MANUFACTURING & AUTOMOTIVE

CONTENT

ABOUT

Our client is a leading provider in the automotive sector, offering a range of services including stolen vehicle tracking, emergency assistance, usage-based insurance, and fleet management.

The project involves the collection of a wide array of data, including geolocation, vehicle position, speed, and acceleration, through advanced electronic devices.

In response to the growing volume and frequency of data collected by on-board sensors, we developed a robust platform for our client capable of storing, streaming, and analyzing vast amounts of vehicle data. This platform, built on technologies such as Apache Kafka, Spark, and Hbase, not only enhanced our client's ability to support insurance companies in the risk assessment process but also enabled the provision of real-time services to customers.

Looking ahead, our client is exploring the potential of incorporating advanced methodologies into their system. This would allow for the implementation of proactive and predictive maintenance for their hardware, such as engines and other critical components

CHALLENGES, SOLUTION & THE RESULTS

CHALLENGES

- Ensuring customer privacy while providing personalized services based on individual data.
- Managing the growing volume and frequency of data collected by on-board sensors.
- Developing a platform capable of storing, streaming, and analyzing vast amounts of data.
- Scaling the architectural structure to optimize data acquisition and processing.

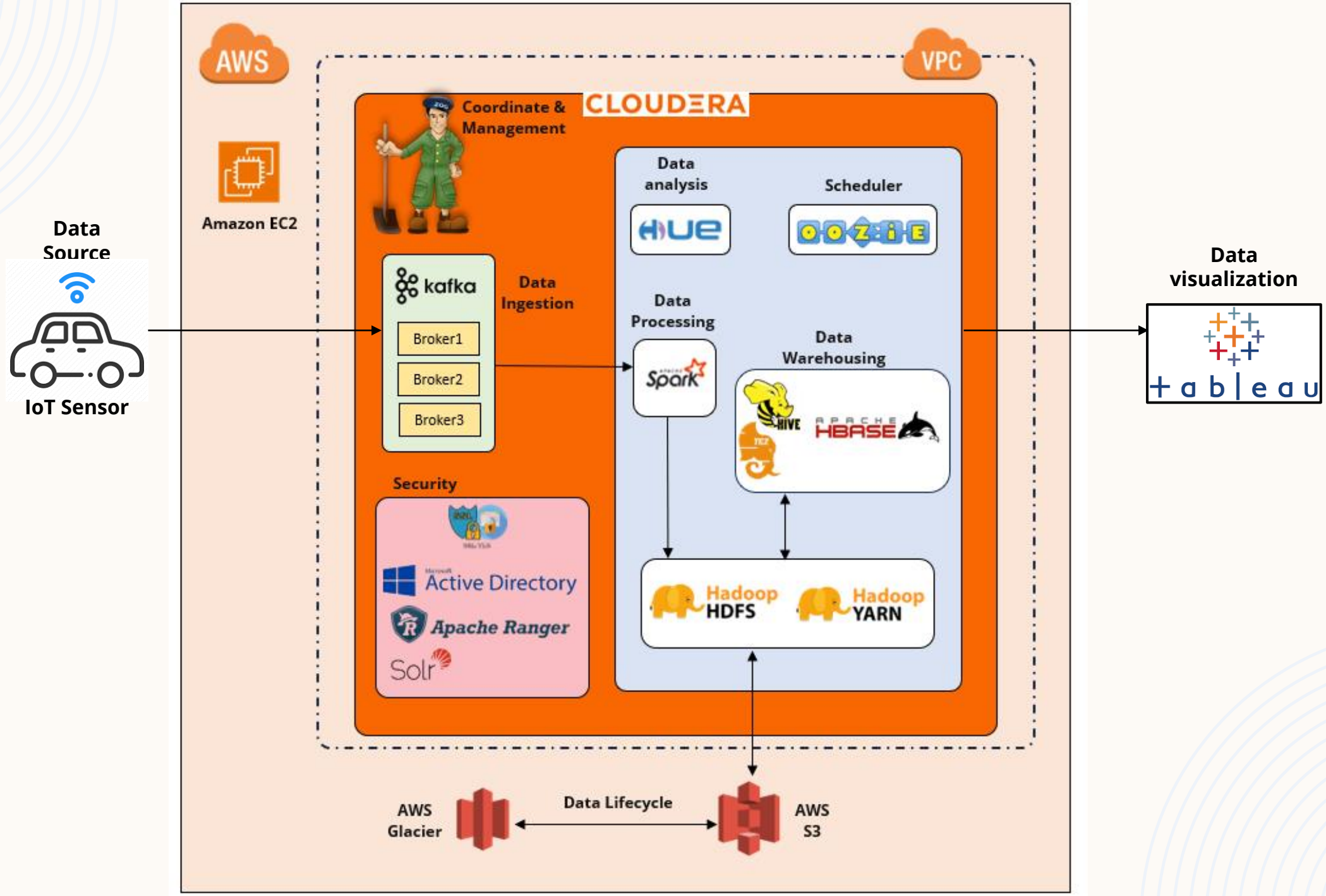
SOLUTION

- Developed a robust platform capable of storing, streaming, and analyzing vast amounts of vehicle data.
- The platform, built on technologies such as Apache Kafka, Spark, and HBase, can process information with latencies of a few seconds, regardless of quantity and frequency.
- The architecture allows for effective management of both real-time processing of data as well as storage for subsequent processing.
- Fast processing of vehicle data enables the gathering of information on available trips, speed data, and geographical information acquired through GPS.

RESULTS

- The project has enabled better support for insurance companies in the risk assessment process and offer real-time services to customers.
- The CDP platform enables the company to meet service levels in the order of seconds required by customers and has laid the foundations for offering new services.
- The new data management strategy gives a competitive advantage and the foundation to position as one of the most innovative players in this market.
- Exploring the inclusion of advanced methodologies to allow the automation of device malfunction notifications and achieve proactive and predictive maintenance for its hardware.
- Predictive maintenance would also avoid incorrect data being collected into the platform by the malfunctioning device and would avoid time-consuming retroactive data clean-up

DATA FLOW DIAGRAM


Data Source
IoT Sensor
AWS
Amazon EC2
VPC
CLOUDERA
Coordinate & Management
Data analysis
Scheduler
HUE
OOZIE
kafka
Broker1
Broker2
Broker3
Data Ingestion
Data Processing
Spark
Data Warehousing
HIVE
APACHE HBASE
Security
Active Directory
Apache Ranger
Solr
Hadoop HDFS
Hadoop YARN
Data visualization
tableau
AWS Glacier
Data Lifecycle
AWS S3

CLUSTER & CAPACITY PLANNING

CLUSTER PLANNING

ENVIRONMENT	OPERATING SYSTEM	LICENSE SOFTWARE	JDK
CDP Private Cloud Base	AWS	RHEL: CentOS-7	Cloudera Support License	Oracle JDK 8

SERVICES

- Cloudera Data Platform (CDP) Private Cloud (PvC) Base version 7.1.7
- Cloudera Manager version 7.4.4
- Apache Kafka: 2.5.0.7.1.7.0-551
- Apache Spark: 2.4.7.7.1.7.0-551
- Apache ZooKeeper: 3.5.5.7.1.7.0-551
- Hive on Tez: 1.0.0
- Hive Meta Store (HMS): 1.0.0
- Apache Hbase: 2.2.3.7.1.7.0-551
- Apache Hadoop (Includes YARN and HDFS): 3.1.1.7.1.7.0-551
- Apache Oozie: 5.1.0.7.1.7.0-551
- Apache Ranger: 2.1.0.7.1.7.0-551
- Apache Solr: 8.4.1.7.1.7.0-551
- Hue: 4.5.0.7.1.7.0-551
- Active Directory
- Tableau
- AWS EC2, EBS, S3 & VPC

CAPACITY PLANNING

Daily Data in Motion:	60	GB
Replication Factor:	3	
Total Data in a Day:	180	GB
Data in a Month:	5.4	TB
Data in Year:	64.8	TB
Data at Rest:	7	TB
Replication of data at rest	21	TB
DFS Data:	85.8	TB
10% Overhead Needed:	8.58	TB
Non DFS Data and Overhead 30% :	25.74	TB
Final Data:	**120.12**	**TB**

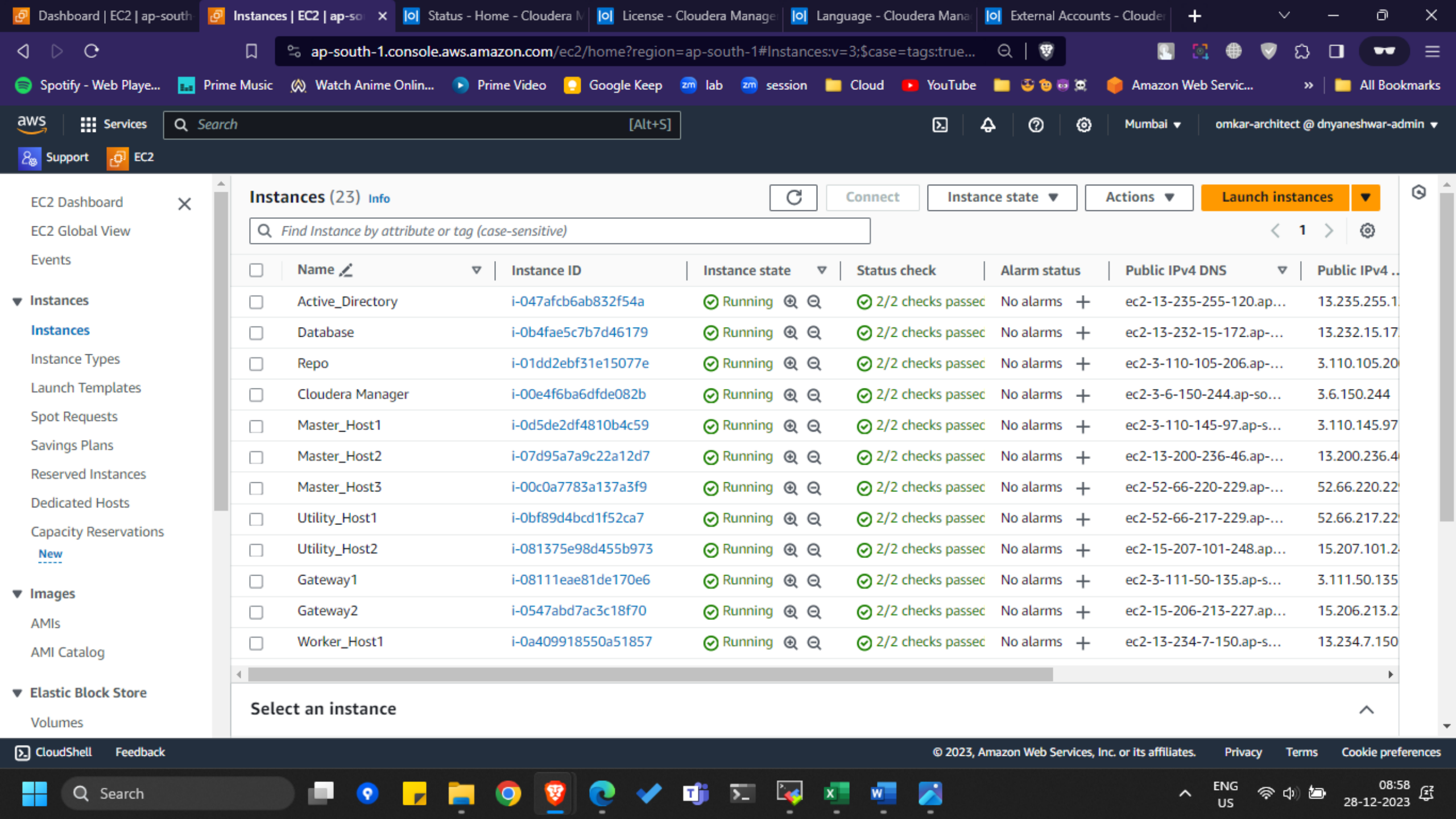
TB per Node	16
Total Data Node	8
10% Data Node Failure	1
Total Data Node	**9**

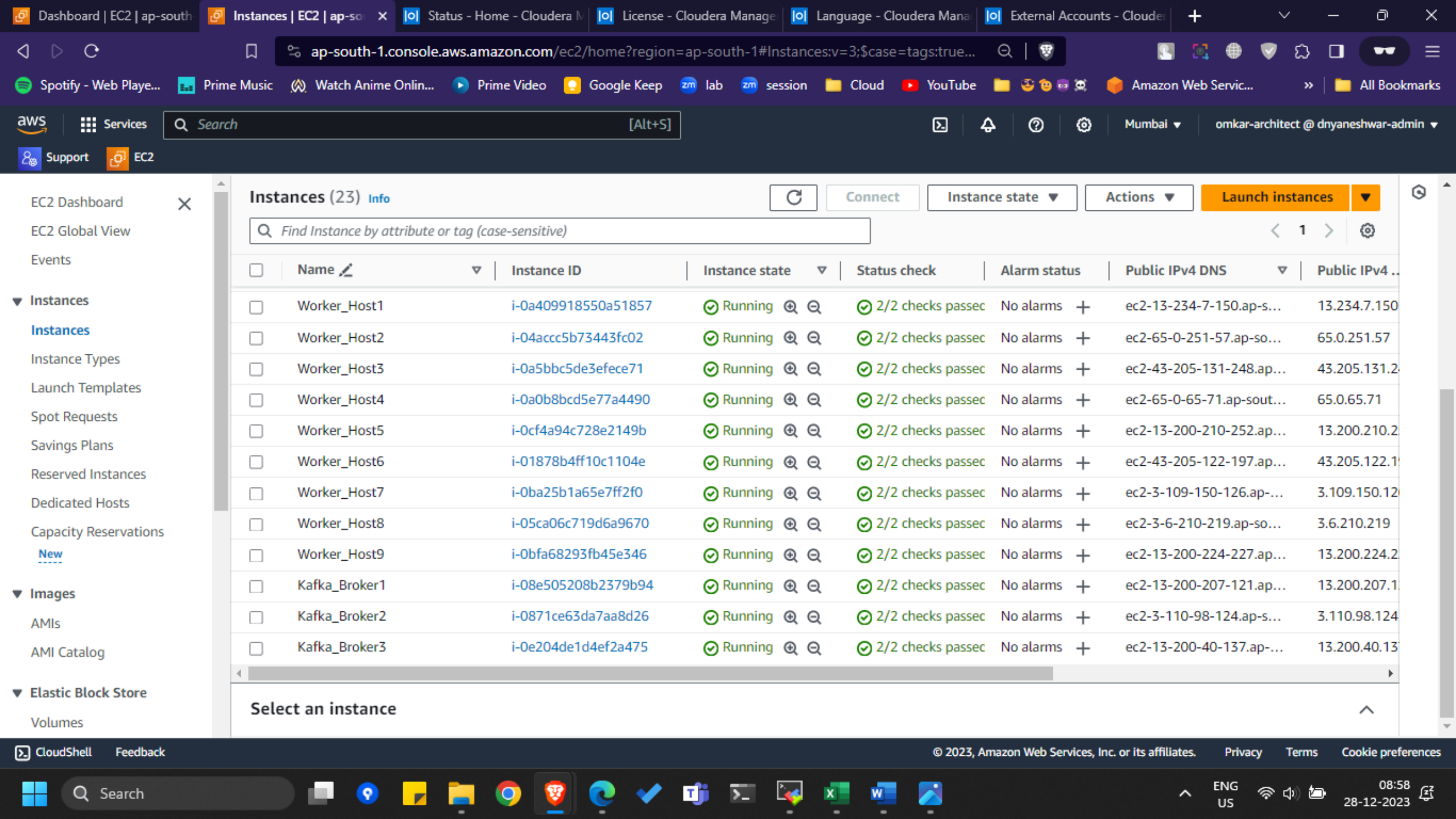
Active directory	1
Master	3
Worker	9
Edge	2
Utility	2
Kafka Broker	3
Total node in Cluster	**20**

CAPACITY PLANNING

Specification	Instance Type	vCPU	Memory	Storage	No. of Nodes
Master Node	r5.4xlarge	16	128	2	3
Worker Node	H1.16xlarge	64	256	16	12
Utility Node	r5.4xlarge	16	128	2	2
Gateway	r5.4xlarge	16	128	2	2

DEPLOYMENT


Dashboard | EC2 | ap-south
Instances | EC2 | ap-so
Status - Home - Cloudera M
License - Cloudera Manage
Language - Cloudera Mana
External Accounts - Cloude
ap-south-1.console.aws.amazon.com/ec2/home?region=ap-south-1#Instances:v=3;$case=tags:true...
Spotify - Web Playe...
Prime Music
Watch Anime Onlin...
Prime Video
Google Keep
lab
session
Cloud
YouTube
Amazon Web Servic...
All Bookmarks
aws
Services
Search
[Alt+S]
Mumbai
omkar-architect @ dnyaneshwar-admin
Support
EC2
EC2 Dashboard
EC2 Global View
Events
Instances
Instances
Instance Types
Launch Templates
Spot Requests
Savings Plans
Reserved Instances
Dedicated Hosts
Capacity Reservations
New
Images
AMIs
AMI Catalog
Elastic Block Store
Volumes
Instances (23) Info
Connect
Instance state
Actions
Launch instances
Find Instance by attribute or tag (case-sensitive)
1
Name
Instance ID
Instance state
Status check
Alarm status
Public IPv4 DNS
Public IPv4 ..
Active_Directory i-047afcb6ab832f54a Running 2/2 checks passec No alarms ec2-13-235-255-120.ap... 13.235.255.1
Database i-0b4fae5c7b7d46179 Running 2/2 checks passec No alarms ec2-13-232-15-172.ap-... 13.232.15.17
Repo i-01dd2ebf31e15077e Running 2/2 checks passec No alarms ec2-3-110-105-206.ap-... 3.110.105.20
Cloudera Manager i-00e4f6ba6dfde082b Running 2/2 checks passec No alarms ec2-3-6-150-244.ap-so... 3.6.150.244
Master_Host1 i-0d5de2df4810b4c59 Running 2/2 checks passec No alarms ec2-3-110-145-97.ap-s... 3.110.145.97
Master_Host2 i-07d95a7a9c22a12d7 Running 2/2 checks passec No alarms ec2-13-200-236-46.ap-... 13.200.236.4
Master_Host3 i-00c0a7783a137a3f9 Running 2/2 checks passec No alarms ec2-52-66-220-229.ap-... 52.66.220.22
Utility_Host1 i-0bf89d4bcd1f52ca7 Running 2/2 checks passec No alarms ec2-52-66-217-229.ap-... 52.66.217.22
Utility_Host2 i-081375e98d455b973 Running 2/2 checks passec No alarms ec2-15-207-101-248.ap... 15.207.101.2
Gateway1 i-08111eae81de170e6 Running 2/2 checks passec No alarms ec2-3-111-50-135.ap-s... 3.111.50.135
Gateway2 i-0547abd7ac3c18f70 Running 2/2 checks passec No alarms ec2-15-206-213-227.ap... 15.206.213.2
Worker_Host1 i-0a409918550a51857 Running 2/2 checks passec No alarms ec2-13-234-7-150.ap-s... 13.234.7.150
Select an instance
CloudShell
Feedback
© 2023, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Search
ENG
US
08:58
28-12-2023


Dashboard | EC2 | ap-south
Instances | EC2 | ap-so
Status - Home - Cloudera M
License - Cloudera Manage
Language - Cloudera Mana
External Accounts - Cloude
ap-south-1.console.aws.amazon.com/ec2/home?region=ap-south-1#Instances:v=3;$case=tags:true...
Spotify - Web Playe...
Prime Music
Watch Anime Onlin...
Prime Video
Google Keep
lab
session
Cloud
YouTube
Amazon Web Servic...
All Bookmarks
aws
Services
Search
[Alt+S]
Mumbai
omkar-architect @ dnyaneshwar-admin
Support
EC2
EC2 Dashboard
EC2 Global View
Events
Instances
Instances
Instance Types
Launch Templates
Spot Requests
Savings Plans
Reserved Instances
Dedicated Hosts
Capacity Reservations
New
Images
AMIs
AMI Catalog
Elastic Block Store
Volumes
Instances (23) Info
Connect
Instance state
Actions
Launch instances
Find Instance by attribute or tag (case-sensitive)
1
Name
Instance ID
Instance state
Status check
Alarm status
Public IPv4 DNS
Public IPv4 ..
Worker_Host1 i-0a409918550a51857 Running 2/2 checks passec No alarms ec2-13-234-7-150.ap-s... 13.234.7.150
Worker_Host2 i-04accc5b73443fc02 Running 2/2 checks passec No alarms ec2-65-0-251-57.ap-so... 65.0.251.57
Worker_Host3 i-0a5bbc5de3efece71 Running 2/2 checks passec No alarms ec2-43-205-131-248.ap... 43.205.131.2
Worker_Host4 i-0a0b8bcd5e77a4490 Running 2/2 checks passec No alarms ec2-65-0-65-71.ap-sout... 65.0.65.71
Worker_Host5 i-0cf4a94c728e2149b Running 2/2 checks passec No alarms ec2-13-200-210-252.ap... 13.200.210.2
Worker_Host6 i-01878b4ff10c1104e Running 2/2 checks passec No alarms ec2-43-205-122-197.ap... 43.205.122.1
Worker_Host7 i-0ba25b1a65e7ff2f0 Running 2/2 checks passec No alarms ec2-3-109-150-126.ap-... 3.109.150.12
Worker_Host8 i-05ca06c719d6a9670 Running 2/2 checks passec No alarms ec2-3-6-210-219.ap-so... 3.6.210.219
Worker_Host9 i-0bfa68293fb45e346 Running 2/2 checks passec No alarms ec2-13-200-224-227.ap... 13.200.224.2
Kafka_Broker1 i-08e505208b2379b94 Running 2/2 checks passec No alarms ec2-13-200-207-121.ap... 13.200.207.1
Kafka_Broker2 i-0871ce63da7aa8d26 Running 2/2 checks passec No alarms ec2-3-110-98-124.ap-s... 3.110.98.124
Kafka_Broker3 i-0e204de1d4ef2a475 Running 2/2 checks passec No alarms ec2-13-200-40-137.ap-... 13.200.40.13
Select an instance
CloudShell
Feedback
© 2023, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Search
ENG
US
08:58
28-12-2023

Dashboard | EC2 | ap-south | Instances | EC2 | ap-south-1 | Status - Home - Cloud | Configuration - Ranger - Cl | Ranger - Sign In | Ranger - Sign In

Not secure | https://3.6.150.244:7183/cmf/home

Spotify - Web Playe... | Prime Music | Watch Anime Onlin... | Prime Video | Google Keep | lab | session | Cloud | YouTube | Amazon Web Servic... | All Bookmarks

CLOUDERA Manager

Search

- Clusters
- Hosts
- Diagnostics
- Audits
- Charts
- Replication
- Administration
- Experiences New

- Parcels
- Running Commands
- Support
- admin

7.4.4

Omkar

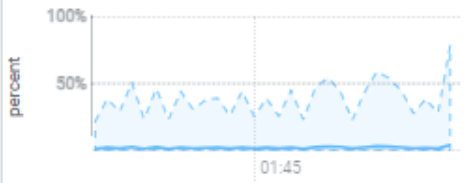
Cloudera Runtime 7.1.7 (Parcels)

20 Hosts	11
CDP-INFRA-SOLR	
HBase	1
HDFS	
Hive	
Hive on Tez	
Hue	1
Kafka	
Oozie	
Ranger	
Ranger KMS	1
Spark	1
Tez	
YARN	
YARN Queue Manager	
ZooKeeper	

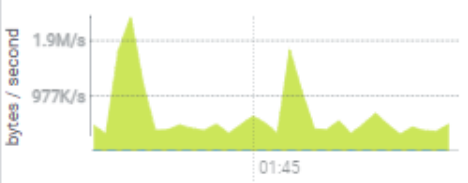
Charts

30m 1h 2h 6h 12h 1d 7d 30d

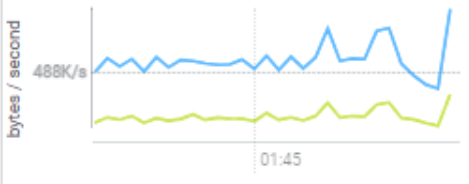
Cluster CPU



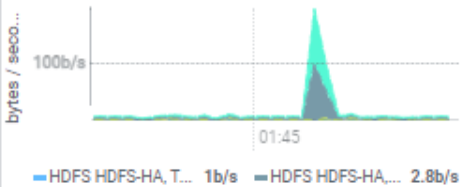
Cluster Disk IO



Cluster Network IO



HDFS IO



Search

ENG US

07:31 28-12-2023

Roles

This table is grouped by hosts having the same roles assigned to them.

Hosts	Count	Roles
ip-172-31-17-226.ap-south-1.compute.internal; ip-172-31-18-[85, 181].ap-south-1.compute.internal; ip-172-31-31-245.ap-south-1.compute.internal	4	RS DN NM
ip-172-31-17-74.ap-south-1.compute.internal	1	FC JN NN YA... YA... HS JHS RM S
ip-172-31-18-12.ap-south-1.compute.internal	1	M FC JN NN S
ip-172-31-18-81.ap-south-1.compute.internal	1	HMS OS RA RU RK... SS
ip-172-31-20-184.ap-south-1.compute.internal; ip-172-31-25-213.ap-south-1.compute.internal; ip-172-31-28-117.ap-south-1.compute.internal	3	KB
ip-172-31-20-32.ap-south-1.compute.internal; ip-172-31-21-238.ap-south-1.compute.internal; ip-172-31-21-9.ap-south-1.compute.internal; ip-172-31-[25, 27]-166.ap-south-1.compute.internal	5	RS DN SS NM
ip-172-31-21-204.ap-south-1.compute.internal	1	G HS2 HS KTR RA G
ip-172-31-22-6.ap-south-1.compute.internal	1	AM AP ES HM RM SM
ip-172-31-29-140.ap-south-1.compute.internal	1	M JN RM S
ip-172-31-30-107.ap-south-1.compute.internal	1	HMS RA RT SS
ip-172-31-30-211.ap-south-1.compute.internal	1	G HS2 RA

Rows per page: 25 | 1 - 11 of 11

Dashboard | EC2 | ap-south-1 | Instances | EC2 | ap-south-1 | Instances - HDFS - Cloudera Ma | Status - Home - Cloudera Manager

Not secure | https://3.6.150.244:7183/cmf/services/120/instances

Spotify - Web Playe... | Prime Music | Watch Anime Onlin... | Prime Video | Google Keep | lab | session | Cloud | YouTube | Amazon Web Servic... | All Bookmarks

CLOUDERA Manager

Search

- Clusters
- Hosts
- Diagnostics
- Audits
- Charts
- Replication
- Administration
- Experiences New

- Parcels
- Running Commands
- Support
- admin

7.4.4

HDFS

Actions

Status | Instances | Configuration | Commands | File Browser | Charts Library | Cache Statistics | Audits | Web UI | Quick Links

Search | Filters | Last Updated: Dec 28, 10:16:26 AM UTC

Filters

- ROLE GROUP
- ROLE TYPE
- HEALTH TEST

Actions for Selected | Migrate Roles | Add Role Instances | Federation and High Availability | Role Groups

Status	Role Type	State	Hostname	Commission State	Role Group
	DataNode	Started	ip-172-31-21-238.ap-south-1.compute.internal	Commissioned	DataNode Default Group
	DataNode	Started	ip-172-31-21-9.ap-south-1.compute.internal	Commissioned	DataNode Default Group
	DataNode	Started	ip-172-31-17-226.ap-south-1.compute.internal	Commissioned	DataNode Default Group
	DataNode	Started	ip-172-31-27-166.ap-south-1.compute.internal	Commissioned	DataNode Default Group
	DataNode	Started	ip-172-31-20-32.ap-south-1.compute.internal	Commissioned	DataNode Default Group
	DataNode	Started	ip-172-31-31-245.ap-south-1.compute.internal	Commissioned	DataNode Default Group
	DataNode	Started	ip-172-31-18-181.ap-south-1.compute.internal	Commissioned	DataNode Default Group
	DataNode	Started	ip-172-31-25-166.ap-south-1.compute.internal	Commissioned	DataNode Default Group
	DataNode	Started	ip-172-31-18-85.ap-south-1.compute.internal	Commissioned	DataNode Default Group
	Failover Controller	Started	ip-172-31-18-12.ap-south-1.compute.internal	Commissioned	Failover Controller Default Group
	Failover Controller	Started	ip-172-31-17-74.ap-south-1.compute.internal	Commissioned	Failover Controller Default Group
	JournalNode	Started	ip-172-31-29-140.ap-south-1.compute.internal	Commissioned	JournalNode Default Group
	JournalNode	Started	ip-172-31-18-12.ap-south-1.compute.internal	Commissioned	JournalNode Default Group
	JournalNode	Started	ip-172-31-17-74.ap-south-1.compute.internal	Commissioned	JournalNode Default Group
	NameNode (Active)	Started	ip-172-31-18-12.ap-south-1.compute.internal	Commissioned	NameNode Default Group
	NameNode (Standby)	Started	ip-172-31-17-74.ap-south-1.compute.internal	Commissioned	NameNode Default Group

Search | ENG US | 03:46 28-12-2023

Dashboard | EC2 | ap-south-1 | Instances | EC2 | ap-south-1 | Instances - YARN - Cloudera Ma | Status - Home - Cloudera Manager

Not secure | https://3.6.150.244:7183/cmf/services/123/instances

Spotify - Web Playe... | Prime Music | Watch Anime Onlin... | Prime Video | Google Keep | lab | session | Cloud | YouTube | Amazon Web Servic... | All Bookmarks

CLOUDERA Manager

Search

- Clusters
- Hosts
- Diagnostics
- Audits
- Charts
- Replication
- Administration
- Experiences New

- Parcels
- Running Commands
- Support
- admin

7.4.4

YARN

Actions

Status | Instances | Configuration | Commands | Applications | Charts Library | Audits | Web UI | Quick Links

Search | Filters | Last Updated: Dec 28, 10:15:52 AM UTC

Filters

- ROLE GROUP
- ROLE TYPE
- HEALTH TEST

Actions for Selected | Add Role Instances | Role Groups

Status	Role Type	State	Hostname	Commission State	Role Group
	JobHistory Server	Started	ip-172-31-17-74.ap-south-1.compute.internal	Commissioned	JobHistory Server Default Group
	NodeManager	Started	ip-172-31-21-238.ap-south-1.compute.internal	Commissioned	NodeManager Default Group
	NodeManager	Started	ip-172-31-21-9.ap-south-1.compute.internal	Commissioned	NodeManager Default Group
	NodeManager	Started	ip-172-31-17-226.ap-south-1.compute.internal	Commissioned	NodeManager Default Group
	NodeManager	Started	ip-172-31-27-166.ap-south-1.compute.internal	Commissioned	NodeManager Default Group
	NodeManager	Started	ip-172-31-20-32.ap-south-1.compute.internal	Commissioned	NodeManager Default Group
	NodeManager	Started	ip-172-31-31-245.ap-south-1.compute.internal	Commissioned	NodeManager Default Group
	NodeManager	Started	ip-172-31-18-181.ap-south-1.compute.internal	Commissioned	NodeManager Default Group
	NodeManager	Started	ip-172-31-25-166.ap-south-1.compute.internal	Commissioned	NodeManager Default Group
	NodeManager	Started	ip-172-31-18-85.ap-south-1.compute.internal	Commissioned	NodeManager Default Group
	ResourceManager (Active)	Started	ip-172-31-29-140.ap-south-1.compute.internal	Commissioned	ResourceManager Default Group
	ResourceManager (Standby)	Started	ip-172-31-17-74.ap-south-1.compute.internal	Commissioned	ResourceManager Default Group

Rows per page: 25 | 1 - 12 of 12

Search | ENG US | 03:46 28-12-2023

CLOUDERA Manager

Search

Clusters
Hosts
Diagnostics
Audits
Charts
Replication
Administration
Experiences New
Parcels
Running Commands
Support
admin

7.4.4

Security

Status | Kerberos Credentials

Auto-TLS is Enabled. Rotate Auto-TLS Certificates | TLS Settings | Security Inspector

This page allows user to enable TLS (Transport Layer Security), Kerberos and HDFS Data At Rest Encryption. After Kerberos is enabled for a cluster, you can also view all the Kerberos related configuration set by the wizard.

Search by cluster name

Last Updated: Dec 28, 2:48:43 PM UTC

Cluster	TLS	Kerberos	HDFS Data At Rest Encryption
Omkar	Enabled Disable TLS	Enabled	Enabled Set up HDFS Data At Rest Encryption

1 - 1 of 1

THANK YOU

Omkar Gaikwad
omkargaikwad2196@gmail.com