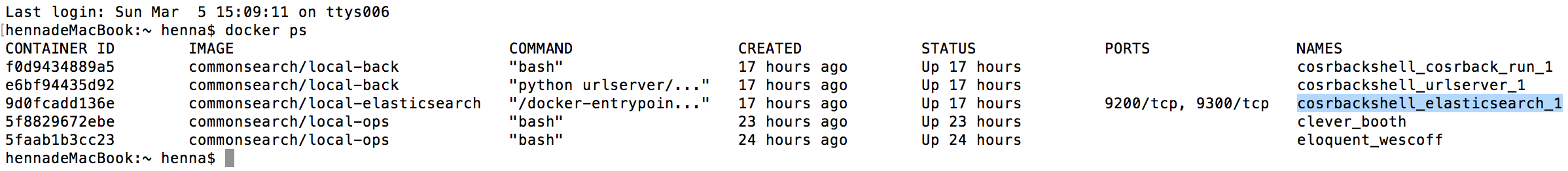
Then we will run Spark, Elastisearch in cosr-back repository.

1. Use Spark in cosr-back repository to narrow down search range. After running following code, the search range will be narrowed, the text on those pages in search range will be indexed by elasticsearch automatically. Also, you can define your own search domain, search key words, and output location in this code. If you haven’t defined output location, there will be no files stored in your local computer though the index files have been already put in elasticsearch.

spark-submit --verbose spark/jobs/pipeline.py –source commoncrawl:limit=2,maxdocs=50 --plugin plugins.filter.All:skip=1 --plugin "plugins.filter.Domains:index\_body=1,domains=.[un.org](http://un.org/) .[undp.org](http://undp.org/)" --plugin "plugins.grep.Words:words=science technology,output=Users/henna/Desktop/out.txt"

You can search key words“science”and “technology”in page body by using such code. Also, you can search title or body as well.

1. To run elastic search on collected search range, we have to get the IP address.



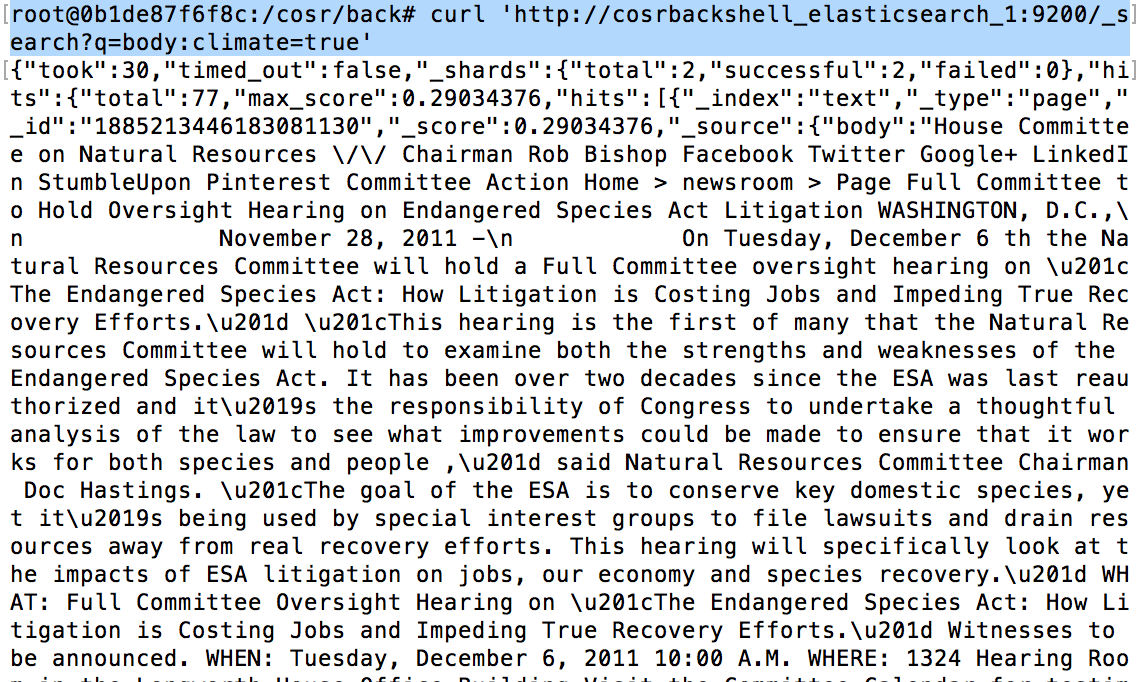
The IP address is “cosrbackshell\_cosrback\_run\_1”

1. After getting the IP address, we can run elasticsearch to get what we want.

First you have to replace the IP address with your IP address.

Then you can define what you want to search. For example, the code below is going to search “climate” in webpage body. Also you can search other words in webpage body or page title.

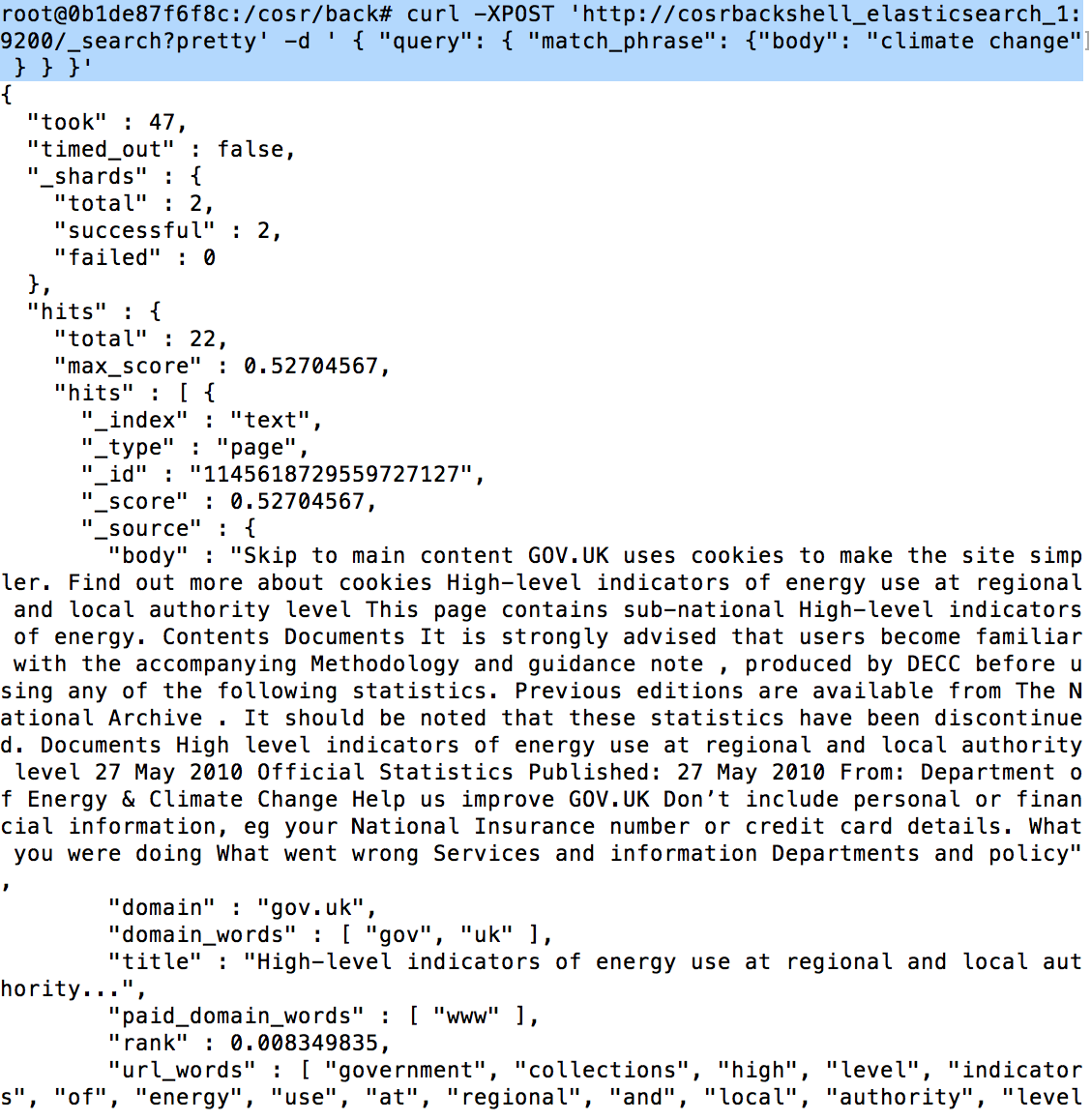
Code: curl 'http://cosrbackshell\_elasticsearch\_1:9200/\_search?q=body:climate=true'



The results are not reader friendly, then we can use such code to make the results much easier to understand.

Code: curl -XPOST 'http://cosrbackshell\_elasticsearch\_1:9200/\_search?pretty' -d ' { "query": { "match\_phrase": {"body": "climate change" } } }'

Therefore, we get 22 result webpages.



after collecting domain, the elasticsearch will index them automatically. After doing this, each webpage will have an index id. After knowing the index id number, you can also search by index id.

Code: curl -XGET 'cosrbackshell\_elasticsearch\_1:9200/text/page/-4274331972783808369?stored\_fields=\*&pretty'

