

Contents

1. About Project
2. Challenges
3. Solution
4. Result
5. Capacity Planning
6. Cluster Planning
7. Project Flow Diagram
8. Cluster Screenshots

About Project

Telefónica Spain is the leading telecommunications company in Spain, offering voice, data, television, and Internet access to consumers and businesses.

Telefónica Spain achieves a 20% increase in customer usage when it deploys a modern data platform from Cloudera.

Challenges

- To handle more than ten billion records a day, the cost was impractical with techniques other than a big data platform
- To gain more customer insight, the organization needed the ability to analyze a wider volume of data, and do so in real time.
- Virtually, all data is received in real time, or almost real time, and volumes are very high because we have almost 40 million connections

Solution

- * Modern Data Platform: Cloudera Enterprise
- * Workloads: Operational Database
- * Key Components: Apache Hive, Apache Flume, Apache HBase, Apache Kafka, Apache Spark, Apache Sqoop, Cloudera Manager
- * Databases: MongoDB.

Result

- 20 percent increase in customer usage, Real-time customer insights and the delivery of more customized experiences
- 39 percent decrease in in-store stock of mobile devices, and a 17 percent increase in store sales
- Double-digit decrease in customer churn rate

Capacity Planning

Total data(Yearly)

Generated data per day = 200 GB

Yearly Data = $200 \times 365 = 73000 \text{ GB} = 73 \text{ TB}$

Total data (RF) = $73000 \text{ GB} \times 3 = 219000 \text{ GB} = 219 \text{ TB}$

10% Overhead space needed = $219 + 21.9(10\% \text{ overhead})$
= 240.9 TB

Data at Rest = $15.44 \text{ TB} \times 3 = 46.32 \text{ TB}$

Data including RF = $240.9 \text{ TB} + 46.32 \text{ TB} = 287.22 \text{ TB}$

30 % NON-DFS = $287.22 \text{ TB} + 86.766 \text{ TB}$
= 373.386 TB = 374 TB

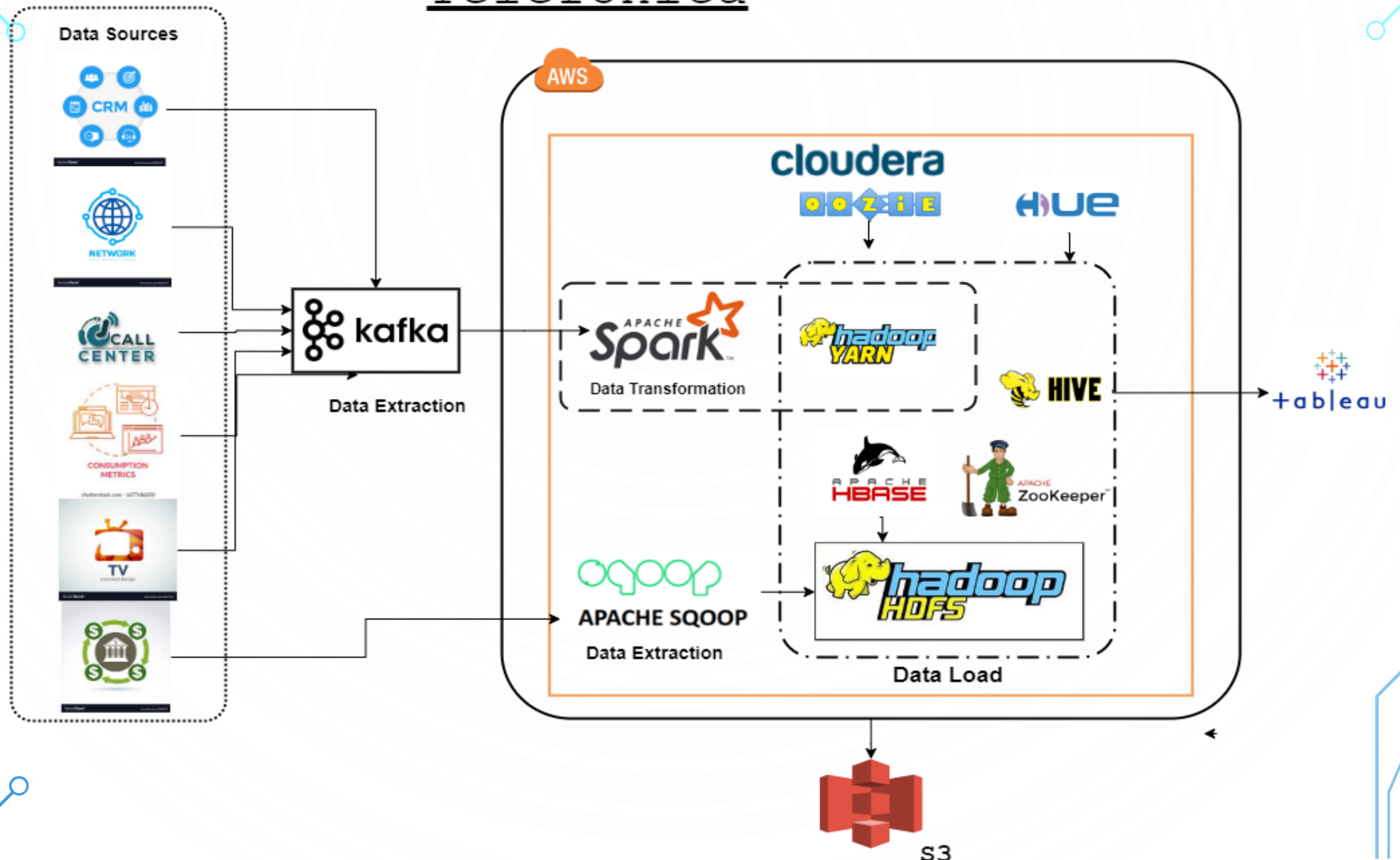
Cluster Planning

Spacification	Nodes	CPU GHz/node	Memory in (RAM)GB	Storage GB/TB	Instance Type
Slave Nodes	36	8	61	8 TB	D2.2xlarge
Master Nodes	03	36	60	8 TB	C4.8xlarge
Utility Nodes	02	08	61	8 TB	R3.2xlarge (mem. Opt)
Edge Nodes	02	16	64	8 TB	m4.4xlarge
Database	01	08	61	8 TB	r3.2xlarge
Active Directory	01	08	61	8 TB	r3.2xlarge
Kafka Broker	03	16	64	8 TB	m4.4xlarge

Versions

CDH	6.3.3
CM	5.16.2
CentOS	7
JDK (Oracle)	1.8.0
HADOOP (hdfs,yarn)	3.0.0
SPARK	2.4.0
HIVE	2.1.1
HUE	4.4.0
OOZIE	5.1.0
ZOOKEEPER	3.4.5
KAFKA	2.2.1
MYSQL	5.7.34
SENTRY	2.1.0
HBASE	2.1.4
SQOOP	1.4.7
FLUME	1.9.0

Telefonica



Status - Home - Cloudera Manag

ec2-18-220-6-150.us-east-2.co...

← → ↻ ⚠ Not secure | 10.0.1.210:7180/cmf/home

Cloudera Manager

Clusters ▾

Hosts ▾

Diagnostics ▾

Audits

Charts ▾

Backup ▾

Administration ▾

Search

📦 📄

Support ▾

admin ▾

Home

Status

All Health Issues

Configuration ▾

All Recent Commands

Switch to Table View

+ Add ▾

🟢 Telefonica

CDH 6.2.0 (Parcels)

🟢 19 Hosts

🟢 HBase

🟢 HDFS

🟢 Hive

🟢 Hue

🟢 Kafka

🟢 Oozie

🟢 Sentry

🟢 Spark

🟡 Sqoop 1 Client

🟢 YARN (MR2 In...

🟢 ZooKeeper

Charts

30m 1h 2h 6h 12h 1d 7d 30d 🖋

HDFS IO

bytes / second

50b/s

11 AM

Total Bytes Read ... 1b/s

Total Bytes Read ... 1b/s

Total Bytes Writ... 2.8b/s

Total Bytes Writ... 2.8b/s

Cluster Network IO

bytes / second

97.7K/s

48.8K/s

11 AM

Total Bytes Re... 50.6K/s

Total Bytes Tr... 82.2K/s

Cluster Disk IO

bytes / second

1.4M/s

977K/s

488K/s

11 AM

Total Disk Byte... 956b/s

Total Disk Byte... 470K/s

Cluster CPU

percent

100%

50%

11 AM

Telefonica, Host CPU Usage Across Hosts 2.1%

Cloudera Management Service

Activate Windows

Go to Settings to activate Windows.

🏠 🔍 📁 📂 🖨 📄 📧 📧

11:16 AM 4/29/2022

Roles

Roles - Cloudera Manager

ec2-18-220-6-150.us-east-2.co...

← → ↻ ⚠ Not secure | 10.0.1.210:7180/cm/hardware/roles#clusterId=1

Cloudera Manager Clusters ▾ Hosts ▾ Diagnostics ▾ Audits Charts ▾ Backup ▾ Administration ▾

Search 📦 🖨 Support ▾ admin ▾

Roles

Hosts	Count	Roles
ip-10-0-1-210.us-east-2.compute.internal	1	HM SM
ip-10-0-1-213.us-east-2.compute.internal	1	HB... HB... M FC JN NN G SS G RM S
ip-10-0-1-240.us-east-2.compute.internal	1	HB... HB... M B FC JN NN G OS G S
ip-10-0-1-33.us-east-2.compute.internal	1	AP ES RM
ip-10-0-1-8.us-east-2.compute.internal	1	HB... HB... M JN G HMS SS HS JHS RM S
ip-10-0-1-[12, 203].us-east-2.compute.internal	2	G HFS G HS2 LB HS KTR G G G G
ip-10-0-2-[96, 144, 217].us-east-2.compute.internal; ip-10-0-3-[8, 100, 129, 159, 237, 243].us-east-2.compute.internal	9	RS DN G G NM
ip-10-0-3-[29, 216-217].us-east-2.compute.internal	3	KB

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

Activate Windows
Go to Settings to activate Windows.

Feedback

11:26 AM
4/29/2022

NameNode HA

Instances - HDFS - Cloudera Manager

Not secure | 10.0.1.210:7180/cm/services/15/instances#

HEALTH TESTS

<input checked="" type="checkbox"/>	DataNode	Started	ip-10-0-3-139.us-east-2.compute.internal	Commissioned	DataNode Default Group
<input type="checkbox"/>	<input checked="" type="checkbox"/> DataNode	Started	ip-10-0-2-144.us-east-2.compute.internal	Commissioned	DataNode Default Group
<input type="checkbox"/>	<input checked="" type="checkbox"/> DataNode	Started	ip-10-0-3-8.us-east-2.compute.internal	Commissioned	DataNode Default Group
<input type="checkbox"/>	<input checked="" type="checkbox"/> DataNode	Started	ip-10-0-2-96.us-east-2.compute.internal	Commissioned	DataNode Default Group
<input type="checkbox"/>	<input checked="" type="checkbox"/> Failover Controller	Started	ip-10-0-1-240.us-east-2.compute.internal	Commissioned	Failover Controller Default Group
<input type="checkbox"/>	<input checked="" type="checkbox"/> Failover Controller	Started	ip-10-0-1-213.us-east-2.compute.internal	Commissioned	Failover Controller Default Group
<input type="checkbox"/>	<input type="checkbox"/> Gateway	N/A	ip-10-0-1-203.us-east-2.compute.internal	Commissioned	Gateway Default Group
<input type="checkbox"/>	<input type="checkbox"/> Gateway	N/A	ip-10-0-1-12.us-east-2.compute.internal	Commissioned	Gateway Default Group
<input type="checkbox"/>	<input checked="" type="checkbox"/> HttpFS	Started	ip-10-0-1-203.us-east-2.compute.internal	Commissioned	HttpFS Default Group
<input type="checkbox"/>	<input checked="" type="checkbox"/> HttpFS	Started	ip-10-0-1-12.us-east-2.compute.internal	Commissioned	HttpFS Default Group
<input type="checkbox"/>	<input checked="" type="checkbox"/> JournalNode	Started	ip-10-0-1-240.us-east-2.compute.internal	Commissioned	JournalNode Default Group
<input type="checkbox"/>	<input checked="" type="checkbox"/> JournalNode	Started	ip-10-0-1-213.us-east-2.compute.internal	Commissioned	JournalNode Default Group
<input type="checkbox"/>	<input checked="" type="checkbox"/> JournalNode	Started	ip-10-0-1-8.us-east-2.compute.internal	Commissioned	JournalNode Default Group
<input type="checkbox"/>	<input checked="" type="checkbox"/> NameNode (Active)	Started	ip-10-0-1-240.us-east-2.compute.internal	Commissioned	NameNode Default Group
<input type="checkbox"/>	<input checked="" type="checkbox"/> NameNode (Standby)	Started	ip-10-0-1-213.us-east-2.compute.internal	Commissioned	NameNode Default Group

Activate Windows
Go to Settings to activate Windows.

25 Per Page

Feedback

11:23 AM
4/29/2022

ResourceManager HA

Instances - YARN (MR2 Included) x + ec2-18-220-6-150.us-east-2.co... - x

← → ↻ ⚠ Not secure | 10.0.1.210:7180/cmfservices/16/instances#filterfreeText=g

Filters

▼ STATUS

● None 2

✓ Good Health 11

► COMMISSION STATE

► MAINTENANCE MODE

► RACK

► ROLE GROUP

► ROLE TYPE

► STATE

► HEALTH TESTS

<input type="checkbox"/>	Role Type	State	Host	Commission State	Role Group
<input type="checkbox"/>	Gateway	N/A	ip-10-0-1-203.us-east-2.compute.internal	Commissioned	Gateway Default Group
<input type="checkbox"/>	Gateway	N/A	ip-10-0-1-12.us-east-2.compute.internal	Commissioned	Gateway Default Group
<input type="checkbox"/>	NodeManager	Started	ip-10-0-3-237.us-east-2.compute.internal	Commissioned	NodeManager Default Group
<input type="checkbox"/>	NodeManager	Started	ip-10-0-2-217.us-east-2.compute.internal	Commissioned	NodeManager Default Group
<input type="checkbox"/>	NodeManager	Started	ip-10-0-3-100.us-east-2.compute.internal	Commissioned	NodeManager Default Group
<input type="checkbox"/>	NodeManager	Started	ip-10-0-3-129.us-east-2.compute.internal	Commissioned	NodeManager Default Group
<input type="checkbox"/>	NodeManager	Started	ip-10-0-3-243.us-east-2.compute.internal	Commissioned	NodeManager Default Group
<input type="checkbox"/>	NodeManager	Started	ip-10-0-3-159.us-east-2.compute.internal	Commissioned	NodeManager Default Group
<input type="checkbox"/>	NodeManager	Started	ip-10-0-2-144.us-east-2.compute.internal	Commissioned	NodeManager Default Group
<input type="checkbox"/>	NodeManager	Started	ip-10-0-3-8.us-east-2.compute.internal	Commissioned	NodeManager Default Group
<input type="checkbox"/>	NodeManager	Started	ip-10-0-2-96.us-east-2.compute.internal	Commissioned	NodeManager Default Group
<input type="checkbox"/>	ResourceManager (Active)	Started	ip-10-0-1-213.us-east-2.compute.internal	Commissioned	ResourceManager Default Group
<input type="checkbox"/>	ResourceManager (Standby)	Started	ip-10-0-1-8.us-east-2.compute.internal	Commissioned	ResourceManager Default Group

10.0.1.210:7180/cmfservices/16/instances/150/status

Activate Windows
Go to Settings to activate Windows

25 Per Page

11:24 AM
4/29/2022

Security

Kerberos Enabled

The screenshot shows the Cloudera Manager web interface. The browser address bar indicates the URL is `10.0.1.210:7180/cmf/cluster/kerberosStatus`. The navigation menu includes Clusters, Hosts, Diagnostics, Audits, Charts, Backup, and Administration. The 'Administration' tab is active, and the 'Security' section is selected, with 'Status' and 'Kerberos Credentials' sub-tabs. Under 'Status', 'TLS Settings' and 'Security Inspector' are visible. A table lists the cluster 'Telefonica' with the status 'Successfully enabled Kerberos.' and a note 'HDFS Data At Rest Encryption is disabled' with a link to 'Set up HDFS Data At Rest Encryption'. The Windows taskbar at the bottom shows the time as 11:25 AM on 4/29/2022.

Status - Security - Cloudera Manager

ec2-18-220-6-150.us-east-2.co...

Not secure | 10.0.1.210:7180/cmf/cluster/kerberosStatus

Cloudera Manager Clusters Hosts Diagnostics Audits Charts Backup Administration

Search Support admin

Security Status Kerberos Credentials

TLS Settings Security Inspector

Cluster		
Telefonica	Successfully enabled Kerberos.	HDFS Data At Rest Encryption is disabled Set up HDFS Data At Rest Encryption

Activate Windows
Go to Settings to activate Windows.

11:25 AM
4/29/2022

Kerberos Credentials - Security - x + ec2-18-220-6-150.us-east-2.co...

Not secure | 10.0.1.210:7180/cm/cluster/credentials

Cloudera Manager Clusters Hosts Diagnostics Audits Charts Backup Administration

Search Support admin

Security Status Kerberos Credentials

g

Generate Missing Credentials Security Inspector Configuration Import Kerberos Account Manager Credentials

Regenerate Selected

<input type="checkbox"/> Principal	Host
<input type="checkbox"/> HTTP/ip-10-0-1-12.us-east-2.compute.internal@HADOOPSECURITY.COM	ip-10-0-1-12.us-east-2.compute.internal
<input type="checkbox"/> HTTP/ip-10-0-1-203.us-east-2.compute.internal@HADOOPSECURITY.COM	ip-10-0-1-203.us-east-2.compute.internal
<input type="checkbox"/> HTTP/ip-10-0-1-213.us-east-2.compute.internal@HADOOPSECURITY.COM	ip-10-0-1-213.us-east-2.compute.internal
<input type="checkbox"/> HTTP/ip-10-0-1-240.us-east-2.compute.internal@HADOOPSECURITY.COM	ip-10-0-1-240.us-east-2.compute.internal
<input type="checkbox"/> HTTP/ip-10-0-1-8.us-east-2.compute.internal@HADOOPSECURITY.COM	ip-10-0-1-8.us-east-2.compute.internal
<input type="checkbox"/> HTTP/ip-10-0-2-144.us-east-2.compute.internal@HADOOPSECURITY.COM	ip-10-0-2-144.us-east-2.compute.internal
<input type="checkbox"/> HTTP/ip-10-0-2-217.us-east-2.compute.internal@HADOOPSECURITY.COM	ip-10-0-2-217.us-east-2.compute.internal
<input type="checkbox"/> HTTP/ip-10-0-2-96.us-east-2.compute.internal@HADOOPSECURITY.COM	ip-10-0-2-96.us-east-2.compute.internal
<input type="checkbox"/> HTTP/ip-10-0-3-100.us-east-2.compute.internal@HADOOPSECURITY.COM	ip-10-0-3-100.us-east-2.compute.internal

Feedback

Activate Windows
Go to Settings to activate Windows.

11:25 AM
4/29/2022

Active Directory

The screenshot displays the AWS Management Console interface for a Server Manager Dashboard. The browser address bar shows the URL `ec2-13-235-70-33.ap-south-1.compute.amazonaws.com`. The dashboard title is "Server Manager ▸ Dashboard".

Left Navigation Panel:

- Dashboard (selected)
- Local Server
- All Servers
- AD CS
- AD DS
- DNS
- File and Storage Services ▸

Dashboard Content:

ROLES AND SERVER GROUPS
Roles: 4 | Server groups: 1 | Servers total: 1

Roles and Server Groups Summary:

Role/Server Group	Count	Manageability	Events	Services	Performance	BPA results
AD CS	1	✓				
AD DS	1	✓				
DNS	1	✓				
File and Storage Services	1	✓				
Local Server	1	✓				
All Servers	1	✓				

Taskbar: The Windows taskbar at the bottom shows the Start button, task view, and several application icons. The system clock indicates 6:56 AM on 3/7/2022.

Active Directory Users and Computers

File Action View Help

Active Directory Users and Computers

hadoopsecurity.local

Builtin

Computers

Domain Controllers

ForeignSecurityPrincipals

hadoop

Managed Service Accounts

sap

Users

Name	Type	Description
cloudera ma...	User	
DcNpRwdbjp	User	
DIEUKNLbl	User	
EkBsWiQcHs	User	
eWavQEhEVW	User	
fmzggeMFqV	User	
gyfsXXiGGY	User	
ICAinrlfsz	User	
mMJxnkVbur	User	
nBkDpdJau	User	
NCewhPCHjV	User	
nnhkbOXkAh	User	
oBgYnnaNRG	User	
OufHuTgCdT	User	
QlIKbpYtah	User	
QljBjLWDuQ	User	
SrNtgCaLcg	User	
ufoGuXikHS	User	
ZyiKyakpUi	User	

TASKS

TASKS

HADOOP-AD	4013	Warning	Microsoft-Windows-DNS-Server-Service	DNS Server	3/6/2022 10:19:09 AM
HADOOP-AD	4013	Warning	Microsoft-Windows-DNS-Server-Service	DNS Server	3/6/2022 10:11:18 AM
HADOOP-AD	414	Warning	Microsoft-Windows-DNS-Server-Service	DNS Server	3/6/2022 9:58:38 AM

AWS Instance

Instances | EC2 Management Console

us-east-2.console.aws.amazon.com/ec2/v2/home?region=us-east-2#Instances:sort=tag:Name

Search for services, features, blogs, docs, and more [Alt+S]

Instances (23) Info

Search

Connect Instance state Actions Launch instances

<input type="checkbox"/>	Name	Instance ID	Instance state	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...	Elas
<input type="checkbox"/>	Active Directory	i-009e1b43b8a602179	Running	2/2 checks passed	No alarms	us-east-2a	-	-	-
<input type="checkbox"/>	Cloudera Manager	i-0e18e595af7d8ffa7	Running	2/2 checks passed	No alarms	us-east-2a	-	-	-
<input type="checkbox"/>	Cloudera-Manager2	i-0dad4f2c2a05bbe06	Running	2/2 checks passed	No alarms	us-east-2a	-	-	-
<input type="checkbox"/>	Gateway-1	i-01e3efcc2740728c5	Running	2/2 checks passed	No alarms	us-east-2a	-	-	-
<input type="checkbox"/>	Gateway-2	i-01685cb41e2f83d9d	Running	2/2 checks passed	No alarms	us-east-2a	-	-	-
<input type="checkbox"/>	kafka	i-0fbdf626ea1b8320	Running	2/2 checks passed	No alarms	us-east-2c	-	-	-
<input type="checkbox"/>	kafka2	i-047e51df032cb8be9	Running	2/2 checks passed	No alarms	us-east-2c	-	-	-
<input type="checkbox"/>	kafka3	i-084df089d3084b70c	Running	2/2 checks passed	No alarms	us-east-2c	-	-	-
<input type="checkbox"/>	Master-1	i-03d9c9b3d04ee924e	Running	2/2 checks passed	No alarms	us-east-2a	-	-	-
<input type="checkbox"/>	Master-2	i-01e6d0b85c1b4148a	Running	2/2 checks passed	No alarms	us-east-2a	-	-	-
<input type="checkbox"/>	Master-3	i-0a1a2a921d0f9e5f7	Running	2/2 checks passed	No alarms	us-east-2a	-	-	-
<input type="checkbox"/>	MIT	i-0b50a35d069494495	Running	2/2 checks passed	No alarms	us-east-2a	-	-	-
<input type="checkbox"/>	Webserver	i-053c6e0fdaddca69f	Running	2/2 checks passed	No alarms	us-east-2a	-	-	-
<input type="checkbox"/>	Worker-Node1	i-014056d33d85cdddb	Running	2/2 checks passed	No alarms	us-east-2b	-	-	-
<input type="checkbox"/>	Worker-Node2	i-0a6c706c42870b580	Running	2/2 checks passed	No alarms	us-east-2b	-	-	-

Select an instance

Activate Windows
Go to Settings to activate Windows.

Feedback Looking for language selection? Find it in the new Unified Settings

© 2022, Amazon Internet Services Private Ltd. or its affiliates. Privacy Terms Cookie preferences

Type here to search

4:43 PM
4/29/2022

Instances | EC2 Management Console

us-east-2.console.aws.amazon.com/ec2/v2/home?region=us-east-2#Instances:sort=tag:Name

Services Search for services, features, blogs, docs, and more [Alt+S]

Ohio babu

Instances (23) Info

Refresh Connect Instance state Actions Launch instances

Search

	Name	Instance ID	Instance state	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...	Ela:
<input type="checkbox"/>	Master-2	i-01e6d0b85c1b4148a	Running	2/2 checks passed	No alarms	us-east-2a	-	-	-
<input type="checkbox"/>	Master-3	i-0a1a2a921d0f9e5f7	Running	2/2 checks passed	No alarms	us-east-2a	-	-	-
<input type="checkbox"/>	MIT	i-0b50a35d069494495	Running	2/2 checks passed	No alarms	us-east-2a	-	-	-
<input type="checkbox"/>	Webserver	i-053c6e0fd9dca69f	Running	2/2 checks passed	No alarms	us-east-2a	-	-	-
<input type="checkbox"/>	Worker-Node1	i-014056d33d85cdddb	Running	2/2 checks passed	No alarms	us-east-2b	-	-	-
<input type="checkbox"/>	Worker-Node2	i-0ae670f642830b689	Running	2/2 checks passed	No alarms	us-east-2b	-	-	-
<input type="checkbox"/>	Worker-Node3	i-0b97c40fbc3f441bc	Running	2/2 checks passed	No alarms	us-east-2b	-	-	-
<input type="checkbox"/>	Worker-Node4	i-021b5bcc6d0eeea8e	Running	2/2 checks passed	No alarms	us-east-2c	-	-	-
<input type="checkbox"/>	Worker-Node5	i-0d6d58995e4bad33a	Running	2/2 checks passed	No alarms	us-east-2c	-	-	-
<input type="checkbox"/>	Worker-Node6	i-0228a0981cb62900d	Running	2/2 checks passed	No alarms	us-east-2c	-	-	-
<input type="checkbox"/>	Worker-Node7	i-0b2512ff5135198db	Running	2/2 checks passed	No alarms	us-east-2c	-	-	-
<input type="checkbox"/>	Worker-Node8	i-0b99d542e6b7f85bc	Running	2/2 checks passed	No alarms	us-east-2c	-	-	-
<input type="checkbox"/>	Worker-Node9	i-096ef43fdb38e2bfe	Running	2/2 checks passed	No alarms	us-east-2c	-	-	-
<input type="checkbox"/>	WorkSpace	i-055cdf4b4a344ce04	Running	2/2 checks passed	No alarms	us-east-2a	ec2-18-220-6-150.us-e...	18.220.6.150	-

Select an instance

Activate Windows
Go to Settings to activate Windows.