

HDFS Shell

\$hdfs dfs -<command> -<option> <URI>

\$hdfs dfs -ls /

URI

hdfs://localhost:8020/user/home

scheme

authority

HDFS path

- Local:

\$hdfs dfs -ls file:///to/path/file3

- HDFS:

\$hdfs dfs -ls hdfs://localhost/to/path/dir

- fs.default.name=hdfs://localhost

\$hdfs dfs -ls /to/path/dir

Команды в shell

- Похожие на команды Unix:

cat, rm, ls, du,

- Поддержка специфичных для HDFS операций

setrep - смена фактора репликации

Команда help

- Вывод списка команд

*\$ hdfs dfs -**help***

- Показать информацию по команде

*\$ hdfs dfs -**help** <command_name>*

Основные команды в shell

ls – листинг директории и статистика файлов

```
[cloudera@localhost ~]$ hdfs dfs -ls /user/cloudera/wordcount/output
Found 3 items
-rw-r--r--      3 cloudera cloudera      0 2014-03-15 11:56
/user/cloudera/wordcount/output/_SUCCESS
drwxr-xr-x    - cloudera cloudera      0 2014-03-15 11:56
/user/cloudera/wordcount/output/_logs
-rw-r--r--      3 cloudera cloudera    31 2014-03-15 11:56
/user/cloudera/wordcount/output/part-00000
```

Основные команды в shell

mkdir – создать директорию

\$hdfs dfs -mkdir /data/new_path

Чтение файлов в shell

cat — вывод источника в stdout

— Весь файл: *\$hdfs dfs -cat /dir/file.txt*

— Полезно вывод перенаправить через pipe в *less*, *head*, *tail* и т.д.

— Получить первые 100 строк из файла:

\$hdfs dfs -cat /dir/file.txt | head -n 100

Чтение файлов в shell

text – аналог ***cat***, но разархивирует архивы:

\$hdfs dfs -cat /dir/file.gz – непонятный текст

\$hdfs dfs -text /dir/file.gz – понятный текст

\$hdfs dfs -cat /dir/file.txt = \$hdfs dfs -text /dir/file.txt

Чтение файлов в shell

tail – выводит последние строки файла

\$hdfs dfs -cat /dir/file.txt | tail – плохо

\$hdfs dfs -tail /dir/file.txt – хорошо

Копирование данных в shell

cp – скопировать файлы из одного места в другое

\$hdfs dfs -cp /dir/file1 /otherDir/file2

! годится только для небольших файлов

Копирование данных в shell

distcp – копирует большие файлы или много файлов

\$hdfs distcp /dir/file1 /otherDir/file2

Копирование данных в shell

mv – перемещение файла из одного места в другое

\$hdfs dfs -mv /dir/file1 /dir2

put (copyFromLocal) – копирование локального файла в HDFS

\$hdfs dfs -put localfile /dir/file

get (copyToLocal) – копирование файла из HDFS в локальную FS

\$hdfs dfs -get /dir/file localfile

Удаление в shell

rm – удалить файл (в корзину)

\$hdfs dfs -rm /dir/file

rm -r – удалить рекурсивно директорию

\$hdfs dfs -rm -r /dir

Статистика в shell

du – размер файла или директории в байтах

\$hdfs dfs -du /dir/

du -h – размер файла или директории в
удобно-читаемом формате

\$hdfs dfs -du -h /dir/

65M /dir

Остальные команды в shell

- Другие команды
 - *chmod, chown, count, test и т.д.*
- Чтобы узнать больше
 - *\$hdfs dfs -help*
 - *\$hdfs dfs -help <command>*

Команда **fsck**

- Проверка неконсистентности файловой системы
- Показывает проблемы
 - Отсутствующие блоки
 - Недореплицированные блоки
- Не устраняет проблем, только информация
 - Namenode попытается автоматически исправить проблемы
- **\$ hdfs fsck <path>**

\$ hdfs fsck /

Права в HDFS

- Ограничения на уровне файла/директории
 - Сходство с моделью прав в POSIX
 - Read (r), Write (w) и Execute (x)
 - Разделяется на пользователя, группу и всех остальных
- Права пользователя определяются исходя из прав той ОС, где он запускает клиентское приложение
- Авторизация через Kerberos

Права в HDFS

```
[cloudera@localhost ~]$ hadoop fs -ls /user/cloudera/wordcount/output
Found 3 items
-rw-r--r--      3 cloudera cloudera      0 2014-03-15 11:56
/user/cloudera/wordcount/output/_SUCCESS
drwxr-xr-x    - cloudera cloudera      0 2014-03-15 11:56
/user/cloudera/wordcount/output/_logs
-rw-r--r--      3 cloudera cloudera    31 2014-03-15 11:56
/user/cloudera/wordcount/output/part-00000
```

Команда DFSAdmin

- Команды для администрирования HDFS
 - *\$hdfs dfsadmin -<command>*
 - Например: *\$hdfs dfsadmin -report*
- **report** – отображает статистику по HDFS
- **safemode** – включение безопасного режима для проведения административных работ
 - Upgrade, backup и т.д.

Балансер HDFS

- Блоки в HDFS могут быть неравномерно распределены по всем Datanode'ам кластера
- Балансер – это утилита, которая автоматически анализирует расположение блоков в HDFS и старается его сбалансировать

\$ hdfs balancer