

Hadoop User Management*

Montasser AKERMI

Last Update: February, 2025

*The course material is hosted at <https://akermi.org/c1/>.

Learning Objectives

Learn how to manage users in Hadoop HDFS: Create a new user in Hadoop HDFS, change access rights and ownership, and manage data in HDFS.

Pre-Lab Questions

- Users in local filesystem vs users in Hadoop cluster?
- Why do we need multi users in a Hadoop cluster?
- Multiple Hadoop namespaces vs multiple Hadoop users?

Create a new user in Hadoop

1. (Optional) Create a new group:
`sudo addgroup classroom`
2. Create new user:
`sudo adduser --ingroup classroom med`
3. Change the access permission of /home/hadoop so that the new user can access and execute Java and Hadoop binary files (java, hadoop, hdfs, yarn, etc.). Also make sure those binary files have the required access permissions:
`chmod o+rx /home/hadoop`
4. Change access permission of /home/hadoop/hdata to 777, this way the new user can read, write, execute under that folder:
`chmod 777 /home/hadoop/hdata`
5. Create a home directory in HDFS for the new user:
`hadoop fs -mkdir -p /user/med`
6. Set ownership of the recently created home directory to the new user:
`hadoop fs -chown med:classroom /user/med`
7. Change permission to 755 to make sure the med can read, write, and execute under that directory:
`hadoop fs -chmod 755 /user/med`
8. Add Java and Hadoop environment variables to /home/med/.bashrc file:

```
### JAVA
export JAVA_HOME=/home/hadoop/jdk1.8.0_92
export PATH=$JAVA_HOME/bin:$PATH
### HADOOP
export HADOOP_HOME=/home/hadoop/hadoop-3.3.4
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export PATH=$PATH:$HADOOP_HOME/bin
```

Testing the new user

- Change to new user:
`su - med`
- Copy a new file from local OS to HDFS:
`hadoop fs -copyFromLocal /home/med/.bashrc /user/med`

Post-Lab Questions

- Why did we need to create a user in the local filesystem first?
- Why add environment variables to `.bashrc`?
- How to change data ownership in HDFS?
- Who is the root in HDFS?