

Exp. No : 1

Installation of Hadoop

1. Download Hadoop Binary file

```
--2024-09-13 14:31:29-- https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz
Resolving downloads.apache.org (downloads.apache.org)... 135.181.214.104, 88.99.208.237, 2a01:4f9:3a:2c57::2, ...
Connecting to downloads.apache.org (downloads.apache.org)|135.181.214.104|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 730107476 (690M) [application/x-gzip]
Saving to: 'hadoop-3.3.6.tar.gz.'

hadoop-3.3.6.tar.gz.1 3K[=>] 26.73M 2.01MB/s eta 4m 19s ^C
--2024-09-13 14:32:08-- https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz
Resolving downloads.apache.org (downloads.apache.org)... 135.181.214.104, 88.99.208.237, 2a01:4f9:3a:2c57::2, ...
Connecting to downloads.apache.org (downloads.apache.org)|135.181.214.104|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 730107476 (690M) [application/x-gzip]
Saving to: 'hadoop-3.3.6.tar.gz'

hadoop-3.3.6.tar.gz 100%[=====] 696.28M 5.17MB/s in 3m 42s
2024-09-13 14:35:50 (3.14 MB/s) - 'hadoop-3.3.6.tar.gz' saved [730107476/730107476]
```

2. Modifying .bashrc file

```
alias ls='ls --color=auto'
#alias dir='dir --color=auto'
#alias vdir='vdir --color=auto'

alias grep='grep --color=auto'
alias fgrep='fgrep --color=auto'
alias egrep='egrep --color=auto'
fi

# colored GCC warnings and errors
#export GCC_COLORS='error=01:31;warning=01:35;note=01:36;caret=01:32;locus=01;quote=01'

# some more ls aliases
alias ll='ls -lF'
alias la='ls -A'
alias l='ls -CF'

# Add an "alert" alias for long running commands. Use like so:
# sleep 10; alert
alias alert='notify-send --urgency=low -t "${ $? }" && echo terminal || echo error)' "${history/tail -n1}sed -e 's/\s/\s[0-9]\{1,*\}\/s/[;A]\s*alert:\/\/\s/'"

# Alias definitions.
# You may want to put all your additions into a separate file like
# ~/.bash_aliases, instead of adding them here directly.
# See /usr/share/doc/bash-doc/examples in the bash-doc package.

if [ -f ~/.bash_aliases ]; then
. ~/.bash_aliases
fi

# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
if [ -f /usr/share/bash-completion/bash_completion ]; then
. /usr/share/bash-completion/bash_completion
elif [ -f /etc/bash_completion ]; then
. /etc/bash_completion
fi
fi

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-and64
export HADOOP_HOME=/home/hadoop/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export HADOOP_YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

3. Updating JAVA_HOME in \$HADOOP_HOME/etc/hadoop/hadoop-env.sh

```

GNU nano 2.9.6 /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh
# Licensed to the Apache Software Foundation (ASF) under one
# or more contributor license agreements. See the NOTICE file
# distributed with this work for additional information
# regarding copyright ownership. The ASF licenses this file
# to you under the Apache License, Version 2.0 (the
# "License"); you may not use this file except in compliance
# with the License. You may obtain a copy of the License at
#
# http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.
#
# Set Hadoop-specific environment variables here.
#
##
## THIS FILE ACTS AS THE MASTER FILE FOR ALL HADOOP PROJECTS.
## SETTINGS HERE WILL BE READ BY ALL HADOOP COMMANDS. THEREFORE,
## ONE CAN USE THIS FILE TO SET YARN, HDFS, AND MAPREDUCE
## CONFIGURATION OPTIONS INSTEAD OF xxx-env.sh.
##
## Precedence rules:
## (yarn-env.sh|hdfs-env.sh) > hadoop-env.sh > hard-coded defaults
## (YARN_xyz|HDFS_xyz) > HADOOP_xyz > hard-coded defaults
##
# Many of the options here are built from the perspective that users
# may want to provide OVERRIDING values on the command line.
# For example:
#
# JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
#
# Therefore, the vast majority (BUT NOT ALL!) of these defaults
# are configured for substitution and not append. If append
# is preferable, modify this file accordingly.
##
# Generic settings for HADOOP
##
# Technically, the only required environment variable is JAVA_HOME.
# All others are optional. However, the defaults are probably not
# preferred. Many sites configure these options outside of Hadoop,
# such as in /etc/profile.d

```

4. Modifying \$HADOOP_HOME/etc/hadoop/core-site.xml

```

GNU nano 6.2 /home/hadoop/hadoop/etc/hadoop/core-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at
  http://www.apache.org/licenses/LICENSE-2.0
  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->
<!-- Put site-specific property overrides in this file. -->
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>

```

5. Modifying \$HADOOP_HOME/etc/hadoop/hdfs-site.xml

```
GNU nano 6.2 /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml *
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the license.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>
  </property>
</configuration>
```

File Name to Write: /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml

H-G Help
 H-Q Cancel
 H-D DOS Format
 H-M Mac Format
 H-A Append
 H-P Prepend
 H-B Backup File
 H-T Browse

6. Modifying \$HADOOP_HOME/etc/hadoop/mapred-site.xml

```
GNU nano 6.2 /home/hadoop/hadoop/etc/hadoop/mapred-site.xml *
<?xml version="1.0" ?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the license.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>yarn.app.mapreduce.am.env</name>
    <value>HADOOP_MAPRED_HOME=$HADOOP_HOME/home/hadoop/hadoop/bin/hadoop</value>
  </property>
  <property>
    <name>mapreduce.map.env</name>
    <value>HADOOP_MAPRED_HOME=$HADOOP_HOME/home/hadoop/hadoop/bin/hadoop</value>
  </property>
  <property>
    <name>mapreduce.reduce.env</name>
    <value>HADOOP_MAPRED_HOME=$HADOOP_HOME/home/hadoop/hadoop/bin/hadoop</value>
  </property>
</configuration>
```

File Name to Write: /home/hadoop/hadoop/etc/hadoop/mapred-site.xml

H-G Help
 H-Q Cancel
 H-D DOS Format
 H-M Mac Format
 H-A Append
 H-P Prepend
 H-B Backup File
 H-T Browse

7. Modifying \$HADOOP_HOME/etc/hadoop/yarn-site.xml

```
GNU nano 6.2 /home/hadoop/hadoop/etc/hadoop/yarn-site.xml *
#!/usr/bin/perl
# Licensed under the Apache License, Version 2.0 (the "License");
# you may not use this file except in compliance with the License.
# You may obtain a copy of the License at
#
# http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License. See accompanying LICENSE file.
<?xml version="1.0"?>
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
</configuration>
```

File Name to Write: /home/hadoop/hadoop/etc/hadoop/yarn-site.xml

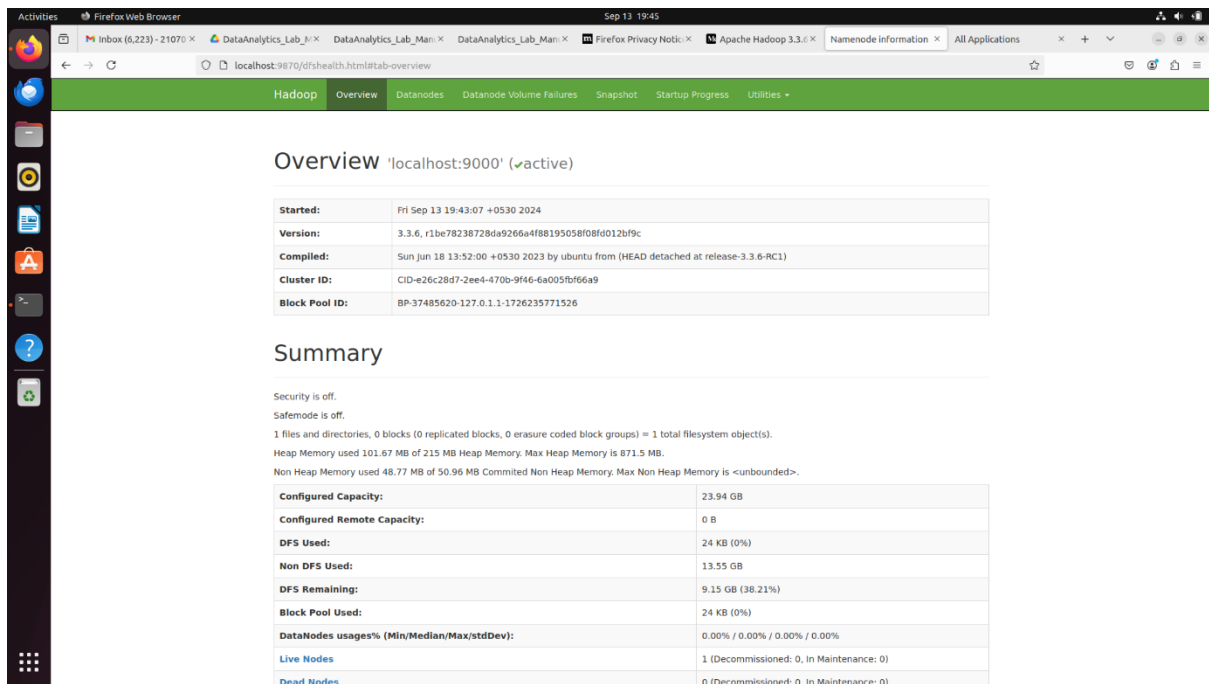
Append Prepend Backup File Browse

Cancel Help DOS Format Mac Format

```
Activities Terminal Sep 13 19:26
hadoop@kiran: ~$

2024-09-13 19:26:11,241 INFO util.GSet: VM type = 64-bit
2024-09-13 19:26:11,244 INFO util.GSet: 2.0% max memory 871.5 MB = 17.4 MB
2024-09-13 19:26:11,244 INFO util.GSet: capacity = 2^21 = 2097152 entries
2024-09-13 19:26:11,290 INFO blockmanagement.BlockManager: Storage policy satisfier is disabled
2024-09-13 19:26:11,290 INFO blockmanagement.BlockManager: dfs.block.access.token.enable = false
2024-09-13 19:26:11,306 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.safeMode.threshold.pct = 0.999
2024-09-13 19:26:11,307 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.safeMode.min.datanodes = 0
2024-09-13 19:26:11,307 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.safeMode.extension = 30000
2024-09-13 19:26:11,308 INFO blockmanagement.BlockManager: defaultReplication = 1
2024-09-13 19:26:11,308 INFO blockmanagement.BlockManager: maxReplication = 512
2024-09-13 19:26:11,308 INFO blockmanagement.BlockManager: minReplication = 1
2024-09-13 19:26:11,308 INFO blockmanagement.BlockManager: maxReplicationStreams = 2
2024-09-13 19:26:11,309 INFO blockmanagement.BlockManager: redundancyCheckInterval = 3000ms
2024-09-13 19:26:11,309 INFO blockmanagement.BlockManager: encryptDataTransfer = false
2024-09-13 19:26:11,309 INFO blockmanagement.BlockManager: maxNumLocksIOLog = 1000
2024-09-13 19:26:11,392 INFO namenode.FSDirectory: GLOBAL serial map: bits=29 maxEntries=536870911
2024-09-13 19:26:11,392 INFO namenode.FSDirectory: USER serial map: bits=24 maxEntries=16777215
2024-09-13 19:26:11,392 INFO namenode.FSDirectory: GROUP serial map: bits=24 maxEntries=16777215
2024-09-13 19:26:11,392 INFO namenode.FSDirectory: XATTR serial map: bits=24 maxEntries=16777215
2024-09-13 19:26:11,414 INFO util.GSet: Computing capacity for map InodeMap
2024-09-13 19:26:11,414 INFO util.GSet: VM type = 64-bit
2024-09-13 19:26:11,415 INFO util.GSet: 1.0% max memory 871.5 MB = 8.7 MB
2024-09-13 19:26:11,415 INFO util.GSet: capacity = 2^20 = 1048576 entries
2024-09-13 19:26:11,415 INFO namenode.FSDirectory: ACLs enabled? true
2024-09-13 19:26:11,415 INFO namenode.FSDirectory: POSIX ACL inheritance enabled? true
2024-09-13 19:26:11,416 INFO namenode.FSDirectory: Xattrs enabled? true
2024-09-13 19:26:11,417 INFO namenode.NameNode: Caching file names occurring more than 10 times
2024-09-13 19:26:11,430 INFO snapshot.SnapshotManager: Loaded config captureOpenFiles: false, skipCaptureAccessTimeOnlyChange: false, snapshotDiffAllowSnapRootDescendant: true, maxSnapshotInit: 65536
2024-09-13 19:26:11,432 INFO snapshot.SnapshotManager: Skiplist is disabled
2024-09-13 19:26:11,444 INFO util.GSet: Computing capacity for map cachedBlocks
2024-09-13 19:26:11,444 INFO util.GSet: VM type = 64-bit
2024-09-13 19:26:11,445 INFO util.GSet: 0.25% max memory 871.5 MB = 2.2 MB
2024-09-13 19:26:11,445 INFO util.GSet: capacity = 2^18 = 262144 entries
2024-09-13 19:26:11,465 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2024-09-13 19:26:11,465 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2024-09-13 19:26:11,465 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2024-09-13 19:26:11,472 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2024-09-13 19:26:11,472 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
2024-09-13 19:26:11,479 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2024-09-13 19:26:11,479 INFO util.GSet: VM type = 64-bit
2024-09-13 19:26:11,479 INFO util.GSet: 0.029999999999999998 max memory 871.5 MB = 267.7 KB
2024-09-13 19:26:11,480 INFO util.GSet: capacity = 2^15 = 32768 entries
2024-09-13 19:26:11,538 INFO namenode.FSImage: Allocated new BlockPoolId: BP-37485620-127.0.1.1-1726235771526
2024-09-13 19:26:11,575 INFO common.Storage: Storage directory /home/hadoop/hadoopdata/hdfs/namenode has been successfully formatted.
2024-09-13 19:26:11,762 INFO namenode.FSImageFormatProtobuf: Saving Image file /home/hadoop/hadoopdata/hdfs/namenode/current/fsimage.cpkt_000000000000000000 using no compression
2024-09-13 19:26:11,883 INFO namenode.FSImageFormatProtobuf: Image file /home/hadoop/hadoopdata/hdfs/namenode/current/fsimage.cpkt_000000000000000000 of size 401 bytes saved in 0 seconds.
2024-09-13 19:26:11,896 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2024-09-13 19:26:11,921 INFO namenode.FSNamesystem: Stopping services started for active state
2024-09-13 19:26:11,922 INFO namenode.FSNamesystem: Stopping services started for standby state
2024-09-13 19:26:11,929 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2024-09-13 19:26:11,929 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at kiran/127.0.1.1
*****/
hadoop@kiran: ~$
```

8. Open <http://localhost:9870>



The screenshot shows a web browser window displaying the Hadoop Overview page. The browser's address bar shows the URL `localhost:9870/dfshealth.html#tab-overview`. The page has a green header with navigation tabs: Hadoop, Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The main content area is titled "Overview 'localhost:9000' (✓active)".

Overview 'localhost:9000' (✓active)

Started:	Fri Sep 13 19:43:07 +0530 2024
Version:	3.3.6, r1be78236728da9266a4f88195058f08fd012bf9c
Compiled:	Sun Jun 18 13:52:00 +0530 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-e26c28d7-2ee4-470b-9f46-6a005fbf66a9
Block Pool ID:	BP-37485620-127.0.1.1-1726235771526

Summary

Security is off.
Safemode is off.
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
Heap Memory used 101.67 MB of 215 MB Heap Memory. Max Heap Memory is 871.5 MB.
Non Heap Memory used 48.77 MB of 50.96 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	23.94 GB
Configured Remote Capacity:	0 B
DFS Used:	24 KB (0%)
Non DFS Used:	13.55 GB
DFS Remaining:	9.15 GB (38.21%)
Block Pool Used:	24 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)

9. Open <http://localhost:8088>

The screenshot shows the Hadoop web interface in a Firefox browser. The address bar indicates the URL is `localhost:8088/cluster`. The page title is "All Applications". On the left, there is a sidebar with a "Cluster" menu containing options like "About", "Nodes", "Node Labels", "Applications", and "Scheduler". The main content area displays various metrics:

- Cluster Metrics:** A table showing 0 Apps Submitted, 0 Apps Pending, 0 Apps Running, 0 Apps Completed, 0 Containers Running, and Used Resources of <memory:0 B, vCores:0>.
- Cluster Nodes Metrics:** A table showing 1 Active Node, 0 Decommissioning Nodes, 0 Decommissioned Nodes, and 0 Lost Nodes.
- Scheduler Metrics:** A table showing Capacity Scheduler, Scheduling Resource Type (memory-mb), and Minimum Allocation (<memory:1024, vCores:1>).
- Applications Table:** A table with columns: ID, User, Name, Application Type, Application Tags, Queue, Application Priority, StartTime, LaunchTime, FinishTime, State, FinalStatus, Running Containers, Allocated CPU V-Cores, and Allocated Memory MB. It shows 0 entries.

A message at the top right says "Screenshot captured. You can paste the image from the clipboard."

10. Open <http://localhost:8042>


The screenshot shows the Hadoop web interface in a Firefox browser. The address bar indicates the URL is `localhost:9870/explorer.html/`. The page title is "Browse Directory". The interface includes a navigation bar with links like "Hadoop", "Overview", "Datanodes", "Datanode Volume Failures", "Snapshot", "Startup Progress", and "Utilities". The main content area displays a directory listing for the root path (/):

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	hadoop	supergroup	0 B	Sep 13 20:00	0	0 B	logs
drwxr-xr-x	hadoop	supergroup	0 B	Sep 13 19:59	0	0 B	test1

Below the table, it says "Showing 1 to 2 of 2 entries". At the bottom, it says "Hadoop, 2023."

NodeManager information

localhost:8042/node



NodeManager information

ResourceManager

NodeManager

- Node Information
- List of Applications
- List of Containers

Tools

NodeManager information	
Total Vmem allocated for Containers	16.80 GB
Vmem enforcement enabled	true
Total Pmem allocated for Container	8 GB
Pmem enforcement enabled	true
Total VCores allocated for Containers	8
Resource types	memory-mb (unit=Mi), vcores
NodeHealthyStatus	true
LastNodeHealthTime	Thu Aug 15 16:52:37 IST 2024
NodeHealthReport	
NodeManager started on	Thu Aug 15 16:48:34 IST 2024
NodeManager Version:	3.3.6 from 1be78238728da9266a4f88195058f08fd012bf9c by ubuntu source checksum d42eb795a5eadb0febf5e44a7f87a9 on 2023-06-18T08:31Z
Hadoop Version:	3.3.6 from 1be78238728da9266a4f88195058f08fd012bf9c by ubuntu source checksum 5652179ad55f76cb287d9c633bb53bbd on 2023-06-18T08:22Z