

Week 1 Assignment

Guruprasad Velikadu Krishnamoorthy

College of Science and Technology, Bellevue University

DSC650-T301: Big Data

Professor. Nasheb Ismaily

November 28, 2023

This document includes the screenshot taken by port-forwarding the interfaces from Google Cloud:

1. Screenshot of HDFS on Port 9870:

Overview 'master:9000' (active)

Started:	Mon Nov 27 21:43:24 -0600 2023
Version:	3.2.3. rabe5358143720085498613d399ba3baf01e0f131
Compiled:	Sat Mar 19 20:18:00 -0500 2022 by ubuntu from branch-3.2.3
Cluster ID:	CID-d81fa9ba-9379-4759-aa7e-150b1d239f5a
Block Pool ID:	BP-1076611287-172.28.1.1-1701142971637

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 100.06 MB of 250.5 MB Heap Memory. Max Heap Memory is 1.72 GB.
Non Heap Memory used 49.38 MB of 50.38 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	96.55 GB
Configured Remote Capacity:	0 B
DFS Used:	48 KB (0%)
Non DFS Used:	16.95 GB
DFS Remaining:	79.57 GB (82.41%)
Block Pool Used:	48 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes:	2 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes:	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes:	0
Entering Maintenance Nodes:	0

2. Screenshot of YARN on Port 8088:

All Applications

Cluster Metrics

Apps Submitted	0	Apps Pending	0	Apps Running	0	Apps Completed	0	Containers Running	0	Used Resources	<memory 0 B, vCores 0>	Total Resources	<memory 0 B, vCores 0>	Reserved Resources	<memory 0 B, vCores 0>	Physical Mem Used %	0	Physical V-Cores Used %	0
----------------	---	--------------	---	--------------	---	----------------	---	--------------------	---	----------------	------------------------	-----------------	------------------------	--------------------	------------------------	---------------------	---	-------------------------	---

Cluster Nodes Metrics

Active Nodes	0	Decommissioning Nodes	0	Decommissioned Nodes	0	Lost Nodes	0	Unhealthy Nodes	0	Rebooted Nodes	0	Shutdown Nodes	0
--------------	---	-----------------------	---	----------------------	---	------------	---	-----------------	---	----------------	---	----------------	---

User Metrics for dr who

Apps Submitted	0	Apps Pending	0	Apps Running	0	Apps Completed	0	Containers Running	0	Containers Pending	0	Containers Reserved	0	Memory Used	0 B	Memory Pending	0 B	Memory Reserved	0 B	V-Cores Used	0	V-Cores Pending	0	V-Cores Reserved	0
----------------	---	--------------	---	--------------	---	----------------	---	--------------------	---	--------------------	---	---------------------	---	-------------	-----	----------------	-----	-----------------	-----	--------------	---	-----------------	---	------------------	---

Scheduler Metrics

Scheduler Type	Fair Scheduler	Scheduling Resource Type	memory-mb (unit-MB), vcores	Minimum Allocation	<memory 512, vCores 1>	Maximum Allocation	<memory 1536, vCores 4>	Maximum Cluster Application Priority	0	Scheduler Busy %	0
----------------	----------------	--------------------------	-----------------------------	--------------------	------------------------	--------------------	-------------------------	--------------------------------------	---	------------------	---

Showing 0 to 0 of 0 entries

3. Screenshot of Spark Master on Port 8080:

spark 3.0.0

Spark Master at spark://master:7077

URL: spark://master:7077
Alive Workers: 2
Cores in use: 5 Total: 0 Used
Memory in use: 13.5 GiB Total: 0.0 B Used
Resources in use:
Applications: 0 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers (2)

Worker ID	Address	State	Cores	Memory	Resources
worker-20231120041104-worker1-0001	worker1-0001	ALIVE	4 (0 Used)	6.7 GiB (0.0 B Used)	
worker-20231120041104-worker2-0002	worker2-0002	ALIVE	4 (0 Used)	6.7 GiB (0.0 B Used)	

Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

Completed Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

4. Screenshot of Spark History Server on Port 18080:

spark 3.0.0

History Server

Event log directory: hdfs://log/spark
Last updated: 2023-11-27 21:46:39
Client local time zone: America/Chicago

No completed applications found!

Did you specify the correct logging directory? Please verify your setting of `spark.history.fs.logDirectory` listed above and whether you have the permissions to access it. It is also possible that your application did not run to completion or did not stop the SparkContext.

[Show incomplete applications](#)

5. Screenshot (page1) of HBase on Port 16010:

The screenshot shows the HBase web interface at <http://localhost:16010/master-status>. The interface includes a navigation bar with links like Home, Table Details, Procedures & Locks, HBase Report, Process Metrics, Local Logs, Log Level, Debug Dump, Metrics Dump, Profiler, and HBase Configuration. The main content area displays the following sections:

- Master master**: A section header for the master node.
- Region Servers**: A table showing the status of region servers. The table has columns: ServerName, Start time, Last contact, Version, Requests Per Second, and Num. Regions. It lists two workers and a total of 2.
- Backup Masters**: A table showing backup masters. It has columns: ServerName, Port, and Start Time. It shows a total of 0.
- Tables**: A section header for tables. It includes a sub-header for User Tables, System Tables, and Snapshots.
- Peers**: A table showing peers. It has columns: Peer Id, Cluster Key, Endpoint, State, Is Serial, Bandwidth, ReplicateAll, Namespaces, Exclude Namespaces, Table Cfs, and Exclude Table Cfs. It shows a total of 0.
- Tasks**: A section header for tasks. It includes a sub-header for Show All Monitored Tasks, Show non-RPC Tasks, Show All RPC Handler Tasks, Show Active RPC Calls, and Show Client Operations. It also has a link to View as JSON.

6. Screenshot (page2) of HBase on Port 16010:

The screenshot shows the HBase web interface at <http://localhost:16010/master-status>, specifically the Tasks and Software Attributes sections.

Tasks

Start Time	Description	State	Status
Tue Nov 28 03:45:31 GMT 2023	Initializing region hbase:meta, 1.1588230740	COMPLETE (since 15sec ago)	Region opened successfully (since 15sec ago)
Tue Nov 28 03:45:32 GMT 2023	Closing region 1588230740	COMPLETE (since 15sec ago)	Closed (since 15sec ago)
Tue Nov 28 03:45:33 GMT 2023	Closing region 15956783653499cd5ee436debb2682	COMPLETE (since 14sec ago)	Closed (since 14sec ago)
Tue Nov 28 03:45:33 GMT 2023	Initializing region master store, 1.15956783653499cd5ee436debb2682	COMPLETE (since 14sec ago)	Region opened successfully (since 14sec ago)
Tue Nov 28 03:45:42 GMT 2023	Closing region 375d3d239a6c2f483761f23023daf60	COMPLETE (since 5sec ago)	Closed (since 5sec ago)
Tue Nov 28 03:45:31 GMT 2023	Master startup	COMPLETE (since 4sec ago)	Calling postStartMaster coprocessors (since 4sec ago)

Software Attributes

Attribute Name	Value	Description
JVM Version	icedTea 1.8.0_275-25.275-b01	JVM vendor and version
HBase Version	2.3.6, revision=74145792620ca6b75146c29ab2726c4643ac9	HBase version and revision
HBase Compiled	Wed Jul 28 22:24:42 UTC 2021, stack	When HBase version was compiled and by whom
HBase Source Checksum	eb11554aeb325ccaf6a9b469ef7ddcd8b0eac35a5377db5c09eac5722007a05ef82349c7afb005ed1abe32e772998b2137194641f5c651d525aa5634	HBase source SHA512 checksum
Hadoop Version	2.10.0, revision=21f110e465e787d9567d6a203b72a6eb9194	Hadoop version and revision
Hadoop Compiled	2019-10-22T21:04Z, jhjung	When Hadoop version was compiled and by whom
Hadoop Source Checksum	762d8877c5ce8c9a2cca5c7e81aa4026	Hadoop source MD5 checksum
ZooKeeper Client Version	3.5.7, revision=f0d452973d373f69c3d8b61d59942dc2c7f660e	ZooKeeper client version and revision hash
ZooKeeper Client Compiled	02/10/2020 11:30 GMT	When ZooKeeper client version was compiled
ZooKeeper Quorum	29012181 29022181 29032181	Addresses of all registered ZK servers. For more, see zk.dump.
ZooKeeper Base Path	/hbase	Root node of this cluster in ZK.

Key to add this cluster as a peer for replication. Use "help 'add_peer'" in the shell for details.

7. Screenshot of Solr on Port 8983:

The screenshot displays the Solr Admin web interface. The top navigation bar includes links for Dashboard, Logging, Security, Core Admin, Java Properties, and Thread Dump. The main content area is divided into several sections: 'Instance' with a 'Start' button, 'Versions' showing three versions of Solr (9.4.0), 'JVM' showing the runtime environment (Edgee Adoptium OpenJDK 64-Bit Server VM 17.0.9 17.0.9+9), 'Runtime' showing processors (4), 'Args' listing various command-line arguments, 'System' showing system metrics (Physical Memory 47.7%, File Descriptor Count 0.0%, JVM Memory 41.1%), and 'Security' showing a warning that security is not enabled. The bottom of the screen shows a Windows taskbar with various application icons and the system clock showing 9:56 PM on 11/27/2023.

8. Screenshot of NIFI on Port 8443:

The screenshot displays the Nifi Flow web interface. The top navigation bar includes links for Navigate and Operate. The main content area is divided into two sections: 'Navigate' showing a search bar and a list of flows, and 'Operate' showing the 'Nifi Flow' process group. The bottom of the screen shows a Windows taskbar with various application icons and the system clock showing 10:03 PM on 11/27/2023.