

Compte rendu BIG DATA

I. Installation et configuration :

1. Vérification d'installation de Hadoop et yarn.

Hadoop ainsi que YARN démarrent et s'exécutent correctement.


HadoopOverviewDatanodesDatanode Volume FailuresSnapshotStartup ProgressUtilities

Overview 'hadoop-master:9000' (active)

Started:	Wed Oct 22 21:53:22 +0100 2025
Version:	3.2.0, re97acb3bd8f3befd27418996fa5d4b50bf2e17bf
Compiled:	Tue Jan 08 07:08:00 +0100 2019 by sunlig from branch-3.2.0
Cluster ID:	CID-d1633c5c-7365-42a0-a12a-142ec9ca0d73
Block Pool ID:	BP-507101907-127.0.0.1-1757609495422

Summary

Security is off.
Safemode is off.
8 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 8 total filesystem object(s).
Heap Memory used 80.07 MB of 252 MB Heap Memory. Max Heap Memory is 1.71 GB.



Cluster

AboutNodesNode LabelsApplicationsNEWNEW SAVINGSUBMITTEDACCEPTEDRUNNINGFINISHEDFAILEDKILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores
0	0	0	0	0	0 B	4 GB	0 B	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Reb
2	0	0	0	0	0


Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:2048, vCores:2>

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU Vcores	Allocated Memory MB	Reserved CPU Vcores	Reserved Memory MB
No data available in table															

Showing 0 to 0 of 0 entries



Application

About Jobs

Tools

JobHistory

Retired Jobs

Show 20 entries

Search:

Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed	Reduces Total	Reduces Completed	Elapsed Time
No data available in table												

Showing 0 to 0 of 0 entries

First Previous Next Last

2. Manipulations HDFS :

Terminal

```

2012-01-01 09:01 Anchorage Music 298.86 MasterCard
2012-01-01 09:01 Pittsburgh Sporting Goods 475.26 Amex
2012-01-01 09:01 Spokane Garden 3.85 Amex
2012-01-01 09:01 Spokane Computers 287.65 MasterCard
2012-01-01 09:01 Fresno CDs 466.64 MasterCard
2012-01-01 09:01 Omaha Baby 255.68 MasterCard
2012-01-01 09:02 Chandler Books 414.08 Cash
2012-01-01 09:02 Minneapolis Computers 182.05 Visa
2012-01-01 09:02 Honolulu Cameras 345.18 Discover
2012-01-01 09:02 Indianapolis Books 135.96 Discover
2012-01-01 09:02 Chandler Books 344.09 Discover
root@hadoop-master:~# hdfs dfs -rm /input/purchases.txt
Deleted /input/purchases.txt
root@hadoop-master:~# hdfs dfs -copyFromLocal /shared_volume/purchases.txt /input
root@hadoop-master:~# hdfs dfs -ls /input
Found 1 items
-rw-r--r-- 2 root supergroup 2549 2025-10-22 21:36 /input/purchases.txt
root@hadoop-master:~#

```

RAM 3.95 GB CPU 4.61% Disk: 4.67 GB used (limit 1006.85 GB)

>_ Terminal

	Name	Container ID	Image	Port(s)	CPU (%)	Last started	Actions
<input type="checkbox"/>	hadoop_project	-	-	-	6.31%	59 minutes ago	
<input type="checkbox"/>	hadoop-mast	55fa85ef20ad	yassern1/h	7077:7077 Show all ports (3)	3.89%	59 minutes ago	
<input type="checkbox"/>	hadoop-slave	f5190a6f6c6f	yassern1/h	8040:8042	1.2%	59 minutes ago	

3. Gérer les droits :

```

root@hadoop-master:~# hdfs dfs -chmod 777 /input/purchases.txt
root@hadoop-master:~# hdfs dfs -chmod ugo-x /input/purchases.txt
root@hadoop-master:~# hdfs dfs -ls /input
Found 1 items
-rw-rw-rw- 2 root supergroup 2549 2025-10-22 21:36 /input/purchases.txt

```

4. Déplacer et copier le fichier purchases.txt dans un autre emplacement HDFS (ex. racine) :

```

root@hadoop-master:~# hdfs dfs -ls -R /
drwxr-xr-x   - root supergroup          0 2025-10-23 01:00 /input
-rw-rw-rw-   2 root supergroup      2549 2025-10-22 21:36 /input/purchases.txt
drwxrwx---   - root supergroup          0 2025-10-02 09:06 /tmp
drwxrwx---   - root supergroup          0 2025-10-02 09:06 /tmp/hadoop-yarn
drwxrwx---   - root supergroup          0 2025-10-02 09:06 /tmp/hadoop-yarn/staging
drwxrwx---   - root supergroup          0 2025-10-02 09:06 /tmp/hadoop-yarn/staging/history
drwxrwx---   - root supergroup          0 2025-10-02 09:06 /tmp/hadoop-yarn/staging/history/done
drwxrwxrwt   - root supergroup          0 2025-10-02 09:06 /tmp/hadoop-yarn/staging/history/done_intermediate

```

```

root@hadoop-master:~# hdfs dfs -cp /input/purchases.txt /purchases.txt
root@hadoop-master:~# hdfs dfs -ls -R /
drwxr-xr-x   - root supergroup          0 2025-10-23 01:00 /input
-rw-rw-rw-   2 root supergroup      2549 2025-10-22 21:36 /input/purchases.txt
-rw-r--r--   2 root supergroup      2549 2025-10-23 01:02 /purchases.txt
drwxrwx---   - root supergroup          0 2025-10-02 09:06 /tmp
drwxrwx---   - root supergroup          0 2025-10-02 09:06 /tmp/hadoop-yarn
drwxrwx---   - root supergroup          0 2025-10-02 09:06 /tmp/hadoop-yarn/staging
drwxrwx---   - root supergroup          0 2025-10-02 09:06 /tmp/hadoop-yarn/staging/history
drwxrwx---   - root supergroup          0 2025-10-02 09:06 /tmp/hadoop-yarn/staging/history/done
drwxrwxrwt   - root supergroup          0 2025-10-02 09:06 /tmp/hadoop-yarn/staging/history/done_intermediate

```

```

root@hadoop-master:~# hdfs dfs -get /purchases.txt /shared_volume/achat.txt
root@hadoop-master:~# hdfs dfs -cp /purchases.txt /input/purchases.txt
cp: '/input/purchases.txt': File exists

```

Dans cette dernière image, nous avons essayé de créer une copie du fichier **purchases.txt** sous un autre nom (**achat.txt**) dans la machine locale. Ensuite, nous avons utilisé une commande pour copier le fichier dans la racine de HDFS, mais comme ce fichier existe déjà dans la racine (selon les étapes précédentes), le système a renvoyé le message **“File exists”**. Ainsi, le fait de copier le fichier dans le dossier **/shared_volume** avec un nom différent **ne change pas le nom du fichier dans HDFS**.

5. télécharger un fichier sur hdfs :

```

root@hadoop-master:~# hdfs dfs -mkdir /web_input
root@hadoop-master:~# hdfs dfs -ls /
Found 1 items
drwxr-xr-x   - root supergroup          0 2025-11-15 15:11 /web_input
root@hadoop-master:~# wget http://www.textfiles.com/etext/FICTION/alice.txt
--2025-11-15 15:16:46-- http://www.textfiles.com/etext/FICTION/alice.txt
Resolving www.textfiles.com (www.textfiles.com)... 208.86.224.90
Connecting to www.textfiles.com (www.textfiles.com)[208.86.224.90]:80... connected.
HTTP request sent, awaiting response... 200 OK

```

```

root@hadoop-master:~# hdfs dfs -put /shared_volume/alice.txt /web_input
root@hadoop-master:~# hdfs dfs -ls /web_input
Found 1 items
-rw-r--r--   2 root supergroup      150886 2025-11-15 15:36 /web_input/alice.txt

```