

Monitoreo de un clúster AWS EMR con Prometheus y Grafana

Volumen raíz de EBS

El volumen raíz de EBS se aplica a los sistemas operativos y las aplicaciones que instale en el clúster. [Restricciones de relación de volumen raíz de EBS](#)

Tamaño (GiB)

15

15- 100 GiB por volumen SSD de uso general (gp3)

IOPS

3000

3000-16000 IOPS por volumen. Elija una relación máxima de 500:1 entre IOPS y el tamaño del volumen.

Rendimiento (MiB/s)

125

125-1000 MiB/s por volumen. Elija una relación máxima de 0.25:1 entre el rendimiento y las IOPS.

▼

Aprovisionamiento y escalado de clústeres - obligatorio

Información

Elija cómo Amazon EMR debe dimensionar su clúster.

Elija una opción

☒

Establecer el tamaño del clúster manualmente

Utilice esta opción si conoce los patrones de la carga de trabajo de antemano.

☐

Utilizar escalado administrado por EMR

Supervise las métricas clave de la carga de trabajo de modo que EMR pueda optimizar el tamaño del clúster y la utilización de los recursos.

☐

Utilizar el escalamiento automático personalizado

Para escalar mediante programación los nodos principales y los nodos de tarea, cree políticas de escalamiento automático personalizadas.

Configuración de aprovisionamiento

Establezca el tamaño del principal grupo de instancias. Amazon EMR intenta aprovisionar esta capacidad al lanzar el clúster.

Nombre	Tipo de instancia	Tamaño de instancia(s)	Utilizar la opción de compra de spot
Central	m4.large	2	<input type="checkbox"/>

▼

Redes - obligatorio

Información

Elija la configuración de red que determina la forma en que usted y otras entidades se comunican con su clúster.

Virtual Private Cloud (VPC)

Información

vpc-0b82f4c0f54bb2915

Examinar

Crear VPC

Resumen

Información

emr-7.8.0

Paquete de aplicaciones

Custom (Hadoop 3.4.1, Hive 3.1.3, Spark 3.5.4)

Configuración del clúster - obligatorio

Grupos de instancias uniformes

Principal (m4.large), Central (m4.large)

Aprovisionamiento y escalado de clústeres - obligatorio

Configuración de aprovisionamiento

Tamaño del núcleo: 2 instancias

Redes - obligatorio

VPC

vpc-0b82f4c0f...

Subred

subnet-06b71e...

Grupos de seguridad de redes principales

❗

Configurar roles de IAM

Debe elegir un rol de servicio y un perfil de instancia antes de crear este clúster.

Elegir roles de IAM

Cancelar

Crear clúster

►

Grupos de seguridad de EC2 (firewall)

►

Pasos (0)

Información

Eliminar

Editar

Agregar

Utilice comandos y scripts para indicar a su clúster dónde encontrar y cómo procesar los datos. Los pasos se ejecutan de forma consecutiva a menos que habilite la opción Simultaneidad.

▼

Terminación del clúster y reemplazo de nodos

Información

Elija la configuración de terminación y proteja su clúster contra un apagado accidental.

Opción de terminación

☐

Terminar manualmente el clúster

☐

Terminar automáticamente el clúster después de que finalice el último paso

☒

Terminar el clúster después del tiempo de inactividad (recomendado)

Tiempo de inactividad

Ingrese el tiempo hasta que el clúster termine.

0 días

04:00:00

Elija una hora mayor a 1 minuto (00:01:00) y menor a 7 días. La hora está en formato hh:mm:ss (24 horas).

☐

Use la protección contra la terminación

Protege al clúster para evitar una terminación accidental. Si está activada, deberá primero desactivar la protección para terminar el clúster. Recomendamos activar la protección frente a terminaciones para los clústeres de larga duración.

Reemplazo de nodos en mal estado - novedad

Información

☒

Activar

Amazon EMR detiene correctamente los procesos en los nodos en mal estado para minimizar la pérdida de datos y las interrupciones del trabajo. Reemplaza rápidamente los nodos en mal estado por nuevas instancias de EC2 para que sus trabajos funcionen sin problemas.

☐

Desactivar

Amazon EMR agrega los nodos en mal estado a una lista de denegación mientras los mantiene en el clúster, lo que le permite tener acceso continuo para solucionar problemas.

►

Acciones de arranque (0)

Información

Eliminar

Editar

Agregar

Utilice las acciones de arranque para instalar software o personalizar la

Resumen

Información

emr-7.8.0

Paquete de aplicaciones

Custom (Hadoop 3.4.1, Hive 3.1.3, Spark 3.5.4)

Configuración del clúster - obligatorio

Grupos de instancias uniformes

Principal (m4.large), Central (m4.large)

Aprovisionamiento y escalado de clústeres - obligatorio

Configuración de aprovisionamiento

Tamaño del núcleo: 2 instancias

Redes - obligatorio

VPC

vpc-0b82f4c0f...

Subred

subnet-06b71e...

Grupos de seguridad de redes principales

❗

Configurar roles de IAM

Debe elegir un rol de servicio y un perfil de instancia antes de crear este clúster.

Elegir roles de IAM

Cancelar

Crear clúster

Elegir una configuración de seguridad o crear una nueva que pueda reutilizarse con otros clústeres.

Configuración de seguridad

Seleccione la configuración del servicio de cifrado, autenticación, autorización y metadatos de instancia del clúster.

Elegir una configuración de seguridad

Examinar

Crear configuración de seguridad

Par de claves de Amazon EC2 para el protocolo SSH al clúster

Información

vockey

X

Examinar

Crear par de claves

▼ Roles de Identity and Access Management (IAM) - obligatorio

Información

Elija o cree un rol de servicio y un perfil de instancia para las instancias de EC2 del clúster.

Rol de servicio de Amazon EMR

Información

El rol de servicio es un rol de IAM que Amazon EMR asume para aprovisionar recursos y realizar acciones de nivel de servicio con otros servicios de AWS.

☒ Elegir un rol de servicio existente

Seleccione un rol de servicio predeterminado o un rol personalizado con políticas de IAM asociadas para que el clúster pueda interactuar con otros servicios de AWS.

☐ Crear un rol de servicio

Deje que Amazon EMR cree un nuevo rol de servicio para que pueda conceder y restringir el acceso a los recursos de otros servicios de AWS.

Rol de servicio

LabRole

⌵

🔑

Perfil de instancia de EC2 para Amazon EMR

El perfil de instancia asigna un rol a cada instancia de EC2 de un clúster. El perfil de instancia debe especificar un rol que pueda acceder a los recursos de los pasos y las acciones de arranque.

☒ Elegir un perfil de instancia existente

Seleccione un rol predeterminado o un perfil de instancia personalizado con políticas de IAM asociadas para que el clúster pueda interactuar con sus recursos de Amazon S3.

☐ Crear un perfil de instancia

Deje que Amazon EMR cree un nuevo perfil de instancia para que pueda especificar un conjunto personalizado de recursos a los que tendrá acceso en Amazon S3.

Perfil de instancia

EMR_EC2_DefaultRole

⌵

🔑

Resumen

Información

emr-7.8.0

Paquete de aplicaciones
Custom (Hadoop 3.4.1, Hive 3.1.3, Spark 3.5.4)

Configuración del clúster - obligatorio

Grupos de instancias uniformes
Principal (m4.large), Central (m4.large)

Aprovisionamiento y escalado de clústeres - obligatorio

Configuración de aprovisionamiento
Tamaño del núcleo: 2 instancias

Redes - obligatorio

VPC
vpc-0b82f4c0f... 🔗

Subred
subnet-06b71e... 🔗

Enlaces de seguridad de redes principales

Cancelar

Crear clúster

Amazon EMR > EMR en EC2: Clústeres > EMR: Monitoring-labd14

EMR-Monitoring-labd14

Se ha actualizado hace menos de un minuto

Terminar

Clonar en AWS CLI

Clonar

▼ Resumen

Información del clúster

ID del clúster
j-XFO23OAT9J41

ARN del clúster
[arn:aws:elasticmapreduce:us-east-1:374864901798:user/j-XFO23OAT9J41](#)

Configuración del clúster
Grupos de instancias

Capacidad
1 Primary (Principal) | 2 Principal | 0 Tarea

Aplicaciones

Versión de Amazon EMR
emr-7.8.0

Aplicaciones instaladas
Hadoop 3.4.1, Hive 3.1.3, Spark 3.5.4

Administración de clústeres

Destino del registro en Amazon S3
[aws-logs-374864901798-us-east-1@elasticmapreduce](#)

IU de aplicación persistente
[Servidor de historial de Spark](#)
[Servidor de línea de tiempo de YARN](#)
[UI de Tez](#)

DNS público del nodo principal
[ec2-3-85-84-75.compute-1.amazonaws.com](#)
[Conectarse al nodo principal mediante SSH](#)
[Conectarse al nodo principal mediante SSM](#)

Estado y hora

Estado
Esperando

Hora de creación
11 de abril de 2025 18:24 (UTC+02:00)

Tiempo transcurrido
3 minutos, 19 segundos

Propiedades

Acciones de arranque

Instancias (hardware)

Pasos

Aplicaciones

Configuraciones

Monitorización

Eventos

Etiquetas (0)

Configuración del grupo de instancias

Información

Opción de escalado de clústeres
Establecer manualmente el tamaño del clúster

Editar opción de escalado de clústeres

Central
Nombre y cantidad máxima de nodos centrales en el clúster
Central | 2 instancias

Grupos de instancias (2)

Información

Terminar la instancia

Cambiar tamaño de grupo de instancias

Agregar grupo de instancias de tareas

2. Conectar al nodo maestro

[illegible]

Parte 2: Configuración de JMX Exporter

1. Instalar JMX Exporter

JMX es una tecnología que viene con Java y que sirve para ver qué está pasando dentro de una aplicación mientras está corriendo. Por ejemplo, puedes saber cuánta memoria está usando, cuántos hilos están activos, cómo va el garbage collector, o incluso métricas personalizadas que tú mismo puedes definir si las programas. El problema es que JMX no es muy amigable cuando quieres usar herramientas modernas de monitoreo como Prometheus.

Ahí entra el **JMX Exporter**, que es como un traductor entre JMX y Prometheus. Básicamente se conecta a las métricas de JMX y las convierte en un formato que Prometheus entiende. Lo puedes correr junto a tu aplicación Java como un agente, y lo que hace es abrir un endpoint HTTP donde se muestran todas las métricas convertidas. Prometheus se conecta a ese endpoint cada cierto tiempo y se trae los datos.

Luego con Prometheus recogiendo todas esas métricas, puedes usar **Grafana** para mostrarlas en dashboards súper visuales: gráficos de uso de memoria, actividad del GC, número de peticiones, errores, etc. Esto te ayuda un montón para saber si tu aplicación va bien o si hay algo raro pasando.

En resumen, el JMX Exporter es la pieza que conecta el mundo Java con el de Prometheus y Grafana. Te permite tener visibilidad completa de lo que pasa dentro de tus aplicaciones Java sin tener que escribir mucho código extra.

```
[hadoop@ip-172-31-30-29 ~]$ wget https://repo1.maven.org/maven2/io/prometheus/jmx/jmx_prometheus_javaagent/0.16.1/jmx_prometheus_javaagent-0.16.1.jar -P /home/hadoop/
--2025-04-14 08:02:36-- https://repo1.maven.org/maven2/io/prometheus/jmx/jmx_prometheus_javaagent/0.16.1/jmx_prometheus_javaagent-0.16.1.jar
Resolving repo1.maven.org (repo1.maven.org)... 146.75.32.209, 2a04:4e42:77::209
Connecting to repo1.maven.org (repo1.maven.org)|146.75.32.209|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 469645 (459K) [application/java-archive]
Saving to: '/home/hadoop/jmx_prometheus_javaagent-0.16.1.jar'

jmx_prometheus_javaagent-0.16 100%[=====] 458.64K --.-KB/s in 0.01s

2025-04-14 08:02:37 (34.3 MB/s) - '/home/hadoop/jmx_prometheus_javaagent-0.16.1.jar' saved [469645/469645]
```

2. Crear el archivo de configuración

```
[hadoop@ip-172-31-30-29 ~]$ nano /home/hadoop/config.yml
[hadoop@ip-172-31-30-29 ~]$ cat config.yml
lowercaseOutputName: true
lowercaseOutputLabelNames: true
rules:
  - pattern: ".*"
[hadoop@ip-172-31-30-29 ~]$
```

3. Configurar el NameNode para usar JMX Exporter

```
[hadoop@ip-172-31-30-29 ~]$ sudo nano /etc/hadoop/conf/hadoop-env.sh

GNU nano 5.8 /etc/hadoop/conf/hadoop-env.sh Modified
export JAVA_LIBRARY_PATH="$JAVA_LIBRARY_PATH:/usr/lib/hadoop-lzo/lib/native"

export HADOOP_CLASSPATH="$HADOOP_CLASSPATH:/usr/share/aws/aws-java-sdk/*"
export HADOOP_CLASSPATH="$HADOOP_CLASSPATH:/usr/share/aws/aws-java-sdk-v2/*"

export HADOOP_CLASSPATH="$HADOOP_CLASSPATH:/usr/share/aws/emr/emrfs/conf:/usr/share/aws/emr/emrfs/lib/*:/usr/share/aws/emr/emrfs/lib/jar/*"

export HADOOP_CLASSPATH="$HADOOP_CLASSPATH:/usr/share/aws/emr/ddb/lib/emr-ddb-hadoop.jar"

export HADOOP_CLASSPATH="$HADOOP_CLASSPATH:/usr/share/aws/emr/goodies/lib/emr-hadoop-goodies.jar"

export HADOOP_CLASSPATH="$HADOOP_CLASSPATH:/usr/share/aws/emr/kinesis/lib/emr-kinesis-hadoop.jar"

# Add CloudWatch sink jar to classpath
export HADOOP_CLASSPATH="$HADOOP_CLASSPATH:/usr/share/aws/emr/cloudwatch-sink/lib/*"

# Add security artifacts to classpath
export HADOOP_CLASSPATH="$HADOOP_CLASSPATH:/usr/share/aws/emr/security/conf:/usr/share/aws/emr/security/lib/*"

export HADOOP_OPTS="$HADOOP_OPTS -server -XX:+ExitOnOutOfMemoryError --add-exports=jdk.compiler/com.sun.tools.javac.util=ALL-UNNAMED"
export JAVA_HOME=/usr/lib/jvm/jre-17
export HADOOP_NAMENODE_HEAPSIZE=1024
export HADOOP_DATANODE_HEAPSIZE=614
export HADOOP_JOB_HISTORYSERVER_HEAPSIZE=2252

export HADOOP_NAMENODE_OPTS="-jjavaagent:/home/hadoop/jmx_prometheus_javaagent0.16.1.jar=12345:/home/hadoop/config.yml"

^G Help      ^O Write Out  ^W Where Is   ^K Cut        ^T Execute    ^C Location   M-U Undo      M-A Set Mark
^X Exit      ^R Read File  ^_ Replace    ^U Paste      ^J Justify    ^_ Go To Line M-E Redo      M-C Copy
```

4. Reiniciar el NameNode

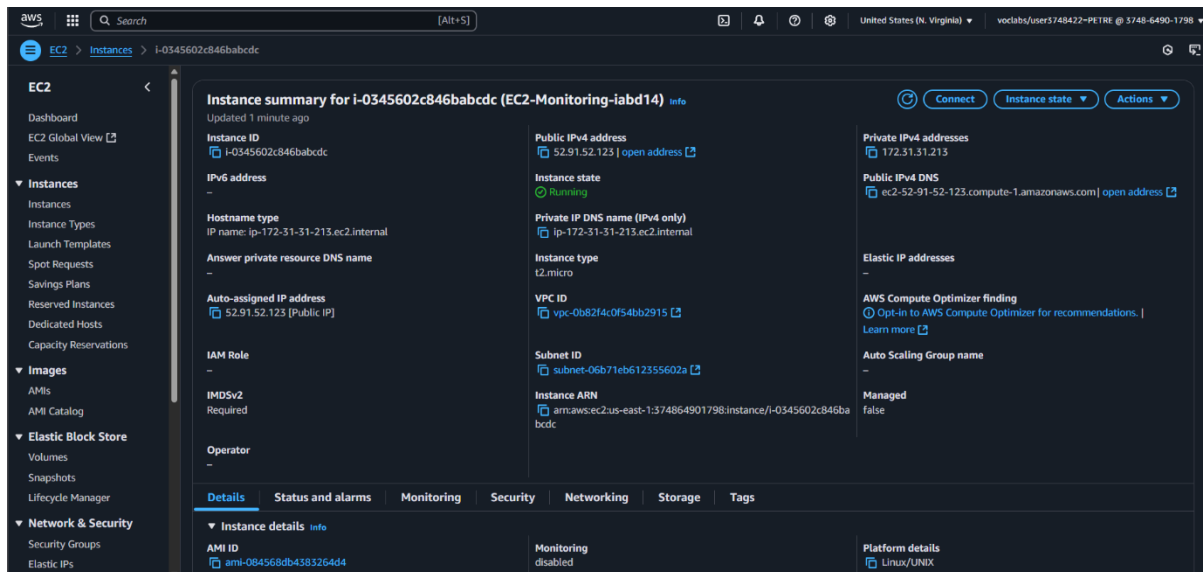
```
[hadoop@ip-172-31-30-29 ~]$ sudo systemctl restart hadoop-hdfs-namenode
```

Comprobamos que está bien configurado con el comando:

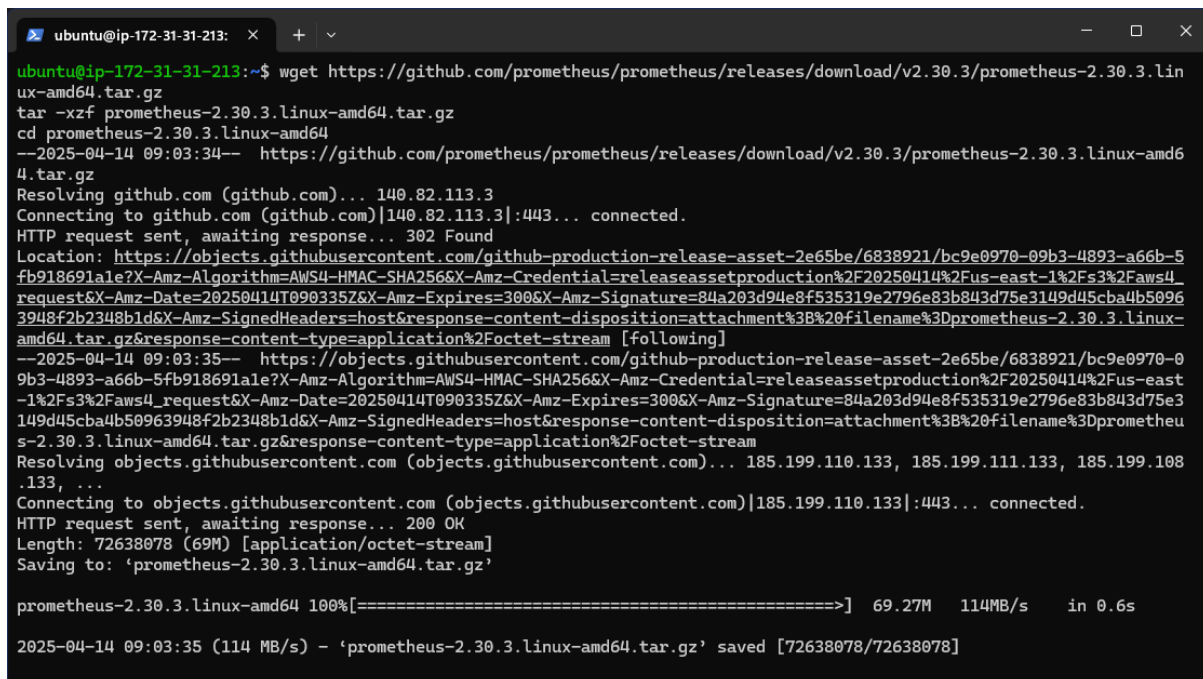
```
hadoop@ip-172-31-30-29:~$ curl http://localhost:12345/metrics
# HELP jvm_memory_objects_pending_finalization The number of objects waiting in the finalizer queue.
# TYPE jvm_memory_objects_pending_finalization gauge
jvm_memory_objects_pending_finalization 0.0
# HELP jvm_memory_bytes_used Used bytes of a given JVM memory area.
# TYPE jvm_memory_bytes_used gauge
jvm_memory_bytes_used{area="heap",} 1.4481744E8
jvm_memory_bytes_used{area="nonheap",} 6.309228E7
# HELP jvm_memory_bytes_committed Committed (bytes) of a given JVM memory area.
# TYPE jvm_memory_bytes_committed gauge
jvm_memory_bytes_committed{area="heap",} 2.4117248E8
jvm_memory_bytes_committed{area="nonheap",} 6.5667072E7
# HELP jvm_memory_bytes_max Max (bytes) of a given JVM memory area.
# TYPE jvm_memory_bytes_max gauge
jvm_memory_bytes_max{area="heap",} 1.073741824E9
jvm_memory_bytes_max{area="nonheap",} -1.0
# HELP jvm_memory_bytes_init Initial bytes of a given JVM memory area.
# TYPE jvm_memory_bytes_init gauge
jvm_memory_bytes_init{area="heap",} 1.30023424E8
jvm_memory_bytes_init{area="nonheap",} 7667712.0
# HELP jvm_memory_pool_bytes_used Used bytes of a given JVM memory pool.
# TYPE jvm_memory_pool_bytes_used gauge
jvm_memory_pool_bytes_used{pool="CodeHeap 'non-nmethods'",} 1416576.0
jvm_memory_pool_bytes_used{pool="Metaspace",} 4.8575416E7
jvm_memory_pool_bytes_used{pool="CodeHeap 'profiled nmethods'",} 6208256.0
jvm_memory_pool_bytes_used{pool="Compressed Class Space",} 5291648.0
jvm_memory_pool_bytes_used{pool="G1 Eden Space",} 0.0
jvm_memory_pool_bytes_used{pool="G1 Old Gen",} 1.40921344E8
jvm_memory_pool_bytes_used{pool="G1 Survivor Space",} 3896096.0
jvm_memory_pool_bytes_used{pool="CodeHeap 'non-profiled nmethods'",} 1600384.0
```

Parte 3: Despliegue de Prometheus y Grafana

1. Crear una instancia EC2 para Prometheus y Grafana



2. Instalar Prometheus



3. Configurar Prometheus

```
ubuntu@ip-172-31-31-213:~/prometheus-2.30.3.linux-amd64$ nano prometheus.yml

GNU nano 7.2 prometheus.yml
- static_configs:
  - targets:
      # - alertmanager:9093

# Load rules once and periodically evaluate them according to the global 'evaluation_interval'.
rule_files:
# - "first_rules.yml"
# - "second_rules.yml"

# A scrape configuration containing exactly one endpoint to scrape:
# Here it's Prometheus itself.
scrape_configs:
# The job name is added as a label 'job=<job_name>' to any timeseries scraped from this config.
- job_name: "prometheus"

# metrics_path defaults to '/metrics'
# scheme defaults to 'http'.

static_configs:
- targets: ["localhost:9090"]

- job_name: 'emr-namenode'
static_configs:
- targets: ['172.31.30.29:12345']

^G Help      ^O Write Out  ^W Where Is   ^K Cut        ^T Execute    ^C Location   M-U Undo      M-A Set Mark
^X Exit      ^R Read File  ^\ Replace    ^U Paste      ^J Justify    ^/_ Go To Line M-E Redo      M-6 Copy
```

Ejecutamos el servicio:

```
ubuntu@ip-172-31-31-213:~/prometheus-2.30.3.linux-amd64$ ./prometheus --config.file=prometheus.yml
level=info ts=2025-04-14T09:12:56.111Z caller=main.go:400 msg="No time or size retention was set so using the default ti
me retention" duration=15d
level=info ts=2025-04-14T09:12:56.112Z caller=main.go:438 msg="Starting Prometheus" version="(version=2.30.3, branch=HEA
D, revision=f29cacc42557f6a8ec30ea9b3c8c089391bd5df)"
level=info ts=2025-04-14T09:12:56.112Z caller=main.go:443 build_context="(go=go1.17.1, user=root@5cfff4265f0e3, date=2021
1005-16:10:52)"
level=info ts=2025-04-14T09:12:56.112Z caller=main.go:444 host_details="(Linux 6.8.0-1024-aws #26-Ubuntu SMP Tue Feb 18
17:22:37 UTC 2025 x86_64 ip-172-31-31-213 (none))"
level=info ts=2025-04-14T09:12:56.112Z caller=main.go:445 fd_limits="(soft=1024, hard=1048576)"
level=info ts=2025-04-14T09:12:56.112Z caller=main.go:446 vm_limits="(soft=unlimited, hard=unlimited)"
level=info ts=2025-04-14T09:12:56.113Z caller=web.go:541 component=web msg="Start listening for connections" address=0.0
.0:9090
level=info ts=2025-04-14T09:12:56.114Z caller=main.go:822 msg="Starting TSDB ..."
level=info ts=2025-04-14T09:12:56.121Z caller=tsdb_config.go:191 component=web msg="TLS is disabled." http2=false
level=info ts=2025-04-14T09:12:56.122Z caller=head.go:479 component=tsdb msg="Replaying on-disk memory mappable chunks i
f any"
level=info ts=2025-04-14T09:12:56.122Z caller=head.go:513 component=tsdb msg="On-disk memory mappable chunks replay comp
leted" duration=1.69µs
level=info ts=2025-04-14T09:12:56.122Z caller=head.go:519 component=tsdb msg="Replaying WAL, this may take a while"
level=info ts=2025-04-14T09:12:56.124Z caller=head.go:590 component=tsdb msg="WAL segment loaded" segment=0 maxSegment=1
level=info ts=2025-04-14T09:12:56.124Z caller=head.go:590 component=tsdb msg="WAL segment loaded" segment=1 maxSegment=1
level=info ts=2025-04-14T09:12:56.124Z caller=head.go:596 component=tsdb msg="WAL replay completed" checkpoint_replay_du
ration=24.897µs wal_replay_duration=1.824885ms total_replay_duration=1.933972ms
level=info ts=2025-04-14T09:12:56.125Z caller=main.go:849 fs_type=EXT4_SUPER_MAGIC
level=info ts=2025-04-14T09:12:56.125Z caller=main.go:852 msg="TSDB started"
level=info ts=2025-04-14T09:12:56.125Z caller=main.go:979 msg="Loading configuration file" filename=prometheus.yml
level=info ts=2025-04-14T09:12:56.128Z caller=main.go:1016 msg="Completed loading of configuration file" filename=promet
heus.yml totalDuration=2.744564ms db_storage=1.016µs remote_storage=1.479µs web_handler=629ns query_engine=955ns scrape=
2.461742ms scrape_sd=39.919µs notify=21.277µs notify_sd=8.572µs rules=1.372µs
```

4. Instalar Grafana

Para instalar Grafana en la misma instancia ejecutamos el bloque de comandos:


```
ubuntu@ip-172-31-31-213: ~/prometheus-2.30.3.linux-amd64$ sudo apt-get update
sudo apt-get install -y apt-transport-https software-properties-common wget
wget -q -O - https://packages.grafana.com/gpg.key | sudo apt-key add -
echo "deb https://packages.grafana.com/oss/deb stable main" | sudo tee -a /etc/apt/sources.list.d/grafana.list
sudo apt-get update
sudo apt-get install -y grafana
Hit:1 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble InRelease
Get:2 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble-updates InRelease [126 kB]
Get:3 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble-backports InRelease [126 kB]
Get:4 http://security.ubuntu.com/ubuntu noble-security InRelease [126 kB]
Get:5 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble/universe amd64 Packages [15.0 MB]
Get:6 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble/universe Translation-en [5982 kB]
Get:7 http://security.ubuntu.com/ubuntu noble-security/main amd64 Packages [741 kB]
Get:8 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble/universe amd64 Components [3871 kB]
Get:9 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble/universe amd64 c-n-f Metadata [301 kB]
Get:10 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble/multiverse amd64 Packages [269 kB]
Get:11 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble/multiverse Translation-en [118 kB]
Get:12 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble/multiverse amd64 Components [35.0 kB]
Get:13 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble/multiverse amd64 c-n-f Metadata [8328 B]
Get:14 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble-updates/main amd64 Packages [991 kB]
Get:15 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble-updates/main Translation-en [219 kB]
Get:16 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble-updates/main amd64 Components [151 kB]
Get:17 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble-updates/main amd64 c-n-f Metadata [13.5 kB]
Get:18 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble-updates/universe amd64 Packages [1052 kB]
Get:19 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble-updates/universe Translation-en [265 kB]
Get:20 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble-updates/universe amd64 Components [367 kB]
Get:21 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble-updates/universe amd64 c-n-f Metadata [26.0 kB]
Get:22 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble-updates/restricted amd64 Packages [892 kB]
Get:23 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble-updates/restricted Translation-en [182 kB]
Get:24 http://us-east-1.ec2.archive.ubuntu.com/ubuntu noble-updates/restricted amd64 Components [212 B]
```

Y ejecutamos el servicio:

```
ubuntu@ip-172-31-31-213: ~/prometheus-2.30.3.linux-amd64$ sudo systemctl start grafana-server
sudo systemctl enable grafana-server
Synchronizing state of grafana-server.service with SysV service script with /usr/lib/systemd/systemd-sysv-install.
Executing: /usr/lib/systemd/systemd-sysv-install enable grafana-server
Created symlink /etc/systemd/system/multi-user.target.wants/grafana-server.service → /usr/lib/systemd/system/grafana-server.service.
```

5. Configurar Grafana

⚠ No es seguro 52.91.52.123:3000/login




Welcome to Grafana

Email or username

Password

Log in

[Forgot your password?](#)

 Grafana

Home

Bookmarks

Starred

Dashboards

Explore

Drilldown New!

Alerting

Connections

Add new connection

Data sources

Administration

General

Stats and license

Default preferences


Settings

Home > Connections > Data sources

Search or jump to...

Data sources

View and manage your connected data source connections



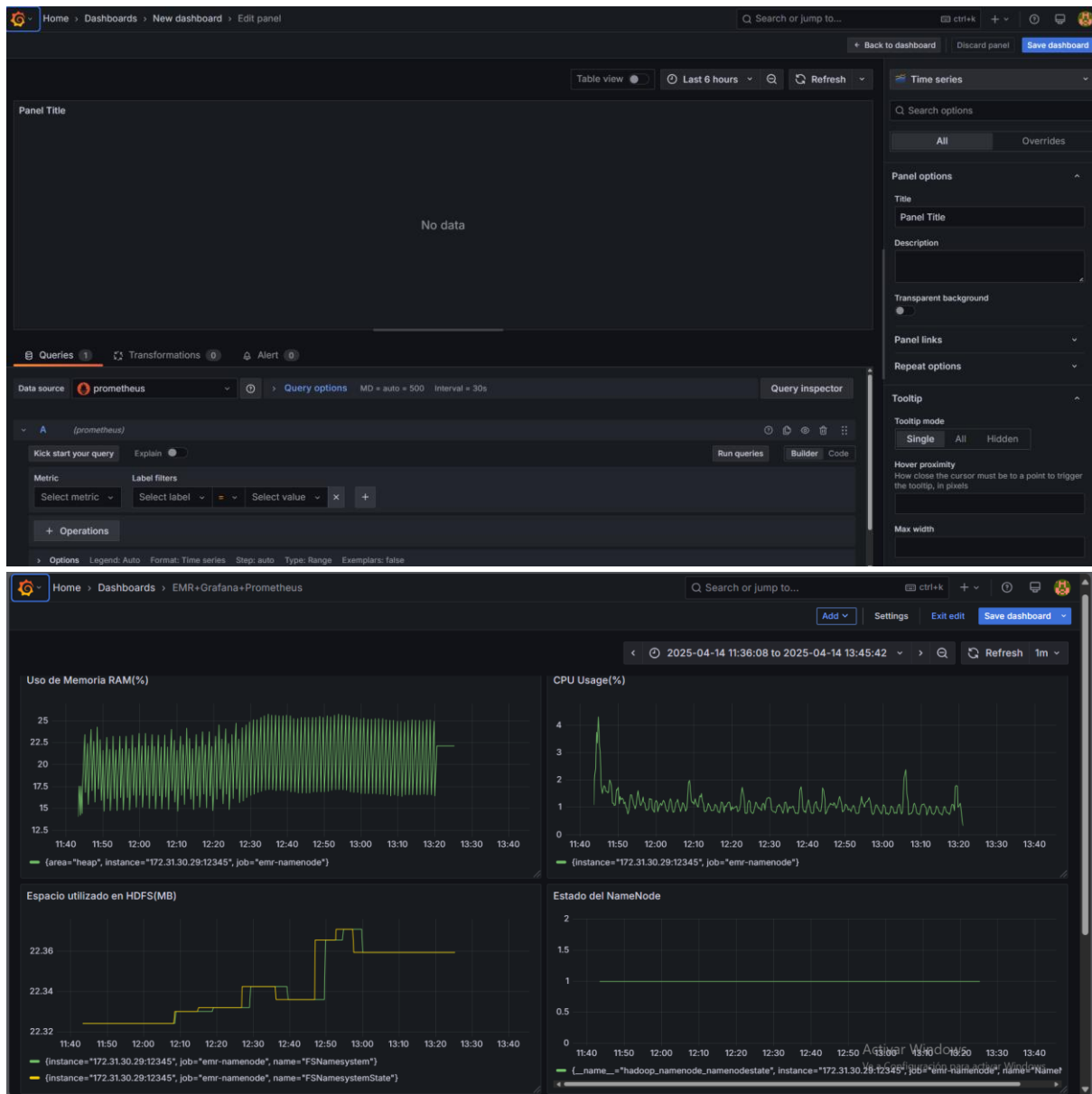
No data sources defined

You can also define data sources through configuration files. [Learn more](#)

Add data source

Parte 4: Visualización de métricas en Grafana

1. Crear un dashboard en Grafana



Para la gráfica:

- Memoria RAM: Para calcular el porcentaje de memoria RAM que se está usando $(\text{jvm_memory_bytes_used}\{\text{area}=\text{"heap"}\} / \text{jvm_memory_bytes_max}\{\text{area}=\text{"heap"}\}) * 100$
- CPU: Para mostrar solo las métricas del EMR $\text{rate}(\text{process_cpu_seconds_total}\{\text{instance}=\text{"172.31.30.29:12345"}\}[1m]) * 100$
- Espacio HDFS: Para mostrar el espacio usado en MB $\text{hadoop_namenode_capacityused} / (1024 * 1024)$
- Estado del namenode: devuelve valor 1 para activo y 0 para standby

hadoop_namenode_namenodestate