



Hadoop Assessment Report

Report Date Range	Start Date	End Date
	28-Jul-2022	04-Aug-2022

Key Metrics

Metrics	Value
Number of Host	6
Number of NameNodes	1
Number of DataNodes	5
Number of EdgeNodes	0
Average Cluster CPU Utilization	2.70%
Average Cluster Memory Utilization	18.54 GB
Hadoop Major Version	Hadoop 3.0.0
Hadoop Minor Version	CDH-6.3.1
Hadoop Distribution	CDH6
HDFS Storage Available	828.22 GB
HDFS Replication Factor	3
Number of tables in Hive	3
Average No. of Yarn Vcores Used	0.00
Average Yarn Memory Used	0.00 GB
HBase version	2.1.0
Spark version	2.4.0
Kafka version	2.2.1
Impala version	3.2.0



Cluster Information

Number of Cluster Configured: 1

Cluster Details :

Cluster Name	Cloudera Version	Health Status
Cluster 1	6.3.1	GOOD_HEALTH

Listing Details for Cluster: Cluster 1

Number of Host: 6

Number of NameNodes: 1

Number of DataNodes: 5

Number of Edge Nodes: 0

Host Details :

Total Cores in the cluster: 48

Total Memory in the cluster: 175.28 GB

Hostname	Host IP	Cores	Memory	Health Status	Distribution
hadoop-cloudera-master-01.cso.jo onix.net	172.16.20.25	8	50.48	GOOD_HEALTH	CentOS Linux 7 (Core)
hadoop-cloudera-worker-01.cso.jo onix.net	172.16.20.24	8	31.26	GOOD_HEALTH	CentOS Linux 7 (Core)
hadoop-cloudera-master-03.cso.j oonix.net	172.16.20.23	8	31.26	GOOD_HEALTH	CentOS Linux 7 (Core)
hadoop-cloudera-worker-02.cso.j oonix.net	172.16.20.21	8	15.51	GOOD_HEALTH	CentOS Linux 7 (Core)
hadoop-cloudera-master-02.cso.jo onix.net	172.16.20.22	8	31.26	GOOD_HEALTH	CentOS Linux 7 (Core)



hadoop-cloudera-worker-03.cso.joonix.net	172.16.20.26	8	15.51	GOOD_HEALTH	CentOS Linux 7 (Core)
------------------------------------------	--------------	---	-------	-------------	-----------------------

MasterNodes Details :

Total Cores Assigned to All the MasterNode: 8

Total Memory Assigned to All the MasterNodes: 50.48 GB

Hostname	Cores	Memory
hadoop-cloudera-master-01.cso.joonix.net	8	50.48 GB

DataNodes Details :

Total Cores Assigned to All the DataNode: 40

Total Memory Assigned to All the DataNodes: 124.80 GB

Hostname	Cores	Memory
hadoop-cloudera-worker-01.cso.joonix.net	8	31.26 GB
hadoop-cloudera-master-03.cso.joonix.net	8	31.26 GB
hadoop-cloudera-worker-02.cso.joonix.net	8	15.51 GB
hadoop-cloudera-master-02.cso.joonix.net	8	31.26 GB
hadoop-cloudera-worker-03.cso.joonix.net	8	15.51 GB

EdgeNodes Details :

Total Cores Assigned to All the EdgeNode: 0

Total Memory Assigned to All the EdgeNodes: 0.00 GB

Clients Installed on Gateway :

Services
spark_on_yarn
hive
sentry
kafka

Services Running in the Cluster :

Service Name	Health Status	Health Concerns
zookeeper	GOOD_HEALTH	
oozie	GOOD_HEALTH	
impala	GOOD_HEALTH	
ks_indexer	GOOD_HEALTH	
solr	GOOD_HEALTH	
hue	GOOD_HEALTH	
spark_on_yarn	GOOD_HEALTH	
hdfs	GOOD_HEALTH	

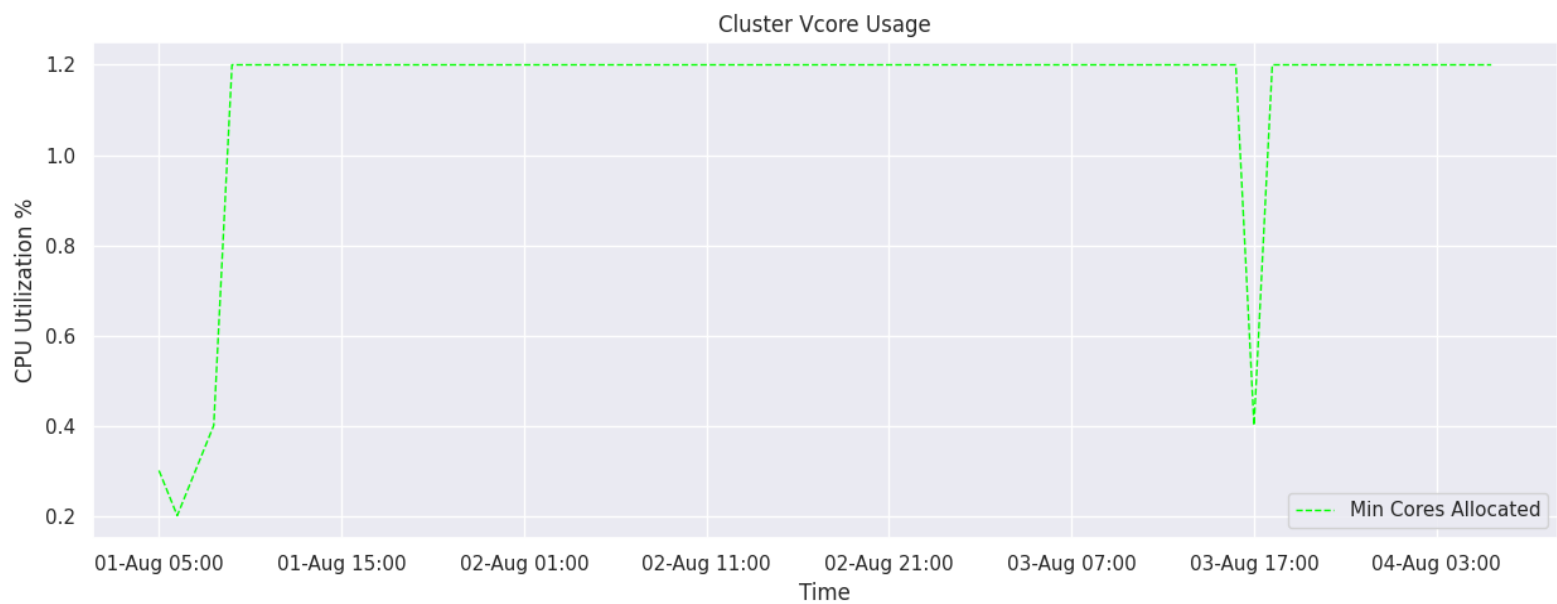
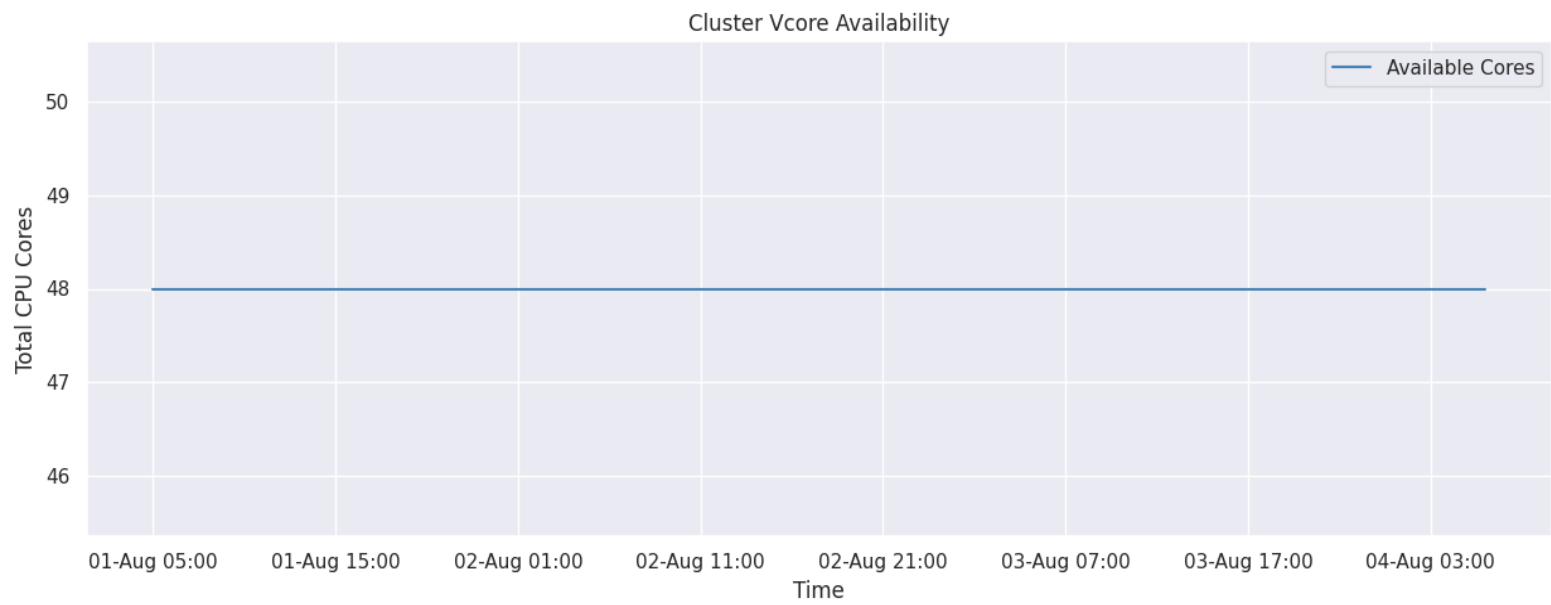


yarn	GOOD_HEALTH	
hbase	GOOD_HEALTH	
hive	GOOD_HEALTH	
sentry	GOOD_HEALTH	
kafka	GOOD_HEALTH	

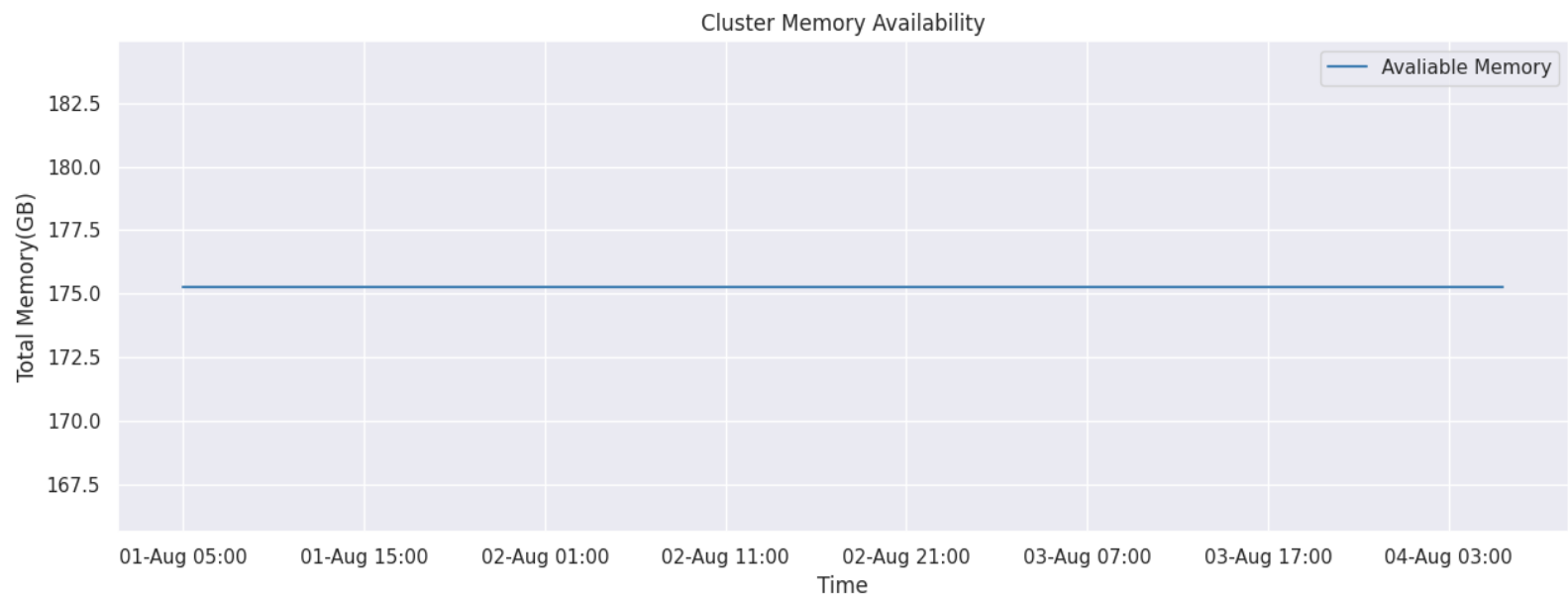


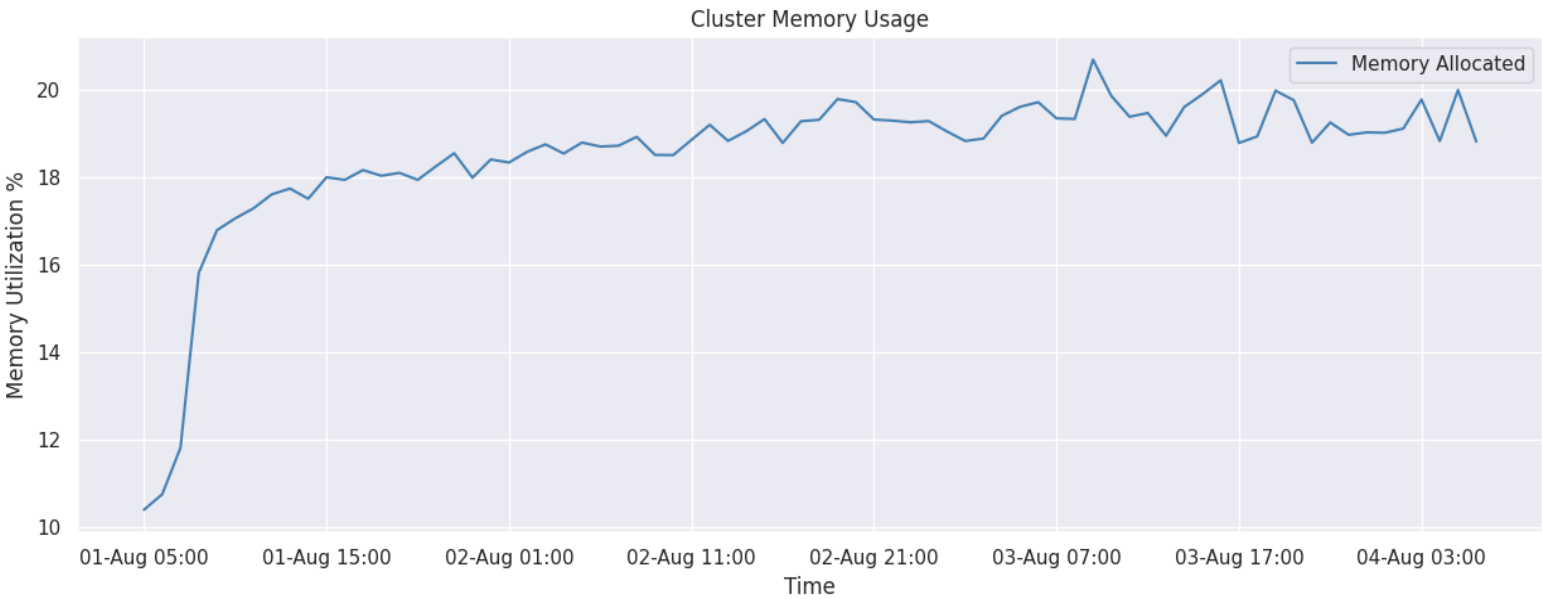
Cluster Metrics

Average Cluster CPU Utilization: 2.70 %



Average Cluster Memory Utilization: 18.54 GB







Hardware and OS Metrics

Types of Servers Details:

Database Server: MYSQL

DNS Server: DNS server is not enabled within machine

Web Server: Web server is not enabled

NTP Server: enabled

Manufacturer and Processor Details:

Manufacturer name: GenuineIntel

Serial Number: 57 06 05 00 FF FB 8B 1F

Family: 6

Model Name: Intel(R) Xeon(R) Platinum 8270 CPU @ 2.70GHz

Microcode: 0x5003302

CPU MHz: 2693.672

CPU Family: 6

Network Interface Card Details:

Network Adapter	IP Address
ens192	172.16.20.25/24
docker0	172.17.0.1/16

List of Hadoop and Non-Hadoop Libraries:

Hadoop Libraries	Non Hadoop Libraries
hadoop	python3.6
zlib	
zstd	
snappy	
lz4	
bzip2	
openssl	
ISA-L	

Programming Softwares:

Python3 installed

Java installed

Scala installed

List of security softwares:

Ranger is not installed

Knox-server is not installed

Splunk is not installed

Nagios is not installed

GRR Rapid response is not installed

MISP is not installed

TheHive is not installed

OSQuery is not installed



Cloudera Navigator is not installed

[Speciality Hardware:](#)

GPU is present



Frameworks and Software Details

Hadoop Major Version: Hadoop 3.0.0

Hadoop Minor Version: CDH-6.3.1

Hadoop Distribution: CDH6

List of Services Installed :

Name	Version
hbase	2.1.0
zookeeper	3.4.5
sentry	2.1.0
kafka	2.2.1
oozie	5.1.0
solr	7.4.0
hive	2.1.1
hue	4.2.0
spark	2.4.0
impala	3.2.0
hadoop-hdfs	3.0.0

Drivers and Connectors:

Salesforce driver not found

Sap driver not found

Drivers
jdbc-stdext-1.8.0.342.b07.jar
jdbc-stdext-1.8.0.jar
jdbc-stdext-3.0.jar
jdbc-stdext.jar
jboss-common-jdbc-wrapper-src.jar
jboss-common-jdbc-wrapper.jar
spring-jdbc-4.3.14.RELEASE.jar
hive-jdbc-1.1.0-cdh5.12.0.jar
openjpa-persistence-jdbc-2.2.2.jar
postgresql-9.0-801.jdbc4.jar
openjpa-jdbc-2.2.2.jar
hive-jdbc-2.1.1-cdh6.3.0.jar
openjpa-jdbc-2.4.2.jar
openjpa-persistence-jdbc-2.4.2.jar
hive-jdbc-1.1.0-cdh5.14.0.jar
hive-jdbc-2.1.1-cdh6.3.0.a9016068a26246ac47c4b2637db33adb.jar
spring-jdbc-4.3.14.RELEASE.7a99635c7ea1aa3632642de14b18022f.jar
jdbcdslog-1.0.6.jar
hive-jdbc-2.1.1-cdh6.3.1.jar
hive-jdbc.jar
hive-jdbc-2.1.1-cdh6.3.1-standalone.jar
mssql-jdbc-6.2.1.jre7.jar
flume-jdbc-channel-1.9.0-cdh6.3.1.jar



flume-jdbc-channel.jar

Installed Connectors

mysql-connector-java.jar

mysql-connector-java-5.1.46-bin.jar

mysql-connector-java-5.1.46.jar

gcs-connector-hadoop3-1.9.10-cdh6.3.1-shaded.jar

gcs-connector-hadoop2-latest.jar



HDFS Section

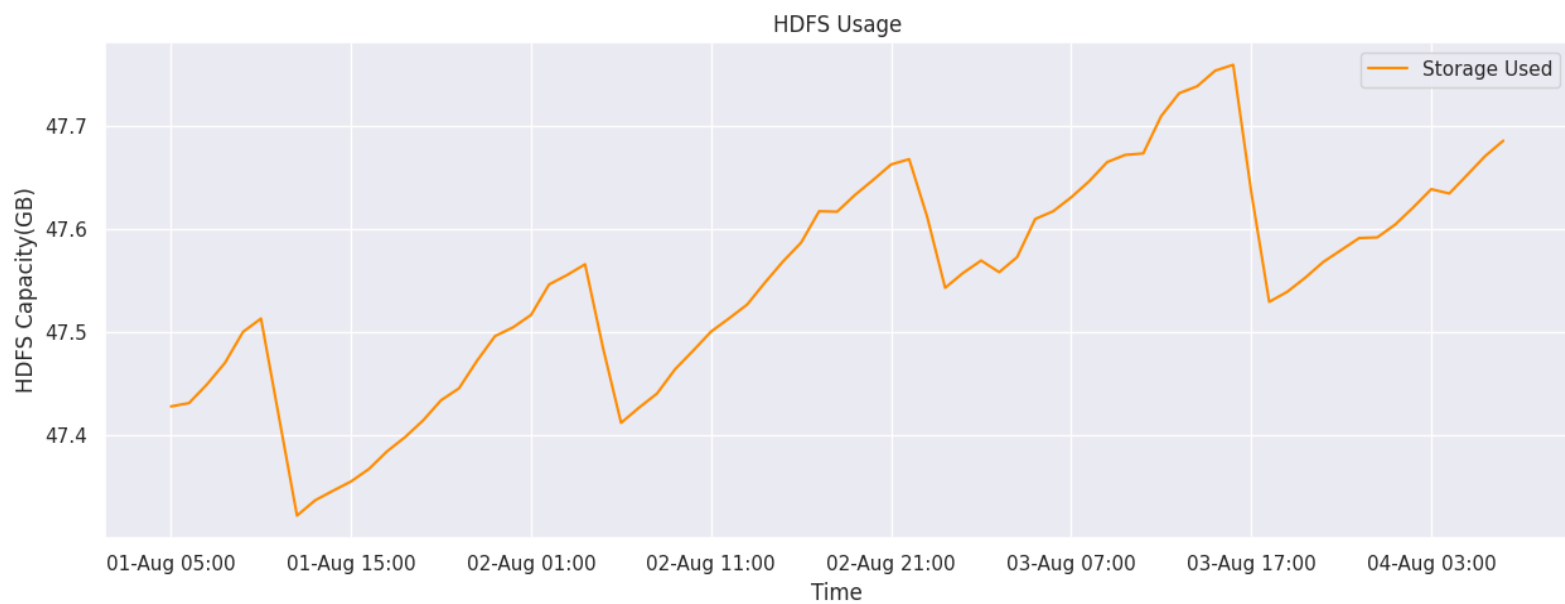
Replication Factor: 3

Trash Interval Setup in the Cluster: Enabled

Compression Technique used: SnappyCodec

HDFS Storage Available: 828.22 GB

HDFS Storage Used: 47.69 GB



HDFS Folder and File Structure:

Folder	Owner	Size	Modified Date	Modified Time	Storage Policy	Permissions	User	User-Group	Other
/hbase	hbase	0	2022-07-31	21:47	unspecified	drwxr-xr-x	rwX	r-x	r-x
/solr	solr	0	2019-10-25	12:58	unspecified	drwxrwxr-x	rwX	rwX	r-x
/tmp	hdfs	0	2019-10-25	12:59	unspecified	drwxrwxrwt	rwX	rwX	rwX
/user	hdfs	0	2019-12-16	14:02	unspecified	drwxr-xr-x	rwX	r-x	r-x

HDFS Data Distribution:

Maximum Size of a file: 35.31 MB

Minimum Size of a file: 0.10 MB

Average Size of a file: 1.38 MB

File Type	Size (mb)
jar	1417.15
csv	16.42
zip	0.56



Hive Metrics

Hive Metastore Details:

Metastore Host: hadoop-cloudera-master-01

Metastore Database: mysql

Metastore Database Name: metastore

Metastore Database Port: 3306

Hive Details:

Number of Databases: 1

Number of tables with partition: 0

Number of tables without partition: 3

Number of Internal Tables: 2

Number of External Tables: 1

Hive Execution Engine: mr

Do tables require Transaction Locking concurrency: No

Hive Interactive Queries: NO

Hot(within 1 day), Warm(from 1 to 3 days), Cold(more than 3 days)



Security

Kerberos Details: Cluster is not kerberized

AD Server Name and Port: LDAP server not present

SSL/TLS Status:

SSL on HDFS is not enabled

SSL on Yarn is not enabled

SSL on Mapreduce is not enabled

Kerberos Status:

keytab not exist

Kerberos on hue is not enabled

Kerberos on mapreduce is not enabled

Kerberos on HDFS is not enabled

Kerberos on yarn is not enabled



Network, Traffic and Monitoring Metrics

Ingress Traffic:

Maximum Incoming Throughput: 13.98 Kbps

Minimum Incoming Throughput: 10.17 Kbps

Average Incoming Throughput: 12.17 Kbps

Current Incoming Throughput: 12.51 Kbps

Egress Traffic:

Maximum Outgoing Throughput: 25.31 Kbps

Minimum Outgoing Throughput: 20.95 Kbps

Average Outgoing Throughput: 23.03 Kbps

Current Outgoing Throughput: 25.27 Kbps

Third Party Monitoring Tools:

Nagios is not Installed

Zabbix is not Installed

Cacti is not Installed

Prometheus server is not enabled

grafana server is not present

ganglia server is not present

check mk server is not present

Orchestration Tools:

oozie is enabled

crontab is installed

airflow is not enabled

Logging Tools:

Datadog is not enabled

Splunk not installed

Newrelic not installed

Elasticsearch not installed

Name	Logs Path
anaconda	/var/log/anaconda
btmp-20220801	/var/log/btmp-20220801
catalogd	/var/log/catalogd
cloudera-scm-agent	/var/log/cloudera-scm-agent



cloudera-scm-alertpublisher	/var/log/cloudera-scm-alertpublisher
cloudera-scm-eventserver	/var/log/cloudera-scm-eventserver
cloudera-scm-firehose	/var/log/cloudera-scm-firehose
cloudera-scm-headlamp	/var/log/cloudera-scm-headlamp
cloudera-scm-server	/var/log/cloudera-scm-server
firewalld	/var/log/firewalld
hadoop-hdfs	/var/log/hadoop-hdfs
hadoop-mapreduce	/var/log/hadoop-mapreduce
hadoop-yarn	/var/log/hadoop-yarn
hbase	/var/log/hbase
hbase-solr	/var/log/hbase-solr
hive	/var/log/hive
httpd	/var/log/httpd
hue	/var/log/hue
hue-httpd	/var/log/hue-httpd
impala-minidumps	/var/log/impala-minidumps
kafka	/var/log/kafka
kafka-logs	/var/log/kafka-logs
mysqld.log	/var/log/mysqld.log
ntpstats	/var/log/ntpstats
oozie	/var/log/oozie
rhsm	/var/log/rhsm
samba	/var/log/samba
sentry	/var/log/sentry
solr	/var/log/solr
spark	/var/log/spark
sssd	/var/log/sssd
statestore	/var/log/statestore
tuned	/var/log/tuned
vmware-imc	/var/log/vmware-imc
vmware-network.1.log	/var/log/vmware-network.1.log
vmware-network.2.log	/var/log/vmware-network.2.log
vmware-network.3.log	/var/log/vmware-network.3.log
vmware-network.4.log	/var/log/vmware-network.4.log
vmware-network.5.log	/var/log/vmware-network.5.log
vmware-network.6.log	/var/log/vmware-network.6.log
vmware-network.7.log	/var/log/vmware-network.7.log
vmware-network.8.log	/var/log/vmware-network.8.log
vmware-network.9.log	/var/log/vmware-network.9.log
vmware-network.log	/var/log/vmware-network.log
vmware-vgauthsvc.log.0	/var/log/vmware-vgauthsvc.log.0
vmware-vmsvc.log	/var/log/vmware-vmsvc.log
vmware-vmsvc-root.log	/var/log/vmware-vmsvc-root.log
vmware-vmtoolsd-root.log	/var/log/vmware-vmtoolsd-root.log
wtmp	/var/log/wtmp
yum.log	/var/log/yum.log
yum.log-20200101	/var/log/yum.log-20200101
yum.log-20220801	/var/log/yum.log-20220801
zookeeper	/var/log/zookeeper



Yarn Metrics

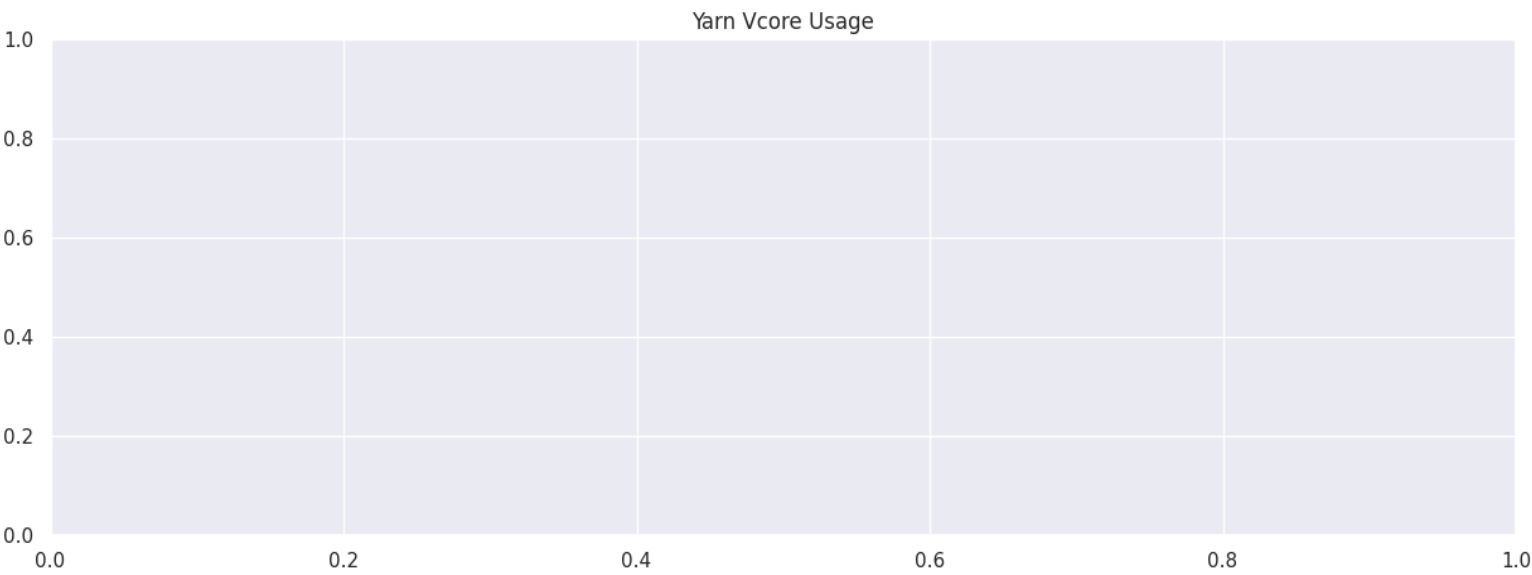
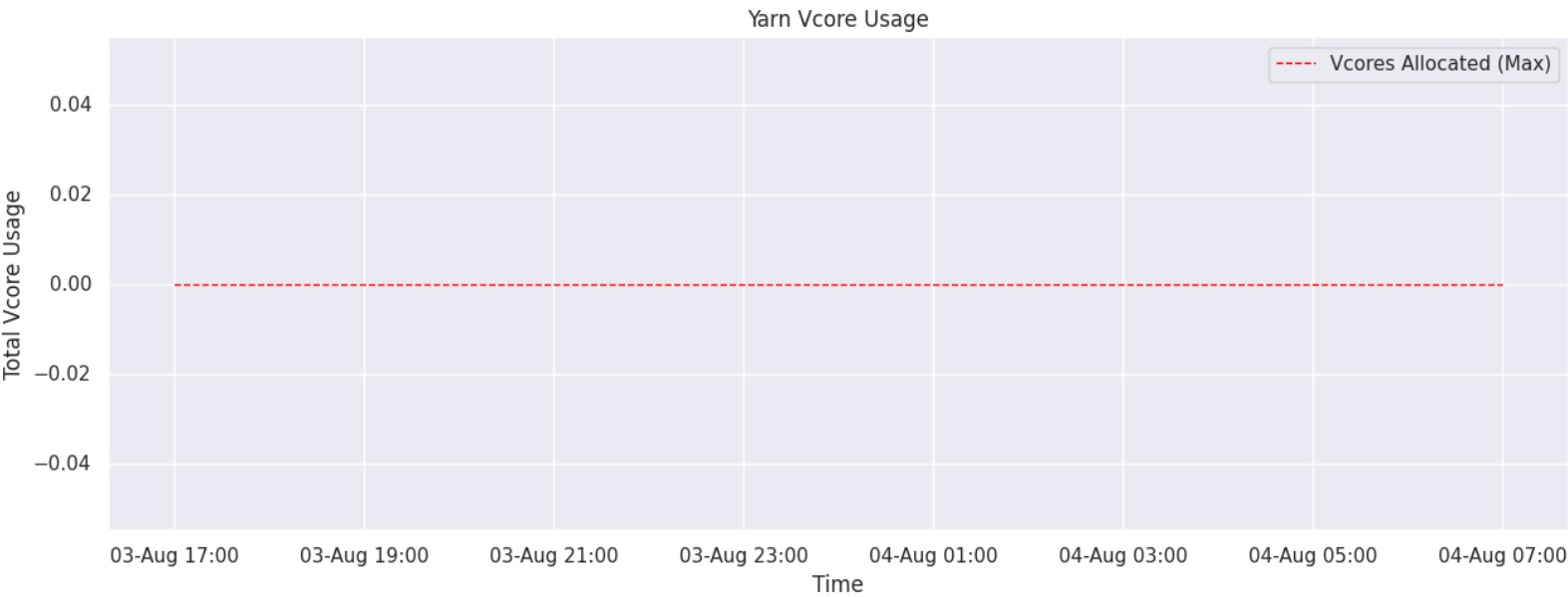
Dynamic resource pool is not Configured

HA Status:

- HDFS: Disabled
- Yarn: Disabled
- Hive: Disabled
- Zookeeper: Disabled

VCore Details:

Total Yarn Vcore: 56.00
Average No. of Vcores Used: 0.00

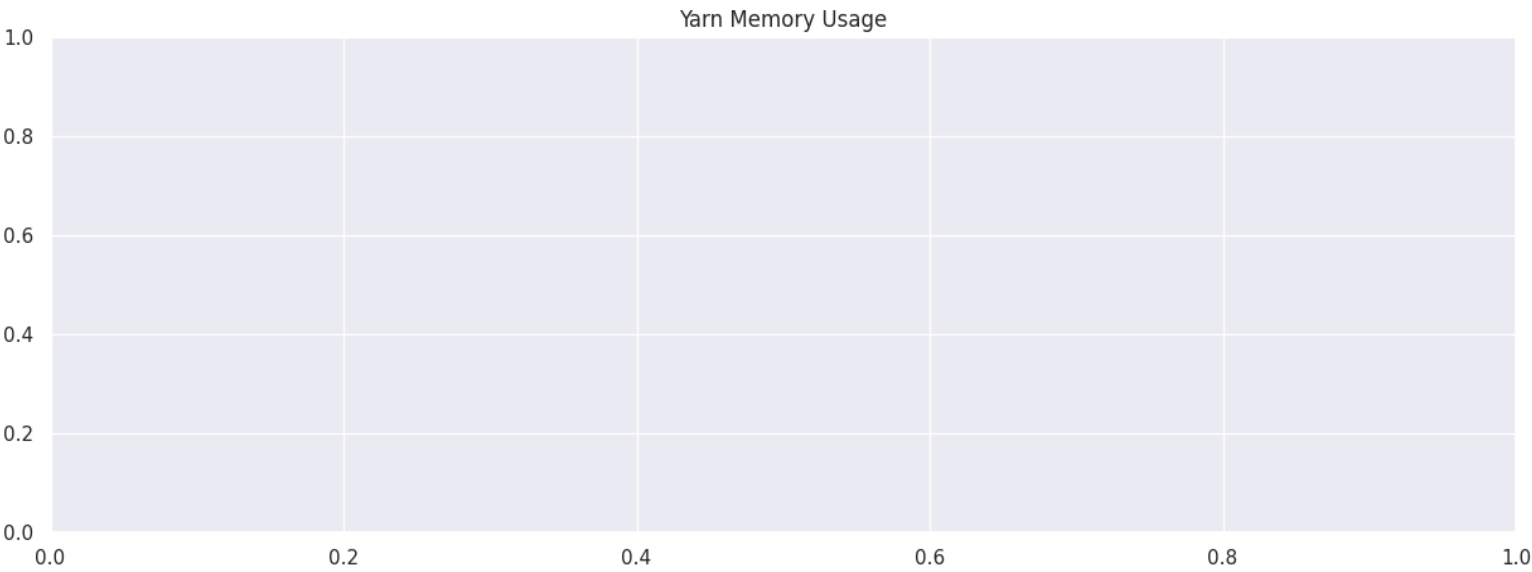
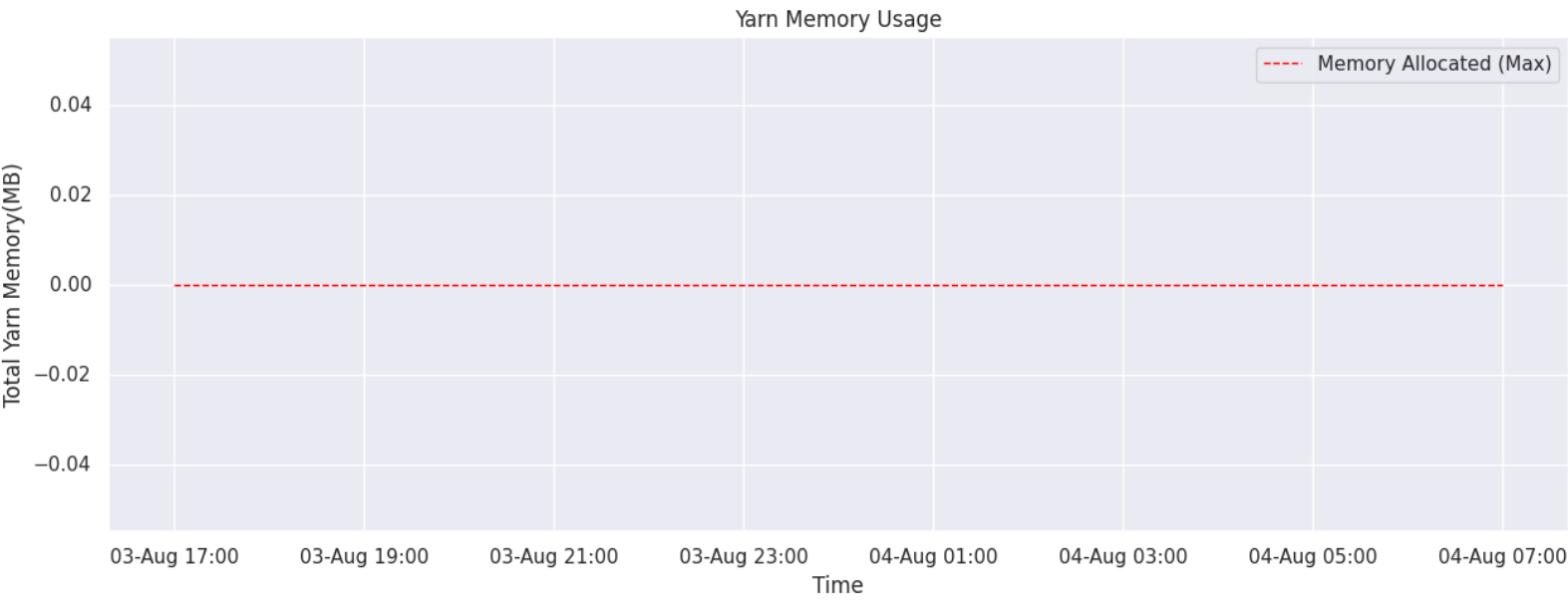




Memory Details:

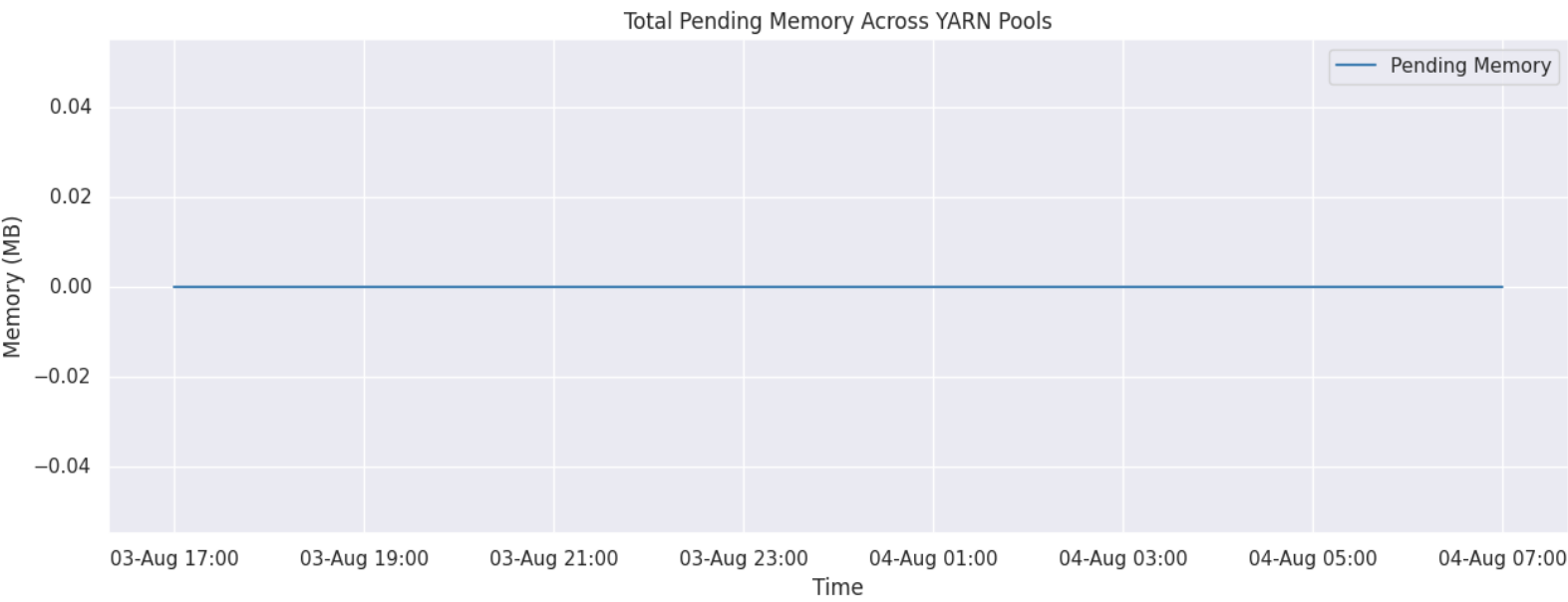
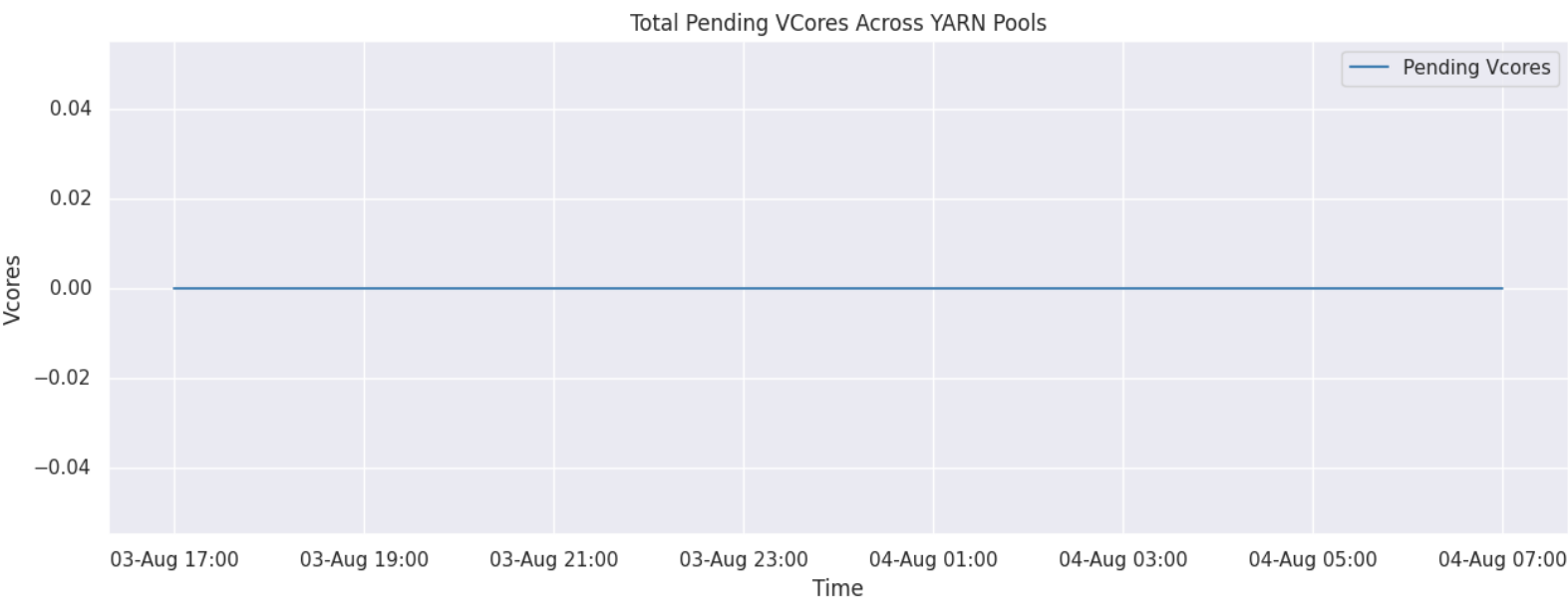
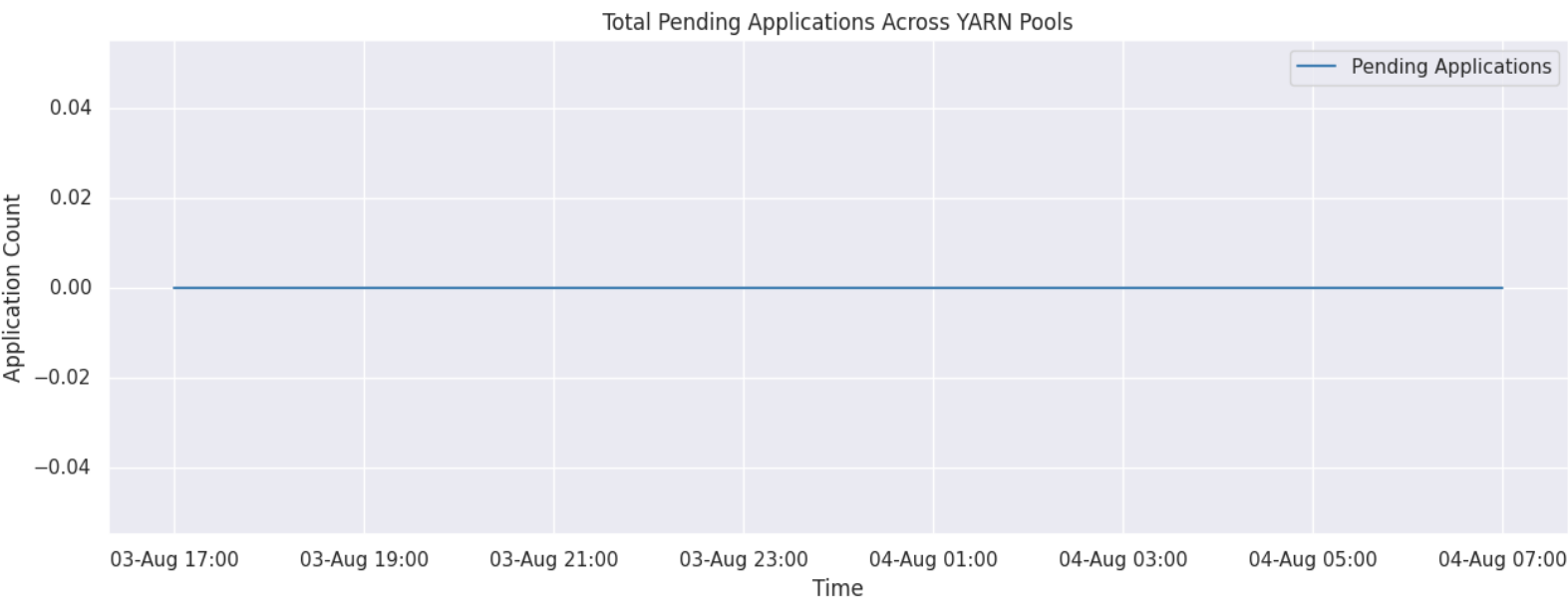
Total Yarn Memory: 15.00 GB

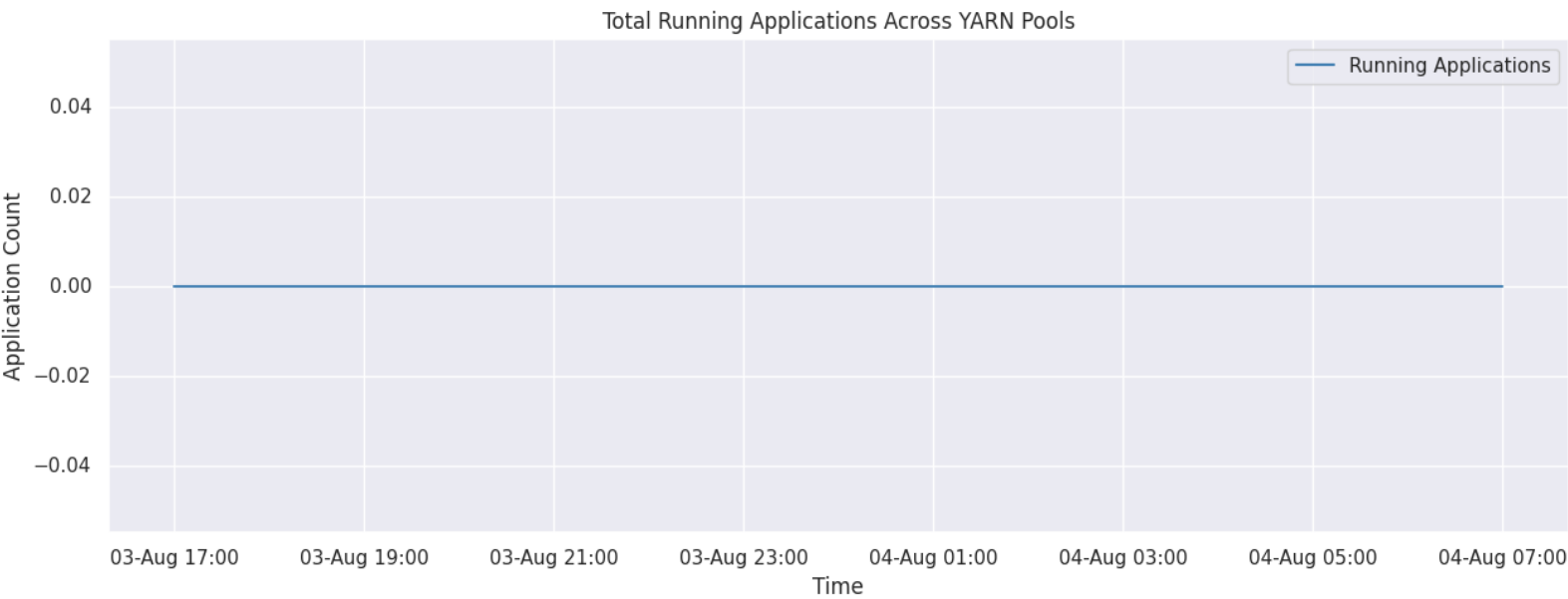
Average Yarn Memory Used: 0.00 MB





Yarn Pending and Running Applications







HBase Metrics

Number of Nodes Serving HBase: 5

Is Hbase replicated to other datacenter: Yes

Do you use Secondary Index on Hbase: Yes

Does Hbase uses Query Engine: Yes

Does Hbase use Apache Phoenix: No

Does Hbase use Triggers and Coprocessors: No

Spark Metrics

Spark Version: 2.4.0-cdh6.3.1

Programming Languages Used By Spark:

Spark Resource Manager: yarn

Dynamic Allocation: Enabled

[Spark Components:](#)

Spark RDD API not found

Spark SQL Dataset API not found

Spark SQL API not found

Spark Dataframe API not found

Spark Streaming API not found

Spark ML Lib API not found

Kafka Metrics

Kafka Retention Period: 168.0 hours

Number of Topics in Kafka: 2

Total Size of Messages in Kafka: 1299.00 KB

Total Number of Messages in Kafka: 75

Cloudera Services

Cloudera Services Used for Data Ingestion: Impala, Hive, Kafka

Impala found with version: 3.2.0

Apache Sentry found with version: 2.1.0

Apache Kudu found with version: 1.10.0