

1. stat

hadoop fs -stat [format] <path> ...

Gibt Statistiken über eine Datei bzw. einen Ordner in einem speziellen Format aus.

Das Format akzeptiert Permissions sowohl in

- octal (%a) (z.B. 0755) als auch in
- symbolic (%A), (z.B. `chmod("g=rw,o=-w", @files)` Gruppe darf lesen/schreiben, andere dürfen nicht schreiben)
- Dateigröße in Blocks (%b),
- Typ (%F),
- Gruppenname des Besitzers (owner) (%g),
- Name (%n),
- Blockgröße (%o),
- Replikation (%r),
- Benutzername des Besitzers (owner) (%u), und
- Änderungsdatum (%y, %Y). %y zeigt das Datum in UTC Format "yyyy-MM-dd HH:mm:ss" und %Y zeigt Millisekunden seit 01.01.1970 (UTC) an.
Wenn das Format nicht spezifiziert ist, wird %y als default genutzt.

Beispiel:

```
hadoop fs -stat "%F %a %u:%g %b %y %n" /file
```

2. getfattr

hadoop fs -getfattr [-R] -n name | -d [-e en] <path>

Zeigt erweiterte Attribute und deren Werte einer Datei oder eines Ordners an.

Optionen:

- R: Rekursive Liste der Attribute aller Dateien und Ordner.
- n name: Kippt den benannten erweiterten Attributwert weg.
- d: Kippt alle erweiterten Attributwerte weg, die in Beziehung zu <path> stehen
- e encoding: Werte nach dem Abrufen kodieren. Valide Kodierungen sind "text", "hex", und "base64". Werte, die als Textzeichenfolgen kodiert sind, werden in doppelte Anführungszeichen (") eingeschlossen und Werte, die als hexadezimal und base64 kodiert sind, mit 0x bzw. 0s vorangestellt.
- path: Pfad zur Datei oder dem Ordner.

Beispiel:

```
hadoop fs -getfattr -d /file
```

```
hadoop fs -getfattr -R -n user.myAttr /dir
```

3. du

hadoop fs -du [-s] [-h] [-x] URI [URI ...]

Zeigt die Größe von Dateien und Ordnern in einem bestimmten Ordner an.

Optionen:

- **s**: führt zu einer aggregierten Zusammenfassung der angezeigten Dateilängen und nicht zu den einzelnen Dateien. Ohne diese Option wird die Berechnung, indem 1-Level tief vom angegebenen Pfad entfernt durchgeführt.
- **h**: formatiert die Dateigröße in ein Lesbares Format (64.0m anstatt 67108864)
- **x**: schließt Snapshots aus der Ergebnisberechnung aus. Ohne die Option -x (Standard) wird das Ergebnis immer aus allen INodes berechnet, einschließlich aller Snapshots unter dem angegebenen Pfad.

Der Befehl gibt drei Spalten im folgendem Format aus:

size | disk_space_consumed_with_all_replicas | full_path_name

Beispiel:

```
hadoop fs -du /user/cloudera/dir1 /user/cloudera/file1 hdfs://nn.example.com/user/cloudera/dir1
```