# BDA Lab 5

## Hadoop

Commands

- ## Check version

suraj@debian:~$ hadoop version

Hadoop 3.3.0

Source code repository https://gitbox.apache.org/repos/asf/hadoop.git -r aa96f1871bfd858f9bac59cf2a81ec470da649af

Compiled by brahma on 2020-07-06T18:44Z

Compiled with protoc 3.7.1

From source with checksum 5dc29b802d6ccd77b262ef9d04d19c4

This command was run using /opt/suraj/hadoop/share/hadoop/common/hadoop-common-3.3.0.jar

- ## List files

suraj@debian:~$ hadoop fs -ls /

Found 2 items

drwxrwxrwt   - hadoop supergroup          0 2021-01-28 07:16 /tmp

drwxr-xr-x   - hadoop supergroup          0 2021-01-28 07:16 /user

- ## Get all dirs recursively

suraj@debian:~$ hadoop fs -ls -R /

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 09:18 /sample

drwxrwxrwt   - hadoop supergroup          0 2021-01-28 07:16 /tmp

drwxrwxrwt   - hadoop supergroup          0 2021-01-28 07:16 /tmp/hadoop-yarn

drwxrwxrwt   - hadoop supergroup          0 2021-01-28 07:16 /tmp/hadoop-yarn/staging

drwxrwxrwt   - hadoop supergroup          0 2021-01-28 07:16 /tmp/hadoop-yarn/staging/history

drwxrwxrwt   - hadoop supergroup          0 2021-01-28 07:16 /tmp/hadoop-yarn/staging/history/done

```
drwxrwxrwt   - hadoop supergroup       0 2021-01-28 07:16 /tmp/hadoop-
yarn/staging/history/done_intermediate

drwx-wx-wx   - hadoop supergroup       0 2021-01-28 07:16 /tmp/hive

drwx-wx-wx   - hadoop supergroup       0 2021-04-19 08:29
/tmp/hive/_resultscache_

drwx------   - hadoop supergroup       0 2021-04-19 07:44
/tmp/hive/_resultscache_/results-2c1256dd-f8e1-432b-992c-5c7a233a2d3b

drwx------   - hadoop supergroup       0 2021-04-19 07:52
/tmp/hive/_resultscache_/results-590dddd7-b941-495c-963a-86e7d1d1f223

drwx------   - hadoop supergroup       0 2021-01-28 07:16
/tmp/hive/_resultscache_/results-b32c20c3-f945-426d-8d8b-6001273e0d0a

drwx------   - hadoop supergroup       0 2021-04-19 08:29
/tmp/hive/_resultscache_/results-fe4f1003-ab40-4551-aa0b-92e73d1eb423

drwxr-xr-x   - hadoop supergroup       0 2021-04-19 08:30
/tmp/hive/hadoop

drwxr-xr-x   - hadoop supergroup       0 2021-04-19 07:53
/tmp/hive/hadoop/0b1e49df-8a70-445a-a2cc-810ccdc50b69

drwxr-xr-x   - hadoop supergroup       0 2021-04-19 07:53
/tmp/hive/hadoop/0b1e49df-8a70-445a-a2cc-810ccdc50b69/_tmp_space.db

drwxr-xr-x   - hadoop supergroup       0 2021-04-19 08:29
/tmp/hive/hadoop/0b299767-ac7b-472e-95e1-57d1a8cc4cd9

drwxr-xr-x   - hadoop supergroup       0 2021-04-19 08:29
/tmp/hive/hadoop/0b299767-ac7b-472e-95e1-57d1a8cc4cd9/_tmp_space.db

drwxr-xr-x   - hadoop supergroup       0 2021-04-19 07:45
/tmp/hive/hadoop/0ef70be7-11b9-4f3c-a0b0-8b5a873f6170

drwxr-xr-x   - hadoop supergroup       0 2021-04-19 07:45
/tmp/hive/hadoop/0ef70be7-11b9-4f3c-a0b0-8b5a873f6170/_tmp_space.db

drwxr-xr-x   - hadoop supergroup       0 2021-04-19 07:52
/tmp/hive/hadoop/150be84f-fc84-4086-b135-20f23dc9f8ad

drwxr-xr-x   - hadoop supergroup       0 2021-04-19 07:52
/tmp/hive/hadoop/150be84f-fc84-4086-b135-20f23dc9f8ad/_tmp_space.db

drwxr-xr-x   - hadoop supergroup       0 2021-04-19 07:53
/tmp/hive/hadoop/16e65b48-317e-452c-960a-07c76ba32205

drwxr-xr-x   - hadoop supergroup       0 2021-04-19 07:53
/tmp/hive/hadoop/16e65b48-317e-452c-960a-07c76ba32205/_tmp_space.db

drwxr-xr-x   - hadoop supergroup       0 2021-04-19 07:45
/tmp/hive/hadoop/20552959-7d4b-4bde-919b-30eaf8df02f8

drwxr-xr-x   - hadoop supergroup       0 2021-04-19 07:45
/tmp/hive/hadoop/20552959-7d4b-4bde-919b-30eaf8df02f8/_tmp_space.db
```

```
drwxr-xr-x   - hadoop supergroup          0 2021-01-28 07:17
/tmp/hive/hadoop/30714089-8c70-49b8-bc74-96c576f6e4a8

drwxr-xr-x   - hadoop supergroup          0 2021-01-28 07:17
/tmp/hive/hadoop/30714089-8c70-49b8-bc74-96c576f6e4a8/_tmp_space.db

drwxr-xr-x   - hadoop supergroup          0 2021-01-28 07:17
/tmp/hive/hadoop/30aac5a7-4828-4f5e-9e56-74d353dd5a6d

drwxr-xr-x   - hadoop supergroup          0 2021-01-28 07:17
/tmp/hive/hadoop/30aac5a7-4828-4f5e-9e56-74d353dd5a6d/_tmp_space.db

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 07:44
/tmp/hive/hadoop/6cf4085c-9b51-4a53-bc6f-cb8f0c53abb8

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 07:44
/tmp/hive/hadoop/6cf4085c-9b51-4a53-bc6f-cb8f0c53abb8/_tmp_space.db

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 07:52
/tmp/hive/hadoop/72876d68-0935-4d5e-9130-e44b2462e2ac

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 07:52
/tmp/hive/hadoop/72876d68-0935-4d5e-9130-e44b2462e2ac/_tmp_space.db

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 08:30
/tmp/hive/hadoop/7455d750-dbc2-47e2-aae0-24d4b053cdbe

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 08:30
/tmp/hive/hadoop/7455d750-dbc2-47e2-aae0-24d4b053cdbe/_tmp_space.db

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 08:30
/tmp/hive/hadoop/bc52d818-0d26-4b82-be67-236820eaeb2c

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 08:30
/tmp/hive/hadoop/bc52d818-0d26-4b82-be67-236820eaeb2c/_tmp_space.db

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 07:44
/tmp/hive/hadoop/bc5b80b7-a7ca-4548-82ec-f157d3ec4f5c

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 07:44
/tmp/hive/hadoop/bc5b80b7-a7ca-4548-82ec-f157d3ec4f5c/_tmp_space.db

drwxr-xr-x   - hadoop supergroup          0 2021-01-28 07:16
/tmp/hive/hadoop/ccca17a7-b635-4040-8611-a1e2fc1bb83c

drwxr-xr-x   - hadoop supergroup          0 2021-01-28 07:16
/tmp/hive/hadoop/ccca17a7-b635-4040-8611-a1e2fc1bb83c/_tmp_space.db

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 08:28
/tmp/hive/hadoop/ee29d0d6-7f1a-49a5-b5c2-7812de099c55

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 08:28
/tmp/hive/hadoop/ee29d0d6-7f1a-49a5-b5c2-7812de099c55/_tmp_space.db

drwxr-xr-x   - hadoop supergroup          0 2021-01-28 07:16
/tmp/hive/hadoop/f895225a-02a0-4a61-98e5-5cb10136addb

drwxr-xr-x   - hadoop supergroup          0 2021-01-28 07:16
/tmp/hive/hadoop/f895225a-02a0-4a61-98e5-5cb10136addb/_tmp_space.db
```

```
drwxr-xr-x   - hadoop supergroup          0 2021-01-28 07:16 /user

drwxr-xr-x   - hadoop supergroup          0 2021-01-28 07:16 /user/hadoop

drwxr-xr-x   - hadoop supergroup          0 2021-01-28 07:16 /user/hive

drwxrwxr-x   - hadoop supergroup          0 2021-01-28 07:16
/user/hive/warehouse
```

- Create dir

```
suraj@debian:~$ hadoop fs -mkdir /sample


suraj@debian:~$ hadoop fs -ls /

Found 3 items

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 09:18 /sample

drwxrwxrwt   - hadoop supergroup          0 2021-01-28 07:16 /tmp

drwxr-xr-x   - hadoop supergroup          0 2021-01-28 07:16 /user
```

- Copy local file to HDFS

```
suraj@debian:~$ touch a.txt

suraj@debian:~$ ls

apps  a.txt  suraj_credentials  htdocs  stack

suraj@debian:~$ hadoop fs -put ~/a.txt /sample/a.txt

suraj@debian:~$ hadoop fs -ls /sample

Found 1 items

-rw-r--r--   1 hadoop supergroup          0 2021-04-19 09:25 /sample/a.txt
```

- Copy HDFS file to local

```
suraj@debian:~$ hadoop fs -get /sample/a.txt ~/a-fromHDFS.txt

suraj@debian:~$ ls

a-fromHDFS.txt  apps  a.txt  suraj_credentials  htdocs  stack
```

- Copy local file to HDFS using copyFromLocal

```
suraj@debian:~$ hadoop fs -copyFromLocal ~/a.txt /sample/a-
copyfromlocal.txt

suraj@debian:~$ hadoop fs -ls /sample
```

```
Found 2 items

-rw-r--r--   1 hadoop supergroup          0 2021-04-19 09:34 /sample/a-
copyfromlocal.txt

-rw-r--r--   1 hadoop supergroup          0 2021-04-19 09:25 /sample/a.txt
```

- Copy HDFS file to local using copyToLocal

```
suraj@debian:~$ hadoop fs -copyToLocal /sample/a.txt ~/a-copytolocal.txt

suraj@debian:~$ ls

a-copytolocal.txt  apps  suraj_credentials  htdocs  stack
```

- Display contents of HDFS file in console

```
suraj@debian:~$ hadoop fs -cat /sample/b.txt

hi how are you

how is your job

how is your family

how is your brother

how is your sister
```

- Copy from a dir to another

```
suraj@debian:~$ hadoop fs -mkdir /sample-copy

suraj@debian:~$ hadoop fs -cp /sample/b.txt /sample-copy/b.txt

suraj@debian:~$ hadoop fs -ls /sample-copy

Found 1 items

-rw-r--r--   1 hadoop supergroup         15 2021-04-19 09:45 /sample-
copy/b.txt
```

- Remove a directory

```
suraj@debian:~$ hadoop fs -ls /

Found 4 items

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 09:41 /sample

drwxr-xr-x   - hadoop supergroup          0 2021-04-19 09:45 /sample-copy

drwxrwxrwt   - hadoop supergroup          0 2021-01-28 07:16 /tmp

drwxr-xr-x   - hadoop supergroup          0 2021-01-28 07:16 /user
```

```
suraj@debian:~$ hadoop fs -rm -r -f /sample-copy

Deleted /sample-copy
```