

Hdfs

Przydatne komendy hdfs:

1. ls - listuje katalog

```
hdfs dfs -ls <path>  
hdfs dfs -ls /user/xyz
```

2. cat - wypisuje zawartość pliku

```
hdfs dfs -cat <path>  
hdfs dfs -cat /user/xyz/file.txt
```

3. mkdir - tworzy katalog

```
hdfs dfs -mkdir <path>  
hdfs dfs -mkdir /user/xyz/subdir
```

4. rm - usuwa plik

```
hdfs dfs -rm <path>  
hdfs dfs -rm /user/xyz/file.txt
```

5. chmod - zmienia prawa dostępu do pliku

```
hdfs dfs -chmod <mode> <path>  
hdfs dfs -chmod 640 /user/xyz/file.txt
```

6. chown - zmienia właściciela pliku

```
hdfs dfs -chown <owner>:<group> <path>  
hdfs dfs -chown xyz:xyz /user/xyz/file.txt
```

7. cp - kopiuje plik na hdfsie

```
hdfs dfs -cp <source> <target>  
hdfs dfs -cp /user/xyz/file /user/xyz/file.copy
```

8. mv - przenosi plik na hdfsie

```
hdfs dfs -mv <source> <target>
hdfs dfs -mv /user/xyz/file /user/xyz/file.new
```

9. put - umieszcza lokalny plik na hdfsie

```
hdfs dfs -put <local_path> <hdfs_path>
hdfs dfs -put file.txt /user/xyz
```

10. get - pobiera plik z hdfs

```
hdfs dfs -get <hdfs_path> <local_path>
hdfs dfs -get /user/xyz/file.txt .
```

11. touchz - tworzy pusty plik

```
hdfs dfs -touchz <path>
hdfs dfs -touchz /usr/xyz/file.txt
```

12. appendToFile - przepisuje plik do końca pliku na hdfsie

```
hdfs dfs -appendToFile <local_path> <hdfs_path>
hdfs dfs -appendToFile abc.txt /user/xyz/file.txt
```

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 z prawami dostępu:

ścieżka	user	grupa	prawa dostępu
/user/<login>	<login>	<login>	755
/user/<login>/inputs	<login>	<login>	755
/user/<login>/outputs	<login>	<login>	755
/user/<login>/alien	<login>	<login>	755

2. wrzucić pliki dump.sql, loremipsum i transaction_logs do /user/<login>/inputs
3. zmienić prawa do odczytu do katalogu /user/<login>/alien tak, żeby tylko 'alien' mógł odczytać jego zawartość
4. spróbuj odczytać ww. katalog jako user '<login>'
5. spróbuj odczytać ww. katalog jako user 'hdfs'