

Hdfs

Przydatne komendy hdfs:

1. ls - listuje katalog
hdfs dfs -ls *ścieżka*
2. cat - wypisuje zawartość pliku
hdfs dfs -cat *ścieżka*
3. mkdir - tworzy katalog
hdfs dfs -mkdir *ścieżka*
4. rm - usuwa plik
hdfs dfs -rm *ścieżka*
5. chmod - zmienia prawa dostępu do pliku
hdfs dfs -chmod *prawa_dostępu* *ścieżka*
6. chown - zmienia właściciela pliku
hdfs dfs -chown *user:grupa* *ścieżka*
7. cp - kopiuje plik na hdfsie
hdfs dfs -cp *źródło cel*
8. mv - przenosi plik na hdfsie
hdfs dfs -mv *źródło cel*
9. put - umieszcza lokalny plik na hdfsie
hdfs dfs -put *ścieżka_lokalna* *ścieżka_na_hdfs*
10. get - pobiera plik z hdfsa
hdfs dfs -get *ścieżka_na_hdfs* *ścieżka_lokalna*
11. touchz - tworzy pusty plik
hdfs dfs -touchz *ścieżka*
12. appendToFile - przepisuje plik do końca pliku na hdfsie
hdfs dfs -appendToFile *ścieżka_lokalna* *ścieżka_na_hdfs*

Jeżeli klaster nie jest zabezpieczony kerberosem, to możemy wykorzystać zmienną środowiskową `HADOOP_USER_NAME`, żeby ustawić użytkownika z którego wykonane zostanie polecenie, np.

```
HADOOP_USER_NAME=hdfs hdfs dfs -mkdir /katalog
```

Odpowiednikiem linuxowego roota na hdfsie jest użytkownik 'hdfs'

Zadania:

1. Utwórz katalogi:
vagrant:vagrant /user/vagrant
vagrant:vagrant /user/vagrant/inputs
vagrant:vagrant /user/vagrant/outputs
flume:flume /user/flume
kursant:kursant /user/kursant
2. wrzucić pliki `dump.sql`, `loremipsum` i `apache_logs` do `/user/vagrant/inputs`
3. zmienić prawa do odczytu do katalogu `/user/kursant` tak, żeby tylko kursant mógł go odczytać
4. spróbuj odczytać katalog jako user 'vagrant'
5. spróbuj odczytać katalog jako user 'hdfs'