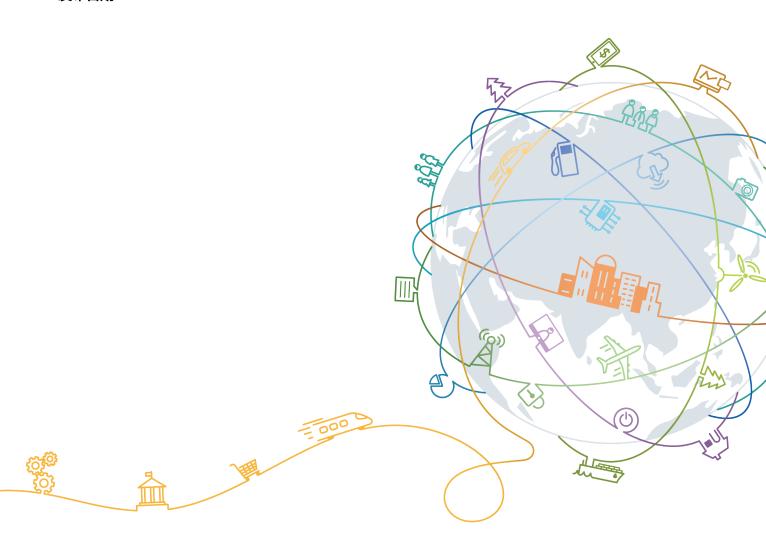
## MapReduce 服务

## 常见问题

**文档版本** 01

发布日期 2020-08-07





#### 版权所有 © 华为技术有限公司 2021。 保留一切权利。

非经本公司书面许可,任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部,并不得以任何形式传播。

#### 商标声明



nuawe和其他华为商标均为华为技术有限公司的商标。 本文档提及的其他所有商标或注册商标,由各自的所有人拥有。

### 注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束,本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定,华为公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因,本文档内容会不定期进行更新。除非另有约定,本文档仅作为使用指导,本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

### 目录

1 产品咨询类	1
1.1 MRS 可以做什么?	1
1.2 MRS 支持什么类型的分布式存储?	1
1.3 如何使用自定义安全组创建 MRS 集群?	1
1.4 如何使用 MRS?	1
1.5 如何保证数据和业务运行安全?	2
1.6 是否可以实现一个大数据的实时处理和存储的实时数据仓库?	2
1.7 什么是区域、可用区?	3
1.8 如何选择区域?	3
1.9 如何选择可用区?	4
1.10 如何获取区域终端节点?	4
1.11 如何配置 Phoenix 连接池?	4
1.12 MRS 是否支持更换网段?	4
1.13 MRS 服务集群节点是否执行降配操作?	4
1.14 Hive 与其他组件有什么关系?	4
1.15 MRS 集群是否支持 Hive on Spark?	5
1.16 Hive 版本之间是否兼容?	5
1.17 MRS 集群哪个版本支持建立 Hive 连接且有用户同步功能?	5
1.18 Spark 和 Hadoop 什么关系?	5
1.19 MRS 集群是否支持 Spark on Hive?	
1.20 数据存储在 OBS 和 HDFS 有什么区别?	5
1.21 Hadoop 压力测试工具如何获取?	6
1.22 Impala 与其他组件有什么关系?	6
1.23 Kudu 和 HBase 间的关系?	6
1.24 MRS 是否支持 Hive on Kudu?	6
1.25 10 亿级数据量场景的解决方案	7
1.26 如何修改 DBService 的 IP?	7
1.27 MRS sudo log 能否清理?	7
2 计费相关	8
2.1 MRS 如何收费?	
2.2 创建 MRS 集群时产品报价如何计费?	8
2.3 MRS 集群中的弹性伸缩功能是如何收费的?	8
2.4 MRS 服务是否支持按需计费转包周期(包年/包月)计费?	9

2.5 集群手动扩容如何计费?	
2.6 MRS 如何续费?	
2.7 Mapreduce 服务集群的 Task 节点如何收费?	
2.8 退订 MRS 服务后,在 ECS 中退订弹性云服务器时报异常什么原因?	9
3 账号密码类	10
3.1 登录 Manager 账号的是什么?	10
- 4 账号权限类	11
- XL	
4.2 如何给新建的账号添加租户管理权限?	
4.3 MRS 服务的 IAM 用户能关联 MRS Manager 用户吗?	
4.4 MRS 集群中如何创建 IAM 委托?	
4.5 如何实现 MRS Manager 和 IAM 的权限粒度为表级别或库级别?	
4.6 如何自定义配置 MRS 服务策略?	
4.7 在 MRS Manager 页面"系统设置"中找不到用户管理,什么原因?	
4.8 IAM 用户同步的具体作用?	
4.9 Hue 有没有配置账号权限的功能?	13
5 客户端使用类	14
<b>5.1 如何使用组件客户端?</b>	
5.2 没有绑定 EIP 是否可以下载客户端?	
5.3 Flume 客户端是否支持 Windows?	
6 Web 页面访问类	
6.1 如何在内网访问 MRS manager 页面?	
6.2 MRS Manager 页面报 OBS 证书即将过期告警,应该如何处理?	
6.3 本地 IP 和华为云不在同一个网段中如何使用 HDFS?	
6.4 修改开源组件 Web 页面会话超时时间	
6.5 MRS 租户管理中的动态资源计划页面无法刷新	
6.6 用户修改域名后无法登录 MRS Manager 页面	
6.7 Kafka Topic 监控页签在 Manager 页面不显示	
7 <u>监控告警</u> 类	
<b>7.1</b> 在 MRS 流式集群中,Kafka topic 监控是否支持发送告警?	
8 性能优化类	
8.1 MRS 集群是否支持重装系统?	
8.2 如何提高集群 Core 节点的资源使用率?	
8.3 如何配置 MRS 默认磁盘使用率?	
8.4 如何关闭防火墙服务以避免对系统造成影响?	
8.5 MRS 集群在不增加 Core 节点的情况下,如何增加节点的 HDFS 容量?	
9 作业开发类	
9.1 如何准备 MRS 的数据源?	
9.2 如何将用户程序提交到 MRS 中?	
9.3 如何调整 JobHistory 进程堆内存?	23

9.4 集群支持提交哪些形式的 Spark 作业?	
9.5 MRS 集群的租户资源最小值改为 0 后,只能同时跑一个 Spark 任务吗?	24
10 集群升级/补丁	25
10.1 MRS 版本如何进行升级?	25
10.2 MRS 是否支持修改版本?	25
10.3 是否可以自主升级 MRS 组件版本?	25
11 周边生态对接使用	26
11.1 MRS 集群监控指标是否可以集成到云监控上?	26
11.2 是否可以使用 MRS 服务对 DLI 业务表进行读写操作?	26
11.3 MRS 集群是否支持 Tableau 连接 MRS 的 Hive?	26
11.4 MRS 建立后同地域同可用区访问 OBS 是内网访问么?是否有限速?	26
11.5 MRS 中可以部署爬虫服务吗?	26
11.6 DWS 和 MRS 服务,是否支持安全删除(删除后防止非法恢复)的能力?	27
11.7 OBS 是否支持 ListObjectsV2 协议?	27
11.8 MRS 的数据存储可以使用并行文件系统的 OBS 吗?	27
11.9 数据湖工厂是否支持对接 MRS 的集群?	27
11.10 为什么 DLF 中建立连接时没有找到 Kerberos 开启的 MRS 服务集群?	27
11.11 如何在 ECS 服务器上用 PySpark 连接内网开启 Kerberos 认证的 MRS Spark 集群?	27
11.12 HBase 同步数据到 CSS 为什么没有映射字段?	27
11.13 Spark 集群能访问 OBS 中的数据吗?	28
11.14 MRS 中的 Spark 组件与 DLI 中的 Spark 有什么区别?	28
11.15 如何在 Flume 组件中把数据从 OBS 导入 HDFS 中?	
11.16 组件对接是否支持外部 KDC?	
11.17 开源 Kylin 3.x 版本与 MRS 1.9.3 版本对接时存在 Jetty 版本的兼容性问题	28
11.18 MRS 导出数据到 OBS 加密桶失败	29
12 集群访问类	30
12.1 MRS 登录集群节点的两种方式能够切换么?	30
12.2 如何远程使用 MRS 集群?	
12.3 本地联调 MRS 集群的 IP 和端口设置方法有什么?	30
12.4 如何获取 ZooKeeper 地址?	30
13 大数据业务开发	32
13.1 MapReduce 服务中的样例工程构建和应用开发有什么区别?是否支持 python 代码?	32
13.2 HBase 支持的压缩算法有哪些?	32
13.3 如何使用 HBase?	32
13.4 MRS 是否支持通过 Hive 的 HBase 外表将数据写入到 HBase?	32
13.5 如何查看 HBase 日志?	32
13.6 开启 Kerberos 认证的集群如何访问 hbase shell?	33
13.7 DynamoDB 的 number 在 Hive 表中用什么类型比较好?	33
13.8 Hive Driver 是否支持对接 dbcp2?	33
13.9 用户 A 如何查看用户 B 创建的 Hive 表?	33
13.10 如何使用本地电脑 IDEA 访问 Hive?	33

13.11 如何在 Windows 下调测 sparkSQL 作业?	33
13.12 如何重置 Kafka 数据?	33
13.13 MRS 是否支持同时运行多个 Flume 任务?	33
13.14 如何修改 FlumeClient 的日志为标准输出日志?	34
13.15 HDFS 如何进行数据均衡?	34
13.16 如何修改 HDFS 的副本数?	34
13.17 如何使用 Python 远程连接 HDFS 的端口?	35
13.18 Hadoop 组件 jar 包位置和环境变量的位置在哪里?	35
13.19 Kudu 支持的压缩算法有哪些?	35
13.20 如何查看 Kudu 日志?	35
13.21 OpenTSDB 是否支持 python 的接口?	35
13.22 Yarn 服务的 container 日志如何查看?	
13.23 MRS 的 Storm 集群提交任务时如何指定日志路径?	36
13.24 Hive 样例工程中的 com.huawei.gaussc10 依赖包在哪里下载?	36
13.25 HDFS 系统支持 OBS 对象存储吗?	36
13.26 kafka 访问协议配置为什么不生效?	36
13.27 HBase 表如何设置和修改数据保留期?	36
13.28 flink run 如何指定 Application-Name?	36
13.29 如何通过 happybase 连接到 MRS 服务的 HBase?	36
13.30 如何更改 Yarn 服务中队列的资源配比?	37
13.31 如何指定 Flink 任务到指定的 Yarn 队列运行中运行?	37
13.32 MRS Kafka 如何查看客户端版本信息?	37
13.33 Kafka 目前支持的访问协议类型有哪些?	37
13.34 presto 如何配置其他数据源?	37
13.35 Hive 查询数据是否支持导出?	37
13.36 MRS 要如何连接 impala-shell	37
13.37 MRS 如何连接 spark-shell	37
13.38 MRS 如何连接 spark-beeline	
13.39 spark job 对应的执行日志保存在哪里?	38
13.40 Hue 下载 excel 无法打开	39
13.41 Yarn 的 ResourceManager 配置是否正常?	40
14 API 使用类	43
14.1 使用调整集群节点接口时参数 node_id 如何配置?	
15 集群管理类	
15.1 如何查看所有集群?	
15.2 如何查看日志信息?	
15.3 如何查看集群配置信息?	
15.4 如何在 MRS 分析集群中安装流式集群的组件?	
15.5 如何在 MRS 集群中安装 Kafka,Flume 组件?	
15.6 如何停止 MRS 集群?	
15.7 升级 Master 节点规格需要关机吗?	
15.8 MRS 支持数据盘扩容吗?	45

ロか
40
46
46
46
46
46
46
46
47
47
47
48
48
48
48
48
48
49
50
52
52

# **1** 产品咨询类

### 1.1 MRS 可以做什么?

基于开源软件Hadoop进行功能增强、Spark内存计算引擎、HBase分布式存储数据库以及Hive数据仓库框架,提供企业级大数据存储、查询和分析的统一平台,帮助企业快速构建海量数据信息处理系统,可解决各大企业的以下需求:

- 海量数据的分析与计算
- 海量数据的存储
- 海量数据流式处理

### 1.2 MRS 支持什么类型的分布式存储?

提供目前主流的Hadoop,目前支持Hadoop 3.1.x版本,并且随社区更新版本。MRS支持的组件版本请参见MRS组件版本一览表。

### 1.3 如何使用自定义安全组创建 MRS 集群?

用户购买集群时,如果选择使用自己创建的安全组,则需要放开9022等端口,或者在 界面上购买集群时,安全组选择"自动创建"。

### 1.4 如何使用 MRS?

MRS是一个在华为云上部署和管理Hadoop系统的服务,一键即可部署Hadoop集群。 MRS提供租户完全可控的企业级大数据集群云服务,轻松运行Hadoop、Spark、 HBase、Kafka、Storm等大数据组件。

MRS使用简单,通过使用在集群中连接在一起的多台计算机,您可以运行各种任务, 处理或者存储(PB级)巨量数据。MRS的基本使用流程如下:

- 开发数据处理程序,MRS的开发指南为用户提供代码示例和教程,帮助您快速开始开发自己的程序并正常运行。
- 2. 上传程序和数据文件到对象存储服务(OBS)中,用户需要先将本地的程序和数据文件上传至OBS中。

- 3. 创建集群,用户可以指定集群类型用于离线数据分析和流处理任务,指定集群中预置的弹性云服务器实例规格、实例数量、数据盘类型(普通IO、高 IO、超高 IO)、要安装的组件(Hadoop、Spark、HBase、Hive、Kafka、Storm等)。用户可以使用引导操作在集群启动前(或后)在指定的节点上执行脚本,安装其他第三方软件或修改集群运行环境等自定义操作。
- 4. 管理作业,MRS为用户提供程序执行平台,程序由用户自身开发,MRS负责程序的提交、执行和监控。
- 5. 管理集群,MRS为用户提供企业级的大数据集群的统一管理平台,帮助用户快速 掌握服务及主机的健康状态,通过图形化的指标监控及定制及时的获取系统的关 键信息,根据实际业务的性能需求修改服务属性的配置,对集群、服务、角色实 例等实现一键启停等操作。
- 6. 删除集群,如果作业执行结束后不需要集群,可以删除MRS集群。集群删除后不再产生费用。

### 1.5 如何保证数据和业务运行安全?

MRS作为一个海量数据管理和分析平台,具备高安全性。主要从以下几个方面保障数据和业务运行安全:

• 网络隔离

整个公有云网络划分为2个平面,即业务平面和管理平面。两个平面采用物理隔离的方式进行部署,保证业务、管理各自网络的安全性。

- 业务平面:主要是集群组件运行的网络平面,支持为用户提供业务通道,对外提供数据存取、任务提交及计算能力。
- 管理平面:主要是公有云管理控制台,用于购买和管理MRS。
- 主机安全

用户可以根据自己业务的需要部署第三方的防病毒软件。针对操作系统和端口部分,MRS提供如下安全措施:

- 操作系统内核安全加固
- 更新操作系统最新补丁
- 操作系统权限控制
- 操作系统端口管理
- 操作系统协议与端口防攻击
- 数据安全

MRS支持数据存储在OBS上,保障客户数据安全。

● 数据完整性

MRS处理完数据后,通过SSL加密传输数据至OBS,保证客户数据的完整性。

# 1.6 是否可以实现一个大数据的实时处理和存储的实时数据仓库?

MRS服务能满足该需求,具体请参考应用场景。

### 1.7 什么是区域、可用区?

我们用区域和可用区来描述数据中心的位置,您可以在特定的区域、可用区创建资源。

- 区域(Region):从地理位置和网络时延维度划分,同一个Region内共享弹性计算、块存储、对象存储、VPC网络、弹性公网IP、镜像等公共服务。Region分为通用Region和专属Region,通用Region指面向公共租户提供通用云服务的Region;专属Region指只承载同一类业务或只面向特定租户提供业务服务的专用Region。
- 可用区(AZ, Availability Zone): 一个AZ是一个或多个物理数据中心的集合, 有独立的风火水电,AZ内逻辑上再将计算、网络、存储等资源划分成多个集群。 一个Region中的多个AZ间通过高速光纤相连,以满足用户跨AZ构建高可用性系统的需求。

图1-1阐明了区域和可用区之间的关系。





目前,华为云已在全球多个地域开放云服务,您可以根据需求选择适合自己的区域和可用区。更多信息请参见**华为云全球站点**。

### 1.8 如何选择区域?

选择区域时,您需要考虑以下几个因素:

● 地理位置

一般情况下,建议就近选择靠近您或者您的目标用户的区域,这样可以减少网络时延,提高访问速度。不过,在基础设施、BGP网络品质、资源的操作与配置等方面,中国大陆各个区域间区别不大,如果您或者您的目标用户在中国大陆,可以不用考虑不同区域造成的网络时延问题。

- 在除中国大陆以外的亚太地区有业务的用户,可以选择"亚太-香港"、"亚太-曼谷"或"亚太-新加坡"区域。
- 在非洲地区有业务的用户,可以选择"南非-约翰内斯堡"区域。
- 在欧洲地区有业务的用户,可以选择"欧洲-巴黎"区域。
- 资源的价格

不同区域的资源价格可能有差异,请参见华为云服务价格详情。

### 1.9 如何选择可用区?

是否将资源放在同一可用区内,主要取决于您对容灾能力和网络时延的要求。

- 如果您的应用需要较高的容灾能力,建议您将资源部署在同一区域的不同可用区内。
- 如果您的应用要求实例之间的网络延时较低,则建议您将资源创建在同一可用区内。

### 1.10 如何获取区域终端节点?

当您通过API使用资源时,您必须指定其区域终端节点。有关华为云的区域和终端节点的更多信息,请参阅**地区和终端节点**。

### 1.11 如何配置 Phoenix 连接池?

Phoenix不支持连接池设置,建议用户自己写代码实现一个管理连接的工具类,模拟连接池,具体可以参考https://stackoverflow.com/questions/35183713/what-is-the-correct-way-to-pool-the-phoenix-query-server-connections。

### 1.12 MRS 是否支持更换网段?

MRS不支持更换网段。

### 1.13 MRS 服务集群节点是否执行降配操作?

MRS服务暂不支持降级配置操作,如果有诉求,建议客户走需求处理。

### 1.14 Hive 与其他组件有什么关系?

● Hive与HDFS间的关系

Hive是Apache的Hadoop项目的子项目,Hive利用HDFS作为其文件存储系统。 Hive通过解析和计算处理结构化的数据,Hadoop HDFS则为Hive提供了高可靠性 的底层存储支持。Hive数据库中的所有数据文件都可以存储在Hadoop HDFS文件 系统上,Hive所有的数据操作也都是通过Hadoop HDFS接口进行。

● Hive与MapReduce间的关系

Hive所有的数据计算都依赖于MapReduce。MapReduce也是Apache的Hadoop项目的子项目,它是一个基于Hadoop HDFS分布式并行计算框架。Hive进行数据分析时,会将用户提交的HiveQL语句解析成相应的MapReduce任务并提交MapReduce执行。

● Hive与DBService间的关系

Hive的MetaStore(元数据服务)处理Hive的数据库、表、分区等的结构和属性信息,这些信息需要存放在一个关系型数据库中,由MetaStore维护和处理。在MRS中,这个关系型数据库由DBService组件维护。

Hive与Spark间的关系

Hive的数据计算也可以运行在Spark上。Spark也是Apache的一个项目,它是基于内存的分布式计算框架。Hive进行数据分析时,会将用户提交的HiveQL语句解析成相应的Spark任务并提交Spark执行。

### 1.15 MRS 集群是否支持 Hive on Spark?

目前仅MRS 1.9.x版本集群支持Hive on Spark。其他版本可使用Hive on Tez替代。

### 1.16 Hive 版本之间是否兼容?

Hive 3.1版本与Hive 1.2版本相比不兼容内容如下:

- 字段类型约束: String不支持转成int
- UDF不兼容: UDF内的Date类型改成Hive内置的
- 索引功能废弃
- 时间函数问题: Hive 3.1版本为UTC时间,Hive 1.2版本为当地时区时间
- 驱动不兼容: Hive 3.1和Hive 1.2版本的idbc驱动不兼容
- Hive 3.1对ORC文件列名大小写,下划线敏感
- Hive 3.1版本列中不能有名为time的列

# 1.17 MRS 集群哪个版本支持建立 Hive 连接且有用户同步功能?

MRS 2.0.5以上版本支持DGC建立hive连接且有IAM用户同步功能。

### 1.18 Spark 和 Hadoop 什么关系?

Spark是和Hadoop数据兼容快速和通用的处理引擎。通过Yarn,Spark可以在Hadoop集群下运行。Spark可以处理在HDFS、HBase、Hive以及Hadoop输入的任何形式的数据。

### 1.19 MRS 集群是否支持 Spark on Hive?

MRS支持Spark on Hive, Spark on Hive参考文档从Hive读取数据再写入HBase。

### 1.20 数据存储在 OBS 和 HDFS 有什么区别?

MRS集群处理的数据源来源于OBS或HDFS,HDFS是Hadoop分布式文件系统(Hadoop Distributed File System),OBS(Object Storage Service)即对象存储服务,是一个基于对象的海量存储服务,为客户提供海量、安全、高可靠、低成本的数据存储能力。MRS可以直接处理OBS中的数据,客户可以基于云管理平台Web界面和OBS客户端对数据进行浏览、管理和使用,同时可以通过REST API接口方式单独或集成到业务程序进行管理和访问数据。

- 数据存储在OBS:数据存储和计算分离,集群存储成本低,存储量不受限制,并 且集群可以随时删除,但计算性能取决于OBS访问性能,相对HDFS有所下降,建 议在数据计算不频繁场景下使用。
- 数据存储在HDFS:数据存储和计算不分离,集群成本较高,计算性能高,但存储量受磁盘空间限制,删除集群前需将数据导出保存,建议在数据计算频繁场景下使用。

### 1.21 Hadoop 压力测试工具如何获取?

请从如下URL中下载: https://github.com/Intel-bigdata/HiBench

### 1.22 Impala 与其他组件有什么关系?

● Impala与HDFS间的关系

Impala默认利用HDFS作为其文件存储系统。Impala通过解析和计算处理结构化的数据,Hadoop HDFS则为Impala提供了高可靠性的底层存储支持。使用Impala将无需移动HDFS中的数据并且提供更快的访问。

• Impala与Hive间的关系

Impala使用Hive的元数据、ODBC驱动程序和SQL语法。与Hive不同,Impala不基于MapReduce算法,它实现了一个基于守护进程的分布式架构,它负责在同一台机器上运行的查询执行的所有方面。因此,它减少了使用MapReduce的延迟,这使Impala比Hive快。

• Impala与MapReduce间的关系

无

● Impala与Spark间的关系

无

● Impala与Kudu间的关系

Kudu与Impala紧密集成,替代Impala+HDFS+Parquet组合。允许使用Impala的SQL语法从Kudu tablets插入、查询、更新和删除数据。此外,还可以用 JDBC或ODBC,Impala作为代理连接Kudu进行数据操作。

• Impala与HBase间的关系

默认的Impala表使用存储在HDFS上的数据文件,这对于使用全表扫描的批量加载和查询是理想的。但是,HBase可以提供对OLTP样式组织的数据的便捷高效查询。

### 1.23 Kudu 和 HBase 间的关系?

Kudu的设计有参考HBase的结构,也能够实现HBase擅长的快速的随机读写、更新功能。二者的结构差别不大,主要差别在于:

- Kudu不依赖Zookeeper,通过自身实现Raft来保证一致性。
- Kudu持久化数据不依赖HDFS,TServer实现数据的强一致性和可靠性。

### 1.24 MRS 是否支持 Hive on Kudu?

MRS不支持Hive on Kudu。

### 目前MRS只支持两种方式访问kudu:

- 通过impala表访问kudu。
- 通过客户端应用程序访问操作kudu表。

### 1.25 10 亿级数据量场景的解决方案

- 有数据更新、联机事务处理OLTP、复杂分析的场景,建议使用云数据库 GaussDB(for MySQL)。
- MRS的Imapal + Kudu也能满足该场景, Impala + Kudu可以在join操作时, 把当前所有的join表都加载到内存中来实现。

### 1.26 如何修改 DBService 的 IP?

MRS不支持修改DBService的IP。

### 1.27 MRS sudo log 能否清理?

MRS sudo log文件是omm用户的操作记录,是为了方便问题的定位 ,可以清理。因为日志占用了一部分存储空间,建议客户可以清除比较久远的操作日志释放资源空间。

- 1. 日志文件较大,可以将此文件目录添加到/etc/logrotate.d/syslog中,让系统做日志老化 ,定时清理久远的日志 。
  - 方法: 更改文件日志目录: sed -i '3 a/var/log/sudo/sudo.log' /etc/logrotate.d/syslog
- 2. 可以根据日志个数和大小进行设置/etc/logrotate.d/syslog,超过设置的日志会自动删除掉。一般默认按照存档大小和个数进行老化的,可以通过size和rotate分别是日志大小限制和个数限制,默认没有时间限制,如需进行时间设置可以增加daily/weekly/monthly指定周期为每天/每周/每月。

# **2** 计费相关

### 2.1 MRS 如何收费?

MRS支持按需购买集群节点和包年/包月购买集群节点,具体的计费详情可通过<mark>价格计算器</mark>查询。

- 包年/包月:根据集群购买时长,一次性支付集群费用。最短时长为1个月,最长时长为1年。
- 按需付费: 节点按实际使用时长计费, 计费周期为一小时。

#### □说明

- 目前除了"启动中"、"失败"和"已删除"状态的集群不收费,其他状态的集群都需要收费。
- 此处费用只是购买集群所需费用,MRS上的数据存储、带宽和流量另外计费。
- 扣费时余额不足先提醒用户续费,在保留期冻结集群资源,续费后再解冻。

### 2.2 创建 MRS 集群时产品报价如何计费?

只选择磁盘个数,没有选择集群虚拟机实例个数时,产品报价是不会产生收费的;需要客户选择实例个数,产品报价会在浏览器下方即时显示出来。

### 2.3 MRS 集群中的弹性伸缩功能是如何收费的?

在创建集群时,Task节点只配置了弹性伸缩的取值范围,而没有配置实例数量时,产品报价不会产生收费。如果客户配置了实例数量,则产品报价会在浏览器下方即时显示出来。

如果在集群使用时,配置的弹性伸缩功能新增了节点,则包年/包月集群和按需计费集 群新增节点的计费都是按实际使用时长计费,计费周期为一小时。

### 2.4 MRS 服务是否支持按需计费转包周期(包年/包月)计 费?

MRS服务暂时不支持按需计费转包周期(包年/包月)计费。客户可以通过对ECS关机的方式来降低消费成本。

### 2.5 集群手动扩容如何计费?

集群扩容收费模式和已有同类型节点保持一致,计费会按照新增节点加上已有节点使用的资源计算。扩容会产生一个新的订单,当扩容成功后才开始计费,详细说明请参考文档**扩容集群**。

### 2.6 MRS 如何续费?

目前MRS提供按需和包年/包月购买方式,按需是每小时扣费,如果余额不足导致欠费,而包年/包月是集群时长用完前需续费。如果您未能续费,华为云不会立即停止您的业务,订单转入保留期,此时集群将终止服务,数据仍然保留。

如需续费,请进入"续费管理"页面进行续费操作。

### 2.7 Mapreduce 服务集群的 Task 节点如何收费?

包年/包月集群和按需计费集群的Task节点的计费模式都是按需计费,即按实际使用时长计费,计费周期为一小时。

### 2.8 退订 MRS 服务后,在 ECS 中退订弹性云服务器时报异常 什么原因?

- 1. 在使用的MRS集群中查询该弹性云服务器的ID确认没有使用。
- 2. 在ECS控制台中找到需要退订的服务器单击"MRS使用中"将机器进行解锁。
- 3. 再次单击"退订",。
- 4. 若依然显示退订异常,请收集弹性云服务器的ID并联系华为云支持人员。

## **3** 账号密码类

## 3.1 登录 Manager 账号的是什么?

系统默认登录Manager的账号为admin,密码为创建集群时用户自己设置的密码。

# **4** 账号权限类

# 4.1 如果不开启 Kerberos 认证,MRS 集群能否支持访问权限细分?

MRS 1.8.0版本之前未开启Kerberos认证的集群不支持访问权限细分。MRS 1.8.0之前版本只有开启Kerberos认证才有角色管理权限,MRS 1.8.0及之后版本的所有集群均拥有角色管理权限。

MRS 2.1.0及之前版本: 在MRS Manager页面选择"系统设置">"配置">"权限配置"查询。

MRS 3.x及之后版本:在FusionInsight Manager页面选择"系统 > 权限"查询。

权限配置的相关使用方法请参考权限管理。

Kerberos集群的使用方式请参考使用Kerberos认证集群。

### 4.2 如何给新建的账号添加租户管理权限?

分析集群和混合集群支持添加租户管理权限,流式集群不支持添加租户管理权限。给 新建账号添加租户管理权限方法如下:

MRS 2.1.0及之前版本:用admin账号登录MRS Manager,在"系统设置">"用户管理"中选择新建的账号,单击"操作"列中的"修改",在"分配角色权限"中单击"选择并绑定角色"。绑定Manager\_tenant角色,则该账号拥有租户管理的查看权限。绑定Manager\_administrator角色,则该账号拥有租户管理的查看和操作权限。

MRS 3.x及之后版本:

**步骤1** 登录FusionInsight Manager, 选择"系统 > 权限 > 用户"。

步骤2 在要修改信息的用户所在行,单击"修改"。

根据实际情况,修改对应参数。

绑定Manager\_tenant角色,则该账号拥有租户管理的查看权限。绑定Manager\_administrator角色,则该账号拥有租户管理的查看和操作权限。

#### □ 说明

修改用户的用户组,或者修改用户的角色权限,最长可能需要3分钟时间生效。

步骤3 单击"确定"完成修改操作。

----结束

### 4.3 MRS 服务的 IAM 用户能关联 MRS Manager 用户吗?

- 1. IAM用户不能关联MRS Manager用户,但是可以同步IAM用户到MRS Manager中,即具有基础的MRS Manager权限,具体请参考IAM用户同步MRS。
- 2. 如果需要IAM用户A也拥有MRS Manager用户B的权限,请登录MRS Manager,选择"系统设置 > 用户管理",修改用户A的绑定的角色,为用户A绑定用户B相同的角色即可。

### 4.4 MRS 集群中如何创建 IAM 委托?

由于华为云各服务之间存在业务交互关系,一些云服务需要与其他云服务协同工作,需要您创建云服务委托,将操作权限委托给该服务,让该服务以您的身份使用其他云服务,代替您进行一些资源运维工作。具体请参考**委托其他云服务管理资源**。配置MRS多用户访问OBS细粒度权限请参考配置MRS多用户访问OBS细粒度权限。

# 4.5 如何实现 MRS Manager 和 IAM 的权限粒度为表级别或库级别?

IAM用户不能关联Manager用户,但是可以同步IAM用户到Manager中,即具有基础的Manager权限,具体请参考IAM用户同步MRS。

用户的表级和库级权限只能在Manager上添加,具体请参考配置Hive权限。

### 4.6 如何自定义配置 MRS 服务策略?

- 1. 在IAM控制台,单击左侧导航栏的"权限",在右上角选择"创建自定义策略"。
- 2. 策略名称: 自定义策略的名称。
- 3. 作用范围:根据服务的属性填写,MRS为项目级服务,选择"项目级服务"。
- 4. 策略配置方式。
  - 可视化视图:通过可视化视图创建自定义策略,无需了解JSON语法,按可视 化视图导航栏选择云服务、操作、资源、条件等策略内容,可自动生成策 略。
  - JSON视图:通过JSON视图创建自定义策略,可以在选择策略模板后,根据 具体需求编辑策略内容;也可以直接在编辑框内编写JSON格式的策略内容。

也可以从"策略内容"区域,单击"从已有策略复制"选择已有策略作为模板进行修改。

- 5. 输入"策略描述"(可选)。
- 6. 单击"确定",自定义策略创建完成。

7. 将新创建的自定义策略授予用户组,使得用户组中的用户具备自定义策略中的权限。

#### □ 说明

详细操作请参考创建自定义策略。

# 4.7 在 MRS Manager 页面"系统设置"中找不到用户管理,什么原因?

该用户没有Manager\_administrator角色权限,所以在MRS Manager页面"系统设置"中不显示"用户管理"。

### 4.8 IAM 用户同步的具体作用?

IAM用户同步是指将绑定MRS相关策略的IAM用户同步至MRS系统中,创建同用户名、不同密码的账号,用于集群管理。同步之后,用户可以使用IAM用户名(密码需要MRS Manager的管理员admin重置后方可使用)登录MRS Manager管理集群。也可以在开启Kerberos认证的集群中,通过界面方式提交作业。具体请参考文档IAM用户同步MRS。

### 4.9 Hue 有没有配置账号权限的功能?

Hue没有配置账号权限的功能,可以通过在Manager的"系统设置"中配置用户角色和用户组来配置账号权限。

## 5 客户端使用类

### 5.1 如何使用组件客户端?

- 1. 以root用户登录任意一个Master节点。
- 2. 执行su omm命令,切换到omm用户。
- 3. 执行cd /opt/client命令,切换到客户端。
- 4. 执行**source bigdata\_env**命令,配置环境变量。 如果当前集群已启用Kerberos认证,执行**kinit** *组件业务用户*认证当前用户。如果 当前集群未启用Kerberos认证,则无需执行此命令。
- 5. 环境变量配置成功后,即可执行组件的客户端命令。例如查看组件的相关信息,可执行HDFS客户端命令**hdfs dfs -ls** /查看HDFS根目录文件。

### 5.2 没有绑定 EIP 是否可以下载客户端?

没有申请弹性公网IP可以下载和更新MRS服务客户端,具体请参考**集群外节点使用**MRS客户端。

### 5.3 Flume 客户端是否支持 Windows?

Flume客户端不支持安装在Windows系统机器中。如需在其他节点使用Flume客户端,请确保该节点与MRS集群在相同的VPC和子网中,具体详情可以参考<mark>虚拟私有云的其他节点使用客户端</mark>。

# **6** Web 页面访问类

### 6.1 如何在内网访问 MRS manager 页面?

当用户的机器和MRS集群不在同一个内网时,可以通过在弹性云服务器控制台中创建一个Windows弹性云服务器来访问MRS Manager,具体请参考**通过Windows弹性云服务器访问MRS Manager**。

# 6.2 MRS Manager 页面报 OBS 证书即将过期告警,应该如何处理?

请参考如何处理集群内部OBS证书过期更新证书。

因为不同节点包含的证书可能不一致,因此需要各个节点单独操作,不建议直接更新一个节点的证书然后拷贝到所有节点。

### 6.3 本地 IP 和华为云不在同一个网段中如何使用 HDFS?

为集群绑定EIP,然后通过EIP访问集群的HDFS,具体请参考**通过弹性公网IP访问Web页面**。

### 6.4 修改开源组件 Web 页面会话超时时间

请合理设置Web页面超时时间,避免由于Web页面长时间暴露造成的信息泄露。

- MRS 2.1及之前版本集群:在集群详情页面,选择"组件管理 > meta > 服务配置",切换"基础配置"为"全部配置",搜索配置项 "http.server.session.timeout.secs",如果有该配置项请参考如下步骤修改,如果没有该配置项则版本不支持动态调整会话时长。
- MRS 3.x及之后版本集群:登录FusionInsight Manager,选择"集群 > 服务 > meta",单击"配置",选择"全部配置"。搜索配置项 "http.server.session.timeout.secs",如果有该配置项请参考如下步骤修改,如果没有该配置项则版本不支持动态调整会话时长。

所有超时时长请设置为统一值,避免时长设置不一致导致的页面实际生效的超时时长 和设置值的冲突。

#### ● 修改Manager页面及认证中心的超时时长

- a. 分别登录集群所有Master节点,在所有Master节点中执行如下修改。
- b. 修改 "/opt/Bigdata/apache-tomcat-7.0.78/webapps/cas/WEB-INF/web.xml"中的 "<session-timeout>20</session-timeout>",其中20为会话超时时间请根据需要修改,单位为分钟,超时时间最长不要超过480分钟。
- c. 修改 "/opt/Bigdata/apache-tomcat-7.0.78/webapps/web/WEB-INF/web.xml"中的 "<session-timeout>20</session-timeout>" 其中20为会话超时时间请根据需要修改,单位为分钟,超时时间最长不要超过480分钟。
- d. 修改 "/opt/Bigdata/apache-tomcat-7.0.78/webapps/cas/WEB-INF/spring-configuration/ticketExpirationPolicies.xml"中的 "p:maxTimeToLiveInSeconds="\${tgt.maxTimeToLiveInSeconds:1200}""和 "p:timeToKillInSeconds="\${tgt.timeToKillInSeconds:1200}"",其中 1200为认证中心的有效时长请根据需要修改,单位为秒,有效时长不要超过 28800秒。
- e. 在主管理节点重启Tomcat节点。
  - i. 在主master节点上用omm用户执行**netstat -anp |grep 28443 |grep** LISTEN | awk '{print \$7}'查询Tomcat的进程号。
  - ii. 执行kill -9 {pid},其中{pid}为上一步中获得的Tomcat进程号。
  - iii. 等待进程自动重启。可以执行netstat -anp | grep 28443 | grep LISTEN 查看进程是否重启成功,如果可以查到进程说明已经重启成功,如果未查到请稍后再次查询。

#### ● 修改开源组件Web页面的超时时间

a. 进入服务全部配置界面。

MRS 2.1及之前版本集群:在集群详情页面,选择"组件管理 > meta > 服务配置"。

MRS 3.x及之后版本集群: 登录FusionInsight Manager, 选择"集群 > 服务 > meta", 单击"配置", 选择"全部配置"。

- b. 根据需要修改"meta"下的"http.server.session.timeout.secs"值,单位为 秒。
- c. 保存配置,不勾选"重新启动受影响的服务或实例"并单击"确定"。 重启会影响业务,建议在业务空闲时执行重启操作。
- d. (可选)若需要使用Spark的Web页面,则需要Spark"全部配置"页面,搜索并修改配置项"spark.session.maxAge"为合适的值,单位为秒。 保存配置,不勾选"重新启动受影响的服务或实例"并单击"确定"。
- e. 重启meta服务及需要使用Web界面的服务,或者在业务空闲时重启集群。 重启会影响业务,建议在业务空闲时执行重启操作,或使用滚动重启功能, 在不影响业务的情况下重启服务,具体请参考**支持滚动重启**。

### 6.5 MRS 租户管理中的动态资源计划页面无法刷新

- 1. 以root用户分别登录Master1和Master2节点。
- 2. 执行ps -ef | grep aos命令检查aos进程号。

- 3. 执行kill -9 aos进程号结束aos进程。
- 4. 等待aos进程自动重启成功,可通过**ps** -**ef** |**grep** aos命令查询进程是否存在,若存在则重启成功,若不存在请稍后再查询。

### 6.6 用户修改域名后无法登录 MRS Manager 页面

### 问题现象

用户修改域名后,通过console页面登录无法登录MRS Manager页面,或者登录MRS Manager页面异常。

### 问题原因

用户修改域名后,没有刷新executor用户的keytab文件,导致executor进程认证失败 后不断循环认证,导致了acs进程内存溢出。

### 解决方案

步骤1 替换executor用户的keytab文件。

- 1. 登录MRS Manager页面,选择"系统 > 用户",在executor用户所在的"操作"列,单击"下载认证凭据",解压后获取keytab文件。
- 2. 使用root用户登录主管理节点,将获取的keytab替换"/opt/executor/webapps/executor/WEB-INF/classes/user.keytab"文件。

步骤2 替换knox用户的keytab和conf文件。

- 1. 登录MRS Manager页面,选择"系统 > 用户",在knox用户所在的"操作"列,单击"下载认证凭据",解压后获取keytab和conf文件。
- 2. 使用root用户登录主管理节点,将获取的keytab替换"/opt/knox/conf/user.keytab"文件。
- 3. 修改/opt/knox/conf/krb5JAASLogin.conf中的principal的值,把域名修改为更改后的域名。
- 4. 将获取的krb5.conf 替换"/opt/knox/conf/krb5.conf"文件。

#### 步骤3 备份原有客户端目录

mv {客户端目录} / opt/client init

**步骤4** 参考更新客户端章节,重新安装客户端。

步骤5 使用root用户登录主备管理节点,执行如下命令,重启knox进程。

su - omm

ps -ef | grep gateway | grep -v grep ( 查找knox进程PID )

kill -9 PID(PID替换为实际的ID,结束knox进程)

/opt/knox/bin/restart-knox.sh ( 启动knox进程 )

步骤6 使用root用户登录主备管理节点,执行如下命令,重启executor进程。

su - omm

netstat -anp | grep 8181 | grep LISTEN ( 查找executor进程PID )

**kill -9** *PID*(PID替换为实际的ID,结束knox进程) **/opt/executor/bin/startup.sh**(启动executor进程) ----**结束** 

### 6.7 Kafka Topic 监控页签在 Manager 页面不显示

步骤1 分别登录集群Master节点,并切换用户为omm。

步骤2 进入目录"/opt/Bigdata/apache-tomcat-7.0.78/webapps/web/WEB-INF/lib/components/Kafka/"。

步骤3 拷贝zookeeper包到该目录cp /opt/share/zookeeper-3.5.1-mrs-2.0/ zookeeper-3.5.1-mrs-2.0.jar ./。

步骤4 重启Tomcat。

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

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

----结束

## 上 监控告警类

# 7.1 在 MRS 流式集群中,Kafka topic 监控是否支持发送告警?

暂不支持Kafka topic监控发送邮件和短信告警,目前用户可以在Manager界面看到告警信息。

# 8 性能优化类

### 8.1 MRS 集群是否支持重装系统?

MRS集群不支持重装系统。

### 8.2 如何提高集群 Core 节点的资源使用率?

- 1. 进入 Yarn服务配置页面:
  - MRS 1.8.10及之前版本,登录MRS Manager页面,具体请参见访问MRS
     Manager,然后选择"服务管理 > Yarn > 服务配置",单击"基础配置"下拉菜单,选择"全部配置"。
  - MRS 1.8.10之后及2.x版本,单击集群名称,登录集群详情页面,选择"组件管理 > Yarn > 服务配置",单击"基础配置"下拉菜单,选择"全部配置"。

### □ 说明

若集群详情页面没有"组件管理"页签,请先完成IAM用户同步(在集群详情页的"概览"页签,单击"IAM用户同步"右侧的"点击同步"进行IAM用户同步)。

- MRS 3.x及后续版本,登录FusionInsight Manager。然后选择"集群 > 待操 作的集群名称 > 服务 > Yarn > 配置 > 全部配置"。
- 2. 搜索并修改"yarn.nodemanager.resource.memory-mb"的值,请根据集群的节点内存实际情况调大该值。
- 3. 保存配置并重启受影响的服务或实例。

### 8.3 如何配置 MRS 默认磁盘使用率?

mapreduce默认磁盘使用率为90%,修改磁盘使用率方法:

MRS 2.1.0及之前版本: 登录MRS Manager页面上,然后在"系统设置 > 阈值配置"中更改告警阈值。

MRS 3.x及之后版本: 登录FusionInsight Manager, 选择"运维 > 告警 > 阈值配置",更改告警阈值。

### 8.4 如何关闭防火墙服务以避免对系统造成影响?

步骤1 以root用户登录集群的各个节点。

**步骤2** 检查防火墙服务是否启动。若防火墙启动了可以执行**步骤3**关闭防火墙。 例如,EulerOS环境下执行**systemctl status firewalld.service**命令。

步骤3 关闭防火墙服务。

例如,EulerOS环境下执行systemctl stop firewalld.service命令。

----结束

# 8.5 MRS 集群在不增加 Core 节点的情况下,如何增加节点的 HDFS 容量?

不增加Core节点的情况下,无法增加节点的HDFS容量。

HDFS服务的容量由集群分配,不会大于Core节点的数据盘空间。创建集群设置Core节点规格时就已确定。后续无法单独更改节点磁盘空间。您可以通过扩容集群增加HDFS服务容量,具体操作请参考**扩容集群**。

# **9** 作业开发类

### 9.1 如何准备 MRS 的数据源?

MRS既可以处理OBS中的数据,也可以处理HDFS中的数据。在使用MRS分析数据前,需要先准备数据。

- 1. 将本地数据上传OBS。
  - a. 登录OBS管理控制台。
  - b. 在OBS上创建userdata桶,然后在userdata桶下创建program、input、output和log文件夹。
    - i. 单击"创建桶",创建一个名称为userdata的桶。
    - ii. 在OBS桶列表中单击桶名称userdata,选择"对象 > 新建文件夹",分别创建program、input、output和log目录。
  - c. 上传数据至userdata桶。
    - i. 进入program文件夹,单击"上传对象"。
    - ii. 单击"添加文件"并选择用户程序。
    - iii. 单击"上传"。
    - iv. 使用同样方式将用户数据文件上传至input目录。
- 2. 将OBS数据导入至HDFS。

当"Kerberos认证"为"关闭",且运行中的集群,可执行将OBS数据导入至HDFS的操作。

- a. 登录MRS管理控制台。
- b. 单击集群名称进入集群详情页面。
- c. 单击"文件管理",选择"HDFS文件列表"。
- d. 进入数据存储目录,如"bd\_app1"。
  - "bd\_app1"目录仅为示例,可以是界面上的任何目录,也可以通过"新建"创建新的目录。
- e. 单击"导入数据",通过单击"浏览"选择OBS和HDFS路径,如**图9-1**所示。

### 图 9-1 导入文件



f. 单击"确定"。 文件上传进度可在"文件操作记录"中查看。

### 9.2 如何将用户程序提交到 MRS 中?

MRS为用户提供程序执行平台,程序由用户自身开发,MRS负责程序的提交、执行和监控。那如何将用户程序提交至MRS中呢?新增作业时,将"执行程序路径"参数配置为用户程序的实际存放路径,详细信息请参考管理作业。

### 9.3 如何调整 JobHistory 进程堆内存?

MRS 2.1.0及之前版本: 登录**MRS Manager**, 选择"服务管理 > Spark > 服务配置 > 全部配置 > JobHistory > 默认"。

MRS 3.x及之后版本: 登录FusionInsight Manager。选择"集群 > 服务 > Spark > 配置 > 全部配置 > JobHistory > 默认"。

将 "SPARK\_DAEMON\_MEMORY"参数的值根据实际情况调大。具体请参考 ALM-43006 JobHistory进程堆内存使用超出阈值。

### 9.4 集群支持提交哪些形式的 Spark 作业?

当前在MRS页面,集群支持提交Spark、Spark Script和Spark SQL形式的Spark作业。

## 9.5 MRS 集群的租户资源最小值改为 0 后,只能同时跑一个 Spark 任务吗?

MRS集群的租户资源最小值改为0后,只能同时跑一个Spark任务。

# 10 集群升级/补丁

### 10.1 MRS 版本如何进行升级?

MRS目前还无法实现低版本到高版本的平滑升级。目前只能重新购买一个新版本的集群,然后将老版本集群的数据迁移到新的集群。

### 10.2 MRS 是否支持修改版本?

MRS不支持修改版本,建议退订集群之后重新创建。

### 10.3 是否可以自主升级 MRS 组件版本?

MRS服务不支持客户自己升级MRS服务组件版本 ,如需要更高的版本的组件请购买相应版本的MRS服务集群,MRS组件版本请参见MRS组件版本一览表。

# **1 1** 周边生态对接使用

### 11.1 MRS 集群监控指标是否可以集成到云监控上?

不可以,MRS集群监控指标为用户的内部监控,属于用户隐私内容,MRS是不感知该内容,也不会集成在云监控上。

### 11.2 是否可以使用 MRS 服务对 DLI 业务表进行读写操作?

如果是数据存放在OBS上,则支持使用MRS服务中的Spark读取DLI表,对表数据进行 灵活处理,结果存入另一张DLI表,如果数据不存放在OBS上,则不支持。

### 11.3 MRS 集群是否支持 Tableau 连接 MRS 的 Hive?

Tableau不支持对接MRS集群的Hive。

# 11.4 MRS 建立后同地域同可用区访问 OBS 是内网访问么?是否有限速?

- 1. MRS建立后同地域同可用区访问OBS是内网访问。
- 2. MRS服务不限速,OBS服务也不限速,客户通过公网访问OBS时,受公网带宽限制。客户通过内网访问OBS时受网卡、磁盘IO及是否有其它进程抢占资源的限制,具体请参考OBS上传下载速率的影响因素有哪些?。

### 11.5 MRS 中可以部署爬虫服务吗?

MRS中不能部署爬虫服务。

# 11.6 DWS 和 MRS 服务,是否支持安全删除(删除后防止非法恢复)的能力?

MRS目前仅支持恢复备份的数据,DWS只能通过快照恢复,即DWS和MRS服务的删除都是安全删除。

### 11.7 OBS 是否支持 ListObjectsV2 协议?

OBS不支持ListObjectsV2协议。

### 11.8 MRS 的数据存储可以使用并行文件系统的 OBS 吗?

MRS服务支持并行文件系统类型的OBS。

### 11.9 数据湖工厂是否支持对接 MRS 的集群?

数据湖工厂(DLF)支持通过V2接口对接MRS集群。若MRS集群不支持V2接口,则只能对接MRS普通集群,不支持对接开启Kerberos认证的安全集群,如果客户一定要使用开启Kerberos认证的MRS集群,建议使用DGC平台对接。

#### □ 说明

- 针对MRS 1.X版本的集群, MRS 1.8.7及之后版本支持V2接口。
- 针对MRS 2.X版本的集群, MRS 2.0.3及之后版本支持V2接口。

# 11.10 为什么 DLF 中建立连接时没有找到 Kerberos 开启的 MRS 服务集群?

DLF不支持对接开启Kerberos认证的MRS集群,如果客户一定要使用开启Kerberos认证的MRS集群,建议使用DGC平台对接。

### 11.11 如何在 ECS 服务器上用 PySpark 连接内网开启 Kerberos 认证的 MRS Spark 集群?

将Spark的spark-defaults.conf文件中

"spark.yarn.security.credentials.hbase.enabled"修改为"true",然后使用**spark-submit --master yarn --keytab keytabfile --principal principal**指定Kerberos认证文件。

### 11.12 HBase 同步数据到 CSS 为什么没有映射字段?

从MRS的HBase同步数据到CSS服务后,整库没有映射字段,要单表才有映射的字段。

### 11.13 Spark 集群能访问 OBS 中的数据吗?

Spark集群可以访问存储在OBS系统中的数据,具体配置请参考配置存算分离集群。

# 11.14 MRS 中的 Spark 组件与 DLI 中的 Spark 有什么区别?

- MRS服务的Spark组件的是建立在客户购买的MRS服务所分配的虚机上,客户可以 根据实际需求调整及优化Spark服务。支持各种接口调用,具体请参考Spark接 口。
- DLI服务的Spark组件是全托管式服务,客户对Spark组件不感知,仅仅可以使用该服务。且接口为封装式接口,具体请参考JDBC API参考

### 11.15 如何在 Flume 组件中把数据从 OBS 导入 HDFS 中?

Flume把数据从OBS导入HDFS中,需要加载hdfs-site.xml和core-site.xml文件到Flume服务端,并重启Flume。

### 11.16 组件对接是否支持外部 KDC?

MRS服务只支持使用自带的内部KDC。

### 11.17 开源 Kylin 3.x 版本与 MRS 1.9.3 版本对接时存在 Jetty 版本的兼容性问题

出于安全考虑,MRS对部分存在严重安全漏洞的开源三方组件进行了版本升级,导致开源Kylin与MRS 1.9.3版本进行对接时存在Jetty版本的兼容性问题。

按以下步骤进行操作可以完成部署和对接。

- 1. 在一个ECS节点进行MRS Client安装,具体请参考集群外节点使用MRS客户端。
- 2. 安装完成后执行如下命令导入环境变量。

source ./bigdata\_envexport HIVE\_CONF=/srv/client/Hive/config/export HCAT HOME=/srv/client/Hive/HCatalog

3. 参照**Kylin官方链接**在MRS Client安装节点上安装Kylin,并指定KYLIN\_HOME。 MRS 1.9.3版本选择Kylin for HBase 1.x版本进行对接。

export KYLIN HOME=/srv/client/apache-kylin-3.0.2-bin-hbase1x

4. 移除Hive客户端目录下/srv/client/Hive/Beeline/lib/目录中jetty相关jar包以避免版本冲突。

#### jetty相关jar包:

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. 启动Kylin服务,观察Kylin日志,正常情况下日志中不会再出现 java.lang.NoSuchMethodException或java.lang.ClassNotFoundException等兼容 性异常 。

#### \$KYLIN\_HOME/bin/kylin.sh start

6. 访问Kylin原生页面http://<hostname>:7070/kylin并运行样例Cube脚本\$ {KYLIN\_HOME}/bin/sample.sh验证kylin是否正常运行。

#### □ 说明

在华为云严选商城中有提供基于Apache Kylin的企业级大数据智能分析平台**Kyligence Enterprise**,您可以根据需要选择使用。

### 11.18 MRS 导出数据到 OBS 加密桶失败

目前仅MRS 1.9.x及之后版本支持OBS加密桶功能,MRS 1.8.x及之前版本不支持OBS加密桶功能。如需使用OBS加密桶请使用MRS 1.9.x及之后版本集群。

# **12** <sub>集群访问类</sub>

## 12.1 MRS 登录集群节点的两种方式能够切换么?

不可以。购买集群时选择登录方式后不能更改登录方式。

## 12.2 如何远程使用 MRS 集群?

远程登录MRS集群,详细内容可参见远程操作指导。

## 12.3 本地联调 MRS 集群的 IP 和端口设置方法有什么?

给需要登录集群的所有节点绑定弹性IP地址,具体请参考**为弹性云服务器申请和绑定 弹性公网IP**。

端口请根据业务实际需要配置,具体请参考开源组件端口列表。

## 12.4 如何获取 ZooKeeper 地址?

- 针对MRS 3.x及之后版本集群,请执行如下步骤获取:
  - a. 登录FusionInsight Manager。
  - b. 选择"集群 > 服务 > ZooKeeper > 实例",获取ZooKeeper实例的"管理 IP"地址。
- 针对MRS 2.1.0及之前版本版本集群,请执行如下步骤获取:
  - a. 在MRS集群详情页的"概览"页签,单击"IAM用户同步"右侧的"单击同步"进行IAM用户同步。
  - b. 选择"组件管理 > ZooKeeper > 实例",获取ZooKeeper实例的"管理IP" 地址。

#### 图 12-1 ZooKeepe 地址



- 针对MRS 2.0.1之前版本集群,请执行如下步骤获取:
  - a. 登录MRS Manager。
  - b. 选择"服务管理 > ZooKeeper > 实例",获取ZooKeeper实例的"管理IP" 地址。

#### 图 12-2 ZooKeepe 地址



# 13 大数据业务开发

## 13.1 MapReduce 服务中的样例工程构建和应用开发有什么区别?是否支持 python 代码?

- 1. MapReduce服务中的样例工程和应用开发没有区别,两者都可以选择。
- 2. MRS服务支持python代码。

## 13.2 HBase 支持的压缩算法有哪些?

HBase目前支持的压缩算法有snappy、lz4和gz。

### 13.3 如何使用 HBase?

使用HBase的相关参考请参见从零开始使用HBase, HBase应用开发。

HBase等各开源组件web站点,请参考**开源组件Web站点**。

## 13.4 MRS 是否支持通过 Hive 的 HBase 外表将数据写入到 HBase?

不支持。Hive on HBase只支持查询,不支持更改数据。

## 13.5 如何查看 HBase 日志?

- 1. 登录集群的Master节点。
- 2. 执行su omm命令, 切换到omm用户。
- 3. 执行cd /var/log/Bigdata/hbase/命令,进入到/var/log/Bigdata/hbase/目录,即 可查看HBase日志信息。

## 13.6 开启 Kerberos 认证的集群如何访问 hbase shell?

需要到MRS Manager页面上下载MRS服务客户端,并在访问节点上安装MRS服务客户端,具体请参考**集群外节点使用MRS客户端**。

## 13.7 DynamoDB 的 number 在 Hive 表中用什么类型比较好?

Hive支持smallint, 推荐使用smallint类型。

## 13.8 Hive Driver 是否支持对接 dbcp2?

Hive driver不支持对接dbcp2数据库连接池。dbcp2数据库连接池调用isValid方法检查连接是否可用,而Hive对于这个方法的实现就是直接抛异常。

## 13.9 用户 A 如何查看用户 B 创建的 Hive 表?

创建一个拥有访问用户B创建的表的角色,创建角色具体操作请参考配置Hive权限,然后为用户A绑定新创建的角色后等待5分钟左右即可访问到用户B创建的表。修改用户绑定的角色信息请参考修改用户信息。

## 13.10 如何使用本地电脑 IDEA 访问 Hive?

访问MRS集群来操作Hive,请参考JDBC客户端运行及结果查询。

## 13.11 如何在 Windows 下调测 sparkSQL 作业?

配置并调测sparkSQL作业请参考准备环境。

### 13.12 如何重置 Kafka 数据?

删除Kafka topic信息即重置Kafka数据,具体命令请参考:

- 删除topic: kafka-topics.sh --delete --zookeeper ZooKeeper集群业务IP.2181/ kafka --topic topicname
- 查询所有topic: kafka-topics.sh --zookeeper ZooKeeper集群业务IP.2181/ kafka --list

执行删除命令后topic数据为空则此topic会立刻被删除,如果有数据则会标记删除,后续Kafka会自行进行实际删除。

## 13.13 MRS 是否支持同时运行多个 Flume 任务?

Flume客户端可以包含多个独立的数据流,即在一个配置文件properties.properties中配置多个Source、Channel、Sink。这些组件可以链接以形成多个流。

#### 例如在一个配置中配置两个数据流,示例如下:

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

## 13.14 如何修改 FlumeClient 的日志为标准输出日志?

- 1. 登录Flume客户端安装节点。
- 2. 进入Flume客户端安装目录,假设Flume客户端安装路径为"/opt/FlumeClient",可以执行以下命令。

cd /opt/FlumeClient/fusioninsight-flume-1.9.0/bin

- 3. 执行./flume-manage.sh stop命令,停止Flume组件。
- 4. 执行**vi /log4j.properties**命令,打开log4j.properties文件,修改 "flume.root.logger"的取值为"\${flume.log.level},console"。
- 5. 执行**vim /flume-manager.sh**命令,打开flume安装目录bin目录下的启动脚本flume-manager.sh。
- 6. 修改flume-manager.sh脚本,注释如下内容。

>/dev/null 2>&1 &

- 7. 执行./flume-manage.sh start命令, 重启FlumeClient。
- 8. 修改完成后,请检查一下docker配置信息是否正确。

## 13.15 HDFS 如何进行数据均衡?

1. 登录集群的Master节点,执行如下命令启动balancer。

source /opt/client/bigdata\_env

sh /opt/client/HDFS/hadoop/sbin/start-balancer.sh -threshold 5

2. 查看日志。

balance任务执行时会在客户端安装目录"/opt/client/HDFS/hadoop/logs"目录下生成名为hadoop-root-balancer-主机名.out日志。

3. (可选)若不想再进行数据均衡,可执行如下命令停止balancer。

source /opt/client/bigdata\_env

sh /opt/client/HDFS/hadoop/sbin/stop-balancer.sh -threshold 5

### 13.16 如何修改 HDFS 的副本数?

- 1. 进入HDFS服务配置页面:
  - MRS 1.8.10及之前版本,登录MRS Manager页面,具体请参见访问MRS Manager,然后选择"服务管理 > HDFS > 服务配置",单击"基础配置"下拉菜单,选择"全部配置"。

MRS 1.8.10之后及2.x版本,单击集群名称,登录集群详情页面,选择"组件管理 > HDFS > 服务配置",单击"基础配置"下拉菜单,选择"全部配置"。

#### □ 说明

若集群详情页面没有"组件管理"页签,请先完成IAM用户同步(在集群详情页的"概览"页签,单击"IAM用户同步"右侧的"点击同步"进行IAM用户同步)。

- MRS 3.x及后续版本,登录FusionInsight Manager。然后选择"集群 > *待操 作的集群名*称 > 服务 > HDFS > 配置 > 全部配置"。
- 2. 搜索并修改"dfs.replication"的值,合理修改这个数值,该参数取值范围为 1~16,重启HDFS实例。

## 13.17 如何使用 Python 远程连接 HDFS 的端口?

HDFS开源3.0.0以下版本的默认端口为50070,3.0.0及以上的默认端口为9870。开源组件的端口号可参考**开源组件端口列表**。

## 13.18 Hadoop 组件 jar 包位置和环境变量的位置在哪里?

hadoopstreaming.jar位置在/opt/share/hadoop-streaming-\*目录下。

#### 山 说明

\*由Hadoop版本决定。

jdk环境变量:/opt/client/JDK/component\_env

Hadoop组件的环境变量位置: /opt/client/HDFS/component\_env

Hadoop客户端路径: /opt/client/HDFS/hadoop

## 13.19 Kudu 支持的压缩算法有哪些?

Kudu目前支持的压缩算法有snappy、lz4和zlib, 默认是lz4。

### 13.20 如何查看 Kudu 日志?

- 1. 登录集群的Master节点。
- 2. 执行su omm命令,切换到omm用户。
- 3. 执行**cd /var/log/Bigdata/kudu/**命令,进入到"/var/log/Bigdata/kudu/"目录,即可查看Kudu日志信息。

## 13.21 OpenTSDB 是否支持 python 的接口?

OpenTSDB基于HTTP提供了访问其的RESTful接口,而RESTful接口本身具有语言无关性的特点,凡是支持HTTP请求的语言都可以对接OpenTSDB,所以OpenTSDB支持python的接口。

## 13.22 Yarn 服务的 container 日志如何查看?

MRS中启动的yarn任务, container的日志目录请参考日志介绍。

## 13.23 MRS 的 Storm 集群提交任务时如何指定日志路径?

客户可以根据自己的需求,修改MRS的流式Core节点上的/opt/Bigdata/MRS\_XXX / 1\_XX \_Supervisor/etc/worker.xml文件,将标签filename的值设定为客户需要的路径,然后在Manager页面重启对应实例。

建议客户尽量不要修改MRS默认的日志配置,可能会造成日志系统异常。

## 13.24 Hive 样例工程中的 com.huawei.gaussc10 依赖包在哪里下载?

MRS目前暂没有com.huawei.gaussc10依赖包,且此为gaussDB依赖包可以不配置。建议用户在构建maven工程时不包含此包即可。

## 13.25 HDFS 系统支持 OBS 对象存储吗?

HDFS系统支持OBS对象存储,具体配置及使用方法请参考obs方式访问OBS。

## 13.26 kafka 访问协议配置为什么不生效?

请参考Kafka安全使用说明章节的"Kafka 访问协议说明"内容并按照用户自己的需求进行配置后方可生效。

## 13.27 HBase 表如何设置和修改数据保留期?

● 创建表时指定 创建t\_task\_log表,列族f, TTL设置86400秒过期 create 't\_task\_log',{NAME => 'f', TTL=>'86400'}

● 在已有表的基础上指定:
disable "t\_task\_log" #禁用表(这个需要停止业务)
alter "t\_task\_log",NAME=>'data',TTL=>'86400' #设置TTL值,作用于列族data enable "t\_task\_log" #恢复表

## 13.28 flink run 如何指定 Application-Name?

flink run 可以指定Application-Name,在flink run 中加入-ynm指定 Application-Name,提交Flink作业请参考<mark>运行Flink作业</mark>。

## 13.29 如何通过 happybase 连接到 MRS 服务的 HBase?

MRS服务集群的HBase服务使用的是thriftserver2,thriftserver1和thriftserver2不能同时并存,但happybase只能通过thriftserver1接口连接到HBase,故建议客户使用

python直接连接到HBase,具体实现请参考demohttps://github.com/huaweicloud/huaweicloud-mrs-example/blob/mrs-1.8/src/hbase-examples/hbase-python-example/DemoClient.py。

## 13.30 如何更改 Yarn 服务中队列的资源配比?

请参考配置资源池的队列容量策略进行修改。

## 13.31 如何指定 Flink 任务到指定的 Yarn 队列运行中运行?

提交任务时使用-qu,--queue <arg>: 指定YARN 队列。flink run参数说民请参见Flink Shell操作命令。

## 13.32 MRS Kafka 如何查看客户端版本信息?

用如下命令 --bootstrap-server 查看新版本客户端信息。

## 13.33 Kafka 目前支持的访问协议类型有哪些?

当前支持4种协议类型的访问: PLAINTEST、SSL、SASL\_PLAINTEXT、SASL\_SSL。

## 13.34 presto 如何配置其他数据源?

presto连接自建MySQL在目录/opt/Bigdata/MRS\_XXX/1\_21\_Coordinator/etc/catalog下配置connector文件即可。

## 13.35 Hive 查询数据是否支持导出?

Hive查询数据支持导出,请参考如下语句进行导出:

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

## 13.36 MRS 要如何连接 impala-shell

请使用impla-shell -i <host:port>选项。要自动连接到特定的Impala数据库,请使用-d <database>选项。例如,如果您的所有Kudu表都位于数据库"impala\_kudu"中,则-d impala\_kudu可以使用此数据库。要退出impala-shell,请使用quit命令。具体请参考从零开始使用Impala。

## 13.37 MRS 如何连接 spark-shell

- 1. 用root用户登录集群Master节点。
- 2. 配置环境变量。

source /opt/client/bigdata\_env

3. 如果当前集群已启用Kerberos认证,执行以下命令认证当前用户。如果当前集群未启用Kerberos认证,则无需执行此命令。

kinit MRS集群用户

#### 例如:

- 开发用户为"机机"用户时请执行: kinit -kt user.keytab sparkuser
- 开发用户为"人机"用户时请执行: kinit sparkuser
- 4. 执行Spark组件的客户端命令。

spark-shell

## 13.38 MRS 如何连接 spark-beeline

- 1. 用root用户登录集群Master节点。
- 2. 配置环境变量。

#### source /opt/client/bigdata\_env

3. 如果当前集群已启用Kerberos认证,执行以下命令认证当前用户。如果当前集群未启用Kerberos认证,则无需执行此命令。

kinit MRS集群用户

#### 例如:

- 开发用户为"机机"用户时请执行: kinit -kt user.keytab sparkuser
- 开发用户为"人机"用户时请执行:kinit sparkuser
- 4. 执行Spark组件的客户端命令。

#### spark-beeline

5. 在spark-beeline中执行命令,例如在obs://mrs-word001/table/目录中创建表test。

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

6. 执行如下命令查询所有表,返回结果中存在表test,即表示访问OBS成功。 show tables;

#### 图 13-1 Spark 验证返回已创建的表名

7. 使用 "Ctrl + C"退出退出spark beeline。

## 13.39 spark job 对应的执行日志保存在哪里?

- spark job没有完成的任务日志保存在Core节点的/srv/BigData/hadoop/data1/nm/containerlogs/
- spark job完成的任务日志保存在HDFS的/tmp/logs/用户名/logs

## 13.40 Hue 下载 excel 无法打开

1. 以root用户登录任意一个Master节点,切换到omm用户。

su - omm

2. 使用如下命令查看当前节点是否为oms主节点。

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

回显active即为主节点,否则请登录另一个Master节点。

#### 图 13-2 oms 主节点



3. 进入"\${BIGDATA\_HOME}/Apache-httpd-xxx/conf"目录。

#### cd \${BIGDATA\_HOME}/Apache-httpd-xxx/conf

4. 打开httpd.conf文件。

#### vim httpd.conf

5. 在文件中搜索21201,并删除文件中的如下内容。proxy\_ip和proxy\_port对应实际 环境中的值。

ProxyHTMLEnable On
SetEnv PROXY\_PREFIX=https://[proxy\_ip]:[proxy\_port]
ProxyHTMLURLMap (https?:\/\/[^:]\*:[0-9]\*.\*) \${PROXY\_PREFIX}/proxyRedirect=\$1 RV

#### 图 13-3 待删除内容

- 6. 退出并保存修改。
- 7. 再次打开httpd.conf文件,搜索**proxy\_hue\_port**,并删除如下内容。
  ProxyHTMLEnable On

SetEnv PROXY\_PREFIX=https://[proxy\_ip]:[proxy\_port]
ProxyHTMLURLMap (https?:\/\/[^:]\*:[0-9]\*.\*) \${PROXY\_PREFIX}/proxyRedirect=\$1 RV

#### 图 13-4 待删除内容

- 8. 退出并保存修改。
- 9. 执行如下命令重启httpd。

#### sh \${BIGDATA\_HOME}/Apache-httpd-xxx/setup/restarthttpd.sh

10. 检查备Master节点上的httpd.conf文件是否已修改,若已修改则处理完成,若未修改,参考上述步骤进行修改备Master节点的httpd.conf文件,无需重启httpd。

## 13.41 Yarn 的 ResourceManager 配置是否正常?

步骤1 登录MRS Manager页面,选择"服务管理 > Yarn > 实例"。

步骤2 分别单击两个ResourceManager名称,选择"更多 > 同步配置",并选择不勾选"重启配置过期的服务或实例。"。

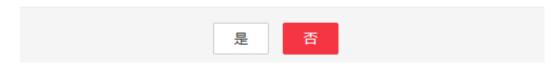
图 13-5 同步配置

## 同步配置



实例配置信息将同步到系统,请重启实例 ResourceManager@172.16.0.129以使配置生效。请确认是否同步 配置?

重启配置过期的服务或实例。



步骤3 单击"是"进行配置同步。

步骤4 以root用户分别登录Master节点。

步骤5 执行cd /opt/Bigdata/MRS\_Current/\*\_\*\_ResourceManager/etc\_UPDATED/命令进入etc\_UPDATED目录。

步骤6 执行grep '\.queues' capacity-scheduler.xml -A2找到配置的所有队列,并检查队列和Manager页面上看到的队列是否——对应。

root-default在页面隐藏,在页面看不到属于正常现象。

×

```
[omm@node-masterIIIZA etc]$
[omm@node-masterIIIZA 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-masterIIIZA etc]$
[omm@node-masterIIIZA etc]$
```

- **步骤7** 执行**grep** '\.capacity</name>' capacity-scheduler.xml -A2找出各队列配置的值,检查每个队列配置的值是否和Manager上看到的一致。并检查所有队列配置的值的总和是否是100。
  - 是,则说明配置正常。
  - 否,则说明配置异常,请执行后续步骤修复。

```
[omm@node-masterlIlZA etc]$
[omm@node-masterlIlZA 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</pre>
<value>100</value>
</property>
<name>yarn.scheduler.capacity.root.testl.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>
</property>
<name>yarn.scheduler.capacity.root.test4.accessible-node-labels.zhaolu.capacity</name>
</property>
<name>yarn.scheduler.capacity.root.test4.capacity</name>
<value>0</value>
</property>
<name>yarn.scheduler.capacity.root.default.capacity</name>
<value>20</value>
```

- 步骤8 登录MRS Manager页面,选择"主机管理"。
- 步骤9 查找主Master节点,主机名称前带实心五角星的Master节点即为主Master节点。
- 步骤10 以root用户登录主Master节点。
- 步骤11 执行su omm切换到omm用户。
- 步骤12 执行sh /opt/Bigdata/om-0.0.1/sbin/restart-controller.sh重启Controller。

请在Manager页面没有其他操作后重启Controller,重启Controller对大数据组件业务 无影响 。 步骤13 重新执行步骤1~步骤7同步ResourceManager的配置并检查配置是否正常。

配置同步完成后Manager页面可能显示配置过期,该显示不影响业务,是由于最新的 配置组件没有加载,待后续组件重启的时会自动加载。

----结束

# **14** API 使用类

## 14.1 使用调整集群节点接口时参数 node\_id 如何配置?

使用调整集群节点接口时,参数node\_id的值固定为node\_orderadd,直接填固定值即可。

## 15 集群管理类

## 15.1 如何查看所有集群?

MRS所有的集群都展示在"集群列表"页面中,进入"集群列表"页面,可查看所有 集群。集群数量较多时,可采用翻页显示,您可以查看任何状态下的集群。

- 现有集群:包括除了"失败"和"已删除"状态以外的所有集群。
- 历史集群:仅包含"已删除"和"失败"状态的集群,目前界面只显示6个月内创建且已删除的集群,若需要查看6个月以前删除的集群,请联系技术支持人员。
- 失败任务管理:仅包含"失败"状态的任务。
  - 集群创建失败的任务
  - 集群删除失败的任务
  - 集群扩容失败的任务
  - 集群缩容失败的任务

## 15.2 如何查看日志信息?

"操作日志"页面记录了用户对集群和作业的操作的日志信息。目前,MRS界面记录的日志信息分为以下几类:

- 集群操作
  - 创建集群、删除集群、扩容集群和缩容集群等操作
  - 创建目录、删除目录和删除文件等操作
- 作业操作: 创建作业、停止作业和删除作业等操作
- 数据操作: IAM用户任务、新增用户、新增用户组等操作

记录用户操作的日志信息如图15-1所示:

#### 图 15-1 日志信息

操作类型	操作IP	撮作内容	时间
集群操作	10.63.167.82	创建d为0bb2a919-666d-40c0-8cb1-a3486431aae6,名字为bigdata_xq318的集群	2016-03-18 17:17:46
集群操作	10.57.99.128	删除id为e92e5dc7-34c1-449d-b353-3651853e7631,名字为bigdata_DVwu的集群	2016-03-10 16:45:24
作业操作	10.63.167.82	提交作业作业id:f591520b-c632-4f33-9d2F063e942c93a2,作业名:distcp,集群id:e92e5dc7-34c1-449d-b353-3651853e7631	2016-03-10 10:26:28
作业操作	10.63.167.82	提交作业作业id:d8a58879-72d4-4ebb-84fb-0eca09b1c981,作业名job_spark,集群id:e92e5dc7-34c1-449d-b353-3651853e7631	2016-03-07 11:02:28
作业操作	10.63.167.82	提交作业作业id:bab88cc1-df9e-4735-b6f8-db190f303295,作业名:mr_01,集群d:e92e5dc7-34c1-449d-b353-3651853e7631	2016-03-07 10:52:37
作业操作	10.63.195.73	提交作业作业id:f346875e-9bd9-42e1-a7ff-422133605b3d,作业名:sparkSql.集群id:e92e5dc7-34c1-449d-b353-3651853e7631	2016-02-23 11:23:22
集群操作	10.63.195.73	创建ld为e92e5dc7-34c1-449d-b353-3651853e7631,名字为bigdata_DVwu的集群	2016-02-23 11:05:24

## 15.3 如何查看集群配置信息?

- 集群创建完成后在MRS控制台单击集群名称进入集群基本信息页面,可以查看到 集群的基本配置信息,包括集群的名称、ID、付费类型、工作地域、创建时间、 Hadoop组件版本及节点的实例规格、容量。其中,节点的实例规格和容量决定了 该集群对数据的分析处理能力。节点实例规格越高,容量越大,集群运行速度越 快,分析处理能力越强,相应的成本也越高。
- 在基本信息页面,单击"点击管理",跳转至MRS集群管理页面。用户可在集群 管理页面查看和处理告警信息、修改集群配置以及升级集群补丁等。

## 15.4 如何在 MRS 分析集群中安装流式集群的组件?

已经创建的集群不支持安装新的组件。如果需要安装流式集群的组件,则需要重新创建流式集群或者混合集群,并且选择所需要的组件。

## 15.5 如何在 MRS 集群中安装 Kafka, Flume 组件?

已经购买的集群不支持安装组件。Kafka和Flume为流式集群的组件,如果要安装 Kafka和Flume组件,则需要购买流式集群或者混合集群并选择该组件。

Kafka和Flume组件,使用方法请参考使用Kafka、使用Flume。

## 15.6 如何停止 MRS 集群?

如果想停止MRS集群,可以在"节点管理"页面,单击各个节点名称,进入"弹性云服务器"页面,选择"关机"即可。

## 15.7 升级 Master 节点规格需要关机吗?

MRS服务集群的Master节点规格升级时会自动触发关机操作,当升级完成后会自动触发开机操作,无需人工干预。

## 15.8 MRS 支持数据盘扩容吗?

MRS暂不支持数据盘扩容,可以参照MRS集群添加新磁盘来扩容磁盘总容量。

## 15.9 现有集群如何增加组件?

已经创建的MRS集群暂不支持组件的添加和移除,建议重新创建MRS集群并包含所需组件。

## 15.10 MRS 集群中安装的组件能否删除?

MRS集群中的组件不可以删除,如果不使用的话可以登录Manager页面在服务管理中 找到对应的组件将其停止。

## 15.11 MRS 集群是否支持更换节点类型?

MRS集群不支持更换节点类型,但是MRS 1.8.2及之后版本集群支持**Master节点规格 升级**。

## 15.12 MRS 是否支持变更 MRS 集群节点?

MRS管理控制台不支持变更集群节点,也不建议用户在ECS管理控制台直接修改MRS 集群节点。如果手动在ECS管理控制台对集群节点执行停止ECS、删除ECS、修改或重 装ECS操作系统,以及修改ECS规格的操作,可能影响集群稳定运行。

如果您对MRS集群节点进行了上述操作,MRS会自动识别并直接删除发生变更的集群 节点。您可以登录MRS管理控制台,通过扩容恢复已经删除的节点。请勿在扩容过程 中对正在扩容的节点进行操作。

## 15.13 集群扩容时 Core 和 Task 节点是否只能增加和现有集群节点同样配置的节点?

集群扩容时只能增加和现有集群节点相同配置的节点,仅当对应节点售罄时,可以选择增加其他配置的节点。扩容集群操作请参考**扩容集群**。

## 15.14 正在使用的 MRS 服务集群是否可以扩展集群规模?

正在使用的集群保证集群没有故障是可以直接扩容操作的,具体请参考扩容集群。

## 15.15 如何取消集群风险告警

- 1. 登录MRS服务控制台。
- 2. 单击集群名称进入集群详情页面。
- 3. 选择"告警管理 > 消息订阅规则"。
- 4. 在待修改的规则所在行的"操作"列单击"编辑",修改"订阅规则"。
- 5. 单击"确定"完成修改。

## 15.16 如何配置引导操作中的参数?

引导操作添加请参考添加引导操作,脚本样例请参考脚本样例。

## 15.17 为什么 MRS 集群显示的资源池内存小于实际集群内存?

在MRS集群中,MRS默认为Yarn服务分配集群内存的50%,用户从逻辑上对Yarn服务的节点按照资源池进行分区管理,所以集群中显示的资源池总内存仅有集群总内存的50%。

## 15.18 如何配置 knox 内存?

步骤1 分别以root用户登录集群Master节点。

步骤2 分别在Master节点执行如下命令打开gateway.sh文件。

su omm

vim /opt/knox/bin/gateway.sh

**步骤3** 将 "APP\_MEM\_OPTS=""" 修改为 "APP\_MEM\_OPTS="-Xms256m -Xmx768m"" 保存并退出文件。

步骤4 分别在Master节点执行如下命令重启konx进程。

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

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

步骤5 执行ps -ef | grep knox命令可查看已设置的内存信息。

图 15-2 knox 内存

omm@node-master!E3Hi =| sp. sef | grop knox | mm. | 11688 | 3 | 15:48 pts/0 00:000:00 | opt/Bigdata/jdkl.8.0\_212//bin/java -0java.library.path=/opt/knox/ext/native | xms256m - xmx768m - jar /opt/knox/bin/gateway.jar | mm. | 23890 | 11554 | 0 | 15:52 pts/0 | 00:00:00 grep --color=auto knox | mm. | mm

----结束

# **16** Kerberos 使用

## 16.1 已购买的 MRS 集群是否支持修改 Kerberos 状态?

MRS服务暂不支持集群创建完成后开启和关闭Kerberos服务,需要在创建MRS服务集群的时候选择开启或者关闭kerberos服务,建议重新创建集群。

## 16.2 如何访问支持 Kerberos 认证的 Manager

- 若集群支持弹性公网IP功能,请参考**访问Manager**。
- 若集群不支持弹性公网IP功能,请参考通过Windows弹性云服务器访问 Manager。

## 16.3 Kerberos 认证服务的端口有哪些?

Kerberos认证服务的端口有21730(TCP)、21731(TCP/UDP)、21732(TCP/UDP)。

## 16.4 如何在运行中的集群中部署 Kerberos 服务?

MRS服务集群不支持客户自定义安装部署Kerberos服务,且部署的Kerberos认证无法在组件之间建立认证方式。故建议客户另外购买一套Kerberos开启的集群,然后进行数据迁移。

## 16.5 开启 Kerberos 认证的集群如何访问 Hive?

- 1. 用root用户登录集群Master节点。
- 2. 配置环境变量。

#### source /opt/client/bigdata\_env

3. 如果当前集群已启用Kerberos认证,执行以下命令认证当前用户,当前用户需要 具有创建Hive表的权限,具体请参见配置Hive权限配置拥有对应权限的角色,并 为用户绑定对应角色。如果当前集群未启用Kerberos认证,则无需执行此命令。

kinit MRS集群用户

例如,kinit hiveuser

4. 执行Hive组件的客户端命令。

#### beeline

5. 在beeline中运行Hive命令,例如:

create table test\_obs(a int, b string) row format delimited fields terminated by "," stored as textfile location "obs://test\_obs";

6. 使用 "Ctrl + C" 退出退出hive beeline。

## 16.6 开启 Kerberos 认证的集群如何访问 Presto?

- 1. 用root用户登录集群Master节点。
- 2. 配置环境变量。

#### source /opt/client/bigdata\_env

- 3. 开启Kerberos认证的安全集群访问Presto。
  - a. 登录Manager创建一个拥有"Hive Admin Privilege"权限的角色,例如 prestorole,创建角色请参考<mark>创建角色</mark>。
  - b. 创建一个属于"Presto"和"Hive"组的用户,同时为该用户绑定3.a中创建的角色,例如presto001,创建用户请参考创建用户。
  - c. 认证当前用户。

#### kinit presto001

- d. 下载用户认证凭据。
  - MRS Manager界面操作:登录MRS Manager,选择"系统设置 > 用户管理",单击新增用户所在行的"更多 > 下载认证凭据"。

#### 图 16-1 下载 Presto 用户认证凭据



- FusionInsight Manager界面操作:登录FusionInsight Manager,选择 "系统 > 权限 > 用户",单击新增用户所在行的"更多 > 下载认证凭 据"。
- e. 解压下载的用户凭证文件,得到"krb5.conf"和"user.keytab"两个文件并放入客户端目录,例如"/opt/client/Presto/"。
- f. 执行如下命令获取用户principal。

#### klist -kt /opt/client/Presto/user.keytab

g. 启用Kerberos认证的集群,执行以下命令连接本集群的Presto Server。

presto\_cli.sh --krb5-config-path {krb5.conf文件路径} --krb5-principal {用户principal} --krb5-keytab-path {user.keytab文件路径} --user {presto用户名}

- krb5.conf文件路径:请替换为**3.e**中设置的文件存放路径,例如"/opt/client/Presto/krb5.conf"。
- user.keytab文件路径:请替换为**3.e**中设置的文件存放路径,例如 "/opt/client/Presto/user.keytab"。
- 用户principal: 请替换为3.f中返回的结果
- presto用户名:请替换为3.b中创建的用户名,例如"presto001"

例如: presto\_cli.sh --krb5-config-path /opt/client/Presto/krb5.conf --krb5-principal prest001@xxx\_xxx\_xxx\_xxx.COM --krb5-keytab-path /opt/client/Presto/user.keytab --user prest0001

h. 在Presto客户端中执行语句创建schema,例如:

CREATE SCHEMA hive.demo01 WITH (location = 'obs://presto-demo002/');

i. 在该schema中建表,该表的数据即存储在OBS桶内,例如:

CREATE TABLE hive.demo01.demo\_table WITH (format = 'ORC') AS SELECT \* FROM tpch.sf1.customer;

#### 图 16-2 安全集群 Presto 验证返回结果

j. 执行exit退出客户端。

## 16.7 Kerberos 认证过期如何处理?

● 对于JAVA应用

在连接HBase、HDFS或者其他大数据组件前,先调用loginUserFromKeytab()创建UGI,然后启动一个定时线程进行检查是否过期并在过期前重新登录。

- 对于shell方式执行的任务
  - a. 先执行kinit命令认证用户。
  - b. 通过操作系统定时任务或者其他定时任务方式定时执行kinit命令认证用户。
  - c. 提交作业执行大数据任务。
- 对于Spark作业

通过spark-shell、spark-submit、spark-sql方式提交作业,可以直接在命令行中指定Keytab和Principal以获取认证,定期更新登录凭证和授权tokens,避免认证过期, 例如:

spark-shell --principal spark2x/hadoop.<系统域名>@<系统域名> --keytab \$ {BIGDATA\_HOME}/FusionInsight\_Spark2x\_8.0.2/install/FusionInsight-Spark2x-2.4.5/keytab/spark2x/SparkResource/spark2x.keytab --master yarn

# **1 7** 元数据管理

## 17.1 Hive 元数据在哪里查看?

- Hive的元数据存放在MRS服务集群的gaussDB中,可以登录到集群的Master节点 上并切换到omm用户,然后执行gsql -p {PORT} -U {USER} -W {PASSWD} -d hivemeta查看。
- Hive元数据存放在外部的关系型数据库存储时,请通过如下步骤获取信息:
  - a. 集群详情页的"数据连接"右侧单击"点击管理"。
  - b. 在弹出页面中查看"数据连接ID"。
  - c. 在MRS控制台,单击"数据连接"。
  - d. 在数据连接列表中根据集群所关联的数据连接ID查找对应数据连接。
  - e. 在对应数据连接的"操作"列单击"编辑",查看该数据连接所连接的RDS 实例及数据库。