

## MapReduce Service

# Preguntas frecuentes

Edición 01  
Fecha 2023-11-18



**Copyright © Huawei Cloud Computing Technologies Co., Ltd. 2025. Todos los derechos reservados.**

Quedan terminantemente prohibidas la reproducción y/o la divulgación totales y/o parciales del presente documento de cualquier forma y/o por cualquier medio sin la previa autorización por escrito de Huawei Cloud Computing Technologies Co., Ltd.

## **Marcas registradas y permisos**



El logotipo HUAWEI y otras marcas registradas de Huawei pertenecen a Huawei Technologies Co., Ltd. Todas las demás marcas registradas y los otros nombres comerciales mencionados en este documento son propiedad de sus respectivos titulares.

## **Aviso**

Es posible que la totalidad o parte de los productos, las funcionalidades y/o los servicios que figuran en el presente documento no se encuentren dentro del alcance de un contrato vigente entre Huawei Cloud y el cliente. Las funcionalidades, los productos y los servicios adquiridos se limitan a los estipulados en el respectivo contrato. A menos que un contrato especifique lo contrario, ninguna de las afirmaciones, informaciones ni recomendaciones contenidas en el presente documento constituye garantía alguna, ni expresa ni implícita.

Huawei está permanentemente preocupada por la calidad de los contenidos de este documento; sin embargo, ninguna declaración, información ni recomendación aquí contenida constituye garantía alguna, ni expresa ni implícita. La información contenida en este documento se encuentra sujeta a cambios sin previo aviso.

## **Huawei Cloud Computing Technologies Co., Ltd.**

Dirección: Huawei Cloud Data Center Jiaoxinggong Road  
Avenida Qianzhong  
Nuevo distrito de Gui'an  
Gui Zhou, 550029  
República Popular China

Sitio web: <https://www.huaweicloud.com/intl/es-us/>

# Índice

<b>1 Descripción de MRS.....</b>	<b>1</b>
1.1 ¿Para qué se utiliza el MRS?.....	1
1.2 ¿Qué tipos de almacenamiento distribuido admite MRS?.....	1
1.3 ¿Cómo creo un clúster MRS mediante un grupo de seguridad personalizado?.....	3
1.4 ¿Cómo uso MRS?.....	3
1.5 Región y AZ.....	4
1.6 ¿Puedo configurar un grupo de conexiones Phoenix?.....	5
1.7 ¿Apoya MRS el cambio del segmento de la red?.....	6
1.8 ¿Puedo degradar las especificaciones de un nodo de clúster MRS?.....	6
1.9 ¿Cuál es la relación entre Hive y otros componentes?.....	6
1.10 ¿Un clúster de MRS soporta Hive en Spark?.....	7
1.11 ¿Cuáles son las diferencias entre las versiones de Hive?.....	7
1.12 ¿Qué versión de clúster de MRS admite la conexión Hive y la sincronización de usuarios?.....	7
1.13 ¿Cuáles son las diferencias entre OBS y HDFS en el almacenamiento de datos?.....	8
1.14 ¿Cómo obtengo la herramienta de prueba de presión de Hadoop?.....	8
1.15 ¿Cuál es la relación entre Impala y otros componentes?.....	8
1.16 Declaración sobre las direcciones IP públicas en el SDK de terceros de código abierto integrado por MRS.....	9
1.17 ¿Cuál es la relación entre Kudu y HBase?.....	9
1.18 ¿Admite MRS la ejecución de Hive en Kudu?.....	10
1.19 ¿Cuáles son las soluciones para procesar mil millones de registros de datos?.....	10
1.20 ¿Puedo cambiar la dirección IP de DBService?.....	10
1.21 ¿Puedo borrar MRS sudo Logs?.....	10
1.22 ¿Cuáles son las restricciones en el tamaño del registro de Storm en un clúster MRS 2.1.0?.....	11
1.23 ¿Qué es Spark ThriftServer?.....	11
1.24 ¿Qué protocolos de acceso admite Kafka?.....	12
1.25 Se notifica el error 408 cuando un nodo MRS accede a OBS (Palabra clave, no enviada a la nube operada conjuntamente).....	12
1.26 ¿Cuáles son las ventajas de la relación de compresión de zstd?.....	12
1.27 ¿Por qué los componentes HDFS, YARN y MapReduce no están disponibles cuando se compra un clúster de MRS?.....	13
1.28 ¿Por qué no está disponible el componente ZooKeeper cuando se compra un clúster MRS?.....	13
1.29 ¿Qué versiones de Python son compatibles con las tareas de Spark en clústeres MRS 3.1.0?.....	13
1.30 ¿Cómo puedo habilitar diferentes programas de servicio para usar diferentes colas de YARN?.....	14
1.31 Diferencias y relaciones entre la consola de gestión de MRS y el Manager de clústeres.....	17

1.32 ¿Cómo desvinculo una EIP del FusionInsight Manager de un clúster MRS?.....	19
1.33 ¿Cuáles son los sistemas operativos de hosts en clústeres MRS de diferentes versiones?.....	20
<b>2 Facturación.....</b>	<b>21</b>
2.1 ¿Cómo se factura MRS?.....	21
2.2 ¿Por qué no se muestra el precio durante la creación del clúster MRS?.....	21
2.3 ¿Cómo se factura el escalado automático de un clúster MRS?.....	22
2.4 ¿Cómo se renueva MRS?.....	22
2.5 ¿Cómo se factura el nodo de tareas en un clúster MRS?.....	22
2.6 ¿Por qué falla mi cancelación de la suscripción a ECS después de cancelar mi suscripción a MRS?.....	23
<b>3 Cuenta y contraseña.....</b>	<b>24</b>
3.1 ¿Qué es la cuenta para iniciar sesión en Manager?.....	24
3.2 ¿Cómo puedo consultar y cambiar el período de validez de la contraseña de una cuenta?.....	24
<b>4 Cuenta y permiso.....</b>	<b>26</b>
4.1 ¿Es compatible un clúster MRS con el control de permisos de acceso si la autenticación Kerberos no está activada?.....	26
4.2 ¿Cómo asigno permiso de gestión de tenant a una cuenta nueva?.....	26
4.3 ¿Cómo personalizo una política de MRS?.....	27
4.4 ¿Por qué no puedo encontrar la gestión de usuarios en la configuración del sistema en MRS Manager?.....	28
4.5 ¿Proporciona Hue la función de configurar los permisos de la cuenta?.....	28
4.6 ¿Por qué no puedo enviar trabajos en la consola después de que mi cuenta de IAM esté asignada con permisos relacionados?.....	28
4.7 ¿Qué debo hacer si se notifica un error que indica una autenticación no válida cuando envío una orden de compra de clúster de MRS?.....	29
<b>5 Uso del cliente.....</b>	<b>30</b>
5.1 ¿Cómo configuro variables de entorno y ejecuto comandos en un cliente de componentes?.....	30
5.2 ¿Cómo desactivo la autenticación de ZooKeeper SASL?.....	30
5.3 Se notifica un error cuando se ejecuta el comando kinit en un nodo cliente fuera de un clúster MRS.....	31
<b>6 Acceso a páginas Web.....</b>	<b>32</b>
6.1 ¿Cómo cambio la duración del tiempo de espera de la sesión para una interfaz de usuario web de componente de código abierto?.....	32
6.2 ¿Por qué no puedo actualizar la página Dynamic Resource Plan en la pestaña de tenant de MRS?.....	35
6.3 ¿Qué hago si la pestaña Kafka Topic Monitoring no está disponible en Manager?.....	35
6.4 ¿Cómo lo hago si se informa de un error o algunas funciones no están disponibles cuando accedo a las interfaces de usuario web de HDFS, Hue, YARN, HetuEngine y Flink?.....	36
6.5 ¿Cómo cambio el modo de acceso a MRS Manager (no para la nube operada conjuntamente)?.....	37
<b>7 Monitoreo de alarmas.....</b>	<b>39</b>
7.1 En un clúster de streaming de MRS, ¿puede la función de monitoreo de Kafka topic enviar notificaciones de alarma?.....	39
7.2 ¿Dónde puedo ver las colas de recursos en ejecución cuando se generan ALM-18022 Recursos de cola de Yarn insuficientes?.....	39
7.3 ¿Cómo entiendo las estadísticas de gráficos de varios niveles en la métrica de solicitudes de operación de HBase?.....	40

<b>8 Ajuste de rendimiento.....</b>	<b>42</b>
8.1 ¿Un clúster MRS admite la reinstalación del sistema?.....	42
8.2 ¿Puedo cambiar el sistema operativo de un clúster MRS?.....	42
8.3 ¿Cómo puedo mejorar la utilización de recursos de los nodos Core en un clúster?.....	42
8.4 ¿Cómo detengo el servicio Firewall?.....	43
<b>9 Desarrollo del trabajo.....</b>	<b>44</b>
9.1 ¿Cómo obtengo mis datos en OBS o HDFS?.....	44
9.2 ¿Qué tipos de trabajos de Spark se pueden enviar en un clúster?.....	45
9.3 ¿Puedo ejecutar varias tareas de Spark al mismo tiempo después de que los recursos mínimos del tenant de un clúster MRS se cambien a 0?.....	45
9.4 ¿Qué hago si no se pueden identificar los parámetros de trabajo separados por espacios?.....	46
9.5 ¿Cuáles son las diferencias entre el modo de client y el modo de cluster de los trabajos de Spark?.....	46
9.6 ¿Cómo puedo ver los registros de trabajos de MRS?.....	46
9.7 ¿Cómo lo hago si se muestra el mensaje "The current user does not exist on MRS Manager. Grant the user sufficient permissions on IAM and then perform IAM user synchronization on the Dashboard tab page."?.....	47
9.8 La ejecución de trabajo de LauncherJob se ha fallado y el mensaje de error "jobPropertiesMap is null." se muestra.....	48
9.9 ¿Cómo lo hago si el estado del trabajo de Flink en la consola MRS es incompatible con el del Yarn?.....	48
9.10 ¿Cómo lo hago si un trabajo de SparkStreaming falla después de haber sido ejecutado docenas de horas y se reporta el error OBS Access 403?.....	48
9.11 ¿Cómo hago si se informa de una alarma que indica que la memoria es insuficiente cuando ejecuto una sentencia SQL en el cliente ClickHouse?.....	48
9.12 ¿Cómo lo hago si se muestra un mensaje de error "java.io.IOException: Connection reset by peer" durante la ejecución de un trabajo de Spark?.....	49
9.13 ¿Cómo hago si se muestra un mensaje de error "requestId=4971883851071737250" cuando un trabajo de Spark accede a OBS?.....	49
9.14 ¿Cómo lo hago si se reporta el error "UnknownScannerException" del trabajo de Spark?.....	50
9.15 ¿Por qué DataArtsStudio ocasionalmente no puede programar trabajos de Spark y la reprogramación también falla?.....	50
9.16 ¿Cómo hago si un trabajo de Flink no se ejecuta y se muestra el mensaje de error "java.lang.NoSuchFieldError: SECURITY_SSL_ENCRYPT_ENABLED"?.....	50
9.17 ¿Por qué no se puede ver el trabajo de Yarn enviado en la interfaz de usuario web?.....	51
9.18 ¿Cómo modifico el HDFS NameSpace (fs.defaultFS) de un clúster existente?.....	51
9.19 ¿Cómo lo hago si YARN detiene la cola de launcher-job debido a un heap size insuficiente cuando envío un trabajo de Flink en el plano de gestión?.....	51
9.20 ¿Cómo lo hago si se muestra el mensaje de error "slot request timeout" cuando envío un trabajo de Flink?.....	52
9.21 Importación y exportación de datos de trabajos de DistCP.....	52
9.22 ¿Cómo puedo ver las sentencias SQL para los trabajos de Hive en la YARN Web UI?.....	52
9.23 ¿Cómo puedo ver los registros de una tarea de Yarn especificada?.....	53
<b>10 Actualización/Instalación de parches de clústeres.....</b>	<b>55</b>
10.1 ¿Cómo actualizo un clúster MRS?.....	55
10.2 ¿Puedo cambiar la versión del clúster MRS?.....	55
<b>11 Interconexión del ecosistema periférico.....</b>	<b>56</b>
11.1 ¿Se puede utilizar MRS para realizar operaciones de lectura y escritura en tablas DLI?.....	56

11.2 ¿OBS soporta el protocolo ListObjectsV2?.....	56
11.3 ¿Se pueden almacenar datos MRS en un sistema de archivos paralelo proporcionado por OBS?.....	56
11.4 ¿Se puede desplegar el servicio Crawler en MRS?.....	56
11.5 ¿DWS y MRS soportan la eliminación segura (evitar la recuperación después de la eliminación)?.....	56
11.6 ¿Por qué no se encuentra el clúster MRS autenticado por Kerberos cuando se configura una conexión desde DLF?.....	57
11.7 ¿Cómo uso PySpark en un ECS para conectarse a un clúster MRS Spark con autenticación de Kerberos habilitada, en la Intranet?.....	57
11.8 ¿Por qué los campos asignados no existen en la base de datos después de que HBase sincronice los datos con CSS?.....	57
11.9 ¿Puede Flume leer datos de OBS?.....	57
11.10 ¿Se puede conectar MRS a un KDC externo?.....	57
11.11 ¿Cómo soluciono el problema de compatibilidad de la versión Jetty en la interconexión de código abierto Kylin 3.x y MRS 1.9.3?.....	57
11.12 ¿Qué sucede si los datos no se exportan desde MRS a un bucket cifrado de OBS?.....	58
11.13 ¿Cómo interconecto MRS con LTS?.....	58
11.14 ¿Cómo instalo HSS en nodos de clúster MRS?.....	61
<b>12 Acceso al clúster.....</b>	<b>62</b>
12.1 ¿Puedo cambiar entre los dos modos de inicio de sesión de MRS?.....	62
12.2 ¿Cómo puedo obtener la dirección IP y el número de puerto de una instancia de ZooKeeper?.....	62
12.3 ¿Cómo accedo a un clúster MRS desde un nodo fuera del clúster?.....	63
<b>13 Desarrollo de servicios de big data.....</b>	<b>66</b>
13.1 ¿Puede MRS ejecutar múltiples tareas de Flume a la vez?.....	66
13.2 ¿Cómo cambio los registros de FlumeClient a los registros estándar?.....	66
13.3 ¿Dónde se almacenan los archivos JAR y las variables de entorno de Hadoop?.....	67
13.4 ¿Qué algoritmos de compresión admite HBase?.....	67
13.5 ¿Puede MRS escribir datos en HBase a través de la tabla externa de HBase de Hive?.....	67
13.6 ¿Cómo veo los registros de HBase?.....	67
13.7 ¿Cómo configuro el TTL para una tabla HBase?.....	67
13.8 ¿Cómo me conecto a HBase de MRS a través de HappyBase?.....	67
13.9 ¿Cómo cambio el número de réplicas HDFS?.....	68
13.10 ¿Cómo modifico la clase de conmutación HDFS activo y en espera?.....	68
13.11 ¿Cuál es el tipo de número recomendado de DynamoDB en las tablas Hive?.....	68
13.12 ¿Se puede interconectar el controlador Hive con DBCP2?.....	69
13.13 ¿Cómo puedo ver la tabla Hive creada por otro usuario?.....	69
13.14 ¿Dónde puedo descargar el paquete de dependencias (com.huawei.gaussc10) en el proyecto de ejemplo de Hive?.....	70
13.15 ¿Puedo exportar el resultado de la consulta de datos de Hive?.....	71
13.16 ¿Cómo lo hago si ocurre un error cuando Hive ejecuta el comando <b>beeline -e</b> para ejecutar varias sentencias?.....	71
13.17 ¿Cómo lo hago si un trabajo "hivesql/hivescript" no se envía después de agregar Hive?.....	71
13.18 ¿Qué pasa si un archivo de Excel descargado en Hue no se puede abrir?.....	72
13.19 ¿Cómo lo hago si las sesiones no se liberan después de que Hue se conecta a HiveServer y se muestra el mensaje de error "ver max user connections"?.....	73

13.20 ¿Cómo se restablecen los datos de Kafka?	74
13.21 ¿Cómo obtengo la versión de cliente de MRS Kafka?	74
13.22 ¿Qué protocolos de acceso son compatibles con Kafka?	74
13.23 ¿Cómo lo hago si se muestra un mensaje de error "Not Authorized to access group xxx" cuando se consume un Kafka topic?	74
13.24 ¿Qué algoritmos de compresión soporta Kudu?	75
13.25 ¿Cómo puedo ver los registros de Kudu?	75
13.26 ¿Cómo manejo las excepciones del servicio Kudu generadas durante la creación de clústeres?	75
13.27 ¿Cuáles son las diferencias entre la construcción de proyectos de ejemplo y el desarrollo de aplicaciones? ¿Es compatible el código Python?	76
13.28 ¿OpenTSDB soporta las API de Python?	76
13.29 ¿Cómo configuro otras fuentes de datos en Presto?	76
13.30 ¿Cómo actualizo el certificado Ranger?	78
13.31 ¿Cómo me conecto a Spark Shell desde MRS?	80
13.32 ¿Cómo me conecto a Spark Beeline desde MRS?	80
13.33 ¿Dónde se almacenan los registros de ejecución de los trabajos de Spark?	81
13.34 ¿Cómo especifico una ruta de acceso de registro al enviar una tarea en un clúster MRS Storm?	81
13.35 ¿Cómo puedo comprobar si la configuración ResourceManager de Yarn es correcta?	82
13.36 ¿Cómo modifico el parámetro allow_drop_detached de ClickHouse?	84
13.37 ¿Cómo lo hago si se informa de una alarma que indica memoria insuficiente durante la ejecución de la tarea de Spark?	85
13.38 ¿Cómo agrego una política de eliminación periódica para evitar registros de tabla de sistema ClickHouse de gran tamaño?	85
13.39 ¿Cómo obtengo un archivo Spark JAR?	86
13.40 ¿Por qué se genera una alarma cuando el proceso NameNode no se reinicia después de modificar el archivo hdfs-site.xml?	86
13.41 Se necesita mucho tiempo para que Spark SQL acceda a las tablas particionadas de Hive antes del inicio de Job	87
13.42 ¿Qué debo hacer si spark.yarn.executor.memoryOverhead no tiene efecto?	87
13.43 ¿Cómo cambio la zona horaria del servicio ClickHouse?	88
13.44 ¿Qué debo hacer si falla la conexión con el servidor ClickHouse? Código de error: 516	89
<b>14 API</b>	<b>90</b>
14.1 ¿Cómo configuro el parámetro node_id cuando uso la API para ajustar nodos de clúster?	90
<b>15 Gestión de clúster</b>	<b>91</b>
15.1 ¿Cómo puedo ver todos los clústeres?	91
15.2 ¿Cómo puedo ver la información de registro?	91
15.3 ¿Cómo puedo ver la información de configuración del clúster?	92
15.4 ¿Cómo agrego servicios a un clúster MRS?	92
15.5 ¿Cómo instalo Kafka y Flume en un clúster de MRS?	92
15.6 ¿Cómo detengo un clúster MRS?	92
15.7 ¿Es necesario apagar un nodo Master antes de actualizar sus especificaciones?	93
15.8 ¿Puedo agregar componentes a un clúster existente?	93
15.9 ¿Puedo eliminar componentes instalados en un clúster MRS?	93

15.10 ¿Puedo cambiar los nodos de clúster MRS en la consola MRS?.....	93
15.11 ¿Cómo puedo proteger las notificaciones de eventos/alarmas de clúster?.....	93
15.12 ¿Por qué la memoria del grupo de recursos mostrada en el cluster MRS es menor que la memoria real del cluster?.....	94
15.13 ¿Cómo configuro la memoria Knox?.....	94
15.14 ¿Qué es la versión de Python instalada para un clúster MRS?.....	94
15.15 ¿Cómo puedo ver el directorio de archivos de configuración de cada componente?.....	95
15.16 ¿Cómo subo un archivo local a un nodo dentro de un clúster?.....	96
15.17 ¿Cómo lo hago si el tiempo en los nodos MRS es incorrecto?.....	96
15.18 ¿Cómo puedo consultar la hora de inicio de un nodo MRS?.....	97
15.19 ¿Cómo lo hago si las relaciones de confianza entre nodos son anormales?.....	97
15.20 ¿Cómo ajusto el tamaño de memoria del proceso manager-executor?.....	98
15.21 ¿Puedo modificar un nodo master en un clúster MRS existente?.....	99
<b>16 Uso de Kerberos.....</b>	<b>100</b>
16.1 ¿Cómo cambio el estado de autenticación de Kerberos de un clúster MRS creado?.....	100
16.2 ¿Cuáles son los puertos del servicio de autenticación de Kerberos?.....	100
16.3 ¿Cómo despliego el servicio Kerberos en un clúster en ejecución?.....	100
16.4 ¿Cómo accedo a Hive en un clúster con autenticación de Kerberos habilitada?.....	100
16.5 ¿Cómo puedo acceder de Presto en un clúster con autenticación de Kerberos habilitada?.....	101
16.6 ¿Cómo accedo a Spark en un clúster con autenticación Kerberos habilitada?.....	102
16.7 ¿Cómo puedo evitar que la autenticación Kerberos caduque?.....	103
<b>17 Gestión de metadatos.....</b>	<b>105</b>
17.1 ¿Dónde puedo ver metadatos de Hive?.....	105



# 1 Descripción de MRS

---

## 1.1 ¿Para qué se utiliza el MRS?

### Pregunta:

¿Qué puede hacer el MapReduce Service (MRS)?

### Respuesta:

MapReduce Service (MRS) es una plataforma de big data de nivel empresarial que le permite crear y operar rápidamente entornos de big data económicos, seguros, de pila completa y nativos de la nube en la nube. Proporciona motores como ClickHouse Spark, Flink, Kafka y HBase, y soporta la convergencia de lago de datos, almacén de datos, inteligencia empresarial (BI) e inteligencia artificial (IA). Totalmente compatible con componentes de código abierto, MRS le ayuda a innovar y expandir rápidamente el crecimiento del servicio.

## 1.2 ¿Qué tipos de almacenamiento distribuido admite MRS?

### Pregunta:

¿Qué tipos de almacenamiento distribuido admite MRS? ¿Qué versiones están disponibles?

### Respuesta:

MRS soporta Hadoop 3.1.x y pronto será compatible con otras versiones principales de Hadoop lanzadas por la comunidad. [Tabla 1-1](#) enumera las versiones de componentes admitidas por MRS.

Para más detalles, consulte [Lista de versiones de componentes MRS](#).

**Tabla 1-1** Versiones de componentes de MRS

Componente	MRS 1.9.2 (Aplicable a MRS 1.9.x)	MRS 3.1.0
Alluxio	2.0.1	N/A
CarbonData	1.6.1	2.0.1
ClickHouse	N/A	21.3.4.25
DBService	1.0.0	2.7.0
Flink	1.7.0	1.12.0
Flume	1.6.0	1.9.0
HBase	1.3.1	2.2.3
HDFS	2.8.3	3.1.1
Hive	2.3.3	3.1.0
Hudi	N/A	0.7.0
Hue	3.11.0	4.7.0
Impala	N/A	3.4.0
Kafka	1.1.0	2.11-2.4.0
KafkaManager	1.3.3.1	N/A
KrbServer	1.15.2	1.17
Kudu	N/A	1.12.1
LdapServer	1.0.0	2.7.0
Loader	2.0.0	N/A
MapReduce	2.8.3	3.1.1
Oozie	N/A	5.1.0
Opentsdb	2.3.0	N/A
Presto	0.216	333
Phoenix (integrated with HBase)	N/A	5.0.0
Ranger	1.0.1	2.0.0
Spark	2.2.2	N/A
Spark2x	N/A	2.4.5
Sqoop	N/A	1.4.7
Storm	1.2.1	N/A

Componente	MRS 1.9.2 (Aplicable a MRS 1.9.x)	MRS 3.1.0
Tez	0.9.1	0.9.2
YARN	2.8.3	3.1.1
ZooKeeper	3.5.1	3.5.6
MRS Manager	1.9.2	N/A
FusionInsight Manager	N/A	8.1.0

## 1.3 ¿Cómo creo un clúster MRS mediante un grupo de seguridad personalizado?

### Pregunta:

¿Cómo creo un clúster MRS mediante un grupo de seguridad personalizado?

### Respuesta:

Puede crear un clúster MRS mediante un grupo de seguridad definido por el usuario de cualquiera de las siguientes maneras:

- Al comprar un clúster y utilizar un grupo de seguridad creado por usted mismo, debe habilitar el puerto 9022.
- Al comprar un clúster, establezca Security Group en Automatically create.

## 1.4 ¿Cómo uso MRS?

MapReduce (MRS) Service es un servicio que puede usar para desplegar y gestionar componentes basados en Hadoop en Huawei Cloud. Le permite desplegar clústeres de Hadoop con unos pocos clics. MRS proporciona clústeres de big data listos para la empresa en la nube. Los tenants pueden controlar completamente los clústeres y ejecutar fácilmente componentes de big data como Hadoop, Spark, HBase, Kafka y Storm en los clústeres.

MRS es fácil de usar. Puede ejecutar varias tareas y procesar o almacenar datos de escala PB mediante equipos conectados en un clúster. Para utilizar MRS, haga lo siguiente:

1. Desarrolle un programa de procesamiento de datos. Para obtener detalles sobre cómo desarrollar rápidamente un programa de este tipo y ejecutarlo correctamente, consulte el código de ejemplo y los tutoriales proporcionados en [Método de construcción de un proyecto de ejemplo de MRS](#).
2. Suba programas locales y archivos de datos a OBS.
3. Cree un clúster. Debe especificar el tipo de clúster (por ejemplo, análisis o streaming) y establecer las especificaciones de instancia de ECS, el número de instancias, el tipo de disco de datos (E/S común, E/S alta y E/S ultraalta) y los componentes que se van a instalar, como Hadoop, Spark, HBase, Hive, Kafka y Storm, en un clúster. Puede utilizar

una acción de arranque para instalar software de terceros o modificar el entorno de ejecución del clúster en un nodo antes o después de iniciar el clúster.

4. Utilice MRS para enviar, ejecutar y monitorear sus programas.
5. Gestione clústeres en MRS Manager, una plataforma de gestión unificada de clústeres de big data a nivel empresarial. Puede obtener información sobre el estado de los servicios y hosts, obtener información crítica del sistema de manera oportuna a partir del monitoreo y la personalización de métricas gráficas, modificar los atributos del servicio en función de los requisitos de rendimiento e iniciar o detener clústeres, servicios e instancias de rol.
6. Termine cualquier clúster de MRS que no necesite una vez completada la ejecución del trabajo. El clúster terminado ya no se factura.

## 1.5 Región y AZ

### Pregunta:

¿Qué son las Regiones y las AZ?

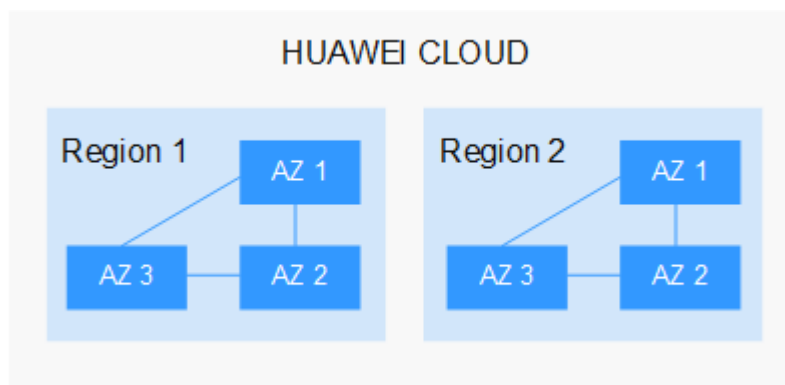
### Respuesta:

Una región y una zona de disponibilidad (AZ) identifican la ubicación de un centro de datos. Puede crear recursos en regiones y zonas de disponibilidad.

- Las regiones se definen en términos de ubicaciones geográficas y latencias de red. Los servicios públicos como Elastic Cloud Server (ECS), Elastic Volume Service (EVS), Object Storage Service (OBS), Virtual Private Cloud (VPC), Elastic IP (EIP) y Image Management Service (IMS) se comparten dentro de la misma región. Las regiones se clasifican en regiones comunes y regiones específicas. Una región común proporciona servicios en la nube comunes disponibles para todos los tenants. Una región dedicada proporciona servicios de un tipo específico o solo para tenants específicos.
- Una AZ contiene uno o más centros de datos físicos. Cada AZ cuenta con instalaciones independientes de electricidad, de refrigeración, de extinción de incendios y a prueba de humedad. La computación, la red, el almacenamiento y otros recursos en una AZ se dividen lógicamente en múltiples clústeres. Las AZ dentro de una región están interconectadas mediante fibras ópticas de alta velocidad para permitirle construir sistemas de alta disponibilidad entre AZ.

**Figura 1-1** muestra la relación entre las regiones y las zonas de disponibilidad.

**Figura 1-1** Regiones y AZ



Huawei Cloud ofrece servicios en todo el mundo. Puede seleccionar una región y AZ según sea necesario.

## ¿Cuáles son las consideraciones para seleccionar una región?

Al seleccionar una región, tenga en cuenta los siguientes factores:

- Localización  
Seleccione una región cercana a usted o a sus usuarios de destino para minimizar la latencia de la red y aumentar la velocidad de acceso. Sin embargo, no hay diferencia clara entre la infraestructura, la calidad de la red BGP y las operaciones y configuraciones de recursos en la misma región/país. Por lo tanto, no es necesario tener en cuenta la latencia de la red al seleccionar una región en un país.
  - Si usted o sus usuarios objetivo se encuentran en el área de Asia Pacífico excluyendo China continental, seleccione la región **CN-Hong Kong, AP-Bangkok** o **AP-Singapore**.
  - Si usted o sus usuarios objetivo están en África, seleccione la región de **AF-Johannesburg**.
  - Si usted o sus usuarios objetivo están en Europa, seleccione la región **EU-Paris**.
- Precio del recurso  
Los precios de los recursos pueden variar en diferentes regiones.

## ¿Cuáles son las consideraciones para seleccionar una zona de disponibilidad?

Tenga en cuenta los requisitos de recuperación ante desastres (DR) y latencia de red al seleccionar una zona de disponibilidad:

- Despliegue recursos en diferentes AZ en la misma región para fines de recuperación ante desastres.
- Despliegue recursos en la misma AZ para una latencia mínima.

## ¿Cómo obtengo un punto de conexión?

Cuando usa recursos mediante una API, debe especificar su punto de conexión regional.

# 1.6 ¿Puedo configurar un grupo de conexiones Phoenix?

### Pregunta:

¿Puedo configurar un grupo de conexiones Phoenix?

### Respuesta:

Phoenix no admite la configuración del grupo de conexiones. Se recomienda escribir código para implementar una clase de herramienta para gestionar conexiones y simular un grupo de conexiones. Para más detalles, consulte <https://stackoverflow.com/questions/35183713/what-is-the-correct-way-to-pool-the-phoenix-query-server-connections>.

## 1.7 ¿Apoya MRS el cambio del segmento de la red?

### Pregunta:

¿Apoya MRS el cambio del segmento de la red?

### Respuesta:

MRS soporta el cambio de segmento de red.

1. En la página de detalles del clúster, haga clic en Switch Subnet a la derecha de Default Effective Subnet.
2. Seleccione otra subred en la VPC a la que pertenece el clúster actual para expandir las direcciones IP de subred disponibles.

La selección de una nueva subred no cambiará las direcciones IP y las subredes de los nodos existentes.

## 1.8 ¿Puedo degradar las especificaciones de un nodo de clúster MRS?

### Pregunta:

¿Puedo degradar las especificaciones de un nodo de clúster MRS?

### Respuesta:

Actualmente, MRS no admite la configuración de degradación.

## 1.9 ¿Cuál es la relación entre Hive y otros componentes?

- Hive y HDFS  
Hive es un proyecto de Apache Hadoop. Hive utiliza Hadoop Distributed File System (HDFS) como su sistema de almacenamiento de archivos. Hive analiza y procesa los datos estructurados almacenados en HDFS. Todos los archivos de datos de la base de datos Hive se almacenan en HDFS, y todas las operaciones de datos en Hive también se realizan mediante API de HDFS.
- Hive y MapReduce  
Todo el cálculo de datos de Hive depende de MapReduce. MapReduce, también un proyecto Apache Hadoop, es un marco de computación paralela basado en HDFS. Durante el análisis de datos, Hive analiza las sentencias HiveQL enviadas por los usuarios en tareas de MapReduce y envía las tareas para que MapReduce lo ejecute.
- Hive y DBService  
MetaStore (servicio de metadatos) de Hive procesa la estructura y la información de atributos sobre las bases de datos, tablas y particiones de Hive que se almacenan en una base de datos relacional. En MRS, DBService mantiene la base de datos relacional.

- Hive y Spark

La computación de datos Hive también se puede implementar en Spark. Spark, también un proyecto de Apache, es un marco de computación distribuida en memoria. Durante el análisis de datos, Hive analiza las sentencias de HiveQL enviadas por los usuarios en las tareas de Spark y envía las tareas para que Spark las ejecute.

## 1.10 ¿Un clúster de MRS soporta Hive en Spark?

### Pregunta:

¿Un clúster de MRS soporta Hive en Spark?

### Respuesta:

- Los clústeres de MRS 1.9.x soporta Hive on Spark.
- Los clústeres de MRS 3 x o posterior admiten Hive on Spark.
- Puede usar Hive on Tez para los clústeres de otras versiones.

## 1.11 ¿Cuáles son las diferencias entre las versiones de Hive?

Hive 3.1 tiene las siguientes diferencias en comparación con Hive 1.2:

- String no se puede convertir a int.
- Las funciones definidas por el usuario (UDF) del tipo **Date** se cambian a UDF integradas en Hive.
- Hive 3.1 ya no proporciona la función de índice.
- Hive 3.1 utiliza las funciones de hora UTC, mientras que Hive 1.2 utiliza la zona horaria local.
- Los controladores JDBC en Hive 3.1 y Hive 1.2 son incompatibles.
- En Hive 3.1, los nombres de columna de los archivos ORC distinguen entre mayúsculas y minúsculas y subrayados.
- Hive 3.1 no permite columnas denominadas **time**.

## 1.12 ¿Qué versión de clúster de MRS admite la conexión Hive y la sincronización de usuarios?

### Pregunta:

¿Qué versión de clúster de MRS admite la conexión Hive y la sincronización de usuarios?

### Respuesta:

MRS 2.0.5 o posterior admite conexiones Hive en DataArts Studio y proporciona la función de sincronización de usuario IAM.

## 1.13 ¿Cuáles son las diferencias entre OBS y HDFS en el almacenamiento de datos?

Los datos procesados por MRS son de OBS o HDFS. OBS es un servicio de almacenamiento basado en objetos que proporciona almacenamiento seguro, confiable y rentable de grandes cantidades de datos. MRS puede procesar datos directamente en OBS. Puede ver, gestionar y usar datos mediante la consola OBS o el cliente OBS. Además, puede utilizar las API de REST de forma independiente o integrar las API en aplicaciones de servicio para gestionar y acceder a los datos.

- Datos almacenados en OBS: El almacenamiento de datos se desacopla de la computación. El costo de almacenamiento del clúster es bajo y la capacidad de almacenamiento no está limitada. Los clústeres se pueden eliminar en cualquier momento. Sin embargo, el rendimiento informático depende del rendimiento de acceso de OBS y es menor que el de HDFS. OBS se recomienda para aplicaciones que no requieren mucho cómputo.
- Datos almacenados en HDFS: El almacenamiento de datos no está desacoplado del proceso. El costo de almacenamiento en clúster es alto y la capacidad de almacenamiento es limitada. El rendimiento informático es alto. Debe exportar datos antes de eliminar clústeres. HDFS se recomienda para escenarios de computación intensiva.

## 1.14 ¿Cómo obtengo la herramienta de prueba de presión de Hadoop?

### Pregunta:

¿Cómo obtengo la herramienta de prueba de presión de Hadoop?

### Respuesta:

Para obtener la herramienta de prueba de presión de Hadoop, visite <https://github.com/Intel-bigdata/HiBench>.

## 1.15 ¿Cuál es la relación entre Impala y otros componentes?

- Impala y HDFS  
Impala utiliza HDFS como su sistema de almacenamiento de archivos. Impala analiza y procesa datos estructurados, mientras que HDFS proporciona almacenamiento subyacente confiable. Impala proporciona acceso rápido a los datos sin mover los datos en HDFS.
- Impala y Hive  
Impala utiliza metadatos Hive, el controlador Open Database Connectivity (ODBC) y la sintaxis SQL. A diferencia de Hive, que es sobre MapReduce, Impala implementa una arquitectura distribuida basada en daemon y maneja todas las ejecuciones de consultas en



el mismo nodo. Por lo tanto, Impala es más rápido que Hive al reducir la latencia causada por MapReduce.

- Impala y MapReduce  
Ninguna
- Impala y Spark  
Ninguna
- Impala y Kudu  
Kudu se puede integrar estrechamente con Impala para reemplazar la combinación de Impala, HDFS y Parquet. Puede insertar, consultar, actualizar y eliminar datos en tabletas Kudu utilizando la sintaxis SQL de Impala. Además, puede usar JDBC u ODBC para conectarse a Kudu para operaciones de datos, usando Impala como el broker.
- Impala y HBase  
Las tablas Impala predeterminadas utilizan archivos de datos almacenados en HDFS, lo que es ideal para la carga por lotes y consulta de escaneo de tablas completas. Sin embargo, HBase proporciona consultas convenientes y eficientes de datos de organización de estilo OLTP.

## 1.16 Declaración sobre las direcciones IP públicas en el SDK de terceros de código abierto integrado por MRS

### Pregunta:

¿Qué es la declaración de dirección IP pública contenida en el SDK de terceros de código abierto integrado con MRS?

### Respuesta:

Los paquetes de terceros de código abierto de los que dependen los componentes de código abierto integrados por MRS contienen ejemplos de uso del SDK. Direcciones IP públicas como 12.1.2.3, 54.123.4.56, 203.0.113.0 y 203.0.113.12 son direcciones IP de ejemplo. MRS no iniciará una conexión a la dirección IP pública ni intercambiará datos con la dirección IP pública.

## 1.17 ¿Cuál es la relación entre Kudu y HBase?

Kudu está diseñado basándose en la estructura de HBase y puede implementar funciones rápidas de lectura/escritura aleatoria y actualización que HBase hace bien. Kudu y HBase son similares en arquitectura. A continuación, se detallan las diferencias:

- HBase utiliza ZooKeeper para garantizar la consistencia de los datos, mientras que Kudu utiliza el algoritmo de consenso de Raft para garantizar la consistencia.
- HBase utiliza HDFS para el almacenamiento de datos resistente, mientras que Kudu utiliza TServer para garantizar una sólida consistencia y confiabilidad de los datos.

## 1.18 ¿Admite MRS la ejecución de Hive en Kudu?

### Pregunta:

¿Admite MRS la ejecución de Hive en Kudu?

### Respuesta:

MRS no soporta Hive en Kudu.

Actualmente, MRS solo admite los dos métodos siguientes para acceder a Kudu:

- Acceso a Kudu a través de tablas de Impala.
- Acceda y opere las tablas Kudu utilizando la aplicación de cliente.

## 1.19 ¿Cuáles son las soluciones para procesar mil millones de registros de datos?

### Pregunta:

¿Cuáles son las soluciones para procesar mil millones de registros de datos?

### Respuesta:

- GaussDB(for MySQL) se recomienda para escenarios, como actualizaciones de datos, procesamiento de transacciones en línea (OLTP) y análisis complejo de mil millones de registros de datos.
- Impala y Kudu en MRS también cumplen con este requisito. Impala y Kudu pueden cargar todas las tablas de join a la memoria en la operación de join.

## 1.20 ¿Puedo cambiar la dirección IP de DBService?

### Pregunta:

¿Puedo cambiar la dirección IP de DBService?

### Respuesta:

MRS no admite el cambio de la dirección IP de DBService.

## 1.21 ¿Puedo borrar MRS sudo Logs?

### Pregunta:

¿Puedo borrar MRS sudo Logs?

## Respuesta:

Los archivos de MRS sudo log registran las operaciones realizadas por el usuario **omm** y son útiles para la localización de fallas.

Puede eliminar los registros de la fecha más temprana para liberar espacio de almacenamiento.

1. Si el archivo de registro es grande, agregue el directorio del archivo de registro a **/etc/logrotate.d/syslog** para permitir que el sistema elimine periódicamente registros.  
Método: Ejecute **sed -i '3 a/var/log/sudo/sudo.log' /etc/logrotate.d/syslog**.
2. Establezca el número máximo y el tamaño de los registros de **/etc/logrotate.d/syslog**. Si el número o el tamaño de los registros supera el umbral, los registros se eliminarán automáticamente. De forma predeterminada, los registros son antiguos según el tamaño y el número de registros archivados. Puede usar **size** y **rotate** para limitar el tamaño y el número de registros archivados, respectivamente. Si es necesario, también puede agregar **daily/weekly/monthly** para especificar la frecuencia con la que se borran los registros.

## 1.22 ¿Cuáles son las restricciones en el tamaño del registro de Storm en un clúster MRS 2.1.0?

### Pregunta:

¿Cuáles son las restricciones en el tamaño del registro de Storm en un clúster MRS 2.1.0?

### Respuesta:

En MRS 2.1.0, el tamaño de los registros de Storm no puede exceder los 20 GB. Si el tamaño de los registros de Storm supera los 20 GB, los registros se eliminarán cíclicamente.

Los registros se almacenan en el disco del sistema y el espacio es limitado. Si los registros necesitan almacenarse durante mucho tiempo, monte los registros.

## 1.23 ¿Qué es Spark ThriftServer?

### Pregunta:

¿Qué es Spark ThriftServer? ¿Para qué se usan las redes sociales?

### Respuesta:

ThriftServer es una interfaz JDBC.

Puede usar JDBC para conectarse a ThriftServer para acceder a los datos de SparkSQL. Por lo tanto, puede ver JDBCServer en los componentes de Spark, pero no en ThriftServer.

## 1.24 ¿Qué protocolos de acceso admite Kafka?

### Pregunta:

¿Qué protocolos de acceso admite Kafka?

### Respuesta:

Kafka admite los siguientes protocolos de acceso: PLAINTEXT, SSL, SASL\_PLAINTEXT y SASL\_SSL.

## 1.25 Se notifica el error 408 cuando un nodo MRS accede a OBS (Palabra clave, no enviada a la nube operada conjuntamente)

### Pregunta:

¿Qué puedo hacer si se notifica el error 408 cuando un nodo MRS accede a OBS?

### Respuesta:

Si se notifica el error 408 cuando un nodo MRS accede a OBS, es necesario cambiar el nombre de dominio de OBS.

Cambie el nombre de dominio OBS al nombre de dominio sufijo con myhuaweicloud.com.

## 1.26 ¿Cuáles son las ventajas de la relación de compresión de zstd?

### Pregunta:

¿Cuáles son las ventajas de la relación de compresión de zstd?

### Respuesta:

La relación de compresión de zstd es dos veces la de orco. zstd es de código abierto.

Para obtener más información, véase <https://github.com/L-Angel/compress-demo>.

CarbonData no soporta lzo. MRS integra zstd.

## 1.27 ¿Por qué los componentes HDFS, YARN y MapReduce no están disponibles cuando se compra un clúster de MRS?

### Pregunta:

¿Por qué no se pueden encontrar HDFS, Yarn y MapReduce al comprar un clúster MRS?

### Respuesta:

Los componentes HDFS, YARN y MapReduce están integrados en Hadoop. Si los tres componentes no están disponibles cuando se compra el clúster MRS, seleccione Hadoop en su lugar. Después de crear un clúster MRS, HDFS, YARN y MapReduce están disponibles en la página **Components**.

## 1.28 ¿Por qué no está disponible el componente ZooKeeper cuando se compra un clúster MRS?

### Pregunta:

¿Por qué no puedo encontrar el componente ZooKeeper al comprar un clúster MRS?

### Respuesta:

- Si crea un clúster de una versión anterior a MRS 3.x, se instala ZooKeeper de forma predeterminada y no se muestra en la GUI.  
Una vez creado el clúster, el componente ZooKeeper estará disponible en la página **Components**.
- Si crea un clúster de MRS 3.x o posterior, el ZooKeeper está disponible en la GUI y está seleccionado de forma predeterminada.

## 1.29 ¿Qué versiones de Python son compatibles con las tareas de Spark en clústeres MRS 3.1.0?

### Pregunta:

¿Qué versiones de Python son compatibles con las tareas de Spark en clústeres MRS 3.1.0?

### Respuesta:

Para los clústeres MRS 3.1.0, se recomienda Python 2.7 o 3.x para las tareas de Spark.

## 1.30 ¿Cómo puedo habilitar diferentes programas de servicio para usar diferentes colas de YARN?

### Pregunta:

¿Cómo puedo habilitar diferentes programas de servicio para usar diferentes colas de YARN?


### Respuesta:

Cree un tenant en Manager.

Para obtener más información, consulte [Adición de un Sub-Tenant](#).

### Procedimiento

**Paso 1** Inicie sesión en el FusionInsight Manager y elija **Tenant Resources**.

**Paso 2** En la lista de tenants de la izquierda, seleccione un tenant principal y haga clic en . En la página para agregar un subtenant, establezca atributos para el subtenant según [Tabla 1-2](#).

**Tabla 1-2** Parámetros de subtenant

Parámetro	Descripción
Cluster	Indica el clúster al que pertenece el tenant principal.
Parent Tenant Resource	Indica el nombre del tenant principal.
Name	<ul style="list-style-type: none"><li>● Indica el nombre del tenant actual. El valor consta de 3 a 50 caracteres, incluidos dígitos, letras y guiones bajos (_).</li><li>● Planifique un nombre de subtenant basado en los requisitos de servicio. El nombre no puede ser el mismo que el de un rol, directorio HDFS o cola de Yarn que existe en el clúster actual.</li></ul>
Tenant Type	Especifica si el tenant es un tenant de hoja. <ul style="list-style-type: none"><li>● Cuando se selecciona <b>Leaf Tenant</b>, el tenant actual es un tenant hoja y no se puede agregar ningún subtenant.</li><li>● Cuando se selecciona <b>Non-leaf Tenant</b>, el tenant actual no es un tenant hoja y se pueden agregar subtenants al tenant actual. Sin embargo, la profundidad del tenant no puede exceder de 5 niveles.</li></ul>

Parámetro	Descripción
Computing Resource	<p>Especifica los recursos de cálculo dinámicos para el tenant actual.</p> <ul style="list-style-type: none"> <li>● Cuando se selecciona <b>Yarn</b>, el sistema crea automáticamente una cola en Yarn y la cola recibe el mismo nombre que el nombre del subtenant. <ul style="list-style-type: none"> <li>– Un tenant de hoja puede enviar trabajos directamente a la cola.</li> <li>– Un tenant que no sea de hoja no puede enviar trabajos directamente a la cola. Sin embargo, Yarn agrega una cola adicional (oculta) llamada <b>default</b> para que el tenant no hoja registre la capacidad de recursos restante del tenant. Los trabajos reales no se ejecutan en esta cola.</li> </ul> </li> <li>● Si <b>Yarn</b> no está seleccionado, el sistema no crea automáticamente una cola.</li> </ul>
Default Resource Pool Capacity (%)	Indica el porcentaje de recursos informáticos utilizados por el tenant actual. El valor base es el total de recursos del tenant principal.
Default Resource Pool Max Capacity (%)	Indica el porcentaje máximo de recursos informáticos utilizados por el tenant actual. El valor base es el total de recursos del tenant principal.
Storage Resource	<p>Especifica los recursos de almacenamiento para el tenant actual.</p> <ul style="list-style-type: none"> <li>● Cuando se selecciona <b>HDFS</b>, el sistema crea automáticamente una carpeta con el nombre del subtenant en el directorio de tenant principal de HDFS.</li> <li>● Cuando <b>HDFS</b> no está seleccionado, el sistema no asigna automáticamente los recursos de almacenamiento.</li> </ul>
Quota	Indica la cuota de archivos y directorios.
Space Quota	<p>Indica la cuota para el espacio de almacenamiento HDFS utilizado por el tenant actual.</p> <ul style="list-style-type: none"> <li>● Si la unidad está establecida en <b>MB</b>, el valor varía entre <b>1</b> y <b>879609302208</b>. Si la unidad está establecida en <b>GB</b>, el valor varía entre <b>1</b> y <b>8589934592</b>.</li> <li>● Este parámetro indica el espacio de almacenamiento HDFS máximo que puede utilizar el tenant, pero no el espacio real utilizado.</li> <li>● Si su valor es mayor que el tamaño del disco físico HDFS, el espacio máximo disponible es el espacio completo del disco físico HDFS.</li> <li>● Si esta cuota es mayor que la cuota del tenant principal, el espacio de almacenamiento real no excede la cuota del tenant principal.</li> </ul>

Parámetro	Descripción
Storage Path	<p>Indica el directorio de almacenamiento de HDFS para el tenant.</p> <ul style="list-style-type: none"> <li>● El sistema crea automáticamente una carpeta con el nombre del subtenant en el directorio del tenant principal de forma predeterminada. Por ejemplo, si el subtenant es <b>ta1s</b> y el directorio principal es <b>/tenant/ta1</b>, la ruta de almacenamiento para el subtenant es entonces <b>/tenant/ta1/ta1s</b>.</li> <li>● La ruta de almacenamiento se puede personalizar en el directorio principal.</li> </ul>
Description	Indica la descripción del tenant actual.

 **NOTA**

Los roles, los recursos informáticos y los recursos de almacenamiento se crean automáticamente cuando se crean los tenants.

- El nuevo rol tiene permisos sobre los recursos informáticos y de almacenamiento. Este rol y sus permisos son controlados automáticamente por el sistema y no se pueden gestionar manualmente eligiendo **System > Permission > Role**. El nombre del rol tiene el formato de *Tenant name\_Cluster ID*. El ID del primer clúster no se muestra de forma predeterminada.
- Cuando utilice este tenant, cree un usuario del sistema y vincule al usuario al rol del tenant. Para obtener más información, consulte [Agregar un usuario y vincularlo a un rol de tenant](#).
- El subtenant puede asignar además los recursos de su tenant principal. La suma de los porcentajes de recursos de subtenants directos bajo un tenant principal en cada nivel no puede exceder 100%. La suma de los porcentajes de recursos de computación de todos los tenants de nivel 1 no puede exceder el 100%.

**Paso 3** Compruebe si el tenant actual necesita estar asociado con recursos de otros servicios.

- En caso afirmativo, vaya a [Paso 4](#).
- Si no, vaya a [Paso 5](#).

**Paso 4** Haga clic en **Associate Service** para configurar otros recursos de servicio utilizados por el tenant actual.

1. Ajusta **Services** a **HBase**.
2. Establezca **Association Type** de la siguiente manera:
  - **Exclusive** indica que los recursos de servicio son utilizados exclusivamente por el tenant y no pueden asociarse con otros tenants.
  - **Shared** indica que los recursos de servicio se pueden compartir con otros tenants.



 **NOTA**

- Solo HBase puede asociarse con un nuevo tenant. Sin embargo, HDFS, HBase y Yarn pueden asociarse con tenants existentes.
  - Para asociar un tenant existente a los recursos de servicio, haga clic en el tenant de destino en la lista de tenant, cambie a la página **Service Associations** y haga clic en **Associate Service** para configurar los recursos que se asociarán al tenant.
  - Para desasociar un tenant existente de los recursos de servicio, haga clic en el tenant de destino en la lista de tenant, cambie a la página **Service Associations** y haga clic en **Delete** en la columna **Operation**. En el cuadro de diálogo que se muestra, seleccione **I have read the information and understand the impact** y haga clic en **OK**.
3. Haga clic en **OK**.

**Paso 5** Haga clic en **OK**. Espere hasta que el sistema muestre un mensaje que indique que el tenant se ha creado correctamente.

----Fin

## 1.31 Diferencias y relaciones entre la consola de gestión de MRS y el Manager de clústeres

### Pregunta:

¿Cuáles son las diferencias y relaciones entre la consola de gestión de MRS y FusionInsight Manager?

### Respuesta:

Puede acceder a Manager desde la consola de gestión de MRS.

Manager se clasifica como MRS Manager y FusionInsight Manager.

- MRS Manager es la página de manager de MRS 2.x o clústeres anteriores.
- FusionInsight Manager es la página de manager de clústeres MRS 3.x o posteriores.

En la siguiente tabla se enumeran las diferencias y relaciones entre la consola de gestión y FusionInsight Manager.

Operaciones comunes	MRS Console	FusionInsight Manager
Cambio de subredes, adición de reglas de grupo de seguridad, control de permisos OBS, gestión de agencias y sincronización de usuarios de IAM	Soportado	No soportado
Adición de grupos de nodos, ampliación, reducción y actualización de especificaciones	Soportado	No soportado

<b>Operaciones comunes</b>	<b>MRS Console</b>	<b>FusionInsight Manager</b>
Aislar hosts, iniciar todos los roles y detener todos los roles	Soportado	Soportado
Descargar el cliente, iniciar servicios, detener servicios y realizar reinicio continuo de servicios	Soportado	Soportado
Consulta del estado de instancia de los servicios, configuración de parámetros y sincronización de configuraciones	Soportado	Soportado
Consulta de alarmas y eventos borrados	Soportado	Soportado
Consulta de la ayuda de alarma	No soportado	Soportado
Establecer umbrales	No soportado	Soportado
Adición de especificaciones de suscripción de mensajes	Soportado	No soportado
Gestión de archivos	Soportado	No soportado
Gestión de trabajos	Soportado	No soportado
Gestión de tenants	Soportado	Soportado
Gestión de etiquetas	Soportado	No soportado
Gestión de permisos (añadir y eliminar usuarios, grupos de usuarios y roles, y cambiar contraseñas)	No soportado	Soportado
Realización de copias de respaldo y restauración	No soportado	Soportado
Auditoría	No soportado	Soportado
Monitoreo de recursos y registro	Soportado	Soportado

## 1.32 ¿Cómo desvinculo una EIP del FusionInsight Manager de un clúster MRS?

### Pregunta:

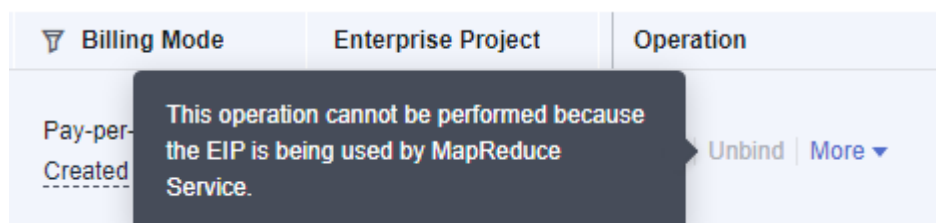
¿Cómo desvinculo una EIP del FusionInsight Manager de un clúster MRS?

### Respuesta:

Después de que una EIP esté vinculada al FusionInsight Manager, no la desvincule. De lo contrario, es posible que otros usuarios no puedan acceder al FusionInsight Manager.

#### NOTA

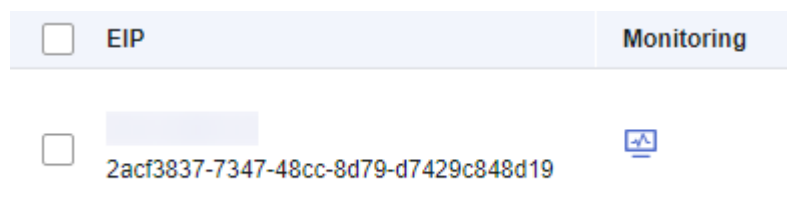
Una vez que una EIP está vinculado al FusionInsight Manager en la consola de MRS, la EIP no se puede separar del FusionInsight Manager en la página **EIPs**.



Se muestra un mensaje que indica que esta operación no se puede realizar porque la EIP está siendo utilizado por MapReduce Service.

Para desvincular una EIP, invoque a la API relacionada del servicio EIP.

**Paso 1** Inicie sesión en la consola de gestión de EIP. En la lista EIP, vea y registre el ID de la EIP deseada.



**Paso 2** Unbind the EIP from API Explorer by referring to [Actualización de una EIP](#).

Por ejemplo, inicie sesión en API Explorer como usuario actual, seleccione una región en función de los requisitos del sitio y establezca **project\_id** en el ID del proyecto de la región actual y **publicip\_id** en el ID de la EIP en el que se va a consultar de forma independiente en [Paso 1](#).

**Paso 3** Una vez que se haya invocada correctamente a la API, inicie sesión en la consola de gestión de EIP. En la lista EIP, puede ver que el EIP está en el estado **Unbound**.

**Paso 4** Inicie sesión en la consola de gestión de MRS, vaya a la pestaña **Dashboard** del clúster y vincule otra EIP al clúster.

----Fin

## 1.33 ¿Cuáles son los sistemas operativos de hosts en clústeres MRS de diferentes versiones?

### Pregunta:

¿Cuáles son los sistemas operativos de hosts en clústeres MRS de diferentes versiones?

### Respuesta:

El sistema operativo de host varía según la versión del clúster. [Tabla 1](#) muestra la relación de asignación.

**Tabla 1-3** Asignaciones entre las versiones de clúster y los sistemas operativos de host

Versión del clúster	x86	Kunpeng (Arm)
MRS 3.2.0-LTS.1	EulerOS 2.10	EulerOS 2.10
MRS 3.1.5	EulerOS 2.9	EulerOS 2.9
MRS 3.1.2-LTS.3	EulerOS 2.9	EulerOS 2.9
MRS 3.1.0	EulerOS 2.5	EulerOS 2.8
MRS 1.9.2	EulerOS 2.2	EulerOS 2.8

# 2 Facturación

---

## 2.1 ¿Cómo se factura MRS?

### Pregunta:

¿Cómo se factura MRS?

### Respuesta:

MRS admite dos modos de facturación: suscripción anual o mensual y pago por uso. Puede calcular su factura utilizando la [Calculadora de precios](#).

- Suscripción anual o mensual: Usted paga por el clúster por año o mes, por adelantado. La duración mínima de uso es de un mes y la duración máxima de uso es de un año.
- Pago por uso: un modo de facturación pospago. Los nodos se cobran por la duración real del uso, con un ciclo de facturación de una hora.

#### NOTA

- Los clústeres en el estado **Starting**, **Failed** o **Terminated** no se cargarán.
- Cuando compra un clúster MRS, el precio que se muestra es solo por el costo del clúster. El uso para el almacenamiento de datos, el ancho de banda y el tráfico en MRS se facturan por separado.

## 2.2 ¿Por qué no se muestra el precio durante la creación del clúster MRS?

### Pregunta:

¿Por qué no se muestra el precio durante la creación del clúster MRS?

### Respuesta:

Si selecciona solo el número de discos, pero no el número de instancias de VM de clúster, no se cobrará la cotización del producto.

La cotización del producto se muestra en la parte inferior del navegador solo después de que el cliente selecciona el número de instancias.

## 2.3 ¿Cómo se factura el escalado automático de un clúster MRS?

### Pregunta:

¿Cómo se factura el escalado automático de un clúster MRS?

### Respuesta:

El precio que se muestra en la parte inferior al comprar un clúster MRS no incluirá la tarifa de escalado automático si especifica solo el rango de escalado automático para los nodos Task sin configurar el recuento de instancias. Si se configura el número de instancias, la cotización del producto se muestra en la parte inferior del navegador.

Si agrega nodos a través de la función de escalado automático, los nodos agregados se facturarán según la duración de uso real por hora independientemente de si el modo de facturación del clúster es anual/mensual o el modo de pago por uso.

## 2.4 ¿Cómo se renueva MRS?

### Pregunta:

¿Cómo se renueva MRS?

### Respuesta:

MRS ofrece dos modos de facturación: pago por uso y suscripción anual/mensual. En el modo de pago por uso, la facturación es por hora y un saldo insuficiente puede causar pagos atrasados. Para una suscripción anual/mensual, debe renovar sus recursos antes de que caduquen. De lo contrario, se le asignará un período de retención a sus recursos. Durante este período, los datos se conservarán, pero los clústeres MRS se detendrán.

Para renovar la suscripción, vaya a la página [Renewals](#).

## 2.5 ¿Cómo se factura el nodo de tareas en un clúster MRS?

### Pregunta:

¿Cómo se cargan los nodos de tarea en el clúster de MapReduce Service (MRS)?

### Respuesta:

Los nodos de tarea de un clúster se facturan en modo de pago por uso, independientemente del modo de facturación del clúster. Es decir, se facturan por hora en función de la duración real del uso.

## 2.6 ¿Por qué falla mi cancelación de la suscripción a ECS después de cancelar mi suscripción a MRS?

### Pregunta:

¿Por qué falla mi cancelación de la suscripción a ECS después de cancelar mi suscripción a MRS?

### Respuesta:

1. Compruebe que el ID de ECS no es utilizado por un clúster MRS en uso.
2. En la consola de ECS, busque el ECS del que desea cancelar la suscripción y haga clic en **Locked by MRS** para desbloquearlo.
3. Haga clic en **Unsubscribe** de nuevo.
4. Si la cancelación de la suscripción sigue fallando, recopila el ECS ID y póngase en contacto con el soporte técnico de Huawei Cloud.

# 3 Cuenta y contraseña

---

## 3.1 ¿Qué es la cuenta para iniciar sesión en Manager?

### Pregunta:

¿Qué es la cuenta para iniciar sesión en Manager?

### Respuesta:

La cuenta predeterminada para iniciar sesión en Manager es **admin** y la contraseña es la que configuraste al crear el clúster.

## 3.2 ¿Cómo puedo consultar y cambiar el período de validez de la contraseña de una cuenta?

### Consultar el período de validez de la contraseña

**Consulta del periodo de validez de la contraseña de un usuario que ejecuta un componente (usuario humano-máquina o usuario máquina-máquina):**

**Paso 1** Inicie sesión en el nodo donde está instalado el cliente como usuario de instalación del cliente.

**Paso 2** Ejecute el siguiente comando para cambiar al directorio del cliente, por ejemplo, **/opt/Bigdata/client**:

```
cd /opt/Bigdata/client
```

**Paso 3** Ejecute el siguiente comando para configurar las variables de entorno:

```
source bigdata_env
```

**Paso 4** Ejecute el siguiente comando e introduzca la contraseña del usuario **kadmin/admin** para iniciar sesión en la consola de kadmin:

```
kadmin -p kadmin/admin
```



 **NOTA**

La contraseña predeterminada del **kadmin/admin** de usuario es **Admin@123**. Cambie la contraseña en su primer inicio de sesión o como se le solicite y mantenga la nueva contraseña segura.

**Paso 5** Ejecute el siguiente comando para ver la información del usuario:

**getprinc** *Internal system username*

Ejemplo: **getprinc user1**

```
kadmin: getprinc user1
.....
Expiration date: [never]
Last password change: Sun Oct 09 15:29:54 CST 2022
Password expiration date: [never]
.....
```

----Fin

**Consulta del periodo de validez de la contraseña de un usuario del sistema operativo:**

**Paso 1** Inicie sesión en cualquier nodo master del clúster como usuario **root**.

**Paso 2** Ejecute el siguiente comando para ver el período de validez de la contraseña (valor de **Password expires**):

**chage -l Username**

Por ejemplo, para ver el período de validez de la contraseña del usuario **root**, ejecute el comando **chage -l root**. La salida de comandos es la siguiente:

```
[root@xxx ~]#chage -l root
Last password change           : Sep 12, 2021
Password expires             : never
Password inactive              : never
Account expires                : never
Minimum number of days between password change : 0
Maximum number of days between password change : 99999
Number of days of warning before password expires : 7
```

----Fin

## Cambio del período de validez de la contraseña

- La contraseña de un usuario máquina-máquina se genera aleatoriamente y nunca caduca de forma predeterminada.
- El período de validez de la contraseña de un usuario humano-máquina se puede cambiar modificando la política de contraseñas en Manager.

Las operaciones son las siguientes:

**Versiones anteriores a MRS 3.x:** Cambie el período de validez de la contraseña de un usuario humano-máquina haciendo referencia a [Modificación de una política de contraseña](#).

**MRS 3.x o posterior:** Cambie el período de validez de la contraseña de un usuario humano-máquina consultando [Configuración de políticas de contraseña](#).

# 4 Cuenta y permiso

---

## 4.1 ¿Es compatible un clúster MRS con el control de permisos de acceso si la autenticación Kerberos no está activada?

Los clústeres MRS con versiones anteriores a 1.8.0 no admiten el control de permisos de acceso si la autenticación Kerberos no está habilitada. Los clústeres MRS con la versión 1.8.0 o posterior admiten la gestión de roles y permisos.

Para el clúster MRS 2.1.0 o anterior, elija **System > Configuration > Permission** en MRS Manager.

Para el clúster MRS 3.x o posterior, elija **System > Permission** en el FusionInsight Manager.

Para obtener más información acerca de cómo configurar permisos, consulte [Gestión de permiso](#).

Para obtener más información acerca de cómo utilizar clústeres con autenticación Kerberos habilitada, consulte [Clústeres con autenticación Kerberos activada](#).

## 4.2 ¿Cómo asigno permiso de gestión de tenant a una cuenta nueva?

Puede asignar permisos de gestión de tenant solo en clústeres de análisis o híbridos, pero no en clústeres de streaming.

Las operaciones varían según la versión del clúster MRS:

**Procedimiento para versiones anteriores a MRS cluster 3.x:**

**Paso 1** Inicie sesión en MRS Manager como usuario **admin**.

**Paso 2** Elija **System > Manage User**. Seleccione la nueva cuenta y haga clic en **Modify** en la columna **Operation**.

**Paso 3** En **Assign Rights by Role**, haga clic en **Select and Add Role**.

- Si vincula el rol **Manager\_tenant** a la cuenta, la cuenta tendrá permiso para ver la información de gestión del tenant.
- Si vincula el rol **Manager\_administrator** a la cuenta, la cuenta tendrá permiso para realizar la gestión del inquilino y ver información relacionada.

**Paso 4** Haga clic en **OK**.

----Fin

#### Procedimiento para MRS cluster 3.x y versiones posteriores:

**Paso 1** Inicie sesión en FusionInsight Manager y elija **System > Permission > User**.

**Paso 2** Localice el usuario y haga clic en **Modify**.

Modifique los parámetros según los requisitos de servicio.

Si vincula el rol **Manager\_tenant** a la cuenta, la cuenta tendrá permiso para ver la información de gestión del tenant. Si vincula el rol **Manager\_administrator** a la cuenta, la cuenta tendrá permiso para realizar la gestión del inquilino y ver información relacionada.

#### **NOTA**

Se tarda unos tres minutos para que la configuración tenga efecto después de que se modifique el permiso de grupo o rol de usuario.

**Paso 3** Haga clic en **OK**.

----Fin

## 4.3 ¿Cómo personalizo una política de MRS?

### Pregunta:

¿Cómo personalizo una política de MRS?

### Respuesta:

1. En la consola de IAM, elija **Permissions** en el panel de navegación y haga clic en **Create Custom Policy**.
2. Establezca un nombre de política en **Policy Name**.
3. Establezca **Scope** en **Project-level service** para MRS.
4. Especifique **Policy View**. Se admiten las siguientes opciones:
  - **Visual editor**: Seleccione las condiciones de solicitud, acciones, recursos y servicios en la nube en el panel de navegación para personalizar la política. No se requiere conocimiento de la sintaxis JSON.
  - **JSON**: Edite las políticas JSON desde cero o basándose en una política existente.También puede hacer clic en **Select Existing Policy/Role** en el área **Policy Content** para seleccionar una política existente como plantilla para la modificación.
5. (Opcional) Introduzca una breve descripción en el área **Description**.
6. Haga clic en **OK**.
7. Adjunte la política a un grupo de usuarios. Los usuarios del grupo heredan los permisos definidos en esta política.

 NOTA

Para obtener más información, consulte [Creación de una política personalizada](#).

## 4.4 ¿Por qué no puedo encontrar la gestión de usuarios en la configuración del sistema en MRS Manager?

### Pregunta:

¿Por qué no puedo encontrar la gestión de usuarios en la configuración del sistema en MRS Manager?

### Respuesta:

Compruebe si tiene el permiso **Manager\_administrator**. Si no tiene este permiso, **Manage User** no estará disponible en la página **System** de MRS Manager.

## 4.5 ¿Proporciona Hue la función de configurar los permisos de la cuenta?

### Pregunta:

¿Proporciona Hue la función de configurar los permisos de la cuenta?

### Respuesta:

El servicio Hue no proporciona la función de configurar permisos de cuenta.

Puede configurar los permisos de cuenta configurando los roles de usuario y los grupos de usuarios en System Settings en FusionInsight Manager. De esta manera, se pueden configurar los permisos de Hue.

## 4.6 ¿Por qué no puedo enviar trabajos en la consola después de que mi cuenta de IAM esté asignada con permisos relacionados?

### Pregunta:

No puedo enviar trabajos en la consola MRS después de que mi cuenta de IAM esté asignada con los permisos **MRS ReadOnlyAccess** y **MRS FullAccess**.

### Respuesta:

La cuenta tiene los permisos **MRS ReadOnlyAccess** y **MRS FullAccess**. Debido a la prioridad del permiso, la cuenta no puede agregar trabajos en la consola MRS.

El grupo al que pertenece la cuenta de IAM tiene los permisos **MRS FullAccess**, **MRS ReadOnlyAccess** y **MRS Administrator**. **MRS FullAccess** y **MRS ReadOnlyAccess** son

permisos detallados, y **MRS Administrator** es una política de control de acceso basado en roles (RBAC). Los permisos detallados tienen prioridades más altas sobre las políticas RBAC. Si se configuran los permisos detallados y las políticas RBAC, los permisos detallados tendrán efecto primero. Los permisos detallados siguen el principio de prioridad de denegación. Por lo tanto, el permiso **MRS ReadOnlyAccess** entra en vigor finalmente. Por lo tanto, se muestra un mensaje que indica que la cuenta no tiene el permiso.

Elimine el permiso **MRS ReadOnlyAccess**, cierre la sesión de la consola e inicie sesión de nuevo en la consola.

## 4.7 ¿Qué debo hacer si se notifica un error que indica una autenticación no válida cuando envío una orden de compra de clúster de MRS?

### Pregunta:

¿Qué puedo hacer si se notifica la autenticación no válida cuando envío un pedido para la compra de un clúster MRS?

### Respuesta:

Cuando presiona F12 para ver la solicitud de red, se indica el código de error 401 para los datos de interfaz de solicitud.

1. Comprueba los registros de solicitudes de la API de backend. La información de alarma es la siguiente: a los usuarios de IAM no se les ha asignado el permiso **mrs:cluster:create** en la política detallada.
2. Los usuarios de IAM del lado del cliente pertenecen a varios grupos de usuarios, y se asignan diferentes políticas de MRS predeterminadas a estos grupos de usuarios. Las políticas de bajos permisos se igualan preferentemente. Esta política no contiene el permiso **mrs:cluster:create**. Como resultado, no se puede enviar la operación de creación de clúster, y se informa del error 401.
3. Después de eliminar el usuario del grupo de usuarios correspondiente a la política de bajo permiso, se puede enviar el pedido de compra del clúster.

# 5 Uso del cliente

---

## 5.1 ¿Cómo configuro variables de entorno y ejecuto comandos en un cliente de componentes?

### Pregunta:

¿Cómo configuro variables de entorno y ejecuto comandos en un cliente de componentes?

### Respuesta:

1. Inicie sesión en cualquier nodo maestro como usuario **root**.
2. Ejecute el comando **su - omm** para cambiar a usuario **omm**.
3. Ejecute el comando **cd *Client installation directory*** para cambiar al cliente.
4. Ejecute el comando **source bigdata\_env** para configurar las variables de entorno.  
Si la autenticación Kerberos está habilitada para el clúster actual, ejecute el comando **kinit *Component service user*** para autenticar al usuario. Si la autenticación Kerberos está deshabilitada, omita este paso.
5. Una vez configuradas las variables de entorno, ejecute el comando de cliente del componente. Por ejemplo, para ver información de componentes, puede ejecutar el comando de cliente de HDFS **hdfs dfs -ls /** para ver el archivo de directorio raíz de HDFS.

## 5.2 ¿Cómo desactivo la autenticación de ZooKeeper SASL?

### Pregunta:

¿Cómo desactivo la autenticación de ZooKeeper SASL?

### Respuesta:

1. Inicie sesión en FusionInsight Manager.
2. Seleccione Cluster > Service > ZooKeeper > Configuration > All Configurations.

3. En el árbol de navegación de la izquierda, elija `quorumpeer > Customization and add zookeeper.sasl.disable = false`.  
Guarde la configuración y reinicie el servicio ZooKeeper.

## 5.3 Se notifica un error cuando se ejecuta el comando `kinit` en un nodo cliente fuera de un clúster MRS

### Síntoma

Después de instalar el cliente en un nodo fuera de un clúster MRS y ejecutar el comando `kinit`, se muestra la siguiente información de error:

```
-bash kinit Permission denied
```

Se muestra la siguiente información de error cuando se ejecuta el comando `java`:

```
-bash: /xxx/java: Permission denied
```

Después de ejecutar el comando `ll /Java installation path/JDK/jdk/bin/java`, se encuentra que el permiso de ejecución del archivo es correcto.

### Localización de fallas

Ejecute el comando `mount | column -t` para comprobar el estado de la partición montada. Se encuentra que el estado de la partición del punto de montaje donde se encuentra el archivo de ejecución de Java es `noexec`. En el entorno actual, el disco de datos donde está instalado el cliente MRS se establece en `noexec`, es decir, se prohíbe la ejecución de archivos binarios. Como resultado, los comandos de Java no se pueden ejecutar.

### Solución

1. Inicie sesión en el nodo donde se encuentra el cliente MRS como usuario `root`.
2. Quite el elemento de configuración `noexec` del disco de datos donde se encuentra el cliente MRS del archivo `/etc/fstab`.
3. Ejecute el comando `umount` para separar el disco de datos y, a continuación, ejecute el comando `mount -a` para volver a montar el disco de datos.

# 6 Acceso a páginas Web

---

## 6.1 ¿Cómo cambio la duración del tiempo de espera de la sesión para una interfaz de usuario web de componente de código abierto?

### Pregunta:

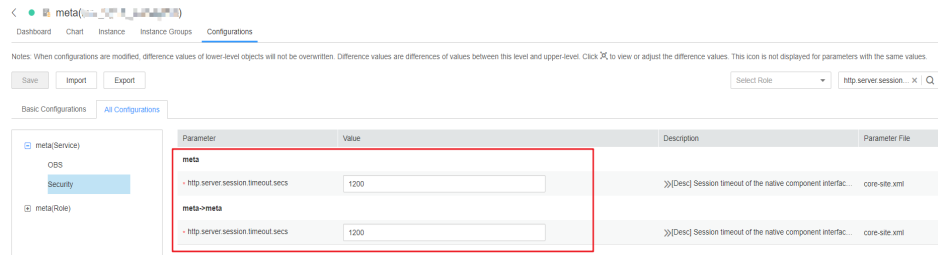
¿Cómo cambio el intervalo de tiempo de espera de sesión en la página Web de un componente de código abierto?

### Respuesta:

Un intervalo de tiempo de espera adecuado de la página web evita la fuga de información causada por la exposición prolongada de la página web.

1. Comprobación de si el clúster admite el ajuste de duración del tiempo de espera de sesión
  - Para versiones de clúster MRS anteriores a 3.x:
    - i. En la página de detalles del clúster, elija **Components > meta > Service Configuration**.
    - ii. Cambie de **Basic** a **All** y busque el **http.server.session.timeout.secs**. Si **http.server.session.timeout.secs** no existe, el clúster no admite el cambio de la duración del tiempo de espera de la sesión. Si el parámetro existe, realice los siguientes pasos para modificarlo.
  - MRS 3.x y posterior: Inicie sesión en FusionInsight Manager y elija **Cluster > Services > meta**. En la página mostrada, haga clic en **Configurations** y seleccione **All Configurations**. Busque el elemento de configuración **http.server.session.timeout.secs**. Si existe este elemento de configuración, realice los siguientes pasos para modificarlo. Si el elemento de configuración no existe, la versión no admite el ajuste dinámico de la duración de la sesión.





Se recomienda establecer todas las duraciones de tiempo de espera de sesión en el mismo valor. De lo contrario, es posible que la configuración de algunos parámetros no surtan efecto debido al conflicto de valores.

2. Modificación de la duración del tiempo de espera en Manager y la página Authentication Center
  - **Para clústeres de versiones anteriores a MRS 3.x:**
    - i. Inicie sesión en cada nodo master del clúster y realice [2.ii](#) a [2.iv](#).
    - ii. Cambie el valor de `<session-timeout>20</session-timeout>` en el archivo `/opt/Bigdata/apache-tomcat-7.0.78/webapps/cas/WEB-INF/web.xml`. `<session-timeout>20</session-timeout>` indica la duración del tiempo de espera de la sesión, en minutos. Cambiarlo en función de los requisitos de servicio. El valor máximo es de 480 minutos.
    - iii. Cambie el valor de `<session-timeout>20</session-timeout>` en el archivo `/opt/Bigdata/apache-tomcat-7.0.78/webapps/web/WEB-INF/web.xml`. `<session-timeout>20</session-timeout>` indica la duración del tiempo de espera de la sesión, en minutos. Cambiarlo en función de los requisitos de servicio. El valor máximo es de 480 minutos.
    - iv. Cambie los valores de `p:maxTimeToLiveInSeconds="{tgt.maxTimeToLiveInSeconds:1200}"` y `p:timeToKillInSeconds="{tgt.timeToKillInSeconds:1200}"` en el archivo `/opt/Bigdata/apache-tomcat-7.0.78/webapps/cas/WEB-INF/spring-configuration/ticketExpirationPolicies.xml`. El valor máximo es de 28,800 segundos.
    - v. Reinicie el nodo Tomcat en el nodo master activo.
      - 1) En el nodo master activo, ejecute el comando `netstat -anp |grep 28443 |grep LISTEN |awk '{print $7}'` como usuario `omm` para consultar el ID de proceso de Tomcat.
      - 2) Ejecute el comando `kill -9 {pid}`, en el que `{pid}` indica el ID de proceso de Tomcat obtenido en [2.v.1](#)).
      - 3) Espere hasta que el proceso se reinicie automáticamente. Puede ejecutar el comando `netstat -anp |grep 28443 |grep LISTEN` para comprobar si el proceso se reinicia correctamente. Si se puede consultar el proceso, el proceso se reinicia correctamente. Si no se puede consultar el proceso, vuelva a consultarlo más tarde.
  - **Para clústeres de MRS 3.x o posterior**
    - i. Inicie sesión en cada nodo master del clúster y realice [2.ii](#) a [2.iii](#) en cada nodo master.
    - ii. Cambie el valor de `<session-timeout>20</session-timeout>` en el archivo `/opt/Bigdata/om-server_XXX/apache-tomcat-XXX/webapps/web/WEB-INF/web.xml`. `<session-timeout>20</session-timeout>` indica la duración del tiempo de espera de la sesión, en minutos. Cambiarlo en función de los requisitos de servicio. El valor máximo es de 480 minutos.

- iii. Agregue **ticket.tgt.timeToKillInSeconds=28800** al archivo **/opt/Bigdata/om-server\_XXX/apache-tomcat-8.5.63/webapps/cas/WEB-INF/classes/config/application.properties**. **ticket.tgt.timeToKillInSeconds** indica el período de validez del centro de autenticación, en segundos. Cambiarlo en función de los requisitos de servicio. El valor máximo es de 28,800 segundos.
  - iv. Reinicie el nodo Tomcat en el nodo master activo.
    - 1) En el nodo master activo, ejecute el comando **netstat -anp |grep 28443 |grep LISTEN |awk '{print \$7}'** como usuario **omm** para consultar el ID de proceso de Tomcat.
    - 2) Ejecute el comando **kill -9 {pid}**, en el que **{pid}** indica el ID de proceso de Tomcat obtenido en [2.iv.1](#).
    - 3) Espere hasta que el proceso se reinicie automáticamente.

Puede ejecutar el comando **netstat -anp |grep 28443 |grep LISTEN** para comprobar si el proceso se reinicia correctamente. Si se muestra el proceso, el proceso se reinicia correctamente. Si el proceso no se muestra, vuelva a consultarlo más tarde.
3. Modificación de la duración del tiempo de espera para una interfaz de usuario web de componentes de código abierto
- a. Acceda a la página **All Configurations**.
    - Para versiones de clúster MRS anteriores a MRS 3.x:

En la página de detalles del clúster, elija **Components > Meta > Service Configuration**.
    - Para la versión 3.x o posterior del clúster MRS:

Inicie sesión en FusionInsight Manager y elija **Cluster > Services > meta**. En la página mostrada, haga clic en **Configurations** y seleccione **All Configurations**.
  - b. Cambie el valor de **http.server.session.timeout.secs** en **meta** según sea necesario. La unidad es segundo.
  - c. Guarde la configuración, anule la selección de **Restart the affected services or instances** y haga clic en **OK**.

Se recomienda realizar el reinicio durante las horas fuera de pico.
  - d. (Opcional) Si necesita usar la interfaz de usuario web de Spark, busque **spark.session.maxAge** en la página **All Configurations** de Spark y cambie el valor (en segundos).

Guarde la configuración, anule la selección de **Restart the affected services or instances** y haga clic en **OK**.
  - e. Reinicie el metaservicio y los componentes en la interfaz de usuario web o reinicie el clúster durante las horas no pico.

Para evitar la interrupción del servicio, reinicie el servicio durante las horas no pico o realice un reinicio continuo. Para obtener más información, consulte [Realización del reinicio continuo](#).

 **NOTA**

La duración del tiempo de espera de la interfaz de usuario web del componente Flink no se puede cambiar.

## 6.2 ¿Por qué no puedo actualizar la página Dynamic Resource Plan en la pestaña de tenant de MRS?

### Pregunta:

¿Qué puedo hacer si no se puede actualizar la página Dynamic Resource Plan en MRS Tenant Management?

### Respuesta:

- Paso 1** Inicie sesión en los nodos Master1 y Master2 como usuario **root**.
- Paso 2** Ejecute el comando **ps -ef |grep aos** para comprobar el ID del proceso AOS.
- Paso 3** Ejecute el comando **kill -9 AOS process ID** para finalizar el proceso AOS.
- Paso 4** Espere hasta que el proceso de AOS se reinicie automáticamente. Puede ejecutar el comando **ps -ef |grep aos** para comprobar si existe el proceso AOS. Si el proceso existe, el reinicio se realiza correctamente. Si el proceso no existe, vuelva a intentarlo más tarde.

----Fin

## 6.3 ¿Qué hago si la pestaña Kafka Topic Monitoring no está disponible en Manager?

### Pregunta:

¿Qué puedo hacer si la página de la pestaña Kafka Topic Monitoring no se muestra en el FusionInsight Manager?

### Respuesta:

- Paso 1** Inicie sesión en los nodos maestros del clúster y cambie a usuario **omm**.
- Paso 2** Vaya al directorio **/opt/Bigdata/apache-tomcat-7.0.78/webapps/web/WEB-INF/lib/components/Kafka/**.
- Paso 3** Copie el paquete ZooKeeper al directorio **cp /opt/share/zookeeper-3.5.1-mrs-2.0/zookeeper-3.5.1-mrs-2.0.jar ./**.
- Paso 4** Reinicie el proceso de Tomcat.

```
sh /opt/Bigdata/apache-tomcat-7.0.78/bin/shutdown.sh
```

```
sh /opt/Bigdata/apache-tomcat-7.0.78/bin/startup.sh
```

----Fin

## 6.4 ¿Cómo lo hago si se informa de un error o algunas funciones no están disponibles cuando accedo a las interfaces de usuario web de HDFS, Hue, YARN, HetuEngine y Flink?

### Pregunta:

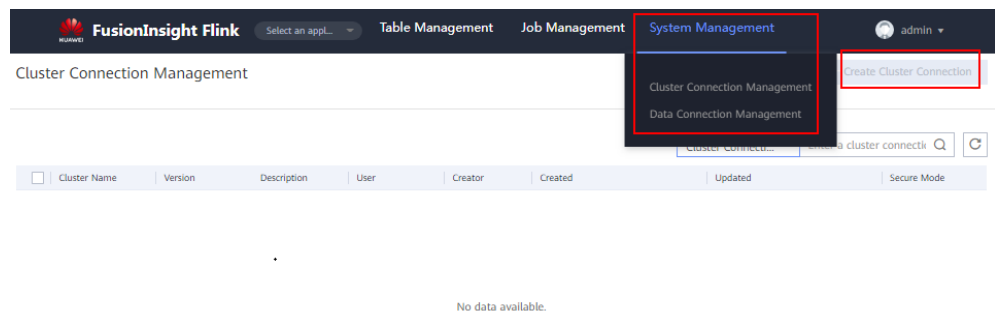
¿Qué puedo hacer si se informa de un error o algunas funciones no están disponibles cuando accedo al WebUI de componentes como HDFS, Hue, Yarn, Flink y HetuEngine?

### Respuesta:

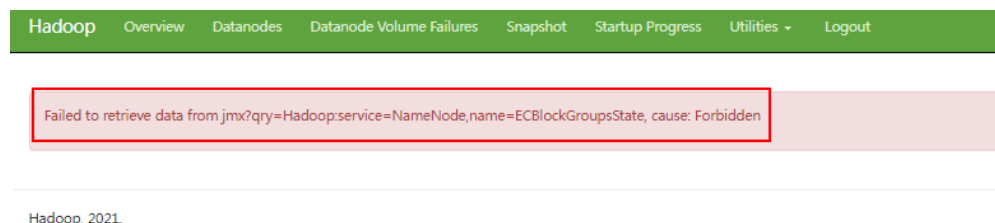
Los usuarios que acceden al WebUI de componentes como HDFS, Hue, Yarn, Flink y HetuEngine no tienen el permiso de gestión de los componentes correspondientes. Como resultado, se informa de un error o algunas funciones no están disponibles.

Ejemplo:

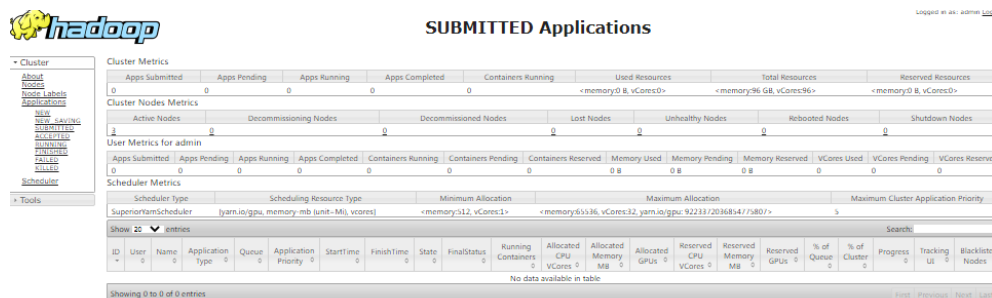
- Después de iniciar sesión en la interfaz de usuario web de Flink como usuario actual, no se puede mostrar parte del contenido y no tiene permiso para crear aplicaciones, conexiones de clúster o conexiones de datos.




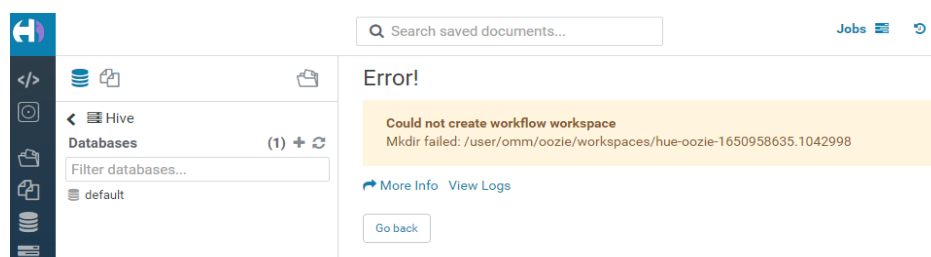
- Después de iniciar sesión en la interfaz de usuario web de HDFS como el usuario actual, se muestra el mensaje de error "Failed to retrieve data from /jmx?qry=java.lang:type=Memory, cause: Forbidden".



- Después de iniciar sesión en la interfaz de usuario web de YARN como usuario actual, no puede ver la información del trabajo.



- Después de iniciar sesión en la interfaz de usuario web de Hue como usuario actual, haga clic en  en el panel de navegación de la izquierda y seleccione **Workflow** se muestra un mensaje de error.




Se recomienda iniciar sesión en las web UI de los componentes como usuario con los correspondientes permisos de gestión. Por ejemplo, puede crear un usuario de servicio que tenga los permisos de gestión en HDFS y puede iniciar sesión en la interfaz de usuario web de HDFS como el usuario creado. Para obtener más información, consulte [Creación de un usuario](#).

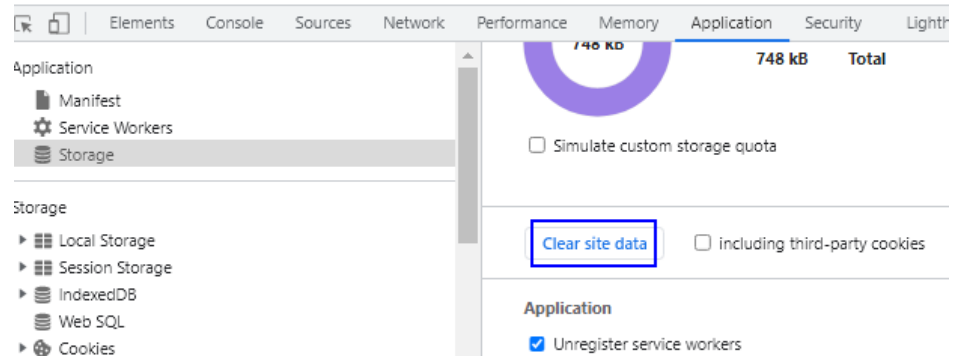
## 6.5 ¿Cómo cambio el modo de acceso a MRS Manager (no para la nube operada conjuntamente)?

### Pregunta:

¿Cómo uso una EIP para acceder a MRS Manager después de acceder a MRS Manager mediante una conexión Direct Connect?

### Respuesta:

- Para MRS 3.x o posterior:  
En la página de pestaña **Dashboard** del clúster, haga clic en  junto a **Access Manager** para cambiar el modo de acceso.
- Para MRS 2.x o anterior:
  - a. En la página del clúster, presione **F12**, elija **Application > Storage** y haga clic en **Clear site data**.



- b. Actualice la página e inicie sesión de nuevo en la página del clúster MRS. En la página de pestaña **Dashboard**, haga clic en **Access Manager** para cambiar el modo de acceso.

# 7 Monitoreo de alarmas

---

## 7.1 En un clúster de streaming de MRS, ¿puede la función de monitoreo de Kafka topic enviar notificaciones de alarma?

### Pregunta:

En un clúster de streaming de MRS, ¿puede la función de monitoreo de Kafka topic enviar notificaciones de alarma?

### Respuesta:

El monitoreo de Kafka topic no puede enviar alarmas por correo electrónico o mensaje SMS. Actualmente, puede ver la información de alarma en Manager.

## 7.2 ¿Dónde puedo ver las colas de recursos en ejecución cuando se generan ALM-18022 Recursos de cola de Yarn insuficientes?

### Pregunta:

¿Dónde puedo ver las colas de recursos en ejecución cuando se generan ALM-18022 Recursos de cola de Yarn insuficientes?

### Respuesta:

Inicie sesión en FusionInsight Manager y elija **Cluster > Services > Yarn**. En el panel de navegación de la izquierda, elija **ResourceManager(Active)** e inicie sesión en la página nativa de Yarn.

Para obtener información detallada, consulte la ayuda en línea.

## 7.3 ¿Cómo entiendo las estadísticas de gráficos de varios niveles en la métrica de solicitudes de operación de HBase?

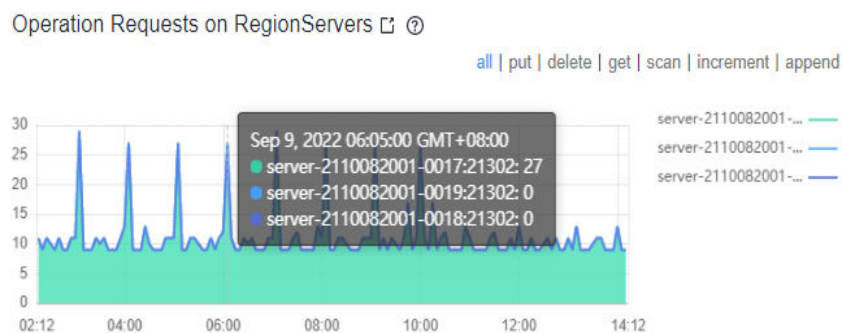
### Pregunta:

¿Cómo entiendo las estadísticas de gráficos de varios niveles en la métrica de solicitudes de operación de HBase?

### Respuesta:

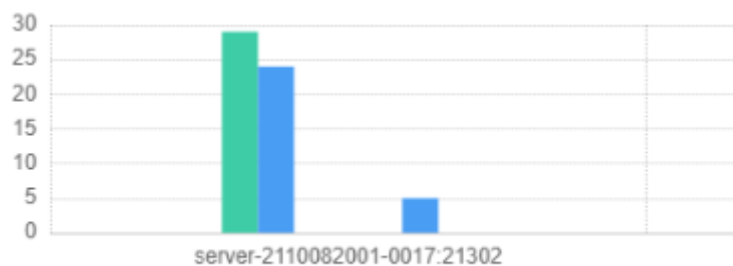
A continuación se utiliza el concepto de monitoreo **Operation Requests on RegionServers** como ejemplo:

1. Inicie sesión en FusionInsight Manager y elija **Cluster > Services > HBase > Resource**. En la página mostrada, puede ver el gráfico **Operation Requests on RegionServers**. Si hace clic en **all** se muestran los 10 RegionServers principales clasificados por el número total de solicitudes de operación en el clúster actual, el intervalo de estadísticas es de 5 minutos.



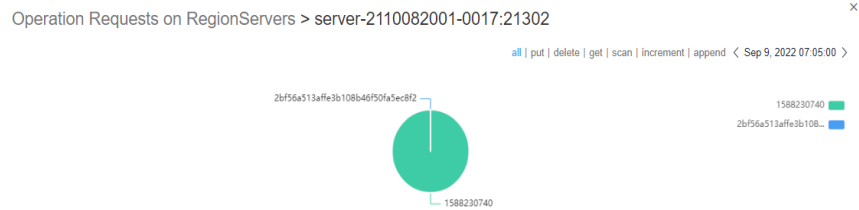
2. Haga clic en un punto del gráfico. Se muestra un gráfico de nivel 2, que muestra el número de solicitudes de operación de todos los RegionServers en los últimos 5 minutos.

### Operation Requests on RegionServers

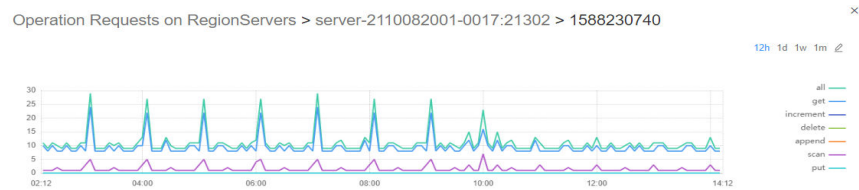




- Haga clic en un gráfico de barras de estadísticas de operaciones. Se muestra un gráfico de nivel 3, que muestra la distribución de las operaciones en cada región dentro del período.



- Haga clic en un nombre de región. Se muestra el diagrama de distribución de las operaciones realizadas cada 5 minutos en las últimas 12 horas. Puede ver el número de operaciones realizadas en el período.



# 8 Ajuste de rendimiento

---

## 8.1 ¿Un clúster MRS admite la reinstalación del sistema?

### Pregunta:

¿Un clúster MRS admite la reinstalación del sistema?

### Respuesta:

Un clúster MRS no admite la reinstalación del sistema.

## 8.2 ¿Puedo cambiar el sistema operativo de un clúster MRS?

### Pregunta:

¿Puedo cambiar el sistema operativo de un clúster MRS?

### Respuesta:

No se puede cambiar el sistema operativo de un clúster MRS.

## 8.3 ¿Cómo puedo mejorar la utilización de recursos de los nodos Core en un clúster?

1. Vaya a la página de configuración del servicio de Yarn.
  - Para los clústeres MRS 1.8.10 o anteriores:  
Inicie sesión en MRS Manager haciendo referencia a [Acceso a MRS Manager](#). Elija **Services > Yarn > Service Configuration** y seleccione **All** en la lista desplegable **Basic**.
  - Para MRS 1.8.10 o posterior y MRS 2.x, haga clic en el nombre del clúster. En la página de detalles del clúster que se muestra, haga clic en la pestaña **Components** y

elija **Yarn**. Haga clic en el **Service Configuration** y seleccione **All** en la lista desplegable **Basic**.

 **NOTA**

Si la pestaña **Components** no está disponible, primero complete la sincronización del usuario de IAM. (En la página **Dashboard**, haga clic en **Synchronize** en el lado derecho de **IAM User Sync** para sincronizar usuarios de IAM.)

- Para la versión 3.x o posterior del clúster MRS:  
Inicie sesión en FusionInsight Manager. Elija **Cluster** > *Name of the desired cluster* > **Services** > **Yarn** > **Configurations** > **All Configurations**.
- 2. Busque **yarn.nodemanager.resource.memory-mb** y aumente el valor en función de la memoria real de los nodos del clúster.
- 3. Guarde el cambio y reinicie los servicios o instancias afectados.

## 8.4 ¿Cómo detengo el servicio Firewall?

### Pregunta:

¿Cómo detengo el servicio Firewall?

### Respuesta:

**Paso 1** Inicie sesión en cada nodo de un clúster como usuario **root**.

**Paso 2** Compruebe si se ha iniciado el servicio de firewall.

Por ejemplo, para comprobar el estado del firewall en EulerOS, ejecute el comando **systemctl status firewalld.service**.

**Paso 3** Detenga el servicio de firewall.

Por ejemplo, para detener el servicio de firewall en EulerOS, ejecute el comando **systemctl stop firewalld.service**.

----Fin

# 9 Desarrollo del trabajo

---

## 9.1 ¿Cómo obtengo mis datos en OBS o HDFS?

MRS puede procesar datos en OBS y HDFS. Puede obtener sus datos en OBS o HDFS de la siguiente manera:

1. Subir datos locales a OBS.
  - a. Inicie sesión en la consola OBS.
  - b. Cree un sistema de archivos paralelo llamado **userdata** en OBS y cree las carpetas **program**, **input**, **output** y **log** en el sistema de archivos.
    - i. Elija **Parallel File System > Create Parallel File System**, y cree un sistema de archivos denominado **userdata**.
    - ii. En la lista del sistema de archivos OBS, haga clic en el nombre del sistema de archivos **userdata**, elija **Files > Create Folder** y cree las carpetas de **program**, **input**, **output**, y **log**.
  - c. Cargue datos al sistema de archivos **userdata**.
    - i. Vaya a la carpeta **program** y haga clic en **Upload File**.
    - ii. Haga clic en **add file** y seleccione un programa de usuario.
    - iii. Haga clic en **Upload**.
    - iv. Cargue el archivo de datos de usuario al directorio **input** utilizando el mismo método.
2. Importar datos OBS a HDFS.

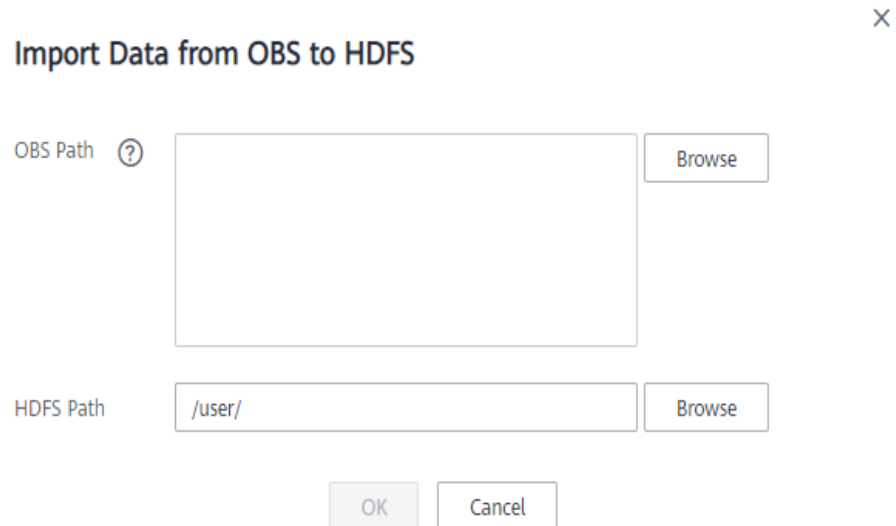
Puede importar datos OBS a HDFS sólo cuando **Kerberos Authentication** está deshabilitado y el clúster se está ejecutando.

  - a. Inicie sesión en la consola de MRS.
  - b. Haga clic en el nombre del clúster.
  - c. En la página que se muestra, seleccione la página de pestaña **Files** y haga clic en **HDFS File List**.
  - d. Seleccione un directorio de datos, por ejemplo, **bd\_app1**.

El directorio **bd\_app1** es solo un ejemplo. Puede utilizar cualquier directorio de la página o crear uno nuevo.

- e. Haga clic en **Import Data** y haga clic en **Browse** para seleccionar una ruta OBS y una ruta HDFS. Consulte [Figura 9-1](#).

**Figura 9-1** Importación de archivos



The screenshot shows a dialog box titled "Import Data from OBS to HDFS". It features a close button (X) in the top right corner. The dialog contains two input fields: "OBS Path" with a help icon (?) and a "Browse" button, and "HDFS Path" with the text "/user/" and a "Browse" button. At the bottom are "OK" and "Cancel" buttons.

- f. Haga clic en **OK**.  
Puede ver el progreso de la carga de archivos en la pestaña **File Operation Records**.

## 9.2 ¿Qué tipos de trabajos de Spark se pueden enviar en un clúster?

### Pregunta:

¿Qué tipos de trabajos de Spark se pueden enviar en un clúster?

### Respuesta:

Los clústeres MRS admiten trabajos de Spark enviados en modo Spark, Spark Script o Spark SQL.

## 9.3 ¿Puedo ejecutar varias tareas de Spark al mismo tiempo después de que los recursos mínimos del tenant de un clúster MRS se cambien a 0?

### Pregunta:

¿Puedo ejecutar varias tareas de Spark al mismo tiempo después de que los recursos mínimos del tenant de un clúster MRS se cambien a 0?

## Respuesta:

Sólo puede ejecutar una tarea de Spark a la vez después de que los recursos mínimos de tenant de un clúster de MRS se cambien a 0.

## 9.4 ¿Qué hago si no se pueden identificar los parámetros de trabajo separados por espacios?

### Pregunta:

¿Se identificarán los parámetros si los parámetros de ejecución de trabajos están separados por espacios?

### Respuesta:

Utilice espacios para separar parámetros. Para evitar que los parámetros se guarden como texto sin formato, agregue un (@) de signo antes de los parámetros, por ejemplo, `@password=admin_123`.

## 9.5 ¿Cuáles son las diferencias entre el modo de client y el modo de cluster de los trabajos de Spark?

Es necesario entender el concepto ApplicationMaster antes de entender las diferencias esenciales entre Yarn-client e Yarn-cluster.

En Yarn, cada instancia de aplicación tiene un proceso ApplicationMaster, que es el primer contenedor iniciado por la aplicación. Interactúa con ResourceManager y solicita recursos. Después de obtener recursos, indica a NodeManager que inicie Container. La diferencia esencial entre los modos Yarn-cluster e Yarn-client radica en el proceso ApplicationMaster.

En el modo Yarn-cluster, Driver se ejecuta en el ApplicationMaster que solicita recursos a Yarn y supervisa el estado de ejecución de un trabajo. Después de que un usuario envíe un trabajo, el cliente se puede detener y el trabajo continúa ejecutándose en Yarn. Por lo tanto, el modo Yarn-cluster no es adecuado para ejecutar trabajos interactivos.

En el modo Yarn-cliente, ApplicationMaster solo solicita al Executor de Yarn. El cliente se comunica con Container solicitado para programar tareas. Por lo tanto, el cliente no puede ser detenido.

## 9.6 ¿Cómo puedo ver los registros de trabajos de MRS?

**Paso 1** En la página **Jobs** de la consola MRS, puede ver los registros de cada trabajo, incluidos los registros de launcherJob y realJob.

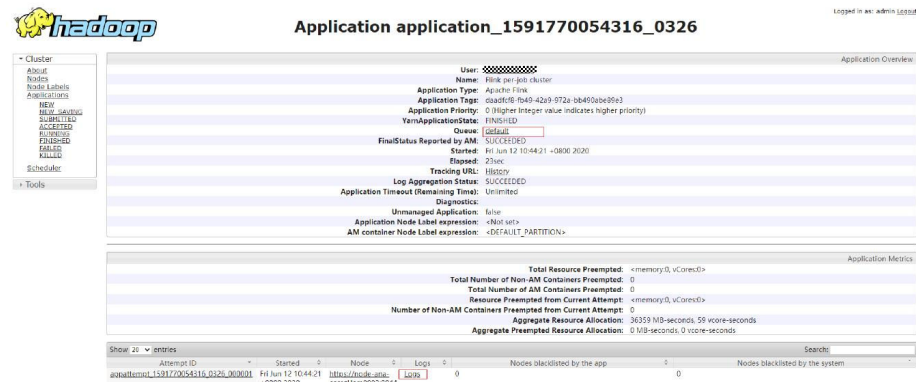
- Por lo general, los registros de errores se imprimen en los trabajos de **stderr** y **stdout** de launcherJob, como se muestra en la siguiente figura:

```

container-localizer-syslog | directory.info | launch_container.sh | prelaunch.err | prelaunch.out | stderr | stdout | syslog
1 org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://hacluster/user/uv-0610-100 already exists
2 at org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:164)
3 at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:288)
4 at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:148)
5 at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1570)
6 at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1567)
7 at java.security.AccessController.doPrivileged(Native Method)
8 at javax.security.auth.Subject.doAs(Subject.java:422)
9 at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1729)
10 at org.apache.hadoop.mapreduce.Job.submit(Job.java:1567)
11 at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1588)
12 at org.apache.hadoop.examples.WordCount.main(WordCount.java:87)
13 at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
14 at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
15 at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
16 at java.lang.reflect.Method.invoke(Method.java:498)

```

- Puede ver los registros de realJob en la interfaz de usuario web de Resource Manager proporcionada por el servicio Yarn en MRS Manager.



**Paso 2** Inicie sesión en el nodo Master del clúster para obtener los archivos de registro de trabajos en **Paso 1**. La ruta de acceso HDFS es `/tmp/logs/{submit_user}/logs/{application_id}`.

**Paso 3** Después de enviar el trabajo, si no se puede encontrar el ID de la aplicación de trabajo en la UI web de Yarn, no se puede enviar el trabajo. Puede iniciar sesión en el nodo Master activo del clúster y ver el registro del proceso de envío de trabajos `/var/log/executor/logs/exe.log`.

----Fin

## 9.7 ¿Cómo lo hago si se muestra el mensaje "The current user does not exist on MRS Manager. Grant the user sufficient permissions on IAM and then perform IAM user synchronization on the Dashboard tab page."?

Si la sincronización de IAM no se realiza cuando se envía un trabajo en un clúster de seguridad, se muestra el mensaje de error "The current user does not exist on MRS Manager. Grant the user sufficient permissions on IAM and then perform IAM user synchronization on the Dashboard tab page."

Antes de enviar un trabajo, en la página **Dashboard**, haga clic en **Synchronize** en el lado derecho de **IAM User Sync** para sincronizar usuarios de IAM.

## 9.8 La ejecución de trabajo de LauncherJob se ha fallado y el mensaje de error "jobPropertiesMap is null." se muestra

La causa del error de launcherJob es que el usuario que envía el trabajo no tiene el permiso de escritura en el directorio `hdfs /mrs/job-properties`.

Este problema se corrige en el parche 2.1.0.6. También puede conceder el permiso de escritura en el directorio `/mrs/job-properties` al usuario sincronizado que envía el trabajo en MRS Manager.

## 9.9 ¿Cómo lo hago si el estado del trabajo de Flink en la consola MRS es incompatible con el del Yarn?

Para ahorrar espacio de almacenamiento, el elemento de configuración de Yarn `yarn.resourcemanager.max-completed-applications` se modifica para reducir el número de registros de trabajos históricos almacenados en Yarn. Los trabajos de Flink son trabajos a largo plazo. El `realJob` todavía se está ejecutando en Yarn, pero el `launcherJob` se ha eliminado. Como resultado, el `launcherJob` no se encuentra en Yarn y el estado del trabajo no se actualiza. Este problema se corrige en el parche 2.1.0.6.

Solución alternativa: Termine el trabajo cuyo `launcherJob` no se puede encontrar. El estado del trabajo enviado más adelante se actualizará.

## 9.10 ¿Cómo lo hago si un trabajo de SparkStreaming falla después de haber sido ejecutado docenas de horas y se reporta el error OBS Access 403?

Cuando un usuario envía un trabajo que necesita leer y escribir OBS, el programa de envío de trabajos agrega la clave de acceso temporal (AK) y la clave secreta (SK) para acceder a OBS de forma predeterminada. Sin embargo, el AK y el SK temporales tienen tiempo de caducidad.

Si desea ejecutar trabajos a largo plazo como Flink y SparkStreaming puede ingresar el AK y el SK en el **Service Parameter** para asegurarse de que los trabajos no fallarán en ejecutarse debido al vencimiento de la clave.

## 9.11 ¿Cómo hago si se informa de una alarma que indica que la memoria es insuficiente cuando ejecuto una sentencia SQL en el cliente ClickHouse?

### Síntoma

El cliente ClickHouse restringe la memoria utilizada por las sentencias GROUP BY. Cuando se ejecuta una sentencia SQL en el cliente ClickHouse, se muestra la siguiente información de error:



```
Progress: 1.83 billion rows, 85.31 GB (68.80 million rows/s., 3.21 GB/s.)  
6%Received exception from server:  
Code: 241. DB::Exception: Received from localhost:9000, 127.0.0.1.  
DB::Exception: Memory limit (for query) exceeded: would use 9.31 GiB (attempt to  
allocate chunk of 1048576 bytes), maximum: 9.31 GiB:  
(while reading column hits):
```

## Solución

- Ejecute el siguiente comando antes de ejecutar una sentencia SQL a condición de que el clúster tenga suficiente memoria:  

```
SET max_memory_usage = 128000000000; #128G
```
- Si no hay suficiente memoria disponible, el ClickHouse le permite desbordar los datos al disco para liberar la memoria: Se recomienda establecer el valor de **max\_memory\_usage** en el doble del tamaño de **max\_bytes\_before\_external\_group\_by**.  

```
set max_bytes_before_external_group_by=200000000000; #20G  
set max_memory_usage=400000000000; #40G
```

## 9.12 ¿Cómo lo hago si se muestra un mensaje de error "java.io.IOException: Connection reset by peer" durante la ejecución de un trabajo de Spark?

### Síntoma

El trabajo de Spark sigue en ejecución y se muestra el mensaje de error "java.io.IOException: Connection reset by peer".

### Solución

Agregue el parámetro **executor.memory Overhead** a los parámetros para enviar un trabajo.

## 9.13 ¿Cómo hago si se muestra un mensaje de error "requestId=4971883851071737250" cuando un trabajo de Spark accede a OBS?

### Síntoma

El mensaje de error "requestId=4971883851071737250" aparece cuando un trabajo de Spark accede a OBS.

### Solución

Inicie sesión en el nodo donde se encuentra el cliente Spark, vaya al directorio **conf** y cambie el valor del parámetro **fs.obs.metrics.switch** en el archivo de configuración **core-site.xml** a **false**.

## 9.14 ¿Cómo lo hago si se reporta el error "UnknownScannerException" del trabajo de Spark?

### Síntoma

Los trabajos de Spark funcionan lentamente. La información de advertencia se imprime en los registros de ejecución y la causa de error es **UnknownScannerException**.

### Solución

Antes de ejecutar un trabajo de Spark, ajuste el valor de **hbase.client.scanner.timeout.period** (por ejemplo, de 60 segundos a 120 segundos).

Inicie sesión en FusionInsight Manager y elija **Cluster > Services > HBase**. Haga clic en **Configurations** luego **All Configurations**, busque **hbase.client.scanner.timeout.period**, y cambie su valor a **120000** (unidad: ms).

## 9.15 ¿Por qué DataArtsStudio ocasionalmente no puede programar trabajos de Spark y la reprogramación también falla?

### Síntoma

DataArtsStudio ocasionalmente no puede programar los trabajos de Spark y la reprogramación también falla. Se muestra la siguiente información de error:

```
Caused by: org.apache.spark.SparkException: Application  
application_1619511926396_2586346 finished with failed status
```

### Solución

Inicie sesión en el nodo donde se encuentra el cliente Spark como usuario **root** y aumente el valor del parámetro **spark.driver.memory** en el archivo **spark-defaults.conf**.

## 9.16 ¿Cómo hago si un trabajo de Flink no se ejecuta y se muestra el mensaje de error "java.lang.NoSuchFieldError: SECURITY\_SSL\_ENCRYPT\_ENABLED"?

### Síntoma

Un trabajo de Flink no se puede ejecutar y se muestra el siguiente mensaje de error:

```
Causado por: java.lang.NoSuchFieldError: SECURITY_SSL_ENCRYPT_ENABLED
```

## Solución

El paquete de dependencias de terceros en el código de cliente entra en conflicto con el paquete de clúster. Como resultado, el trabajo no se envía al clúster MRS. Necesita modificar el paquete de dependencias, establecer el alcance del paquete Hadoop de código abierto y el paquete Flink en el archivo POM a **provide** y empaquetar y ejecutar el trabajo de nuevo.

## 9.17 ¿Por qué no se puede ver el trabajo de Yarn enviado en la interfaz de usuario web?

Después de crear un trabajo de Yarn, no se puede ver si inicia sesión en la web UI como usuario **admin**.

- El usuario **admin** es un usuario de la página de gestión del clúster. Compruebe si el usuario tiene el permiso **supergroup**. Generalmente, solo el usuario con el permiso **supergroup** puede ver trabajos.
- Inicie sesión en Yarn como el usuario que envía trabajos para ver trabajos en Yarn. No vea los trabajos con el usuario **admin**.

## 9.18 ¿Cómo modifico el HDFS NameSpace (fs.defaultFS) de un clúster existente?

Puede modificar o agregar el NameSpace HDFS (fs.defaultFS) del clúster modificando los archivos **core-site.xml** y **hdfs-site.xml** del cliente. Sin embargo, no se recomienda realizar esta operación en el servidor.

## 9.19 ¿Cómo lo hago si YARN detiene la cola de launcher-job debido a un heap size insuficiente cuando envío un trabajo de Flink en el plano de gestión?

### Síntoma

La cola de launcher-job se detiene por YARN cuando se envía un trabajo de Flink en el plano de gestión.

### Solución

Aumente el heap size de la cola de launcher-job.

1. Inicie sesión en el nodo OMS activo como usuario **omm**.
2. Cambie el valor de **job.launcher.resource.memory.mb** de **/opt/executor/webapps/executor/WEB-INF/classes/servicebroker.xml** a **2048**.
3. Ejecute el comando **sh /opt/executor/bin/restart-executor.sh** para reiniciar el proceso del ejecutor.

## 9.20 ¿Cómo lo hago si se muestra el mensaje de error "slot request timeout" cuando envío un trabajo de Flink?

### Síntoma

Cuando se envía un job de Flink, JobManager se inicia correctamente. Sin embargo, TaskManager permanece en el estado inicial hasta el tiempo de espera. Se muestra la siguiente información de error:

```
org.apache.flink.runtime.jobmanager.scheduler.NoResourceAvailableException: Could not allocate the required slot within slot request timeout. Please make sure that the cluster has enough resources
```

### Causas posibles

1. Los recursos en la cola YARN son insuficientes. Como resultado, TaskManager no puede iniciarse.
2. Sus archivos JAR entran en conflicto con los del entorno. Puede ejecutar el programa WordCount para determinar si se produce el problema.
3. Si el clúster está en modo de seguridad, es posible que el certificado SSL de Flink esté configurado incorrectamente o haya caducado.

### Solución

1. Agregar recursos a la cola YARN.
2. Excluir las dependencias de Flink y Hadoop en sus archivos JAR para que Flink y Hadoop puedan depender solo de los archivos JAR en el entorno.
3. Volver a configurar el certificado SSL de Flink. Para obtener más información, consulte [Uso de Flink desde principio](#).

## 9.21 Importación y exportación de datos de trabajos de DistCP

- ¿Un trabajo de DistCP compara la coherencia de los datos durante la importación y exportación de datos?  
No. Los trabajos de DistCP sólo copian datos pero no los modifican.
- Cuando se exportan datos desde un trabajo de DistCP, si ya existen algunos archivos en OBS, ¿cómo procesará el trabajo los archivos?  
Los trabajos de DistCP sobrescribirán los archivos en OBS.

## 9.22 ¿Cómo puedo ver las sentencias SQL para los trabajos de Hive en la YARN Web UI?

**Paso 1** Inicie sesión en FusionInsight Manager como usuario del servicio.

**Paso 2** Elija **Cluster > Services > Yarn**. Haga clic en **ResourceManager(xxx,Active)** junto a **ResourceManager WebUI**.

**Paso 3** En la interfaz de usuario web de YARN, haga clic en el ID del trabajo que se va a ver.

The screenshot shows the Hadoop YARN web interface. On the left is a navigation menu with options like 'Cluster', 'About Nodes', 'Node Labels', 'Applications', 'Scheduler', and 'Tools'. The main area displays 'Cluster Metrics' and 'User Metrics for testc'. Below these is a table of applications. The application with ID 'application\_...\_0004' is highlighted with a red box. The table columns include ID, User, QueueUser, Name, Application Type, Application Tags, Queue, Application Priority, StartTime, LaunchTime, and FinishTime.

**Paso 4** Haga clic en **ApplicationMaster** o **History** junto a **Tracking URL**.

The screenshot shows the details of the application 'application\_...\_0004'. The 'Tracking URL' section contains a link labeled 'History' which is highlighted with a red box. Other details include User: hiveuser, Application Type: insert into user\_info partition(year='...B') (Stage-1), YarnApplicationState: FINISHED, and FinalStatus Reported by AM: SUCCEEDED.

**Paso 5** Elija **Configuration** en el panel de navegación izquierdo y busque el parámetro **hive.query.string** en la esquina superior derecha para consultar el correspondiente HiveSQL.

The screenshot shows the configuration for MapReduce Job 'job\_...\_0004'. The left navigation menu has 'Configuration' selected. The main area shows a table of configuration key-value pairs. In the top right corner, there is a search box with a red border and a magnifying glass icon.

----Fin

## 9.23 ¿Cómo puedo ver los registros de una tarea de Yarn especificada?

**Paso 1** Inicie sesión en el nodo maestro activo como usuario **root**.

**Paso 2** Ejecute el siguiente comando para inicializar variables de entorno:

`sourceClient installation directory/bigdata_env`

**Paso 3** Si la autenticación de Kerberos está habilitada para el clúster, ejecute el siguiente comando para autenticar al usuario. Si la autenticación Kerberos está deshabilitada, omita este paso.

`kinit MRS cluster user`

**Paso 4** Ejecute el siguiente comando para obtener los registros de una tarea especificada:

**yarn logs -applicationId** *application\_ID* del trabajo que desea ver

**---Fin**

# 10 Actualización/Instalación de parches de clústeres

---

## 10.1 ¿Cómo actualizo un clúster MRS?

- Clústeres no LTS  
Un clúster existente que no sea LTS no puede actualizarse a una versión posterior. Puede crear un clúster de la nueva versión y migrar datos del clúster antiguo al nuevo clúster.
- Clústeres MRS LTS  
Los clústeres MRS LTS se pueden actualizar a una versión LTS de destino. Póngase en contacto con nuestros ingenieros de O&M y compre servicios profesionales para la actualización del clúster.

## 10.2 ¿Puedo cambiar la versión del clúster MRS?

MRS no admite el cambio de versiones de clústeres o componentes existentes. Puede crear un clúster de la nueva versión para probar las características actualizadas.

# 11 Interconexión del ecosistema periférico

---

## 11.1 ¿Se puede utilizar MRS para realizar operaciones de lectura y escritura en tablas DLI?

Si ha almacenado datos en OBS, puede usar Spark en MRS para leer tablas de Data Lake Insight (DLI), procesar de forma flexible los datos de la tabla y guardar el resultado en otra tabla DLI. Si no ha almacenado datos en OBS, no puede usar MRS para leer o escribir tablas DLI.

## 11.2 ¿OBS soporta el protocolo ListObjectsV2?

OBS no admite el protocolo ListObjectsV2.

## 11.3 ¿Se pueden almacenar datos MRS en un sistema de archivos paralelo proporcionado por OBS?

MRS admite el almacenamiento en un sistema de archivos paralelo proporcionado por OBS.

## 11.4 ¿Se puede desplegar el servicio Crawler en MRS?

El servicio Crawler no se puede desplegar en MRS.

## 11.5 ¿DWS y MRS soportan la eliminación segura (evitar la recuperación después de la eliminación)?

Tanto MRS como Data Warehouse Service (DWS) admiten la eliminación segura. MRS admite la restauración solo a partir de los datos de copia de respaldo. DWS solo admite la restauración a partir de instantáneas.



## 11.6 ¿Por qué no se encuentra el clúster MRS autenticado por Kerberos cuando se configura una conexión desde DLF?

Data Lake Factory (DLF) no se puede interconectar con un clúster MRS que tenga habilitada la autenticación de Kerberos. Si desea utilizar clústeres MRS con la autenticación de Kerberos habilitada, utilice la plataforma DataArts Studio.

## 11.7 ¿Cómo uso PySpark en un ECS para conectarse a un clúster MRS Spark con autenticación de Kerberos habilitada, en la Intranet?

Cambie el valor de `spark.yarn.security.credentials.hbase.enabled` en el archivo `spark-defaults.conf` de Spark a `true` y use el comando `spark-submit --master yarn --keytab keytabfile --principal principal` para especificar el archivo de autenticación de Kerberos.

## 11.8 ¿Por qué los campos asignados no existen en la base de datos después de que HBase sincronice los datos con CSS?

Después de sincronizar los datos de HBase de MRS con Cloud Search Service (CSS), la base de datos no tendrá campos asignados. Sólo las tablas tienen campos asignados.

## 11.9 ¿Puede Flume leer datos de OBS?

De forma predeterminada, Flume no puede leer datos de OBS. Si tiene requisitos de servicio, debe personalizar las fuentes para implementar esta función.

## 11.10 ¿Se puede conectar MRS a un KDC externo?

MRS solo admite el centro de distribución de claves (KDC) integrado.

## 11.11 ¿Cómo soluciono el problema de compatibilidad de la versión Jetty en la interconexión de código abierto Kylin 3.x y MRS 1.9.3?

Por motivos de seguridad, MRS ha actualizado algunos componentes de terceros de código abierto que tienen graves vulnerabilidades de seguridad. La actualización causa el problema de compatibilidad de la versión Jetty cuando Kylin de código abierto se interconecta con MRS 1.9.3.

Realice las siguientes operaciones para resolver el problema:

1. Instale el cliente MRS en un nodo ECS. Para obtener más información, consulte [Uso de un cliente MRS en nodos fuera de un clúster](#). Esta operación utiliza la ruta de instalación del cliente MRS `/srv/client/` como ejemplo.
2. Una vez completada la instalación, ejecute los siguientes comandos para importar la variable de entorno de cliente MRS `bigdata_env` y las variables de entorno `HIVE_CONF` y `HCAT_HOME` requeridas por Kylin.

```
source /srv/client/bigdata_env
export HIVE_CONF=/srv/client/Hive/config/
export HCAT_HOME=/srv/client/Hive/HCatalog
```

3. Instale Kylin en el nodo donde está instalado el cliente MRS y especifique `KYLIN_HOME`. Para obtener más información, consulte [el sitio web oficial de Kylin](#). Para MRS 1.9.3, seleccione Kylin para HBase 1.x para interconexión.
4. Quite los paquetes Jetty .jar del directorio `/srv/client/Hive/Beeline/lib/` en el directorio del cliente Hive para evitar conflictos de versiones.

```
Paquetes de Jetty .jar:
javax-websocket-server-impl-9.4.26.v20200117.jar
websocket-server-9.4.26.v20200117.jar
jetty-all-9.4.26.v20200117-uber.jar
jetty-runner-9.4.26.v20200117.jar
apache-jsp-9.4.26.v20200117.jar
```

5. Inicie el servicio Kylin y compruebe los registros de Kylin. En casos normales, los registros no contienen errores de compatibilidad, como `java.lang.NoSuchMethodException` y `java.lang.ClassNotFoundException`.
6. Acceda a la página nativa de Kylin en `http://<hostname>:7070/kylin` y ejecute el script Cube de ejemplo `${KYLIN_HOME}/bin/sample.sh` para comprobar si Kylin se está ejecutando correctamente.

```
$KYLIN_HOME/bin/kylin.sh start
```

## 11.12 ¿Qué sucede si los datos no se exportan desde MRS a un bucket cifrado de OBS?

MRS admite la encriptación de bucket OBS desde la versión 1.9.x. Si desea utilizar buckets de OBS cifrados, utilice la versión de clúster de MRS 1.9.x o posterior.

## 11.13 ¿Cómo interconecto MRS con LTS?

### Prerrequisitos

Ha obtenido un par AK/SK al referirse a [¿Cómo obtengo un par AK/SK?](#)

### Procedimiento

1. Instale ICAgent en el nodo MRS. Para obtener más información, consulte [Instalación de ICAgent \(Huawei Cloud Hosts\)](#)

 **NOTA**

Para la primera instalación, instale primero un servidor y, a continuación, instale por lotes todos los demás hosts de la misma manera.

2. Cree un grupo de hosts y agregue el host ICAgent instalado en **1** al grupo de hosts.
  - a. Inicie sesión en la consola de Log Tank Service (LTS) y elija **Host Management** en el panel de navegación izquierdo. En la página mostrada, haga clic en **Create Host Group** en la esquina superior derecha.
  - b. En el panel deslizable **Create Host Group** que se muestra, escriba un nombre de grupo host en el campo **Host Group** y establezca **Host OS** en **Windows**.
  - c. En la lista de hosts, seleccione uno o varios hosts que desea agregar al grupo y haga clic en **OK**.
    - Puede filtrar hosts por nombre de host o dirección IP de host. También puede hacer clic en **Search by Host IP Address** e introducir varias direcciones IP de host en el cuadro de búsqueda que se muestra para buscar coincidencias.
    - Si los hosts deseados no están en la lista, haga clic en **Install ICAgent**. En el panel de deslizable **Install ICAgent** que se muestra, instale ICAgent en los hosts según se le indique.
3. Cree un grupo de log.
  - a. Inicie sesión en la consola LTS, elija **Log Management** en el panel de navegación de la izquierda y haga clic en **Create Log Group** en la esquina superior derecha.
  - b. En el panel deslizable **Create Log Group** que se muestra, escriba un nombre de grupo de registro. Un nombre de grupo de registro:
    - Puede contener solo letras, números, guiones (-), guiones bajos (\_), y puntos (.).
    - No puede comenzar con un punto (.) o guión bajo (\_), y no puede terminar con un punto (.).
    - Puede contener de 1 a 64 caracteres.
  - c. Establezca **Log Retention Duration** en 1 a 30 días. Si no se especifica este parámetro, los registros se conservan durante 7 días de forma predeterminada. Puede cambiar la duración de retención del registro después de crear el grupo de registros.

LTS ofrece una cuota gratuita de 500 MB por mes para lectura/escritura de registros, retención e indexación. Cuando se agote la cuota gratuita, se te facturará el exceso de uso sobre una base de pago por uso.
  - d. Haga clic en **OK**.
    - En la lista de grupos de registros de la página **Log Management**, puede ver detalles del grupo de registros, incluidos el nombre del grupo de registros, la duración de retención del registro, el tiempo de creación, el tipo de creación y las etiquetas. La duración de la retención del registro y las etiquetas se pueden modificar. Haga clic en **Modify** en la columna **Operation** del grupo de registros para modificar la duración.
    - Haga clic en el nombre del grupo de registro para acceder a la página de detalles de flujo de registro.
4. Cree un flujo de registro.
  - a. En la consola LTS, haga clic en el botón de expansión de un grupo de registros.

- b. Haga clic en **Create Log Stream**. En el panel deslizante **Create Log Stream** que se muestra, escriba un nombre de flujo de registro. Un nombre de flujo de registro:
  - Puede contener solo letras, números, guiones (-), guiones bajos (\_), y puntos (.).
  - No puede comenzar con un punto (.) o guión bajo (\_), y no puede terminar con un punto (.)
  - Puede contener de 1 a 64 caracteres.
- c. Seleccione un proyecto empresarial. Puede hacer clic en **View Enterprise Projects** para ver todos los proyectos de empresa.
- d. Haga clic en **OK**.

En la página de flujo de registro, puede ver detalles del flujo de registro, incluidos el nombre del flujo de registro, el proyecto de empresa, la duración de retención del registro, el tiempo de creación y el tipo de creación.

 **NOTA**

Puede configurar diferentes flujos de registro para diferentes componentes.

5. Agregue hosts para la ingestión de registros.
  - a. En el panel de navegación izquierdo de la consola LTS, seleccione **Log Ingestion**.
  - b. En la página de la pestaña **All** que se muestra, haga clic en **Elastic Cloud Server (ECS)**.
  - c. Añada la siguiente configuración a **Collection Path**:

Ruta de registro de tarea YARN:

```
/srv/BigData/*/nm/containerlogs/**/container-localizer-syslog  
/srv/BigData/*/nm/containerlogs/**/directory.info  
/srv/BigData/*/nm/containerlogs/**/launch_container.sh  
/srv/BigData/*/nm/containerlogs/**/prelaunch.err  
/srv/BigData/*/nm/containerlogs/**/prelaunch.out  
/srv/BigData/*/nm/containerlogs/**/stderr  
/srv/BigData/*/nm/containerlogs/**/stdout  
/srv/BigData/*/nm/containerlogs/**/syslog*  
/srv/BigData/*/nm/containerlogs/**/*.*.log
```

 **NOTA**

Configuración de ruta:

- Los registros se pueden recopilar de forma recursiva. Un (\*\*) de doble asterisco puede representar hasta 5 niveles de directorio en una ruta.
- Puede utilizar un asterisco (\*) como comodín para la coincidencia difusa. El comodín (\*) puede representar uno o más caracteres de un directorio o nombre de archivo.
- Si la ruta de recopilación se establece en un directorio (como **/var/logs/**), solo se recopilan los archivos **.log**, **.trace** y **.out** del directorio.

**Registros de componente:**

Todos los registros de servicio HDFS:

```
/var/log/Bigdata/audit/hdfs/jn/hdfs-audit-journalnode.log  
/var/log/Bigdata/audit/hdfs/jn/SecurityAuth.audit  
/var/log/Bigdata/hdfs/*/
```

Configure las rutas de recopilación de registros para otros componentes haciendo referencia a los registros de servicio HDFS.

## 11.14 ¿Cómo instalo HSS en nodos de clúster MRS?

Puede instalar Host Security Service (HSS) en los nodos del clúster MRS si los agentes HSS admiten las versiones del sistema operativo de los nodos.

Realice los siguientes pasos para instalar HSS:

1. Compruebe si los agentes HSS admiten las versiones del sistema operativo de los nodos del clúster MRS.

Para consultar las versiones del sistema operativo, realice las siguientes operaciones:

- a. Inicie sesión en la consola de gestión de MRS.
- b. Seleccione **Clusters > Active Clusters**, seleccione un clúster en ejecución y haga clic en el nombre del clúster para acceder a su página de detalles.
- c. Haga clic en la pestaña **Nodes**. Haga clic en el grupo de nodos **master** o **core** y haga clic en cualquier nodo de la lista de nombres de nodo para acceder a la página de información básica del servicio en la nube.
- d. Haga clic en **Basic Information** para ver la imagen en la información de ECS y confirme la versión de EulerOS en la imagen.

Por ejemplo, **EulerOS\_2.5\_x86\_64** indica que la versión del sistema operativo es EulerOS 2.5 (64 bits).

- e. Compruebe si las versiones del sistema operativo de los nodos MRS son compatibles con los agentes HSS.

Para obtener más información acerca de los sistemas operativos compatibles con los agentes HSS, consulte [Sistemas operativos compatibles](#).

- En caso afirmativo, vaya a **2** para instalar agentes HSS.
- Si no, HSS no se puede instalar.

2. Compre cuotas HSS e inicie sesión en cada nodo del clúster MRS para instalar agentes HSS. Una vez instalados los agentes, puede habilitar HSS.

### **NOTA**

Asegúrese de haber adquirido HSS en la región de nodo del clúster MRS y de haber utilizado el paquete de instalación o el comando de instalación en la región para instalar agentes HSS en los nodos.

Para obtener detalles acerca de cómo comprar cuotas de HSS, instalar agentes y habilitar HSS, consulte [Pasos iniciales con HSS](#).

# 12 Acceso al clúster

## 12.1 ¿Puedo cambiar entre los dos modos de inicio de sesión de MRS?

No. Puede seleccionar el modo de inicio de sesión al crear el clúster. No puede cambiar el modo de inicio de sesión después de crear el clúster.

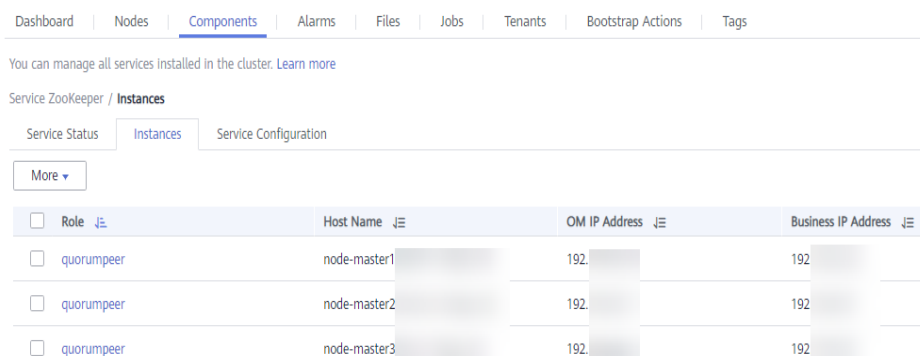
## 12.2 ¿Cómo puedo obtener la dirección IP y el número de puerto de una instancia de ZooKeeper?

Puede obtener la dirección IP y el número de puerto de una instancia de ZooKeeper a través de la consola MRS o FusionInsight Manager.

Método 1: Obtención de la dirección IP y el número de puerto de un ZooKeeper a través de la consola MRS

1. En la página **Dashboard**, haga clic en **Synchronize** en el lado derecho de **IAM User Sync** para sincronizar usuarios de IAM.
2. Haga clic en la pestaña **Components** y elija **ZooKeeper**. En la página mostrada, haga clic en **Instances** para ver la dirección IP empresarial de una instancia de ZooKeeper.

**Figura 12-1** Direcciones ZooKeeper



The screenshot shows the 'Components' page in the MRS console, specifically the 'Instances' tab for the 'Service ZooKeeper'. It displays a table with the following data:

<input type="checkbox"/>	Role	Host Name	OM IP Address	Business IP Address
<input type="checkbox"/>	quorumpeer	node-master1	192.	192
<input type="checkbox"/>	quorumpeer	node-master2	192.	192
<input type="checkbox"/>	quorumpeer	node-master3	192.	192

3. Haga clic en la pestaña **Service Configuration**. En la página mostrada, busque el parámetro **clientPort** para ver el número de puerto de la instancia ZooKeeper.

Método 2: Obtención de la dirección IP y el número de puerto de un ZooKeeper a través del FusionInsight Manager

1. Inicie sesión en FusionInsight Manager. Para obtener más información, consulte [Acceso a Manager](#).
2. Realice las siguientes operaciones para obtener la dirección IP y el número de puerto de una instancia de ZooKeeper.
  - Para grupos de MRS 3.x o anterior
    - i. Elija **Services > ZooKeeper**. En la página mostrada, haga clic en la pestaña **Instance** para ver la dirección IP empresarial de una instancia de ZooKeeper.

**Figura 12-2** Direcciones ZooKeeper

<input type="checkbox"/> Role	Host Name	OM IP Address	Business IP Address
<input type="checkbox"/> quorumpeer	node-master1	192.	192
<input type="checkbox"/> quorumpeer	node-master2	192.	192
<input type="checkbox"/> quorumpeer	node-master3	192.	192

- ii. Haga clic en la pestaña **Service Configuration**. En la página mostrada, busque el parámetro **clientPort** para ver el número de puerto de la instancia ZooKeeper.
- Para clústeres de MRS 3.x o posterior
  - i. Elija **Cluster > Services > ZooKeeper**. En la página mostrada, haga clic en la pestaña **Instance** para ver la dirección IP empresarial de una instancia de ZooKeeper.
  - ii. Haga clic en la pestaña **Configurations**. En la página mostrada, busque el parámetro **clientPort** para ver el número de puerto de la instancia ZooKeeper.

## 12.3 ¿Cómo accedo a un clúster MRS desde un nodo fuera del clúster?

### Creación de un ECS de Linux fuera del clúster para acceder al clúster MRS

**Paso 1** Cree un ECS fuera del clúster. Para obtener más información, consulte [Compra de un ECS](#).

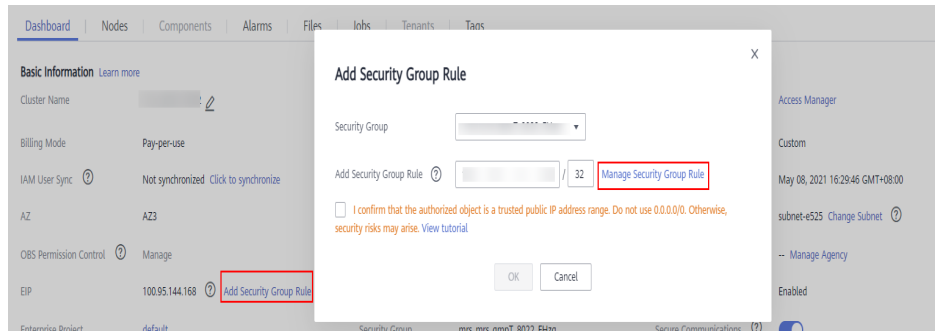
Establezca **AZ**, **VPC** y **Security Group** del ECS en los mismos valores que los del clúster al que se va a acceder.

**Paso 2** En la consola de gestión de VPC, solicite una EIP y vincúlela al ECS.

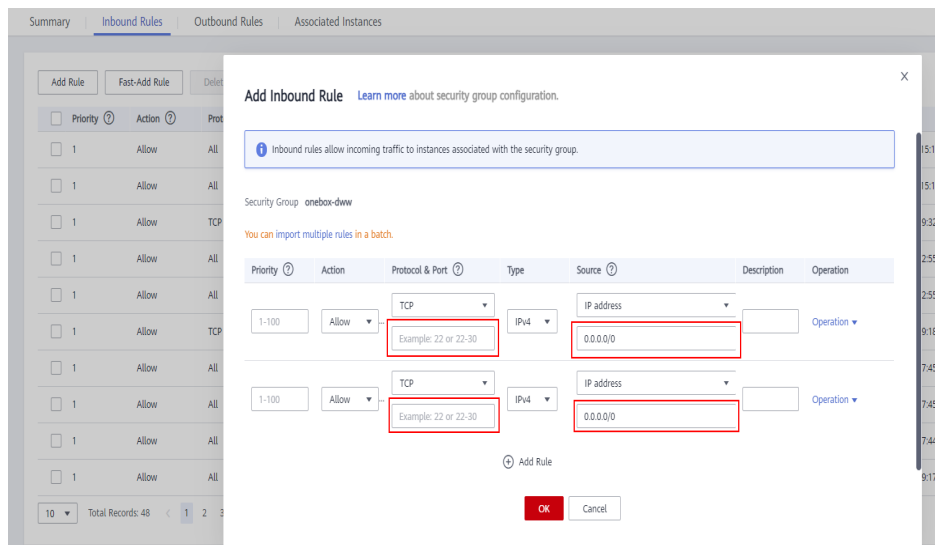
Para obtener más información, consulte [Asignación de un EIP](#).

**Paso 3** Configure reglas de grupo de seguridad para el clúster.

1. En la página de pestaña **Dashboard**, haga clic en **Add Security Group Rule**. En el cuadro de diálogo **Add Security Group Rule** que se muestra, haga clic en **Manage Security Group Rule**.



- Haga clic en la pestaña **Inbound Rules** y, a continuación, haga clic en **Add Rule**. En el cuadro de diálogo **Add Inbound Rule**, configure la dirección IP del ECS y habilite todos los puertos.



- Después de agregar la regla de grupo de seguridad, puede descargar e instalar el cliente en el ECS. Para obtener más información, consulte [Instalación de un cliente](#).
- Utilice el cliente.

Inicie sesión en el nodo cliente como usuario de instalación del cliente y ejecute el siguiente comando para cambiar al directorio de cliente:

```
cd /opt/hadoopclient
```

Ejecute el siguiente comando para cargar variables de entorno:

```
source bigdata_env
```

Si la autenticación de Kerberos está habilitada para el clúster, ejecute el siguiente comando para autenticar al usuario. Si la autenticación de Kerberos está deshabilitada para el clúster actual, la autenticación no es necesaria.

```
kinit MRS cluster user
```

Ejemplo:

```
kinit admin
```

Ejecute el comando de cliente de un componente.

Ejemplo:

Ejecute el siguiente comando para ver los archivos en el directorio raíz de HDFS:

```
hdfs dfs -ls /
```



```
Found 15 items
drwxrwx--x - hive      hive      0 2021-10-26 16:30 /apps
drwxr-xr-x - hdfs     hadoop   0 2021-10-18 20:54 /datasets
drwxr-xr-x - hdfs     hadoop   0 2021-10-18 20:54 /datastore
drwxrwx---+ - flink    hadoop   0 2021-10-18 21:10 /flink
drwxr-x--- - flume    hadoop   0 2021-10-18 20:54 /flume
drwxrwx--x - hbase    hadoop   0 2021-10-30 07:31 /hbase
...
```

----Fin

## Creación de un ECS de Windows fuera del clúster para acceder al MRS Manager

Cree un ECS de Windows en la consola de ECS para acceder a MRS Manager. Para obtener más información, consulte [Acceso al MRS Manager mediante un Windows ECS](#).

# 13 Desarrollo de servicios de big data

---

## 13.1 ¿Puede MRS ejecutar múltiples tareas de Flume a la vez?

El cliente Flume admite múltiples flujos de datos independientes. Puede configurar y vincular múltiples Source, Channel, y sink en el archivo de configuración **properties.properties**. Estos componentes se pueden vincular para formar múltiples flujos.

A continuación se muestra un ejemplo de configuración de dos flujos de datos en un archivo de configuración:

```
server.sources = source1 source2
server.sinks = sink1 sink2
server.channels = channel1 channel2

#dataflow1
server.sources.source1.channels = channel1
server.sinks.sink1.channel = channel1

#dataflow2
server.sources.source2.channels = channel2
server.sinks.sink2.channel = channel2
```

## 13.2 ¿Cómo cambio los registros de FlumeClient a los registros estándar?

1. Inicie sesión en el nodo donde está instalado el cliente Flume.
2. Vaya al directorio de instalación del cliente de Flume, por ejemplo, **/opt/FlumeClient**. Ejecute el siguiente comando:  
**cd /opt/FlumeClient/fusioninsight-flume-1.9.0/bin**
3. Ejecute el comando **./flume-manage.sh stop force** para detener FlumeClient.
4. Ejecute el comando **vi./conf/log4j.properties** para abrir el archivo **log4j.properties** y cambiar el valor **flume.root.logger** a **\${flume.log.level},console**.
5. Ejecute el comando **./flume-manage.sh start force** para reiniciar FlumeClient.
6. Después de la modificación, compruebe si la configuración de Docker es correcta.

## 13.3 ¿Dónde se almacenan los archivos JAR y las variables de entorno de Hadoop?

- **hadoopstreaming.jar**: `/opt/share/hadoop-streaming-*` (\* indica la versión de Hadoop.)
- Variables de entorno JDK: `/opt/client/JDK/component_env`
- Variables de entorno Hadoop: `/opt/client/HDFS/component_env`
- Cliente de Hadoop: `/opt/client/HDFS/hadoop`

## 13.4 ¿Qué algoritmos de compresión admite HBase?

HBase soporta los algoritmos de compresión Snappy, LZ4 y gzip.

## 13.5 ¿Puede MRS escribir datos en HBase a través de la tabla externa de HBase de Hive?

No. Hive en HBase solo admite consultas de datos.

## 13.6 ¿Cómo veo los registros de HBase?

1. Inicie sesión en el nodo master del clúster como usuario **root**.
2. Ejecute el comando `su - omm` para cambiar a usuario **omm**.
3. Ejecute el comando `cd /var/log/Bigdata/hbase/` para ir al directorio `/var/log/Bigdata/hbase/` y ver los registros de HBase.

## 13.7 ¿Cómo configuro el TTL para una tabla HBase?

- Establezca el tiempo de vida (TTL) al crear una tabla:  
Cree la tabla `t_task_log`, establezca la familia de columnas en `f` y establezca el TTL en **86400** segundos.

```
create 't_task_log',{NAME => 'f', TTL=>'86400'}
```

- Establezca el TTL para una tabla existente:  
`alter "t_task_log",NAME=>'data',TTL=>'86400' # Set the TTL value for the column family data.`

## 13.8 ¿Cómo me conecto a HBase de MRS a través de HappyBase?

ThriftServer1 y ThriftServer2 no pueden coexistir. El servicio HBase del clúster MRS utiliza ThriftServer2. Sin embargo, el HappyBase solo puede conectarse a HBase a través de la interfaz ThriftServer1. Por lo tanto, use Python para conectarse directamente a HBase. Para obtener más información, véase <https://github.com/huaweicloud/huaweicloud-mrs-example/blob/mrs-1.8/src/hbase-examples/hbase-python-example/DemoClient.py>.

## 13.9 ¿Cómo cambio el número de réplicas HDFS?

1. Vaya a la página de configuración del servicio HDFS.
  - Para clústeres MRS 1.8.10 o anteriores  
Inicie sesión en MRS Manager (consulte [Acceso a MRS Manager](#)), elija **Services > HDFS > Service Configuration** y seleccione **All** en la lista desplegable **Basic**.
  - Para MRS 1.8.10 o posterior y MRS 2.x, haga clic en el nombre del clúster. En la página de detalles del clúster que se muestra, elija **Components > HDFS > Service Configuration** y seleccione **All** en la lista desplegable **Basic**.

### NOTA

Si la pestaña **Components** no está disponible, primero complete la sincronización del usuario de IAM. (En la página **Dashboard**, haga clic en **Synchronize** en el lado derecho de **IAM User Sync** para sincronizar usuarios de IAM.)

- Para la versión 3.x o posterior del clúster MRS:  
Inicie sesión en FusionInsight Manager y elija **Cluster > Name of the desired cluster > Services > HDFS > Configurations > All Configurations**.
2. Busque **dfs.replication** y cambie el valor (rango de valores: 1 a 16), y reinicie la instancia HDFS.

## 13.10 ¿Cómo modifico la clase de conmutación HDFS activo y en espera?

Si la clase **org.apache.hadoop.hdfs.server.namenode.ha.AdaptiveFailoverProxyProvider** no está disponible cuando un clúster de MRS 3.x se conecta a NameNodes mediante HDFS, la causa es que la clase de conmutación HDFS activa/en espera del clúster está configurada incorrectamente. Para resolver el problema, realice las siguientes operaciones:

- Método 1: Agregue el paquete **hadoop-plugins-xxx.jar** al directorio **classpath** o **lib** de su programa.  
El paquete **hadoop-plugins-xxx.jar** se almacena en el directorio del cliente HDFS, por ejemplo, **\$HADOOP\_HOME/share/hadoop/common/lib/hadoop-plugins-8.0.2-302023.jar**.
- Método 2: Cambie el elemento de configuración de HDFS a la clase de código abierto correspondiente, como se muestra a continuación:  
`dfs.client.failover.proxy.provider.hacluster=org.apache.hadoop.hdfs.server.namenode.ha.ConfiguredFailoverProxyProvider`

## 13.11 ¿Cuál es el tipo de número recomendado de DynamoDB en las tablas Hive?

Se recomienda **smallint**.

## 13.12 ¿Se puede interconectar el controlador Hive con DBCP2?

El controlador Hive no se puede interconectar con el grupo de conexiones de base de datos DBCP2. El grupo de conexiones de base de datos DBCP2 invoca el método `isValid` para comprobar si hay una conexión disponible. Sin embargo, Hive lanza directamente una excepción al implementar este método.

## 13.13 ¿Cómo puedo ver la tabla Hive creada por otro usuario?

Versiones anteriores a MRS 3.x:

1. Inicie sesión en MRS Manager y elija **System > Permission > Manage Role**.
2. Haga clic en **Create Role** y defina **Role Name** y **Description**.
3. En la tabla **Permission**, elija **Hive > Hive Read Write Privileges**.
4. En la lista de base de datos, haga clic en el nombre de la base de datos donde se almacena la tabla creada por el usuario B. Se muestra la tabla.
5. En la columna **Permission** de la tabla creada por el usuario B, seleccione **SELECT**.
6. Haga clic en **OK** y vuelva a la página **Role**.
7. Elija **System > Manage User**. Localice la fila que contiene el usuario A, haga clic en **Modify** para enlazar el nuevo rol al usuario A y haga clic en **OK**. Después de unos 5 minutos, el usuario A puede acceder a la tabla creada por el usuario B.

MRS 3.x o posterior:

1. Inicie sesión en FusionInsight Manager y elija **Cluster > Services**. En la página que se muestra, elija **Hive**. En la página mostrada, elija **More** y compruebe si **Enable Ranger** está atenuado.
  - En caso afirmativo, vaya a **9**.
  - En caso negativo, realice **2 a 8**.
2. Inicie sesión en FusionInsight Manager y elija **System > Permission > Role**.
3. Haga clic en **Create Role** y defina **Role Name** y **Description**.
4. En la tabla **Configure Resource Permission**, elija *Name of the desired cluster* > **Hive > Hive Read Write Privileges**.
5. En la lista de base de datos, haga clic en el nombre de la base de datos donde se almacena la tabla creada por el usuario B. Se muestra la tabla.
6. En la columna **Permission** de la tabla creada por el usuario B, seleccione **Select**.
7. Haga clic en **OK** y vuelva a la página **Role**.
8. Elija **Permission > User**. En la página **Local User** que se muestra, busque la fila que contiene el usuario A, haga clic en **Modify** en la columna **Operation** para enlazar el nuevo rol al usuario A y haga clic en **OK**. Después de unos 5 minutos, el usuario A puede acceder a la tabla creada por el usuario B.
9. Realice los siguientes pasos para agregar la política de permisos de acceso de Ranger de Hive:

- a. Inicie sesión en FusionInsight Manager como administrador de Hive y elija **Cluster > Services**. En la página que se muestra, elija **Ranger**. En la página mostrada, haga clic en la URL junto a **Ranger WebUI** para ir a la página de gestión de Ranger.
  - b. En la página de inicio, haga clic en el nombre del complemento del componente en el área **HADOOP SQL**, por ejemplo, **Hive**.
  - c. En la página de pestaña **Access**, haga clic en **Add New Policy** para agregar una política de control de permisos de Hive.
  - d. En el cuadro de diálogo **Create Policy** que se muestra, establezca los siguientes parámetros:
    - **Policy Name**: Introduzca un nombre de política, por ejemplo, **table\_test\_hive**.
    - **database**: Ingrese o seleccione la base de datos donde se almacena la tabla creada por el usuario B, por ejemplo, **default**.
    - **table**: Ingrese o seleccione la tabla creada por el usuario B, por ejemplo, **test**.
    - **column**: Ingrese y seleccione una columna, por ejemplo, **\***.
    - En el área **Allow Conditions**, haga clic en **Select User** y seleccione el usuario A, haga clic en **Add Permissions** y seleccione **select**.
    - Haga clic en **Add**.
10. Realice los siguientes pasos para agregar la política de permisos de acceso de Ranger de HDFS:
- a. Inicie sesión en FusionInsight Manager como usuario **rangeradmin** y elija **Cluster > Services**. En la página que se muestra, elija **Ranger**. En la página mostrada, haga clic en la URL junto a **Ranger WebUI** para ir a la página de gestión de Ranger.
  - b. En la página de inicio, haga clic en el nombre del complemento del componente en el área **HDFS** por ejemplo, **hacluster**.
  - c. Haga clic en **Add New Policy** para agregar una política de control de permisos de HDFS.
  - d. En el cuadro de diálogo **Create Policy** que se muestra, establezca los siguientes parámetros:
    - **Policy Name**: Introduzca un nombre de política, por ejemplo, **tablehdfs\_test**.
    - **Resource Path**: Establezca este parámetro en la ruta de acceso HDFS donde se almacena la tabla creada por el usuario B, por ejemplo, **"/user/hive/warehouse/Database name/Table name"**.
    - En el área **Allow Conditions**, seleccione el usuario A para **Select User** y haga clic en **Add Permissions** en la columna **Permissions** y seleccione **Read** y **Execute**.
    - Haga clic en **Add**.
11. Vea información básica sobre la política en la lista de políticas. Después de que la política entre en vigor, el usuario A puede ver la tabla creada por el usuario B.

## 13.14 ¿Dónde puedo descargar el paquete de dependencias (com.huawei.gaussc10) en el proyecto de ejemplo de Hive?

MRS no tiene el paquete de dependencias **com.huawei.gaussc10**, que es un paquete de dependencias de GaussDB y no necesita ser configurado. Excluya este paquete cuando se construye un proyecto Maven.

## 13.15 ¿Puedo exportar el resultado de la consulta de datos de Hive?

Ejecute la siguiente sentencia para exportar el resultado de la consulta de los datos de Hive:

```
insert overwrite local directory "/tmp/out/" row format delimited fields terminated by "\t" select * from table;
```

## 13.16 ¿Cómo lo hago si ocurre un error cuando Hive ejecuta el comando `beeline -e` para ejecutar varias sentencias?

Cuando Hive de MRS 3.x ejecuta el comando `beeline -e " use default;show tables;"`, se muestra el siguiente mensaje de error: Error while compiling statement: FAILED: ParseException line 1:11 missing EOF at ';' near 'default' (state=42000,code=40000).

Soluciones:

- Método 1: Reemplace el comando `beeline -e " use default;show tables;"` por `beeline --entirelineascommand=false -e "use default;show tables;"`.
- Método 2:
  - a. En el directorio `/opt/Bigdata/client/Hive` del cliente Hive, cambie `export CLIENT_HIVE_ENTIRELINEASCOMMAND=true` en el archivo `component_env` a `export CLIENT_HIVE_ENTIRELINEASCOMMAND=false`.

Figura 13-1 Cambio del archivo `component_env`



```
PATH_NEW="echo $PATH | sed "s|/opt/Bigdata/client/Hive/Beeline/bin:||g" | sed "s|/opt/Bigdata/client/Hive/Beeline/bin:||g"
PATH_NEW="echo $PATH_NEW | sed "s|/opt/Bigdata/client/Hive/HCatalog/bin:||g" | sed "s|/opt/Bigdata/client/Hive/HCatalog/bin:||g"
export PATH=/opt/Bigdata/client/Hive/Beeline/bin:/opt/Bigdata/client/Hive/HCatalog/bin:$PATH_NEW
export CLIENT_HIVE_URI=jdbc:hive2://192.168.0.88:2181,192.168.0.9:2181,192.168.0.258:2181/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2
export HIVE_HOME=/opt/Bigdata/client/Hive/Beeline
export HIVE_LIB=/opt/Bigdata/client/Hive/Beeline/Lib
export HCAT_CONF_DIR=/opt/Bigdata/client/Hive/HCatalog/conf/
export CLIENT_HIVE_ENTIRELINEASCOMMAND=false
```

- b. Ejecute el siguiente comando para verificar la configuración:  
`source /opt/Bigdata/client/bigdata_env`  
`beeline -e " use default;show tables;"`

## 13.17 ¿Cómo lo hago si un trabajo "hivesql/hivescript" no se envía después de agregar Hive?

Este problema se produce porque el permiso `MRS CommonOperations` enlazado al grupo de usuarios al que pertenece el usuario que envía el trabajo no incluye el permiso Hive después de sincronizarse con Manager. Para solucionar este problema, realice las siguientes operaciones:

1. Añada el servicio Hive.

2. Inicie sesión en la consola de IAM y cree un grupo de usuarios. La política enlazada al grupo de usuarios es la misma que la del grupo de usuarios al que pertenece el usuario que envía el trabajo.
3. Agregue el usuario que envía el trabajo al nuevo grupo de usuarios.
4. Actualice la página de detalles del clúster en la consola MRS. El estado de la sincronización de usuario de IAM es **Not synchronized**.
5. Haga clic en **Synchronize** a la derecha de **IAM User Sync**. Vuelva a la página anterior. En el panel de navegación de la izquierda, elija **Operation Logs** y compruebe si el usuario ha cambiado.
  - En caso afirmativo, vuelva a enviar el trabajo Hive.
  - Si no, compruebe si todas las operaciones anteriores están completadas.
    - En caso afirmativo, póngase en contacto con el personal de O&M.
    - Si no, ejecute el trabajo de Hive una vez completadas las operaciones anteriores.

## 13.18 ¿Qué pasa si un archivo de Excel descargado en Hue no se puede abrir?

### NOTA

Esta sección solo se aplica a versiones anteriores a MRS 3.x.

1. Inicie sesión en un nodo Master como usuario **root** y cambie a usuario **omm**.  
**su - omm**
2. Compruebe si el nodo actual es el nodo OMS activo.

**sh \${BIGDATA\_HOME}/om-0.0.1/sbin/status-oms.sh**

Si se muestra **active** en la salida del comando, el nodo es el nodo activo. De lo contrario, inicie sesión en el otro nodo Master.

Figura 13-2 Nodo OMS activo

NodeName	HostName	HAVersion	StartTime	HAActive	HAAllResOK	HARunPhase
192-168-67-136			2022-09-27 20:23:41	Active	normal	Active
192-168-67-142			2022-09-27 20:24:39	standby	normal	Deactivated

NodeName	ResName	ResStatus	ResHAStatus	ResType
192-168-67-136	acs	Normal	Normal	Single_active
192-168-67-136	oss	Normal	Normal	Single_active
192-168-67-136	cep	Normal	Normal	Single_active
192-168-67-136	controller	Normal	Normal	Single_active
192-168-67-136	disaster	Normal	Normal	Single_active
192-168-67-136	floatip	Normal	Normal	Single_active
192-168-67-136	fs	Normal	Normal	Single_active
192-168-67-136	gaussDB	Active_normal	Normal	Active_standby
192-168-67-136	heartbeatcheck	Normal	Normal	Single_active
192-168-67-136	httpd	Normal	Normal	Single_active
192-168-67-136	iam	Normal	Normal	Single_active
192-168-67-136	ntp	Active_normal	Normal	Active_standby
192-168-67-136	okerberos	Active_normal	Normal	Active_standby
192-168-67-136	oldap	Active_normal	Normal	Active_standby
192-168-67-136	pms	Normal	Normal	Single_active
192-168-67-136	tomcat	Normal	Normal	Single_active
192-168-67-142	acs	Stopped	Normal	Single_active
192-168-67-142	oss	Stopped	Normal	Single_active
192-168-67-142	cep	Stopped	Normal	Single_active
192-168-67-142	controller	Stopped	Normal	Single_active
192-168-67-142	disaster	Stopped	Normal	Single_active
192-168-67-142	floatip	Stopped	Normal	Single_active
192-168-67-142	fs	Stopped	Normal	Single_active
192-168-67-142	gaussDB	Stopped_normal	Normal	Single_active
192-168-67-142	heartbeatcheck	Stopped	Normal	Single_active
192-168-67-142	httpd	Stopped	Normal	Single_active
192-168-67-142	iam	Stopped	Normal	Single_active
192-168-67-142	ntp	Standby_normal	Normal	Active_standby
192-168-67-142	okerberos	Standby_normal	Normal	Active_standby
192-168-67-142	oldap	Standby_normal	Normal	Active_standby
192-168-67-142	pms	Stopped	Normal	Single_active
192-168-67-142	tomcat	Stopped	Normal	Single_active

3. Vaya al directorio **{BIGDATA\_HOME}/Apache-httpd-\*/conf**.  
**cd \${BIGDATA\_HOME}/Apache-httpd-\*/conf**
4. Abra el archivo **httpd.conf**.  
**vim httpd.conf**
5. Busque **21201** en el archivo y elimine el siguiente contenido del archivo. (Los valores de *proxy\_ip* y *proxy\_port* son los mismos que en el entorno real.)



```
ProxyHTMLEnable On
SetEnv PROXY_PREFIX=https://[proxy_ip]:[proxy_port]
ProxyHTMLURLMap (https?:\//\[^\:]*:[0-9]*.*) ${PROXY_PREFIX}/proxyRedirect=$1 RV
RV
```

Figura 13-3 Contenido que se eliminará



```
494 <VirtualHost *:80>
495   ServerName https://192.168.0.175:8080
496   SSLProxyEngine On
497   ProxyRequests Off
498   Traceable Off
499   ProxyTimeout 1200
500   RewriteEngine On
501   ProxyHTMLEnable On
502   # LogLevel: warn: rewrite:trace6
503   RewriteMap proxylist dbm:/opt/bigdata/Apache-httpd-2.4.26/conf/proxylist.dbm
504
505   SetEnv PROXY_PREFIX=https://192.168.0.175:20026
506   ProxyHTMLURLMap (https?:\//\[^\:]*:[0-9]*.*) ${PROXY_PREFIX}/proxyRedirect=$1 RV
507
508   RewriteRule ^(\//.*)$ ${proxylist:Hue}$1 [E=TARGET_PATH:$1.L.P]
509
510   Header edit Location "(?!(https://[cas_ip]:[cas_port]|https://192.168.0.175:20009|https://192.168.0.175:8080|https://192.168.0.175:8080)$ https://192.168.0.175:8080)$1
511   ProxyPassReverseCookiePath / / interpolate
512
```

6. Guarde la modificación y salga.
7. Vuelva a abrir el archivo **httpd.conf**, busque **proxy\_hue\_port** y elimine el siguiente contenido:

```
ProxyHTMLEnable On
SetEnv PROXY_PREFIX=https://[proxy_ip]:[proxy_port]
ProxyHTMLURLMap (https?:\//\[^\:]*:[0-9]*.*) ${PROXY_PREFIX}/proxyRedirect=$1 RV
RV
```

Figura 13-4 Contenido que se eliminará



```
494 <VirtualHost *:80>
495   ServerName https://[proxy_ip]:[proxy_hue_port]
496   SSLProxyEngine On
497   ProxyRequests Off
498   Traceable Off
499   ProxyTimeout 1200
500   RewriteEngine On
501   ProxyHTMLEnable On
502   # LogLevel: warn: rewrite:trace6
503   RewriteMap proxylist dbm:[httpd_home]/conf/proxylist.dbm
504
505   SetEnv PROXY_PREFIX=https://[proxy_ip]:[proxy_port]
506   ProxyHTMLURLMap (https?:\//\[^\:]*:[0-9]*.*) ${PROXY_PREFIX}/proxyRedirect=$1 RV
507
508   RewriteRule ^(\//.*)$ ${proxylist:Hue}$1 [E=TARGET_PATH:$1.L.P]
509
510   Header edit Location "(?!(https://[cas_ip]:[cas_port]|https://[proxy_ip]:[proxy_hue_port]|https://192.168.0.175:20009|https://192.168.0.175:8080|https://192.168.0.175:8080)$ https://[proxy_ip]:[proxy_hue_port]$1
511   ProxyPassReverseCookiePath / / interpolate
512
```

8. Guarde la modificación y salga.
9. Ejecute el siguiente comando para reiniciar el proceso **httpd**:  
**sh \${BIGDATA\_HOME}/Apache-httpd-\*/setup/restarthttpd.sh**
10. Compruebe si el archivo **httpd.conf** del nodo Master en espera está modificado. Si se modifica el archivo, no se requiere ninguna otra acción. Si el archivo no se modifica, modifique el archivo **httpd.conf** en el nodo Master en espera de la misma manera. No es necesario reiniciar el proceso **httpd**.
11. Descargue el archivo de Excel de nuevo. Puede abrir el archivo correctamente.

## 13.19 ¿Cómo lo hago si las sesiones no se liberan después de que Hue se conecta a HiveServer y se muestra el mensaje de error "ver max user connections"?

Versiones aplicables: MRS 3.1.0 y anteriores

1. Modifique el siguiente archivo en los dos nodos de Hue:  
`/opt/Bigdata/FusionInsight_Porter_8.*/install/FusionInsight-Hue-*/hue/apps/ beeswax/src/beeswax/models.py`

2. Cambie las configuraciones en las líneas 396 y 404.

Cambie **q Changed = self.filter(owner=user, application=application).exclude(guid='').exclude(secret='')** a **q = self.filter(owner=user, application=application).exclude(guid=None).exclude(secret=None)**.

```

394 def get_session(self, user, application='beeswax', filter_open=True):
395     try:
396         q = self.filter(owner=user, application=application).exclude(guid='').exclude(secret='')
397         if filter_open:
398             q = q.filter(status_code=0)
399         return q.latest("last_used")
400     except:
401         return None
402
403 def get_sessions(self, user, application='beeswax', filter_open=True):
404     q = self.filter(owner=user, application=application).exclude(guid='').exclude(secret='')
405     if filter_open:
406         q = q.filter(status_code=0)
407     q = q.order_by("-last_used")
    
```

```

394 def get_session(self, user, application='beeswax', filter_open=True):
395     try:
396         q = self.filter(owner=user, application=application).exclude(guid=None).exclude(secret=None)
397         if filter_open:
398             q = q.filter(status_code=0)
399         return q.latest("last_used")
400     except:
401         return None
402
403 def get_sessions(self, user, application='beeswax', filter_open=True):
404     q = self.filter(owner=user, application=application).exclude(guid=None).exclude(secret=None)
405     if filter_open:
406         q = q.filter(status_code=0)
407     q = q.order_by("-last_used")
    
```

## 13.20 ¿Cómo se restablecen los datos de Kafka?

Puede restablecer los datos de Kafka eliminando topics de Kafka.

- Eliminar un topic: **kafka-topics.sh --delete --zookeeper ZooKeeper Cluster service IP address:2181/kafka --topic topicname**
- Consultar todos los topics: **kafka-topics.sh --zookeeper ZooKeeper cluster service IP address:2181/kafka --list**

Después de ejecutar el comando de eliminación, los topics vacíos se eliminarán inmediatamente. Si un topic tiene datos, el topic será marcado para su eliminación y será eliminado por Kafka más adelante.

## 13.21 ¿Cómo obtengo la versión de cliente de MRS Kafka?

Ejecute el comando **--bootstrap-server** para consultar la información sobre el cliente.

## 13.22 ¿Qué protocolos de acceso son compatibles con Kafka?

Kafka admite PLAINTEXT, SSL, SASL\_PLAINTEXT y SASL\_SSL.

## 13.23 ¿Cómo lo hago si se muestra un mensaje de error "Not Authorized to access group xxx" cuando se consume un Kafka topic?

Este problema se debe al conflicto entre la autenticación de Ranger y la autenticación de ACL de un clúster. Si un clúster de Kafka utiliza ACL para el control de acceso de permisos y la autenticación de Ranger está habilitada para el componente de Kafka, todas las autenticaciones del componente son gestionadas por Ranger. Los permisos establecidos por el complemento de autenticación original no son válidos. Como resultado, la autorización de ACL no tiene efecto. Puede desactivar la autenticación Ranger de Kafka y reiniciar el servicio Kafka para corregir la falla. El procedimiento es el siguiente:

1. Inicie sesión en FusionInsight Manager y elija **Cluster > Services > Kafka**.
2. En la esquina superior derecha de la página **Dashboard**, haga clic en **More** y elija **Disable Ranger**. En el cuadro de diálogo que se muestra, escriba la contraseña y haga clic en **OK**. Una vez que la operación se haya realizado correctamente, haga clic en **Finish**.
3. En la esquina superior derecha de la página **Dashboard**, haga clic en **More** y elija **Restart Service** para reiniciar el servicio Kafka.

## 13.24 ¿Qué algoritmos de compresión soporta Kudu?

Kudu es compatible con **Snappy**, **LZ4** y **zlib**. **LZ4** se utiliza de forma predeterminada.

## 13.25 ¿Cómo puedo ver los registros de Kudu?

1. Inicie sesión en el nodo Master del clúster.
2. Ejecute el comando **su - omm** para cambiar a usuario **omm**.
3. Ejecute el comando **cd /var/log/Bigdata/kudu/** para ir al directorio **/var/log/Bigdata/kudu/** y ver los registros de Kudu.

## 13.26 ¿Cómo manejo las excepciones del servicio Kudu generadas durante la creación de clústeres?

### Consulta de los registros de excepciones del servicio Kudu

1. Inicie sesión en la consola de MRS.
2. Haga clic en el nombre del clúster.
3. En la página mostrada, elija **Components > Kudu > Instances** y busque la dirección IP de la instancia anormal.

Si la pestaña **Components** no está disponible, primero complete la sincronización del usuario de IAM. (En la página **Dashboard**, haga clic en **Synchronize** en el lado derecho de **IAM User Sync** para sincronizar usuarios de IAM.)

4. Inicie sesión en el nodo donde reside la instancia anormal y vea el registro de Kudu.

```
cd /var/log/Bigdata/Kudu
[root@node-master1AERu kudu]# ls
healthchecklog  runninglog  startlog
```

Puede encontrar los registros de comprobación de estado de Kudu en el directorio **healthchecklog**, los registros de inicio en el directorio **startlog** y los registros de ejecución del proceso Kudu en el directorio **runninglog**.

```
[root@node-master1AERu logs]# pwd
/var/log/Bigdata/kudu/runninglog/master/logs
[root@node-master1AERu logs]# ls -al
kudu-master.ERROR  kudu-master.INFO  kudu-master.WARNING
```

Los registros de ejecución se clasifican en tres tipos: **ERROR**, **INFO**, y **WARNING**. Cada tipo de registros de ejecución se registra en el archivo correspondiente. Puede ejecutar el comando **cat** para ver los registros de ejecución de cada tipo.

## Manejo de excepciones de servicio Kudu

El archivo `/var/log/Bigdata/kudu/runninglog/master/logs/kudu-master.INFO` contiene la siguiente información de error:

```
"Unable to init master catalog manager: not found: Unable to initialize catalog manager: Failed to initialize sys tables async: Unable to load consensus metadata for tablet 000000000000000000000000: xxx"
```

Si esta excepción ocurre cuando el servicio Kudu se instala por primera vez, el servicio KuduMaster no se inicia. La incoherencia de datos provoca el error de inicio. Para resolver el problema, realice los siguientes pasos para borrar los directorios de datos y reinicie el servicio Kudu. Si el servicio Kudu no se instala por primera vez, borrar los directorios de datos provocará la pérdida de datos. En este caso, migre datos y borre el directorio de datos.

1. Busque los directorios de datos `fs_data_dir`, `fs_wal_dir` y `fs_meta_dir`.  

```
find /opt -name master.gflagfile
```

```
cat /opt/Bigdata/FusionInsight_Kudu_*/*_KuduMaster/etc/master.gflagfile | grep fs_
```
2. En la página de detalles del clúster, elija **Components > Kudu** y haga clic en **Stop Service**.
3. Borre los directorios de datos Kudu en todos los nodos KuduMaster y KuduTserver. El siguiente comando utiliza dos discos de datos como ejemplo.  

```
rm -Rvf /srv/Bigdata/data1/kudu, rm -Rvf /srv/Bigdata/data2/kudu
```
4. En la página de detalles del clúster, elija **Components > Kudu** y elija **More > Restart Service**.
5. Compruebe el estado del servicio Kudu y los registros.

## 13.27 ¿Cuáles son las diferencias entre la construcción de proyectos de ejemplo y el desarrollo de aplicaciones? ¿Es compatible el código Python?

- El proyecto de muestra y el desarrollo de aplicaciones en MRS son los mismos. Puede seleccionar cualquiera de ellos.
- MRS soporta código Python.

## 13.28 ¿OpenTSDB soporta las API de Python?

OpenTSDB soporta las API de Python. OpenTSDB proporciona RESTful APIs basadas en HTTP que son independientes del lenguaje. Cualquier idioma que admita solicitudes HTTP puede interconectarse con OpenTSDB.

## 13.29 ¿Cómo configuro otras fuentes de datos en Presto?

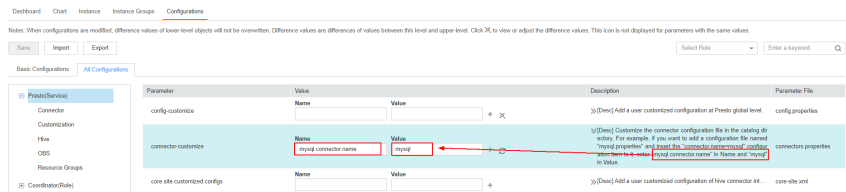
En esta sección, MySQL se utiliza como ejemplo.

- Para los clústeres MRS 1.x y 3.x, haga lo siguiente:
  - a. Inicie sesión en la consola de gestión de MRS.

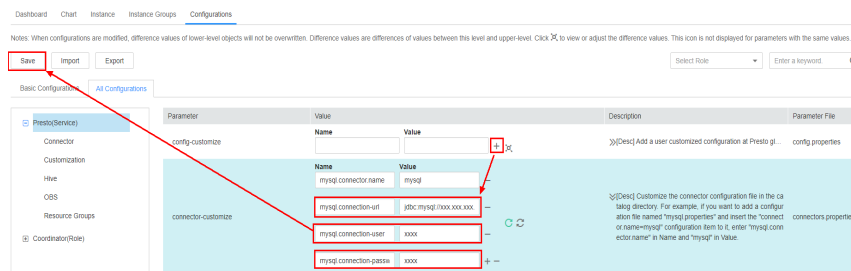
- b. Haga clic en el nombre del clúster para ir a su página de detalles.
- c. Haga clic en la pestaña **Components** y, a continuación, haga **Presto** en la lista de componentes. En la página que se muestra, haga clic en la pestaña **Configurations** y luego en la subpestaña **All Configurations**.
- d. En la página de configuración de Presto que se muestra, busque **connector-customize**.
- e. Establezca **Name** y **Value** de la siguiente manera:

**Name:** mysql.connector.name

**Value:** mysql



- f. Haga clic en el signo más (+) para agregar tres campos más y establezca **Name** y **Value** de acuerdo con la siguiente tabla. A continuación, haga clic en **Save**.



Nombre	Valor	Descripción
mysql.connection-url	jdbc:mysql:// xxx.xxx.xxx.xxx:3306	Grupo de conexiones de base de datos
mysql.connection-user	xxxx	Nombre de usuario de la base de datos
mysql.connection-password	xxxx	Contraseña de base de datos

- g. Reinicie el servicio Presto.
- h. Ejecute el siguiente comando para conectarse al servidor Presto del clúster:  
**presto\_cli.sh --krb5-config-path** {krb5.conf path} **--krb5-principal** {User principal} **--krb5-keytab-path** {user.keytab path} **--user** {presto username}
- i. Inicie sesión en Presto y ejecute el comando **show catalogs** para comprobar si se puede consultar la lista de orígenes de datos mysql de Presto.

```
[root@node-master2uoHG bin]# ./presto_cli.sh
--server http://152.157.136.20
show catalogs:
Catalog
-----
hive
jmx
mysql
system
tpcds
tpch
(6 rows)

Query 20220422_121338_00002_ra2vb, FINISHED, 3 nodes
Splits: 53 total, 53 done (100.00%)
0:00 [0 rows, 0B] [0 rows/s, 0B/s]
```

Ejecute el comando **show schemas from mysql** para consultar la base de datos de MySQL.

- Para los clústeres de *x* de MRS 2., haga lo siguiente:
  - a. Cree el archivo de configuración **mysql.properties** que contiene el siguiente contenido:  
connector.name=mysql  
connection-url=jdbc:mysql://mysqlIp:3306  
connection-user=Username  
connection-password=Password

 **NOTA**

- **mysqlIp** indica la dirección IP de la instancia MySQL, que debe poder comunicarse con la red MRS.
  - El nombre de usuario y la contraseña son los que se usan para iniciar sesión en la base de datos MySQL.
- b. Cargue el archivo de configuración al directorio **/opt/Bigdata/MRS\_Current/1\_14\_Coordinator/etc/catalog/** en el nodo master (donde reside la instancia de Coordinador) y al directorio **/opt/Bigdata/MRS\_Current/1\_14\_Worker/etc/catalog/** en el nodo core (dependiendo del directorio real en el clúster) y cambie el grupo de propietarios de archivos a **omm:wheel**.
  - c. Reinicie el servicio Presto.

## 13.30 ¿Cómo actualizo el certificado Ranger?

Se usa MRS 1.9.3 como ejemplo. Reemplácelo con la versión real del clúster. Después de actualizar el certificado, borre manualmente la alarma que indica que el archivo de certificado no es válido o está a punto de caducar.

 **NOTA**

Después de actualizar el certificado Ranger, su período de validez es de 10 años.

Después de que el certificado de Ranger caduque, la interfaz de usuario web de Ranger sigue siendo accesible, pero se mostrará un mensaje que indica que el certificado no es de confianza cuando acceda a la interfaz de usuario web.

- Si Ranger no está instalado en el clúster, inicie sesión en cada nodo master y ejecute el siguiente comando para cambiar el nombre del archivo de certificado:  
**mv /opt/Bigdata/MRS\_1.9.3/install/MRS-Ranger-1.0.1/ranger/ranger-1.0.1-admin/ranger-admin-keystore.jks /opt/Bigdata/MRS\_1.9.3/install/MRS-Ranger-1.0.1/ranger/ranger-1.0.1-admin/ranger-admin-keystore.jks\_bak**

- Si Ranger se ha instalado en el clúster, actualice el certificado de la siguiente manera:
  - a. Descargue **MRS\_1.9\_Patch\_UpdateRangerJks\_All\_20210203.tar.gz** del bucket obs-patch y súbalo al directorio **/tmp** en el nodo donde se ejecuta la instancia activa de RangerAdmin del clúster.

En MRS Manager, elija **Service > Ranger > Instance** y obtenga la dirección IP del nodo donde se ejecuta la instancia activa de RangerAdmin.

- CN-Hong Kong: [https://mrs-patch-ap-southeast-1.obs.ap-southeast-1.myhuaweicloud.com/MRS\\_Common\\_Script/MRS\\_1.9\\_Patch\\_UpdateRangerJks\\_All\\_20210203.tar.gz](https://mrs-patch-ap-southeast-1.obs.ap-southeast-1.myhuaweicloud.com/MRS_Common_Script/MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz)
  - AP-Bangkok: [https://mrs-patch-ap-southeast-2.obs.ap-southeast-2.myhuaweicloud.com/MRS\\_Common\\_Script/MRS\\_1.9\\_Patch\\_UpdateRangerJks\\_All\\_20210203.tar.gz](https://mrs-patch-ap-southeast-2.obs.ap-southeast-2.myhuaweicloud.com/MRS_Common_Script/MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz)
  - AP-Singapore: [https://mrs-patch-ap-southeast-3.obs.ap-southeast-3.myhuaweicloud.com/MRS\\_Common\\_Script/MRS\\_1.9\\_Patch\\_UpdateRangerJks\\_All\\_20210203.tar.gz](https://mrs-patch-ap-southeast-3.obs.ap-southeast-3.myhuaweicloud.com/MRS_Common_Script/MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz)
  - LA-Sao Paulo: [https://mrs-container1-patch-sa-brazil-1.obs.myhuaweicloud.com/MRS\\_Common\\_Script/MRS\\_1.9\\_Patch\\_UpdateRangerJks\\_All\\_20210203.tar.gz](https://mrs-container1-patch-sa-brazil-1.obs.myhuaweicloud.com/MRS_Common_Script/MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz)
  - LA-Ciudad de México: [https://mrs-container1-patch-na-mexico-1.obs.myhuaweicloud.com/MRS\\_Common\\_Script/MRS\\_1.9\\_Patch\\_UpdateRangerJks\\_All\\_20210203.tar.gz](https://mrs-container1-patch-na-mexico-1.obs.myhuaweicloud.com/MRS_Common_Script/MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz)
- b. Inicie sesión en el nodo donde se encuentra la instancia activa de RangerAdmin y ejecute los siguientes comandos:

```
cd /tmp
chmod 700 MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz
chown omm:wheel MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz
su - omm
cd /tmp
tar -zxvf MRS_1.9_Patch_UpdateRangerJks_All_20210203.tar.gz
```
  - c. Replace the certificate files.

```
cd updateRangerJks
sh updateRangerJks.sh #{IP address of the active master node} #{IP address of the active RangerAdmin node} #{Certificate password}
```

#### NOTA

- Este script reiniciará el proceso del controller. Durante el proceso de reinicio, es posible que no se vea la página MRS Manager.
  - Obtenga la dirección IP del nodo master activo de **Hosts** en MRS Manager.
  - Para obtener la dirección IP del nodo RangerAdmin activo, elija **Services > Ranger > Instances** en MRS Manager.
  - *#{Certificate password}* es una contraseña definida por el usuario.
- d. Inicie sesión en la consola de MRS.
  - e. Elija **Clusters > Active Clusters** y haga clic en un nombre de clúster para ir a la página de detalles del clúster.
  - f. Elija **Components > Ranger > Service Configuration** y modifique la configuración del RangerAdmin.

- i. Busque el **policymgr\_https\_keystore\_password** y cambie su valor a la contraseña del certificado introducida en **c**, es decir, *\${Certificate password}*. Se recomienda copiar y pegar la contraseña. Si las contraseñas son diferentes, Ranger no se reiniciará.
- ii. Guarde la configuración y realice un reinicio continuo de RangerAdmin.
- g. Compruebe que puede iniciar sesión en la interfaz de usuario web de RangerAdmin.
  - i. Elija **Components > Ranger > Service Status**. Haga clic en **Ranger Summary** y haga clic en **RangerAdmin** correspondiente a **Ranger Web UI**.
  - ii. En la página de inicio de sesión de la interfaz de usuario web de Ranger, el nombre de usuario predeterminado para el clúster MRS 1.9.2 es **admin** y la contraseña es **admin@12345**. El nombre de usuario predeterminado para el clúster MRS 1.9.3 o posterior es **admin** y la contraseña **ranger@A1!**. Después de iniciar sesión en la interfaz de usuario web de Ranger por primera vez, cambie la contraseña y manténgala segura.
- h. Inicie sesión en el nodo donde se encuentra la instancia RangerAdmin y elimine los archivos temporales.

```
rm -rf /tmp/updateRanger.Jks
rm -rf /tmp/updateRanger.Jks.tar.gz
```

Para un clúster con una topología personalizada, si las instancias master y RangerAdmin activas no están en el mismo nodo, inicie sesión en el nodo master activo y elimine los archivos temporales.

## 13.31 ¿Cómo me conecto a Spark Shell desde MRS?

1. Inicie sesión en el nodo master del clúster como usuario **root**.
2. Ejecute el siguiente comando para configurar las variables de entorno:

```
source Client installation directory/bigdata_env
```
3. Si la autenticación de Kerberos está habilitada para el clúster, autentifique al usuario. Si la autenticación Kerberos está deshabilitada, omita este paso.  
Comando: **kinit MRS cluster user**  
Ejemplo:
  - Si el usuario es un usuario máquina-máquina, ejecute **kinit -kt user.keytab sparkuser**.
  - Si el usuario es un usuario hombre-máquina, ejecute **kinit sparkuser**.
4. Ejecute el siguiente comando para conectarse a Spark shell:

```
spark-shell
```

## 13.32 ¿Cómo me conecto a Spark Beeline desde MRS?

1. Inicie sesión en el nodo master del clúster como usuario **root**.
2. Ejecute el siguiente comando para configurar las variables de entorno:

```
source Client installation directory/bigdata_env
```
3. Si la autenticación de Kerberos está habilitada para el clúster, autentifique al usuario. Si la autenticación Kerberos está deshabilitada, omita este paso.



Comando: **kinit MRS cluster user**

Ejemplo:

- Si el usuario es un usuario máquina-máquina, ejecute **kinit -kt user.keytab sparkuser**.
- Si el usuario es un usuario hombre-máquina, ejecute **kinit sparkuser**.

4. Ejecute el siguiente comando para conectarse a Spark Beeline:

**spark-beeline**

5. Ejecute comandos en Spark Beeline. Por ejemplo, cree la tabla **test** en el directorio **obs://mrs-word001/table/**.

**create table test(id int) location 'obs://mrs-word001/table/';**

6. Consultar todas las tablas.

**show tables;**

Si se muestra el **test** de tabla en la salida del comando, se accede a OBS correctamente.

**Figura 13-5** Nombre de tabla devuelto

```
0: jdbc:hive2://ha-cluster/default> create table test(id int) location 'obs://mrs-word001/table/';
+-----+
| Result |
+-----+
+-----+
No rows selected (2.515 seconds)
0: jdbc:hive2://ha-cluster/default> show tables;
+-----+
| database | tableName | isTemporary |
+-----+
| default  | test      | false       |
| default  | test_obs  | false       |
+-----+
2 rows selected (0.127 seconds)
```

7. Presione **Ctrl+C** para salir de Spark beeline.

## 13.33 ¿Dónde se almacenan los registros de ejecución de los trabajos de Spark?

- Los registros de los trabajos de Spark sin terminar se almacenan en el directorio **/srv/BigData/hadoop/data1/nm/containerlogs/** del nodo Core.
- Los registros de los trabajos de Spark terminados se almacenan en el directorio **/tmp/logs/username/logs** de HDFS.

## 13.34 ¿Cómo especifico una ruta de acceso de registro al enviar una tarea en un clúster MRS Storm?

Puede modificar el archivo **/opt/Bigdata/MRS\_XXX/1\_XX\_Supervisor/etc/worker.xml** en el nodo Core de streaming de MRS, establecer el valor de **filename** en la ruta de acceso y reiniciar la instancia correspondiente en Manager.

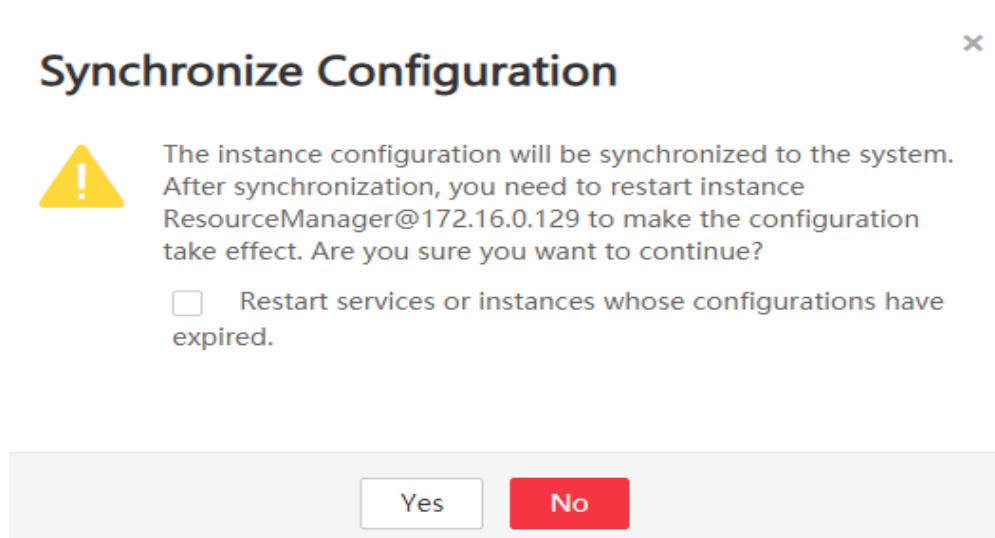
Se recomienda no modificar la configuración de registro predeterminada de MRS. De lo contrario, el sistema de registro puede volverse anormal.

## 13.35 ¿Cómo puedo comprobar si la configuración ResourceManager de Yarn es correcta?

**Paso 1** Inicie sesión en MRS Manager y elija **Services > Yarn > Instance**.

**Paso 2** Sincronice la configuración entre los dos nodos ResourceManager. Realice los siguientes pasos en cada nodo ResourceManager: Haga clic en el nombre del nodo ResourceManager y elija **More > Synchronize Configuration**. En el cuadro de diálogo que aparece, anule la selección **Restart services or instances whose configurations have expired** y haga clic en **Yes**.

Figura 13-6 Cuadro de diálogo Synchronize Configuration



**Paso 3** Haga clic en **Yes** para sincronizar la configuración.

**Paso 4** Inicie sesión en los nodos Master como usuario **root**.

**Paso 5** Ejecute el comando `cd /opt/Bigdata/MRS_Current/*_*_ResourceManager/etc_UPDATED/` para ir al directorio `etc_UPDATED`.

**Paso 6** Ejecute el comando `grep '\.queues' capacity-scheduler.xml -A2` para mostrar todas las colas configuradas y comprobar si las colas son coherentes con las que se muestran en Manager.

**root-default** está oculto en la página Manager.

```
[omm@node-master111ZA etc]$  
[omm@node-master111ZA etc]$ grep '\.queues' capacity-scheduler.xml -A2  
<name>yarn.scheduler.capacity.root.queues</name>  
<value>default,root-default,launcher-job,test1,test2,test3,test4</value>  
</property>  
[omm@node-master111ZA etc]$  
[omm@node-master111ZA etc]$
```

**Paso 7** Ejecute el comando `grep '\.capacity</name>' capacity-scheduler.xml -A2` para mostrar el valor de cada cola y comprobar si el valor de cada cola es el mismo que el mostrado en Manager. Compruebe si la suma de los valores configurados para todas las colas es **100**.

- Si la suma es **100**, la configuración es correcta.

- Si la suma no es **100**, la configuración es incorrecta. Realice los siguientes pasos para rectificar la falla.

```
[omm@node-master117A etc]$
[omm@node-master117A etc]$ grep '\.capacity</name>' capacity-scheduler.xml -A2
<name>yarn.scheduler.capacity.root.root-default.accessible-node-labels.zhaolu.capacity</name>
<value>0.0</value>
</property>
--
<name>yarn.scheduler.capacity.root.launcher-job.capacity</name>
<value>10</value>
</property>
--
<name>yarn.scheduler.capacity.root.accessible-node-labels.zhaolu.capacity</name>
<value>100</value>
</property>
--
<name>yarn.scheduler.capacity.root.test1.capacity</name>
<value>10</value>
</property>
--
<name>yarn.scheduler.capacity.root.test2.capacity</name>
<value>10</value>
</property>
--
<name>yarn.scheduler.capacity.root.test3.capacity</name>
<value>10</value>
</property>
--
<name>yarn.scheduler.capacity.root.capacity</name>
<value>100</value>
</property>
--
<name>yarn.scheduler.capacity.root.root-default.capacity</name>
<value>40.0</value>
</property>
--
<name>yarn.scheduler.capacity.root.test4.accessible-node-labels.zhaolu.capacity</name>
<value>100</value>
</property>
--
<name>yarn.scheduler.capacity.root.test4.capacity</name>
<value>0</value>
</property>
--
<name>yarn.scheduler.capacity.root.default.capacity</name>
<value>20</value>
</property>
[omm@node-master117A etc]$
```

- Paso 8** Inicie sesión en MRS Manager y seleccione **Hosts**.
- Paso 9** Determinar el nodo Master activo. El nombre de host del nodo Master activo comienza con un pentágono sólido.
- Paso 10** Inicie sesión en el nodo maestro activo como usuario **root**.
- Paso 11** Ejecute el comando **su - omm** para cambiar a usuario **omm**.
- Paso 12** Ejecute el comando **sh /opt/Bigdata/om-0.0.1/sbin/restart-controller.sh** para reiniciar el controller cuando no se esté realizando ninguna operación en Manager.
- El reinicio del controller no afectará a los servicios de componentes de big data.
- Paso 13** Repita **Paso 1** a **Paso 7** para sincronizar las configuraciones de ResourceManager y comprobar si las configuraciones son correctas.

Si la última configuración no se ha cargado después de completar la sincronización de configuración, se mostrará un mensaje en la página Manager indicando que la configuración ha caducado. Sin embargo, esto no afectará a los servicios. La última configuración se cargará automáticamente cuando el componente se reinicie.

----Fin

## 13.36 ¿Cómo modifico el parámetro `allow_drop_detached` de ClickHouse?

**Paso 1** Inicie sesión en el nodo donde se encuentra el cliente ClickHouse como usuario `root`.

**Paso 2** Ejecute los siguientes comandos para ir al directorio de instalación del cliente y establecer las variables de entorno:

```
cd /opt/Client installation directory
```

```
source bigdata_env
```

**Paso 3** Si la autenticación de Kerberos está habilitada para el clúster, ejecute el siguiente comando para autenticar al usuario. Si la autenticación Kerberos está deshabilitada, omita este paso.

```
kinit MRS cluster user
```

### NOTA

El usuario debe tener los permisos de administrador ClickHouse.

**Paso 4** Ejecute el comando `clickhouse client --host 192.168.42.90 --secure -m`, en el que `192.168.42.90` indica la dirección IP del nodo de instancia ClickHouseServer. La salida de comandos es la siguiente:

```
[root@server-2110082001-0017 hadoopclient]# clickhouse client --host
192.168.42.90 --secure -m
ClickHouse client version 21.3.4.25.
Connecting to 192.168.42.90:21427.
Connected to ClickHouse server version 21.3.4 revision 54447.
```

**Paso 5** Ejecute el siguiente comando para establecer el valor del parámetro `allow_drop_detached`, por ejemplo, `1`:

```
set allow_drop_detached=1;
```

**Paso 6** Ejecute el siguiente comando para consultar el valor del parámetro `allow_drop_detached`:

```
SELECT * FROM system.settings WHERE name = 'allow_drop_detached';
```

```
server-2110081635-0001 :) SELECT * FROM system.settings WHERE name = 'allow_drop_detached';
SELECT *
FROM system.settings
WHERE name = 'allow_drop_detached'
Query id: 8211d1ff-5717-49af-929f-8e4170c6e1d1
+----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| name          | value | changed | description          | min | max | readonly | type |
+----+-----+-----+-----+-----+-----+-----+-----+
| allow_drop_detached | 1     | 1       | Allow ALTER TABLE ... DROP DETACHED PART[ITION] ... queries | NULL | NULL | 0         | Bool |
+----+-----+-----+-----+-----+-----+-----+-----+
1 rows in set. Elapsed: 0.004 sec.
```

**Paso 7** Ejecute el comando `q;` para salir del cliente ClickHouse.

----Fin

## 13.37 ¿Cómo lo hago si se informa de una alarma que indica memoria insuficiente durante la ejecución de la tarea de Spark?

### Síntoma

Cuando se ejecuta una tarea de Spark, se informa de una alarma que indica una memoria insuficiente. El ID de alarma es 18022. Como resultado, no se puede usar memoria disponible.

### Procedimiento

Establezca los parámetros del ejecutor en el script SQL para limitar el número de núcleos y la memoria de un ejecutor.

Por ejemplo, la configuración es la siguiente:

```
set hive.execution.engine=spark;  
set spark.executor.cores=2;  
set spark.executor.memory=4G;  
set spark.executor.instances=10;
```

Cambie los valores de los parámetros según sea necesario.

## 13.38 ¿Cómo agrego una política de eliminación periódica para evitar registros de tabla de sistema ClickHouse de gran tamaño?

### Síntoma

Los registros de ClickHouse generados en la tabla del sistema son demasiado grandes. Eliminar registros a la vez lleva mucho tiempo.

### Procedimiento de manejo

Para eliminar los registros periódicamente, agregue una política de eliminación. En el cliente ClickHouse, ejecute la siguiente sentencia para establecer el TTL para la tabla del sistema:

```
alter table system.Table name modify TTL event_date + INTERVAL Number of days for storing data day;
```

 **ATENCIÓN**

Esta sentencia solo se utiliza en un nodo SQL para configurar el TTL de la tabla del sistema. Si el TTL necesita ser configurado en todos los nodos, ejecute esta sentencia en cada nodo. No ejecute la sentencia con **on cluster** en cada nodo. De lo contrario, el ClickHouse seguirá funcionando.

Ejecute la sentencia anterior durante las horas fuera de pico. Esta operación puede llevar mucho tiempo si hay una gran cantidad de datos.

## 13.39 ¿Cómo obtengo un archivo Spark JAR?

Huawei proporciona Huawei Mirrors para que descargue todos los archivos JAR de dependencias de proyectos de ejemplo, pero necesita descargar los archivos JAR de dependencias de código abierto desde el repositorio central de Maven u otros repositorios personalizados.

 **NOTA**

Asegúrese de que se cumplen las siguientes condiciones antes de utilizar una herramienta de desarrollo en un entorno local para descargar archivos JAR de dependencia:

- La red local es normal.  
Abra un navegador e introduzca el Huawei Mirrors URL en el cuadro de dirección para comprobar si su acceso es normal. Si el acceso es anormal, compruebe si la red local es accesible.
- Compruebe si el proxy está habilitado para la herramienta de desarrollo. Si está habilitado, desactívelo realizando las siguientes operaciones.

Tome IntelliJ IDEA 2020.2 como ejemplo. Elija **File > Settings > Appearance & Behavior > System Settings > HTTP Proxy**, seleccione **No proxy** y haga clic en **OK** para guardar la configuración.

Para obtener más información sobre cómo configurar un espejo de código abierto, consulte [Configuración de imágenes de código abierto de Huawei](#).

## 13.40 ¿Por qué se genera una alarma cuando el proceso NameNode no se reinicia después de modificar el archivo hdfs-site.xml?

Si no reinicia el proceso NameNode después de modificar el parámetro **dfs.NameNode.checkpoint.period**, es posible que se notifique falsamente una alarma. En este caso, debe reiniciar el proceso NameNode tan pronto como sea posible.

## 13.41 Se necesita mucho tiempo para que Spark SQL acceda a las tablas particionadas de Hive antes del inicio de Job

### Síntoma

Cuando se utiliza Spark SQL para acceder a tablas particionadas de Hive almacenadas en OBS, la velocidad de acceso es lenta y se invoca a un gran número de API de consulta de OBS.

Ejemplo de SQL:

```
select a,b,c from test where b=xxx
```

### Localización de fallas

De acuerdo con la configuración, la tarea debe analizar solo la partición cuya b sea xxx. Sin embargo, los registros de tareas muestran que la tarea analiza todas las particiones y luego calcula los datos cuya b es xxx. Como resultado, el cálculo de la tarea es lento. Además, se envía un gran número de solicitudes OBS porque todos los archivos necesitan ser analizados.

De forma predeterminada, la optimización del plan de ejecución basada en estadísticas de partición está habilitada en MRS, lo que equivale a la ejecución automática de Analyze Table. (El método de configuración predeterminado es establecer **spark.sql.statistics.fallBackToHdfs** en **true**. Puede establecer este parámetro en **false**.) Después de activar esta función, las estadísticas de partición de tabla se analizan durante la ejecución SQL y se utilizan como estimación de costos en el plan de ejecución. Por ejemplo, las pequeñas tablas identificadas durante la evaluación de costes se difunden a cada nodo en la memoria para operaciones de unión, reduciendo significativamente el tiempo de mezclado. Esta función optimiza enormemente el rendimiento en escenarios de unión, pero aumenta el número de invocaciones de OBS.

### Procedimiento

Establezca el siguiente parámetro en Spark SQL y, a continuación, ejecute la sentencia SQL:

```
set spark.sql.statistics.fallBackToHdfs=false;
```

Alternativamente, ejecute el comando **--conf** para establecer este parámetro en **false** antes del inicio.

```
--conf spark.sql.statistics.fallBackToHdfs=false
```

## 13.42 ¿Qué debo hacer si spark.yarn.executor.memoryOverhead no tiene efecto?

### Síntoma

La memoria de sobrecarga del ejecutor debe ajustarse para las tareas de Spark. El parámetro **spark.yarn.executor.memoryOverhead** se establece en **4096**. Sin embargo, el valor predeterminado **1024** se utiliza para aplicar recursos durante el cálculo real.

## Localización de fallas

En Spark 2.3 y versiones posteriores, utilice el nuevo parámetro **spark.executor.memoryOverhead** para establecer la memoria de sobrecarga del ejecutor. Si se establecen los parámetros antiguos y nuevos, el valor de **spark.yarn.executor.memoryOverhead** no tiene efecto y se utiliza el valor de **spark.executor.memoryOverhead**.

Lo mismo sucede si usa **spark.driver.memoryOverhead** para configurar la memoria de sobrecarga del controlador.

## Procedimiento

Utilice el nuevo parámetro:

```
spark.executor.memoryOverhead=4096
```

## 13.43 ¿Cómo cambio la zona horaria del servicio ClickHouse?

### Síntoma

Cuando se insertan datos en ClickHouse, la zona horaria es incorrecta. La diferencia horaria es de ocho horas.

## Localización de fallas

La zona horaria de Clickhouse es Europe/Moscow por defecto. Necesita cambiarla a la zona horaria local.


## Procedimiento

**Paso 1** Inicie sesión en el Administrador de FusionInsight, elija **Cluster > Services > ClickHouse**, y haga clic en **Configurations** y, a continuación, en **All Configurations**.

**Paso 2** Busque **clickhouse-config-customize** en el cuadro de búsqueda de la izquierda y agregue el nombre de parámetro **timezone** y el valor **Asia/Shanghai**.

### NOTA

El valor de **timezone** debe ser la zona horaria local. **Asia/Shanghai** se utiliza como ejemplo.

Parameter	Value				
ClickHouse->ClickHouseServer					
 clickhouse-config-customize	<table border="1"><thead><tr><th>Name</th><th>Value</th></tr></thead><tbody><tr><td>timezone</td><td>Asia/Shanghai</td></tr></tbody></table>	Name	Value	timezone	Asia/Shanghai
Name	Value				
timezone	Asia/Shanghai				

----Fin



## 13.44 ¿Qué debo hacer si falla la conexión con el servidor ClickHouse? Código de error: 516

### Síntoma

Cuando se ejecuta el comando **ClickHouse client** para conectarse al servidor ClickHouse, se notifica el siguiente error:

```
ClickHouse exception, code: 516, host: 192.168.0.198, port: 8443; Code: 516,
e.displayText() = DB::Exception: clickDevelopuser: Authentication failed:
password is incorrect or there is no user with such name
```

### Localización de fallas

El nombre de usuario o contraseña utilizado para conectar el servidor ClickHouse es incorrecto.

### Procedimiento

Introduzca el nombre de usuario o contraseña correctos. Para obtener más información, consulte [Uso de ClickHouse desde principio](#).

# 14 API

---

## 14.1 ¿Cómo configuro el parámetro `node_id` cuando uso la API para ajustar nodos de clúster?

Cuando usas la API para ajustar nodos de clúster, el valor de `node_id` se fija en `node_orderadd`.

# 15 Gestión de clúster

---

## 15.1 ¿Cómo puedo ver todos los clústeres?

Puede ver todos los clústeres MRS en la página **Clusters**. Puede ver clústeres en diferentes estados.

- **Active Clusters**: todos los clústeres excepto los clústeres en los estados **Failed** y **Terminated**.
- **Cluster History**: clústeres en el estado **Terminated**. Solo se muestran los clústeres terminados en los últimos seis meses. Si desea ver los clústeres finalizados hace más de seis meses, póngase en contacto con los ingenieros de soporte técnico.
- **Failed Tasks**: tareas en estado **Failed**. Las tareas fallidas incluyen lo siguiente:
  - Tasks failed to create clusters
  - Tasks failed to terminate clusters
  - Tasks failed to scale out clusters
  - Tasks failed to scale in clusters

## 15.2 ¿Cómo puedo ver la información de registro?

Puede ver los registros de operaciones de clústeres y trabajos en la página **Operation Logs**. Los registros de operaciones MRS registran las siguientes operaciones:

- Operaciones de clúster
  - Crear, finalizar y ampliar o reducir clústeres
  - Crear directorios y eliminar directorios o archivos
- Operaciones de trabajos: Crear, detener y eliminar trabajos
- Operaciones de datos: Tareas de usuario de IAM, agregar usuarios y agregar grupos de usuarios

**Figura 15-1** muestra los registros de operaciones.

**Figura 15-1** Información de registros

Operation Type	Operator IP Address	Operation Description	Time
Cluster		Create id is: aae6 and name as: bigdata_xq318 cluster	2016-03-18 17:17:46
Cluster		Delete the id for 7631 name for bigdata_DVwu cluster	2016-03-10 16:45:24
Job		createJob.jobId: 93a2.jobName:distcp_clusterid: 7631	2016-03-10 10:26:28
Job		createJob.jobId: c981.jobName:job_spark_clusterid: 7631	2016-03-07 11:02:28

## 15.3 ¿Cómo puedo ver la información de configuración del clúster?

- Después de crear un clúster, haga clic en el nombre del clúster en la consola MRS. En la página mostrada, puede ver información de configuración básica sobre el clúster, incluidos el nombre, ID, modo de facturación, zona de disponibilidad, tiempo de creación, versión del componente Hadoop, así como las especificaciones de instancia y la capacidad de nodo. Las especificaciones de instancia y la capacidad de nodo determinan la capacidad de análisis y procesamiento de datos. Las especificaciones de instancia más altas y una mayor capacidad permiten un procesamiento de datos más rápido a un costo más alto.
- En la página de información básica, haga clic en **Access Manager** para acceder la página de gestión de clúster de MRS. En MRS Manager, puede ver y manejar alarmas y modificar la configuración del clúster.

## 15.4 ¿Cómo agrego servicios a un clúster MRS?

No puede instalar nuevos componentes para un clúster creado cuya versión sea MRS 3.1.0 o anterior. Para instalar componentes de clúster de streaming, debe crear un clúster de streaming o un clúster híbrido e instalar los componentes necesarios.

Puede agregar componentes a clústeres personalizados MRS 3.1.2-LTS.3 o posteriores. Para obtener más información, consulte [Gestión de servicios](#).

## 15.5 ¿Cómo instalo Kafka y Flume en un clúster de MRS?

No puede instalar los componentes Kafka y Flume para un clúster creado de MRS 3.1.0 o anterior. Kafka y Flume son componentes para un clúster de streaming. Para instalar Kafka y Flume, cree un clúster de streaming o híbrido e instale Kafka y Flume.

Puede agregar componentes a clústeres personalizados MRS 3.1.2-LTS.3 o posteriores. Para obtener más información, consulte [Gestión de servicios](#).

Para obtener más información sobre cómo usar Kafka y Flume, consulte [Uso de Kafka](#) y [Uso de Flume](#).

## 15.6 ¿Cómo detengo un clúster MRS?

Para detener un clúster MRS, detenga cada nodo del clúster en el ECS. Haga clic en el nombre de cada nodo de la página de pestaña **Nodes** para ir a la página **Elastic Cloud Server** y haga clic en **Stop**.

## 15.7 ¿Es necesario apagar un nodo Master antes de actualizar sus especificaciones?

El nodo Master se apaga automáticamente durante la actualización de especificaciones y se inicia automáticamente después de la actualización. No se requiere intervención manual.

## 15.8 ¿Puedo agregar componentes a un clúster existente?

No puede agregar ni quitar ningún componente a y desde un clúster creado de MRS 3.1.0. Sin embargo, puede crear un clúster MRS que contenga los componentes necesarios.

Puede agregar componentes a clústeres personalizados MRS 3.1.2-LTS.3 o posteriores. Para obtener más información, consulte [Gestión de servicios](#).

## 15.9 ¿Puedo eliminar componentes instalados en un clúster MRS?

No puede eliminar ningún componente de un clúster MRS creado de MRS 3.1.0. Si no se requiere un componente, inicie sesión en MRS Manager y detenga el componente en la página **Services**.

Puede agregar componentes a clústeres personalizados MRS 3.1.2-LTS.3 o posteriores. Para obtener más información, consulte [Gestión de servicios](#).

## 15.10 ¿Puedo cambiar los nodos de clúster MRS en la consola MRS?

No puede cambiar los nodos del clúster de MRS en la consola de MRS. También se recomienda no cambiar los nodos del clúster de MRS en la consola de ECS. Detener o eliminar manualmente un ECS, modificar o reinstalar el sistema operativo ECS o modificar las especificaciones de ECS para un nodo de clúster en la consola ECS afectarán a la estabilidad del clúster.

Si se elimina un ECS, se modifica o reinstala el sistema operativo ECS, o se modifican las especificaciones de ECS en la consola ECS, MRS identificará y eliminará automáticamente el nodo. Puede iniciar sesión en la consola MRS y restaurar el nodo eliminado a través de escalado horizontal. No realice operaciones en los nodos que se están escalando.

## 15.11 ¿Cómo puedo proteger las notificaciones de eventos/ alarmas de clúster?

1. Inicie sesión en la consola de MRS.
2. Haga clic en el nombre del clúster.
3. En la página mostrada, elija **Alarms >Notification Rules**.

4. Busque la fila que contiene la regla que desea modificar, haga clic en **Edit** en la columna **Operation** y anule la selección de los niveles de gravedad de alarma o evento.
5. Haga clic en **OK**.

## 15.12 ¿Por qué la memoria del grupo de recursos mostrada en el cluster MRS es menor que la memoria real del cluster?

En un clúster MRS, MRS asigna el 50% de la memoria del clúster a Yarn de forma predeterminada. Los nodos Yarn se gestionan lógicamente por grupo de recursos. Por lo tanto, la memoria total del grupo de recursos mostrado en el clúster es solo el 50% de la memoria total del clúster.

## 15.13 ¿Cómo configuro la memoria Knox?

**Paso 1** Inicie sesión en un nodo Master del clúster como usuario **root**.

**Paso 2** Ejecute el siguiente comando en el nodo Master para abrir el archivo **gateway.sh**:

```
su omm
```

```
vim /opt/knox/bin/gateway.sh
```

**Paso 3** Cambie **APP\_MEM\_OPTS=""** a **APP\_MEM\_OPTS="-Xms256m -Xmx768m"** y guarde el archivo y salga.

**Paso 4** Ejecute el siguiente comando en el nodo Master para reiniciar el proceso Knox:

```
sh /opt/knox/bin/gateway.sh stop
```

```
sh /opt/knox/bin/gateway.sh start
```

**Paso 5** Repita los pasos anteriores en cada nodo Master.

**Paso 6** Ejecute el comando **ps -ef |grep Knox** para comprobar la memoria configurada.

Figura 15-2 memoria Knox



```
omm@node-master1EJH1:~$ ps -ef |grep Knox
omm      11688      1   3 15:48 pts/0    00:00:08 /opt/Bigdata/jdk1.8.0_212/bin/java -Djava.library.path=/opt/knox/ext/native -Xms256m -Xmx768m -jar /opt/knox/bin/gateway.jar
omm      29369 11350   0 15:52 pts/0    00:00:00 grep --color=auto Knox
omm@node-master1EJH1:~$
```

----Fin

## 15.14 ¿Qué es la versión de Python instalada para un clúster MRS?

Inicie sesión en un nodo master como usuario **root** y ejecute el comando **Python3** para consultar la versión de Python.

**Tabla 15-1** Versiones de Python soportadas por MRS

Versión de clúster de MRS	Versión de Python
MRS 3.1.0	Python 3.8.0
MRS 3.0.5	Python 3.7.0
MRS 3.0.2	Python 3.7.0
MRS 2.1.1	Python 3.6.8
MRS 2.1.0	Python 3.6.8
MRS 1.9.3	Python 3.6.8

## 15.15 ¿Cómo puedo ver el directorio de archivos de configuración de cada componente?

Las rutas del archivo de configuración de los componentes de uso común son las siguientes:

Componente	Directorio de archivos de configuración
ClickHouse	<i>Client installation directory/ClickHouse/clickhouse/config</i>
Flink	<i>Client installation directory/Flink/flink/conf</i>
Flume	<i>Client installation directory/fusioninsight-flume-xxx/conf</i>
HBase	<i>Client installation directory/HBase/hbase/conf</i>
HDFS	<i>Client installation directory/HDFS/hadoop/etc/hadoop</i>
Hive	<i>Client installation directory/Hive/config</i>
Hudi	<i>Client installation directory/Hudi/hudi/conf</i>
Kafka	<i>Client installation directory/Kafka/kafka/config</i>
Loader	<ul style="list-style-type: none"> <li>● <i>Client installation directory/Loader/loader-tools-xxx/loader-tool/conf</i></li> <li>● <i>Client installation directory/Loader/loader-tools-xxx/schedule-tool/conf</i></li> <li>● <i>Client installation directory/Loader/loader-tools-xxx/shell-client/conf</i></li> <li>● <i>Client installation directory/Loader/loader-tools-xxx/sqoop-shell/conf</i></li> </ul>
Oozie	<i>Client installation directory/Oozie/oozie-client-xxx/conf</i>
Spark2x	<i>Client installation directory/Spark2x/spark/conf</i>
Yarn	<i>Client installation directory/Yarn/config</i>

Componente	Directorio de archivos de configuración
ZooKeeper	<i>Client installation directory/Zookeeper/zookeeper/conf</i>

## 15.16 ¿Cómo subo un archivo local a un nodo dentro de un clúster?

- Paso 1** Inicie sesión en la consola de MRS.
- Paso 2** En el panel de navegación de la izquierda, elija **Clusters > Active Clusters** y haga clic en el nombre del clúster de destino para ir a su página de detalles.
- Paso 3** En la página **Nodes**, haga clic en un nombre de nodo para iniciar sesión en la consola de ECS.
- Paso 4** Vincule una dirección IP elástica al nodo del clúster. Para obtener más información, consulte [Asignación de un EIP](#).
- Paso 5** Cargue el archivo local al nodo del clúster. Para obtener más información, consulte [¿Cómo puedo cargar un archivo a un ECS?](#)

----Fin

## 15.17 ¿Cómo lo hago si el tiempo en los nodos MRS es incorrecto?

- Si el tiempo en un nodo dentro del clúster es incorrecto, inicie sesión en el nodo y rectifique el error de [2](#).
  - Si el tiempo en un nodo dentro del clúster es diferente al de un nodo fuera del clúster, inicie sesión en el nodo y rectifique el error de [1](#).
1. Ejecute el comando `vi /etc/ntp.conf` para editar el archivo de configuración del cliente NTP, agregue las direcciones IP del nodo master en el clúster MRS y comente la dirección IP de otros servidores.

```
server master1_ip prefer
server master2_ip
```

**Figura 15-3** Adición de las direcciones IP del nodo de master

```
# Use public servers from the pool.ntp.org project.
# Please consider joining the pool (http://www.pool.ntp.org/join.html).
#server 0.centos.pool.ntp.org iburst
#server 1.centos.pool.ntp.org iburst
#server 2.centos.pool.ntp.org iburst
#server 3.centos.pool.ntp.org iburst
#server 4.centos.pool.ntp.org iburst
server 10.9.2.38 prefer
server 10.9.2.39
#broadcast 192.168.1.255 autokey # broadcast server
```

2. Ejecute el comando `service ntpd stop` para detener el servicio NTP.
3. Ejecute el comando `/usr/sbin/ntpdate IP address of the active master node` para sincronizar manualmente la hora.



4. Ejecute el comando `service ntpd start` o `systemctl restart ntpd` para iniciar el servicio NTP.
5. Ejecute el comando `ntpstat` para comprobar el resultado de la sincronización de tiempo:

## 15.18 ¿Cómo puedo consultar la hora de inicio de un nodo MRS?

Inicie sesión en el nodo de destino y ejecute el siguiente comando para consultar la hora de inicio:

```
date -d "$(awk -F. '{print $1}' /proc/uptime) second ago" +"%Y-%m-%d %H:%M:%S"
```

```
[root@server-2110082001-0018 ~]#date -d "$(awk -F. '{print $1}' /proc/uptime) second ago" +"%Y-%m-%d %H:%M:%S"
2021-12-13 15:56:23
```

## 15.19 ¿Cómo lo hago si las relaciones de confianza entre nodos son anormales?

Si se informa de "ALM-12066 Fallas de confianza mutua entre nodos" en Manager o no hay relación de confianza SSH entre nodos, rectifique el error realizando las siguientes operaciones:

1. Ejecute el comando `ssh-add -l` en ambos nodos del clúster de confianza para comprobar si hay identidades.

```
omm@node-group-2eU40 ~]$
omm@node-group-2eU40 ~]$
omm@node-group-2eU40 ~]$
omm@node-group-2eU40 ~]$ ll .ssh/
total 32
srw-----. 1 omm wheel    0 Dec 29 14:17 agent.pid
-rw-----. 1 omm wheel 12901 Mar  9 14:48 authorized_keys
-rw-----. 1 omm wheel   54 Sep 24 11:42 config
-rw-----. 1 omm wheel 1766 Sep 24 11:43 id_rsa
-rw-----. 1 omm wheel  402 Sep 24 11:42 id_rsa.pub
-rw-----. 1 omm wheel   88 Jun  8 2020 id_rsa.sha256
omm@node-group-2eU40 ~]$
omm@node-group-2eU40 ~]$ ssh-add -l
The agent has no identities.
omm@node-group-2eU40 ~]$
omm@node-group-2eU40 ~]$
omm@node-group-2eU40 ~]$
omm@node-group-2eU40 ~]$ vim /var/log/Bigdata/nodeagent/
agentlog/  alarmlog/  monitorlog/ scriptlog/
omm@node-group-2eU40 ~]$ vim /var/log/Bigdata/nodeagent/scriptlog/
agent_alarm_py.log          install.log
agent_alarm_py.log.1       installntp.log
```

2. Si no se muestra ninguna identidad, ejecute el comando `ps -ef|grep ssh-agent` para encontrar el proceso ssh-agent, elimine el proceso y espere a que el proceso se reinicie automáticamente.

```
omm@node-group-2eU40 ~]$
omm@node-group-2eU40 ~]$ ssh-add -l
The agent has no identities.
omm@node-group-2eU40 ~]$
omm@node-group-2eU40 ~]$ ps -ef|grep ssh-agent
omm 18729 1 0 14:53 ? 00:00:00 ssh-agent -a /home/omm/.ssh/agent.pid
omm 25098 1 0 14:54 ? 00:00:00 bash /opt/Bigdata/om-agent/nodeagent/bin/ssh-agent-monitor-startup.sh
omm 25206 25098 0 14:54 ? 00:00:00 bash /opt/Bigdata/om-agent/nodeagent/bin/ssh-agent-monitor.sh
omm 27201 4913 0 14:54 pts/0 00:00:00 grep --color=auto ssh-agent
omm@node-group-2eU40 ~]$
omm@node-group-2eU40 ~]$ ssh-add -l
```

3. Ejecute el comando `ssh-add -l` para comprobar si se han agregado las identidades. En caso afirmativo, ejecute manualmente el comando `ssh` para comprobar si la relación de confianza es normal.

```
omm 22276 4913 0 14:53 pts/0 00:00:00 grep --color=auto ssh-agent
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ssh-add -l
The agent has no identities.
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ps -ef|grep ssh-agent
omm 18729 1 0 14:53 ? 00:00:00 ssh-agent -a /home/omm/.ssh/agent.pid
omm 25098 1 0 14:54 ? 00:00:00 bash /opt/Bigdata/om-agent/nodeagent/bin/ssh-agent-monitor-startup.sh
omm 25286 25098 0 14:54 ? 00:00:00 bash /opt/Bigdata/om-agent/nodeagent/bin/ssh-agent-monitor.sh
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ssh-add -l
2048 SHA256:uChnRUBhh1Hyxpf0Z1bS0zym1KXMIaFyvn0IMpiZjg /home/omm/.ssh/id_rsa (RSA)
[omm@node-group-2eU40 ~]$
[omm@node-group-2eU40 ~]$ ssh 10.33.109.226
Warning: Permanently added '10.33.109.226' (ECDSA) to the list of known hosts.
```

4. Si existen identidades, compruebe si el archivo de **authorized\_keys** en el directorio / **home/omm/.ssh** contiene la información en el archivo **id\_rsa.pub** en **/home/omm/.ssh** del nodo del mismo nivel. Si no, agregue manualmente la información sobre el nodo del mismo nivel.
5. Compruebe si los permisos en los archivos del directorio **/home/omm/.ssh** son correctos.
6. Compruebe el archivo **/var/log/Bigdata/nodeagent/scriptlog/ssh-agent-monitor.log**.
7. Si se elimina el directorio **home** del usuario **omm**, póngase en contacto con el personal de soporte de MRS.

## 15.20 ¿Cómo ajusto el tamaño de memoria del proceso manager-executor?

### Síntoma

El proceso **manager-executor** desarrollado por Huawei se ejecuta en el nodo Master1 o Master2 en el clúster MRS en modo activo/en espera. Este proceso se utiliza para encapsular las operaciones del plano de gestión y control de MRS en el clúster de MRS, como el envío de trabajos, informes de latidos, determinados informes de alarmas, así como la creación, la ampliación y reducción de clústeres. Al enviar trabajos en el plano de gestión y control de MRS, la memoria Executor puede resultar insuficiente a medida que aumentan las tareas o el número de tareas simultáneas aumenta. Como resultado, el uso de la CPU es alto y el proceso Executor experimenta errores fuera de memoria (OOM).

### Procedimiento

1. Inicie sesión en el nodo Master1 o Master2 como usuario **root** y ejecute el siguiente comando para cambiar a usuario **omm**:  
**su - omm**
2. Ejecute el siguiente comando para modificar el script **catalina.sh**. Específicamente, busque **JAVA\_OPTS** en el script, busque los elementos de configuración similares a **JAVA\_OPTS="-Xms1024m -Xmx4096m** y cambie los valores de los elementos a los deseados, y guarde la modificación.

**vim /opt/executor/bin/catalina.sh**

```
JAVA_OPTS="-Xms1024m -Xmx4096m"
JAVA_OPTS="-Xms1024m -Xmx4096m"
LOG4J_PROPERTIES_PATH=${CATALINA_HOME}/lib/log4j.properties
CATALINA_OPTS="-XX:+PrintGC -XX:+PrintGCDetails -XX:+PrintGCTimeStamps -XX:+PrintGCApplicationStoppedTime \
-XX:+PrintHeapAtGC -Xloggc:/var/log/executor/logs/gc.log -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 \
-XX:GCLogFileSize=20M -XX:OnOutOfMemoryError='kill -9 %p' -XX:+HeapDumpOnOutOfMemoryError \
-XX:HeapDumpPath=/var/log/executor/logs/executor-dump.hprof"
```

3. El proceso **manager-executor** solo se ejecuta en el nodo Master1 o Master2 en modo activo/en espera. Compruebe si existe en el nodo antes de reiniciarlo.



# 16 Uso de Kerberos

---

## 16.1 ¿Cómo cambio el estado de autenticación de Kerberos de un clúster MRS creado?

No puede cambiar el servicio Kerberos después de crear un clúster MRS.

## 16.2 ¿Cuáles son los puertos del servicio de autenticación de Kerberos?

El servicio de autenticación de Kerberos utiliza los puertos 21730 (TCP), 21731 (TCP/UDP) y 21732 (TCP/UDP).

## 16.3 ¿Cómo despliego el servicio Kerberos en un clúster en ejecución?

El clúster MRS no admite la instalación y el despliegue de Kerberos personalizadas, y la autenticación de Kerberos no se puede configurar entre componentes. Para habilitar la autenticación Kerberos, debe crear un clúster con Kerberos habilitado y migrar datos.

## 16.4 ¿Cómo accedo a Hive en un clúster con autenticación de Kerberos habilitada?

1. Inicie sesión en el nodo master del clúster como usuario **root**.
2. Ejecute el siguiente comando para configurar las variables de entorno:  
**source /opt/client/bigdata\_env**
3. Si la autenticación de Kerberos está habilitada para el clúster actual, ejecute el siguiente comando para autenticar al usuario:  
**kinit MRS cluster user**  
Ejemplo: **kinit hiveuser**

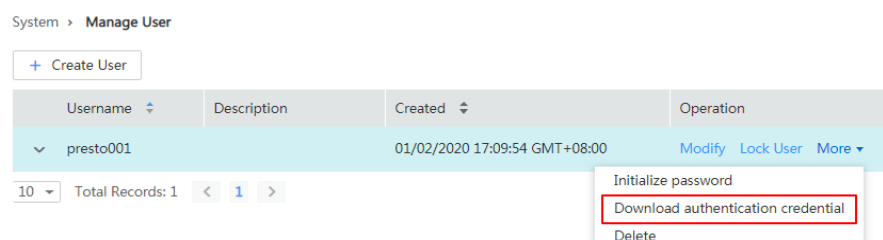
El usuario actual debe tener permiso para crear tablas Hive. Para obtener más información acerca de cómo configurar los permisos de Hive y vincular los roles al usuario, consulte [Configuración de permisos de Hive](#).

4. Ejecute el comando cliente del componente Hive.  
**beeline**
5. Ejecute el comando Hive en Beeline, por ejemplo:  
**create table test\_obs(a int, b string) row format delimited fields terminated by ","  
stored as textfile location "obs://test\_obs";**
6. Presione **Ctrl+C** para salir de Hive beeline.

## 16.5 ¿Cómo puedo acceder de Presto en un clúster con autenticación de Kerberos habilitada?

1. Inicie sesión en el nodo master del clúster como usuario **root**.
2. Ejecute el siguiente comando para configurar las variables de entorno:  
**source /opt/client/bigdata\_env**
3. Acceda a Presto en un clúster con la autenticación de Kerberos habilitada.
  - a. Inicie sesión en MRS Manager y cree un rol con el permiso **Hive Admin Privilege**, por ejemplo, **prestorerole**. Para obtener más información sobre cómo crear un rol, consulte [Creación de un Rol](#).
  - b. Cree un usuario, por ejemplo, **presto001**, que pertenece a los grupos **Presto** y **Hive** y vincúlelo al rol creado en el **3.a**. Para obtener más información sobre cómo crear un usuario, vea [Creación de un usuario](#).
  - c. Autentique usuario **presto001**.  
**kinit presto001**
  - d. Descargue la credencial de autenticación de usuario.
    - Operaciones en MRS Manager: Inicie sesión en MRS Manager y elija **System > Manage User**. Busque la fila que contiene el nuevo usuario, haga clic en **More** y seleccione **Download authentication credential**.

**Figura 16-1** Descargar la credencial de autenticación de usuario de Presto



- Operaciones en FusionInsight Manager:  
Inicie sesión en FusionInsight Manager, seleccione **System > Permission > User**. En la página mostrada, busque la fila que contiene el usuario, elija **More > Download Authentication Credential**.
- e. Descomprima el archivo de credenciales de usuario descargado y guarde los archivos **krb5.conf** y **user.keytab** obtenidos en el directorio del cliente, por ejemplo, **/opt/client/Presto/**.

- f. Ejecute el siguiente comando para obtener el principal de usuario:  
**klist -kt /opt/client/Presto/user.keytab**
- g. Ejecute el siguiente comando para conectarse al servidor Presto del clúster:  
**presto\_cli.sh --krb5-config-path {krb5.conf file path} --krb5-principal {User's principal} --krb5-keytab-path {user.keytab file path} --user {presto username}**
- **krb5.conf file path**: ruta de archivo establecida en 3.e, por ejemplo, /opt/client/Presto/krb5.conf.
  - **user.keytab file path**: ruta de archivo establecida en 3.e, por ejemplo, /opt/client/Presto/user.keytab.
  - **User's principal**: principal obtenido en 3.f.
  - **presto username**: usuario creado en 3.b, por ejemplo, presto001.
- Ejemplo: **presto\_cli.sh --krb5-config-path /opt/client/Presto/krb5.conf --krb5-principal presto001@xxx\_xxx\_xxx\_xxx.COM --krb5-keytab-path /opt/client/Presto/user.keytab --user presto001**
- h. En el cliente de Presto, ejecute la siguiente sentencia para crear un esquema:  
**CREATE SCHEMA hive.demo01 WITH (location = 'obs://presto-demo002/');**
- i. Cree una tabla en el esquema. Los datos de la tabla se almacenan en el bucket OBS, como se muestra en el siguiente ejemplo:  
**CREATE TABLE hive.demo01.demo\_table WITH (format = 'ORC') AS SELECT \* FROM tpch.sf1.customer;**

Figura 16-2 Resultado devuelto

```

[root@node-master2]# presto_cli.sh --krb5-config-path /opt/client/Presto/krb5.conf --krb5-principal presto001@6585C37_1776_488E_87E0_D90C42998A1.COM --krb5-keytab-path /opt/client/Presto/user.keytab --user presto001
--krb5-remote-service-name HTTP --server https://192.168.3.22:7021 --krb5-keytab-path /opt/client/Presto/user.keytab --krb5-principal presto001@6585C37_1776_488E_87E0_D90C42998A1.COM --krb5-config-path /opt/client/Presto/krb5.conf --user presto001
presto> CREATE SCHEMA hive.demo01 WITH (location = 'obs://mrs-word001/presto-demo02/');
presto> CREATE TABLE hive.demo01.demo_table WITH (format = 'ORC') AS SELECT * FROM tpch.sf1.customer;
CREATE TABLE: 10000 rows
Query 20191223_185909_00006_17qgh_ FINISHED, 2 nodes
Spills: 42 total, 42 Done (100.0%)
0/1 [100% rows] [13.76 rows/s] [0/0]

```

- j. Ejecute **exit** para salir del cliente.

## 16.6 ¿Cómo accedo a Spark en un clúster con autenticación Kerberos habilitada?

1. Inicie sesión en el nodo master del clúster como usuario **root**.
2. Ejecute el siguiente comando para configurar las variables de entorno:  
**source /opt/client/bigdata\_env**
3. Si la autenticación de Kerberos está habilitada para el clúster actual, ejecute el siguiente comando para autenticar al usuario.  
**kinit MRS cluster user**  
Ejemplo:  
Si el usuario de desarrollo es un usuario máquina-máquina, ejecute **kinit -kt user.keytab sparkuser**.  
Si el usuario de desarrollo es un usuario hombre-máquina, ejecute **kinit sparkuser**.
4. Ejecute el siguiente comando para conectarse a Spark Beeline:  
**spark-beeline**
5. Ejecute comandos en Spark Beeline. Por ejemplo, cree la tabla **test** en el directorio **obs://mrs-word001/table/**.

**create table test(id int) location 'obs://mrs-word001/table/';**

6. Ejecute el siguiente comando para consultar todas las tablas. Si se muestra la tabla **test** en la salida del comando, el acceso a OBS se realiza correctamente.

**show tables;**

**Figura 16-3** Nombre de tabla devuelta

```
0: jdbc:hive2://ha-cluster/default> create table test(id int) location 'obs://mrs-word001/table/';
+-----+
| Result |
+-----+
No rows selected (2.515 seconds)
0: jdbc:hive2://ha-cluster/default> show tables;
+-----+
| database | tableName | isTemporary |
+-----+
| default  | test      | false       |
| default  | test_obs  | false       |
+-----+
2 rows selected (0.127 seconds)
```

7. Presione **Ctrl+C** para salir de Spark Beeline.

## 16.7 ¿Cómo puedo evitar que la autenticación Kerberos caduque?

- Aplicaciones de Java:

Antes de conectar a HBase, HDFS u otros componentes de big data, invoque a `loginUserFromKeytab` para crear un UGI. A continuación, inicie un subproceso programado para comprobar periódicamente si la autenticación Kerberos caduca. Inicie sesión de nuevo en el sistema antes de que caduque la autenticación Kerberos.

```
private static void startCheckKeytabTgtAndReLoginJob() {
//The credential is checked every 10 minutes, and updated before the
expiration time.
    ThreadPool.updateConfigThread.scheduleWithFixedDelay(() -> {
        try {

UserGroupInformation.getLoginUser().checkTGTAndReLoginFromKeytab();
            logger.warn("get tgt:{}",
UserGroupInformation.getLoginUser().getTGT());
                logger.warn("Check Kerberos Tgt And ReLogin From Keytab
Finish.");
        } catch (IOException e) {
            logger.error("Check Kerberos Tgt And ReLogin From Keytab
Error", e);
        }
        }, 0, 10, TimeUnit.MINUTES);
    logger.warn("Start Check Keytab TGT And ReLogin Job Success.");
}
```

- Tareas ejecutadas en modo shell:

- a. Ejecute el comando **kinit** para autenticar al usuario.
- b. Cree una tarea programada del sistema operativo o cualquier otra tarea programada para ejecutar el comando **kinit** para autenticar al usuario periódicamente.
- c. Envíe trabajos para ejecutar tareas de big data.

- Ttrabajos de Spark:

Si envía trabajos con `spark-shell`, `spark-submit` o `spark-sql`, puede especificar **Keytab** y **Principal** en el comando para realizar la autenticación y actualizar periódicamente las credenciales de inicio de sesión y los tokens de autorización para evitar el vencimiento de la autenticación.

Ejemplo:

```
spark-shell --principal spark2x/hadoop.<System domain name>@<System domain name> --keytab ${BIGDATA_HOME}/FusionInsight_Spark2x_8.1.0.1/install/FusionInsight-Spark2x-2.4.5/keytab/spark2x/SparkResource/spark2x.keytab --master yarn
```



# 17 Gestión de metadatos

## 17.1 ¿Dónde puedo ver metadatos de Hive?

- Los metadatos de Hive se almacenan en la base de datos de GaussDB del clúster MRS. Puede realizar los siguientes pasos para ver los metadatos:
  - a. Inicie sesión en el nodo activo de DBServer como usuario **root**.
  - b. Inicie sesión en la base de datos de Hive:

```
su - omm  
source $DBSERVER_HOME/.dbservice_profile  
gsql -p 20051 -U USER -W PASSWD -d hivemeta
```
  - c. Vea todas las tablas de metadatos de la base de datos de metadatos de Hive:

```
\d+
```
  - d. Ejecute **q** para salir de la página para ver la tabla de metadatos.
  - e. Vea datos en la tabla de metadatos:

```
select * from Table name;
```

### NOTA

Para consultar la dirección IP del nodo DBServer activo, inicie sesión en FusionInsight Manager, elija **Cluster > Services > DBService** y haga clic en la pestaña **Instance**.

Puede ver el nombre de usuario predeterminado (USUARIO) y la contraseña inicial (PASSWD) de la base de datos DBservice en [Cuentas de usuario](#).

- Si los metadatos de Hive se almacenan en una base de datos relacional externa, realice los siguientes pasos:
  - a. En la página de **Dashboard** del clúster, haga clic en **Manage** a la derecha de **Data Connection**.
  - b. En la página mostrada, obtenga el valor de **Data Connection ID**.
  - c. En la consola MRS, haga clic en **Data Connections**.
  - d. En la lista de conexiones de datos, localice la conexión de datos en función del ID de conexión de datos obtenido en **b**.
  - e. Haga clic en **Edit** en la columna **Operation** de la conexión de datos.

El **RDS Instance** y el **Database** indican la base de datos relacional en la que se almacenan los metadatos de Hive.