阿里云
E-MapReduce

集群规划与配置

文档版本：20190522

为了无法计算的价值 | 阿里云

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的”现状“、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含”阿里云”、Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	**禁止:** 重置操作将丢失用户配置数据。
	该类警示信息可能导致系统重大变更甚至故障，或者导致人身伤害等结果。	**警告:** 重启操作将导致业务中断，恢复业务所需时间约10分钟。
	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	**说明:** 您也可以通过按Ctrl + A选中全部文件。
>	多级菜单递进。	设置 > 网络 > 设置网络类型
粗体	表示按键、菜单、页面名称等UI元素。	单击 **确定**。
`courier 字体`	命令。	执行 `cd /d C:/windows` 命令，进入Windows系统文件夹。
##	表示参数、变量。	`bae log list --instanceid` *`Instance_ID`*
[]或者[a|b]	表示可选项，至多选择一个。	`ipconfig` *`[-all|-t]`*
{}或者{a|b}	表示必选项，至多选择一个。	`swich` *`{stand | slave}`*

目录

法律声明.....I
通用约定.....I
1 集群规划.....1
1.1 角色授权.....1
1.2 用户管理.....3
1.3 实例类型.....4
1.4 组件角色部署.....5
1.5 Gateway 实例.....11
1.6 ECS 实例说明.....11
1.7 存储说明.....13
1.8 本地盘机型支持.....14
1.9 经典网络与 VPC 互访.....15
1.10 E-MapReduce 集群容灾能力.....15
2 集群配置.....17
2.1 创建集群.....17
2.2 集群列表与详情.....22
2.3 扩容集群.....27
2.4 释放集群.....29
2.5 集群续费管理.....30
2.6 SSH 登录集群.....33
2.7 服务列表.....39
2.8 集群脚本.....41
2.9 访问链接与端口.....42
2.10 安全组.....47
2.11 创建 Gateway.....48
2.12 集群资源管理.....53
2.13 配置弹性伸缩.....55
2.13.1 弹性伸缩概述.....55
2.13.2 按时间伸缩规则配置.....56
2.13.3 按负载伸缩规则配置.....58
2.13.4 弹性伸缩抢占式实例.....61
2.13.5 弹性伸缩记录.....62
2.14 配置专有网络.....62
2.15 MetaService.....64
2.16 元数据管理.....66
2.16.1 Hive 元数据管理.....67
2.16.2 Kafka 元数据管理.....73
3 变更配置.....75
3.1 升级节点配置.....75
3.2 磁盘扩容.....77

3.3 切换支付类型.....80
3.4 移除异常节点.....80
3.5 多机器组.....81
4 第三方软件.....83
4.1 引导操作.....83
4.2 软件配置.....85
4.3 组件参数配置.....86

1 集群规划

1.1 角色授权

在用户开通 E-MapReduce 服务时，需要授予一个名称为“AliyunEMRDefaultRole”的系统默认角色给 E-MapReduce 的服务账号，当且仅当该角色被正确授予后，E-MapReduce 才能正常地调用相关服务（ECS，OSS 等），创建集群以及保存日志等。

> **注意:**
>
> 首次开通 E-MapReduce 服务时，必须使用主账号完成角色授权流程，否则子账号和主账号均不能使用 E-MapReduce 。

角色授权流程

1. 当用户创建集群或创建按需执行计划时，如果没有正确地给 E-MapReduce 的服务账号授予默认角色，则会看到如下提示，此时需单击前往 RAM 进行授权，进行角色授权。

2. 单击同意授权，将默认角色 AliyunE-MapReduceDefaultRole 授予给 E-MapReduce 的服务账号。

3. 当完成以上授权步骤后，用户需刷新 E-MapReduce 的控制台，然后就可以进行操作了。如果您想查看 AliyunE-MapReduceDefaultRole 相关的详细策略信息，可以登录 RAM 的控制台查看，也可以单击查看链接。

默认角色包含的权限内容

默认角色 AliyunEMRDefaultRole 包含的权限信息如下：

- ECS 相关权限：

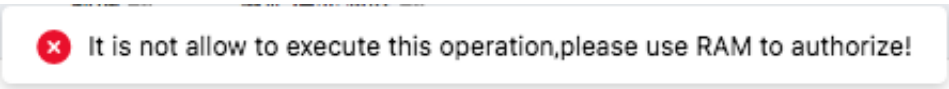
权限名称（Action）	权限说明
ecs:CreateInstance	创建 ECS 实例
ecs:RenewInstance	ECS 实例续费
ecs:DescribeRegions	查询 ECS 地域信息
ecs:DescribeZones	查询 Zone 信息
ecs:DescribeImages	查询镜像信息
ecs:CreateSecurityGroup	创建安全组
ecs:AllocatePublicIpAddress	分配公网 IP
ecs:DeleteInstance	删除机器实例
ecs:StartInstance	启动机器实例
ecs:StopInstance	停止机器实例
ecs:DescribeInstances	查询机器实例
ecs:DescribeDisks	查询机器相关磁盘信息
ecs:AuthorizeSecurityGroup	设置安全组入规则
ecs:AuthorizeSecurityGroupEgress	设置安全组出规则
ecs:DescribeSecurityGroupAttribute	查询安全组详情
ecs:DescribeSecurityGroups	查询安全组列表信息

- OSS 相关权限：

权限名称（Action）	权限说明
oss:PutObject	上传文件或文件夹对象
oss:GetObject	获取文件或文件夹对象
oss:ListObjects	查询文件列表信息

子账户授权

为确保子账号能正常使用 E-MapReduce 控制台的功能，需要使用主账号登录RAM 控制台给子账号授予 AliyunEMRFullAccess 或 AliyunEMRDevelopAccess 策略。否则会有如下错误提示：



说明：

- AliyunEMRFullAccess 策略授予子账号的权限是管理 E-MapReduce 的权限，主要包括对 E-MapReduce 的所有资源的所有操作权限；
- AliyunEMRDevelopAccess 策略授予子账号的权限是E-MapReduce开发者权限，与 AliyunEMRFullAccess 策略相比，不授予集群的创建和释放等操作权限。

  关于 RAM 子账号，策略，角色等，请参考RAM 产品介绍

  ；

1.2 用户管理

用户管理是在指定集群上创建相关服务所需的账号，目前支持 Knox 账号和 Kerberos 账号。

创建 RAM 账号

使用 RAM，您可以创建、管理用户账号（比如员工、系统或应用程序），并可以控制这些用户账号对您名下资源具有的操作权限。创建 RAM 账号的具体操作步骤如下：

1. 登录阿里云 E-MapReduce 控制台，进入集群列表页面。
2. 单击集群 ID 右侧的管理。
3. 在左侧导航栏中单击用户管理。
4. 单击右上角创建 RAM 子账号跳转到 RAM 访问控制页面。
5. 通过 RAM 访问控制您可以新建用户、用户组，添加权限等，请参考访问控制文档。

添加 Knox 账号

您可以为当前阿里云账号的 RAM 子账号开通 Knox 账号或 Linux 账号，具体操作步骤如下：

1. 进入用户管理页面，在页面上选择需要添加到集群的账号，单击操作列的设置Knox密码。
2. 在添加 Knox用户的对话框中设置密码，然后单击确认。
3. 刷新用户管理页面，当Knox账号列的状态显示已同步时，表示 Knox 账号添加成功。

   添加成功后，您即可使用账号名称及上一步中设置的密码登录 Knox。

关于 Knox 的详细使用说明请参见 *Knox* 使用说明

删除 Knox 账号

1. 进入用户管理页面，在页面上选择需要从集群删除的账号，单击操作列的删除 Knox 密码。
2. 刷新用户管理页面，当 Knox 账号列的状态显示未同步时，表示 Knox 账号删除成功。

Kerberos 账号

高安全模式启动后，高安全集群中的各组件会通过 Kerberos 进行认证，详细信息请参考 *Kerberos* 简介。

常见问题

- 不同集群不能共享 Knox 账号，即：在cluster-1上添加 Knox 账号 A 之后，并不会共享给 cluster-2，若想在 cluster-2 上使用 Knox 账号 A，需要在 cluster-2 上重新添加账号 A。原因是 Knox 账号是创建在集群中的，每个集群的 Knox 账号不互通。
- 添加 Knox 账号过程中，若显示同步失败，请单击重试重新添加。
- 添加 Knox 账号过程中，多次重试仍然失败，单击左侧集群与服务管理，检查 ApacheDS 是否已停止，若已停止则启动 ApacheDS 后，再回到用户管理进行重试。

1.3 实例类型

E-MapReduce 集群中由多个不同实例类型的节点组成，他们分别是主实例节点（master），核心实例节点（core）和计算实例节点（task）。

不同的实例在部署的时候会部署完全不同的服务进程，以完成完全不同的任务。举例来说，我们会在主实例节点（Master）上部署 Hadoop HDFS 的 NameNode 服务，Hadoop YARN 的 ResourceManager 服务，而在核心实例节点（Core）上部署 DataNode 服务，Hadoop YARN 的 NodeManager 服务，在计算实例节点（task）。顾名思义，只进行计算，部署 Hadoop YARN 的 NodeManager 服务，不部署任何 HDFS 相关的服务。

在创建集群的时候需要确定对应的三种实例类型的 ECS 规格，相同实例类型的 ECS 在同一个实例组内。并且可以在后期通过扩容来扩容对应实例组内的机器数量（主实例组除外）。

> 说明:
> 计算实例节点（task）从 3.2.0 及以后版本开始支持。

主实例节点（Master）

主实例是集群服务部署管控等组件的节点，举例来说，Hadoop YARN 的 ResourceManager 就部署在主实例节点上。您可以通过 *SSH* 的方式连接到主实例上，通过软件的 Web UI 来查看集

群上的服务的运行情况。同时，当需要进行快速的测试或者运行作业的时候，也可以登录到主实例上，通过命令行直接提交作业。当集群开启了高可用的时候会有 2 个主实例节点（默认只有 1 个）。

核心实例节点（Core）

核心实例是被主实例管理的实例节点。上面会运行 Hadoop HDFS 的 Datanode 服务，并保存所有的数据。同时也会部署计算服务，比如 Hadoop YARN 的 NodeManager 服务来执行计算任务。为满足存储数据量或者是计算量上升的需要，核心实例可以随时进行扩容，不影响当前集群的正常运行。核心实例可以使用多种不同的存储介质来保存数据。参考本地盘和块存储。

计算实例节点（Task）

计算实例是专门负责计算的实例节点，是一个可选的实例类型。如果核心实例的计算能力足够的情况下，可以不使用计算实例。计算实例可以在任何时候快速的为集群增加额外的计算能力，如 Hadoop 的 MapReduce tasks， Spark executors 等。在计算实例上不会保存 HDFS 的数据，因此在计算实例上不运行 Hadoop HDFS 的 Datanode 服务。计算实例可以随时的新增和减少，都不会影响到现有集群的运行。计算实例节点的减少可能会引起 MapReduce 和 Spark 的作业的失败，能否成功取决于该计算服务的重试容错能力。

1.4 组件角色部署

本节介绍服务进程在 E-MapReduce 不同集群中各节点的部署情况。

E-MapReduce 集群创建时，会在实例类型中部署不同的服务角色，例如 Hadoop HDFS 中的 NameNode 部署在 Master 节点中。对于大规模集群，可以通过引导操作实现自定义的部署和分配，例如独立部署 ZooKeeper 集群，采用独立的阿里云 RDS 作为 Hive MetaStore 存储。默认的详细服务部署信息如下：

Hadoop 集群

Hadoop 集群服务组件具体部署信息

- 必选服务

组件名	主实例节点	核心实例节点
HDFS	- KMS - SecondaryNameNode - HttpFS - HDFS Client - NameNode	- DataNode - HDFS Client

组件名	主实例节点	核心实例节点
YARN	- ResourceManager - App Timeline Server - JobHistory - WebAppProxyServer - Yarn Client	- Yarn Client - NodeManager
Hive	- Hive MetaStore - HiveServer2 - Hive Client	Hive Client
GANGLIA	- Gmond - Httpd - Gmetad - Ganglia Client	- Gmond - Ganglia Client
HUE	Hue	N/A
SPARK	- Spark Client - SparkHistory - ThriftServer	Spark Client
ZEPPLIN	Zeppelin	N/A
TEZ	- Tomcat - Tez Client	Tez Client
SQOOP	Sqoop Client	Sqoop Client
PIG	Pig Client	Pig Client
HAPROXY	HAProxy	N/A
APACHEDS	ApacheDS	N/A
KNOX	Knox	N/A

- 可选服务

组件名	主实例节点	核心实例节点
Flume	- Flume Agent - Flume Client	- Flume Agent - Flume Client
LIVY	Livy	N/A
SUPERSET	Superset	N/A

组件名	主实例节点	核心实例节点
FLINK	- FlinkHistoryServer - Flink Client	Flink Client
RANGER	- RangerPlugin - RangerAdmin - RangerUserSync	RangerPlugin
STORM	- Storm Client - UI - Nimbus - Logviewer	- Storm Client - Supervisor - Logviewer
PHOENIX	Phoneix Client	Phoneix Client
HBASE	- HMaster - HBase Client - ThriftServer	- HBase Client - HRegionServer
ZOOKEEPER	- ZooKeeper follower - ZooKeeper Client	- ZooKeeper leader - ZooKeeper follower - ZooKeeper Client
OOZIE	Oozie	N/A
PRESTO	- Presto Client - PrestoMaster	- Presto Client - PrestoWorker
IMPALA	- Impala Runtime and Shell - Impala Catalog Server - Impala StateStore Server	- Impala Runtime and Shell - Impala Daemon Server - Impala StateStore Server

Druid 集群

Druid 集群服务组件具体部署信息

- 必选服务

组件名	主实例节点	核心实例节点
DRUID	- Druid Client - Coordinator - Overlord - Broker	- MiddleManager - Historical - Druid Client
HDFS	- KMS - SecondaryNameNode - HttpFS - HDFS Client - NameNode	- DataNode - HDFS Client
GANGLIA	- Gmond - Httpd - Gmetad - Ganglia Client	- Gmond - Ganglia Client
ZOOKEEPER	- ZooKeeper follower - ZooKeeper Client	- ZooKeeper leader - ZooKeeper follower - ZooKeeper Client

- 可选服务

组件名	主实例节点	核心实例节点
YARN	- ResourceManager - App Timeline Server - JobHistory - WebAppProxyServer - Yarn Client	- Yarn Client - NodeManager
SUPERSET	Superset	N/A

Kafka 集群

Kafka 集群服务组件具体部署信息

- 必选服务

组件名	主实例节点	核心实例节点
KAFKA-MANAGER	Kafka Manager	N/A

组件名	主实例节点	核心实例节点
KAFKA	- Kafka Client - KafkaMetadataMonitor - Kafka Rest Proxy - Kafka Broker controller - Kafka Schema Registry	- Kafka Broker broker - Kafka Client
GANGLIA	- Gmond - Httpd - Gmetad - Ganglia Client	- Gmond - Ganglia Client
ZOOKEEPER	- ZooKeeper follower - ZooKeeper Client	- ZooKeeper leader - ZooKeeper follower - ZooKeeper Client

- 可选服务

组件名	主实例节点	核心实例节点
RANGER	- RangerPlugin - RangerUserSync - RangerAdmin	RangerPlugin
KNOX	Knox	N/A
APACHEDS	ApacheDS	N/A

Zookeeper 集群

Zookeeper 集群服务组件具体部署信息

组件名	主实例节点	核心实例节点
GANGLIA	N/A	· Gmond · Httpd · Gmetad · Ganglia Client
ZOOKEEPER	N/A	· ZooKeeper leader · ZooKeeper Client · ZooKeeper follower

Data Science 集群

Data Science 集群服务组件具体部署信息

- 必选服务

组件名	主实例节点	核心实例节点
HDFS	- KMS - SecondaryNameNode - HttpFS - HDFS Client - NameNode	- DataNode - HDFS Client
JUPYTER	Jupyter	Jupyter
GANGLIA	- Gmond - Httpd - Gmetad - Ganglia Client	- Gmond - Ganglia Client
ZOOKEEPER	- ZooKeeper follower - ZooKeeper Client	- ZooKeeper leader - ZooKeeper follower - ZooKeeper Client
ANALYTICS-ZOO	- Analytics-Zoo-Python - Analytics-Zoo-Scala	- Analytics-Zoo-Python - Analytics-Zoo-Scala
KNOX	Knox	N/A
APACHEDS	ApacheDS	N/A
TENSORFLOW-ON-YARN	- TensorFlow-On-YARN - TensorFlow-On-YARN-Gateway - TensorFlow-On-YARN-History-Server	- TensorFlow-On-YARN-Client - TensorFlow-On-YARN-Gateway
TENSORFLOW	TensorFlow	TensorFlow
ZEPPELIN	Zeppelin	N/A
SPARK	- Spark Client - ThriftServer - SparkHistory	Spark Client

组件名	主实例节点	核心实例节点
YARN	- ResourceManager - App Timeline Server - JobHistory - WebAppProxyServer - Yarn Client	- Yarn Client - NodeManager

- 可选服务

组件名	主实例节点	核心实例节点
HUE	Hue	RangerPlugin
HIVE	- HiveServer2 - Hive MetaStore - Hive Client	Hive Client

1.5 Gateway 实例

Gateway 一般为独立的一个集群，由多台相同配置的节点组成。

在创建 Gateway 集群时，可以关联到一个已经存在的 Hadoop 集群上，该集群上会部署 Hadoop（HDFS+YARN）、Hive、Spark、Sqoop、Pig 等客户端，方便对集群进行操作。这样做的好处是：Gateway 可以作为一个独立的提交点，不会占用 Hadoop 集群的资源，尤其是在 Master 提交的方式，可以提高 Master 节点的稳定性。

您可以创建多个不同的 Gateway 集群，来给不同的用户使用，让他们可以使用各自独有的环境配置来满足不同的业务需求。

1.6 ECS 实例说明

本节将介绍 ECS 实例类型以及实例类型适用的场景。

EMR 目前支持的 ECS 实例类型

- 通用型

  vCPU 与 Memory 比为 1：4，如 32 核 128 GiB，使用云盘作为存储。

- 计算型

  vCPU 与Memory比为 1：2，如 32 核 64 GiB，使用云盘作为存储，提供了更多的计算资源。

- 内存型

  vCPU 与 Memory 比为 1：8，如 32 核 256 GiB，使用云盘作为存储，提供了更多的内存资源。

- 大数据型

  使用本地 SATA 盘作为数据存储，拥有很高的存储性价比，是大数据量（TB 级别的数据量）场景下的推荐机型。

  > 说明:
  > 当前只有 Hadoop，Data Science，Druid 类型集群支持的 Core 节点支持大数据型实例。Zookeeper 和 Kafka 类型集群不支持 Core 节点。

- 本地 SSD 型

  使用本地 SSD 盘，拥有极高的本地 IOPS 和吞吐能力。

- 共享型（入门级）

  共享 CPU 的机型，在大计算量的场景下，稳定性不够。入门级学习使用，不推荐企业客户使用。

- GPU

  使用 GPU 的异构机型，可以用来运行机器学习等场景。

实例类型适用场景

- Master 主实例

  适合通用型或内存型实例，数据直接使用阿里云的云盘来保存，有三个备份的保证，数据高可靠。

- Core 核心实例

  小数据量（TB 级别以下）或者是使用 OSS 作为主要的数据存储时、可以使用通用型、计算型或内存型。当数据量较大时（10 TB 或以上），推荐使用大数据机型，以获得极高的性价比。当 Core 核心实例使用本地盘时，数据可靠性得不到保证，会由 EMR 平台来进行维护和保证。

- Task 计算实例

  作为集群的计算能力的补充，可以使用除大数据型以外的所有的机型。目前本地 SSD 型尚未支持，后续会加入到 Task 中。

1.7 存储说明

在节点上存在两种角色的磁盘，一类是系统盘，用来安装操作系统。一类是数据盘，用来保存数据。

系统盘默认都是一块，而数据盘可以有很多块，目前上限可以一个节点挂 16 块。每一块都可以有不同的配置、类型、和容量。EMR 默认使用 SSD 云盘作为集群的系统盘。EMR 默认挂载 4 块云盘，目前的内网带宽的情况下 4 块云盘是比较合理的配置。

云盘与本地盘

有两种类型的磁盘可以用作数据的存储：

- 云盘

  包括 SSD 云盘、高效云盘、普通云盘。

  特点是，磁盘并不直接挂载在本地的计算节点上，通过网络访问远端的一个存储节点。每一份数据在后端都有两个实时备份，一共三份数据。所以当一份数据损坏的时候（磁盘损坏，不是用户自己的业务上的破坏），会自动的使用备份数据恢复。

- 本地盘

  包括大数据型的 SATA 本地盘，和本地 SSD 盘。

  直接挂载在计算节点上的磁盘，拥有超过云盘的性能表现。使用本地盘的时候不能选择数量，只能使用默认配置好的数量，和线下物理机一样，数据没有后端的备份机制，需要上层的软件来保证数据可靠性。

适用的场景

在 EMR 中，所有云盘和本地盘都会在节点释放的时候清除数据，磁盘无法独立的保存下来，并再次使用。Hadoop HDFS 会使用所有的数据盘作为数据存储。 Hadoop YARN 也会使用所有的数据盘作为计算的临时存储。

当业务数据量并不太大（TB 级别以下）的时候，可以使用云盘，IOPS 和吞吐相比本地盘都会小些。数据量大的时候，推荐都使用本地盘，EMR 会来维护本地盘的数据可靠性。如果发现在使用中明显的吞吐量不够用，可以切换到本地盘的存储上。

OSS

在 EMR 中可以将 OSS 作为 HDFS 使用。 用户可以非常方便的读写 OSS，所有使用 HDFS 的代码也可以简单的修改就能访问 OSS 上的数据了。

比如：

Spark中读取数据

```
sc.textfile("hdfs://user/path")
```

替换存储类型 hdfs -> oss

```
sc.textfile("oss://user/path")
```

对于 MR 或者 Hive 作业也是一样，HDFS 命令可以直接操作 OSS 数据

```
hadoop fs -ls oss://bucket/path
hadoop fs -cp hdfs://user/path  oss://bucket/path
```

这个过程，您不需要输入 AK 和 endpoint，EMR 都会自动替用户使用当前集群所有者的信息补全。

OSS 的 IOPS 不高，在一些需要高 IOPS 的场景，不适合使用，比如流式计算 Spark Streaming 或 HBase。

1.8 本地盘机型支持

阿里云为了满足大数据场景下的存储的需求，在云上推出了本地盘的机型：D1 系列。

D1 系列提供了本地盘而非云盘作为存储。解决了之前使用云盘产生多份冗余数据导致的成本高的问题，数据的传输不再需要全部通过网络，从而提高了磁盘的吞吐能力。同时还能发挥 Hadoop 的就近计算的优势。

相比于使用云盘的方式，本地盘机型极大的提高了存储性能，并降低了存储单价，达到和线下物理机几乎相同的成本。

本地盘机型在提供了大量的优势的情况下，也带来了一个问题：数据可靠性。对于云盘来说，由于有阿里云默认的磁盘多备份策略，所以用户完全感知不到磁盘的损坏，由云盘自动保证数据可靠。当使用了本地盘以后数据可靠性就需要由上层的软件来保证。同时，如果有磁盘与节点的故障情况，也需要进行人工的运维处理。

EMR + D1 方案

EMR 产品针对本地盘机型，如 D1，推出了一整套的自动化运维方案，帮助您方便可靠的使用本地盘机型，不需要关心整个运维的过程的同时，做到数据高可靠，服务高可用。

主要的一些点如：

- 强制节点的高可靠分布
- 本地盘与节点的故障监控
- 数据迁移时机自动决策

- 自动的故障节点迁移与数据平衡
- 自动的 HDFS 数据检测
- 网络拓扑调优

EMR 通过整个后台的管控系统的自动化运维，协助用户更好的使用本地盘机型，实现高性价比的大数据系统。

> 说明:
> 如需使用 D1 机型搭建 Hadoop 集群，请工单联系我们协助操作。

1.9 经典网络与 VPC 互访

本节介绍如何使经典网络的 ECS 可以和 VPC 下的 E-MapReduce(以下简称 EMR）集群网络互访。

ClassicLink

目前阿里云存在两种网络类型，一个是经典网络，一个是 VPC。很多用户的业务系统因为历史的原因还会在经典网络中，而 EMR 集群是在 VPC 中。为了解决这个问题，阿里云推出了 *ClassLink方案*。

大致步骤如下，详细的请参考上面的 ClassLink 文档:

1. 首先按照上面文档中指定网段创建 vswitch。
2. 在创建集群的时候，请使用该网段的 vswitch 来部署 EMR 集群。
3. 在 ECS 控制台将对应的经典网络节点连接到这个 VPC。
4. 设置安全组访问规则。

通过以上操作就可以让经典网络的 ECS 和 VPC 下的集群互访了。

1.10 E-MapReduce 集群容灾能力

本文将介绍 E-MapReduce 集群数据容灾和服务容灾能力。

数据容灾

Hadoop 分布式文件系统 (HDFS) 将每一个文件的数据进行分块存储，同时每一个数据块又保存有多个副本(系统默认为每一个数据块存放 3 个副本)，尽量保证这些数据块副本分布在不同的机架之上（在大多数情况下，副本系数是 3，HDFS 的存放策略是将一个副本存放在本地机架节点上，一个副本存放在同一个机架的另一个节点上，最后一个副本放在不同机架的节点上）。

HDFS 会定期扫描数据副本，若发现数据副本发生丢失，则会快速的进行数据的复制以保证副本的数量。若发现节点丢失，则节点上的所有数据也会快速的进行复制恢复。在阿里云上，如果是使用云盘的技术，则在后台每一个云盘都会对应三个数据副本，当其中的任何一个出现问题时，副本数据都会自动进行切换并恢复，以保证数据的可靠性。

Hadoop HDFS 是一个经历了长时间考验且具有高可靠性的数据存储系统，已经能够实现海量数据的高可靠性存储。同时基于云上的特性，也可以在 OSS 等服务上进行数据的额外备份，来达到更高的数据可靠性。

服务容灾

Hadoop 的核心组件都会进行 HA 的部署，即有至少 2 个节点的服务互备，如 YARN，HDFS，Hive Server，Hive Meta，以保证在任何时候，其中任何一个服务节点挂掉时，当前的服务节点都能自动的进行切换，保证服务不会受到影响。

2 集群配置

2.1 创建集群

本文介绍创建 E-MapReduce（以下简称 EMR） 集群的详细操作步骤和相关配置。

进入创建集群页面

1. 登录阿里云 E-MapReduce 控制台。
2. 完成 RAM 授权，操作步骤请参见角色授权。
3. 在上方选择所在的地域（Region），所创建集群将会在对应的地域内，一旦创建后不能修改。
4. 单击创建集群，进行创建。

创建集群流程

> **注意:**
> 集群除了名字以外，一旦创建完成就无法被修改。所以在创建时请仔细确认需要的配置。

要创建集群，您需要继续完成以下 3 个步骤：

1. 软件配置

配置项说明：

- 产品版本：EMR 产品的主要版本，代表了一整套的开源软件环境，它会定时的根据内部组成软件的升级进行升级。一般如果 Hadoop 相关的软件有进行升级，EMR 也会升级，这个时候就会升级这个主版本号。低版本的集群无法自动的升级到一个高版本上。
- 集群类型：目前的EMR提供了：
  - Hadoop 集群，提供半托管的 Hadoop、Hive、Spark 离线大规模分布式数据存储和计算，SparkStreaming、Flink、Storm 流式数据计算，Presto、Impala 交互式查询，Oozie、Pig 等 Hadoop 生态圈的组件，具体的组件信息可以在选择界面的列表中查看。
  - Kafka 集群，是半托管分布式的、高吞吐量、高可扩展性的消息系统。提供一套完整的服务监控体系，保障集群稳定运行，用户无需部署运维，更专业、更可靠、更安全。广泛用于日志收集、监控数据聚合等场景，支持离线或流式数据处理、实时数据分析等。
  - Druid 集群，提供半托管式实时交互式分析服务，大数据查询毫秒级延迟，支持多种数据摄入方式。可与 EMR Hadoop、EMR Spark、OSS、RDS 等服务搭配组合使用，构建灵活稳健的实时查询解决方案。
  - Data Science 集群，主要面向大数据+ AI 场景，提供了 Hive、Spark 离线大数据 ETL，TensorFlow 模型训练，用户可以选择 CPU+GPU 的异构计算框架，利用英伟达 GPU 对部分深度学习算法进行高性能计算。
- 必选服务：展示选择的集群类型下的所有的软件组件列表，包括名称和版本号。
- 可选服务：根据需求，您可选择不同的组件，被选中的组件会默认启动相关的服务进程。

> 说明：
> 您选择的组件越多，对您机器的配置要求就越高，否则很可能无法有足够的资源来运行这些服务。

- 高级设置
  - Kerberos 集群模式：是否开启集群的 Kerberos 认证功能。一般的个人用户集群无需该功能，默认关闭。
  - 软件自定义配置：可以指定一个 json 文件对集群中的基础软件例如 Hadoop、Spark、Hive 等进行配置，详细使用方法请参见软件配置。

2. 硬件配置

配置项说明：

- 付费配置
  - 付费类型：包年包月是一次性支付一个长期的费用，价格相对来说会比较便宜，特别是包三年的时候折扣会很大。按量付费是根据实际使用的小时数来支付费用，每个小时计一次费用。适合与短期的测试或者是灵活的动态任务，价格相对来说会贵一些。
    - 按量付费
    - 包年包月
      - 付费时长：您可选择购买 1 个月、2 个月、3 个月、6 个月、9 个月、1 年、2 年、3 年。包年包月如果购买一年则会在原价（12个月）的基础上打 85 折。
      - 自动续费：到期前7天执行自动续费操作，续费时长为一个月。
- 网络配置
  - 可用区：选择集群所在的可用区（Zone），可用区为在同一地域下的不同物理区域，可用区之间内网互通。不同的可用区会有不同的机型和磁盘。在每个 Region 内存在多个可用区。可用区在物理上属于不同的区域，一般来说如果需要较好的网络，推荐您选择相同的可用区，但是这样也会使创建集群失败的风险增大，因为单个可用区的存库不一定那么充足。如果需要大量的机器可以工单咨询我们。
  - 网络类型：默认使用专有网络（VPC），专有网络需要额外提供所属 VPC 以及子网（交换机），若还未创建，可前往 VPC控制台进行创建。E-MapReduce 专有网络详细使用说明查看专有网络。
  - VPC：选择在该地域的 VPC。如没有，单击创建 VPC / 子网(交换机）前往新建。
  - 交换机：选择在对应的 VPC 下的在对应可用区的交换机，如果在这个可用区没有可用的交换机，那么就需要前往去创建一个新的使用。
  - 安全组名称：集群所属的安全组。这里只展示用户在 E-MapReduce 产品中创建的安全组，目前尚不支持选择在 E-MapReduce 外创建的安全组。如果需要新建安全组，可以输入新的安全组的名字完成新建。长度限制为 2-64 个字符，以大小写字母或中文开头，可使用中文、字母、数字、“-”和“_”。
- 集群节点配置
  - 高可用：打开后，Hadoop 集群会有 2 个 master 来支持 ResourceManager 和 NameNode 的高可用。HBase 集群原来就支持高可用，只是另一个节点用其中一个

core 节点来充当，如果打开高可用，会独立使用一个 master 节点来支持高可用，更加的安全可靠。默认为非高可用模式，master节点数量为1。

- 节点类型：
  - Master 主实例节点，主要负责 ResourceManager，NameNode 等控制进程的部署。
  - Core 核心实例节点，主要负责集群所有数据的存储，可以按照需要进行扩容。
  - Task 纯计算节点，不保存数据。调整集群的计算力使用。
- 节点配置：不同规格的机型的选择。各个机型有各自比较适用的场景，可以根据需要选择。
- 数据盘类型：集群的节点使用的数据盘类型，数据盘有 3 种类型，普通云盘、高效云盘和 SSD 云盘，根据不同机型和不同的 Region，会有不同。当用户选择不同的区的时候，该区支持什么盘，下拉框就会展示什么类型的盘。数据盘默认设置为随着集群的释放而释放。本地盘的计算节点，磁盘是默认选定的，无法修改。
- 数据盘容量：目前推荐的集群容量最小是 40 G 单机，最大可以到 32 T 单节点。本地盘的容量是默认的，无法调整。
- 实例数量：需要的总的节点的台数。一个集群至少需要 3 台实例（高可用集群需增加 1 个 Master 节点，至少 4 台）。

3. 基础配置

配置项说明：

- 基础信息
  - 集群名称：集群的名字，长度限制为 1-64 个字符，仅可使用中文、字母、数字、“-”和“_”。
  - 远程登录：是否打开安全组 22 端口，默认开启。
  - 密钥对：关于密钥对的使用，请参考 *SSH* 密钥对。
  - 登录密码：设置 master 节点的登录密码。8 - 30 个字符，且必须同时包含大写字母、小写字母、数字和特殊字符!@#$%^&*。
- 高级设置
  - 统一 Meta 数据库：Hive 使用统一的集群外部的meta数据库，集群释放后 Meta 信息仍然存在。推荐先关闭。
  - 添加 Knox 用户：添加用户访问开源大数据软件 Web UI 的账号。
  - 权限设置：通过 RAM 角色为在集群上运行的应用程序提供调用其他阿里云服务所需的必要权限。无需调整，使用默认即可。
    - 服务角色：这个是用户将权限授予EMR服务，允许 EMR 代表用户调用其他阿里云的服务，例如 ECS 和 OSS。
    - ECS应用角色：这个是当用户的程序在 EMR 计算节点上运行的时候，可以不填写阿里云的 AccessKey 来访问相关的云服务，例如 OSS。EMR 会自动的申请一个临时 AccessKey 来授权这次访问。而这个 AccessKey 的权限控制将由这个角色来控制。
  - 引导操作（可选）： 您可以在集群启动 Hadoop 前执行您自定义的脚本，详细使用说明请参见引导操作。

配置清单和集群费用

页面右边会显示您所创建集群的配置清单以及集群费用。根据付费类型的不同，会展示不同的价格信息。按量付费集群显示每小时费用，包年包月显示总费用。

确认创建

当所有的信息都有效填写以后，创建按钮会亮起，确认无误后单击创建将会创建集群。

说明:

- 若是按量付费集群，集群会立刻开始创建。页面会返回概览页，就能看到在集群概览中有一个初始化中的集群。请耐心等待，集群创建会需要几分钟时间。完成之后集群的状态会切换为空闲。

- 若是包年包月集群，则会先生成订单，在支付完成订单以后集群才会开始创建。

登录Core节点

登录Core节点，请按照以下步骤操作：

1. 在Master节点上切换到hadoop账号。

```
su hadoop
```

2. 免密码SSH登录到对应的Core节点。

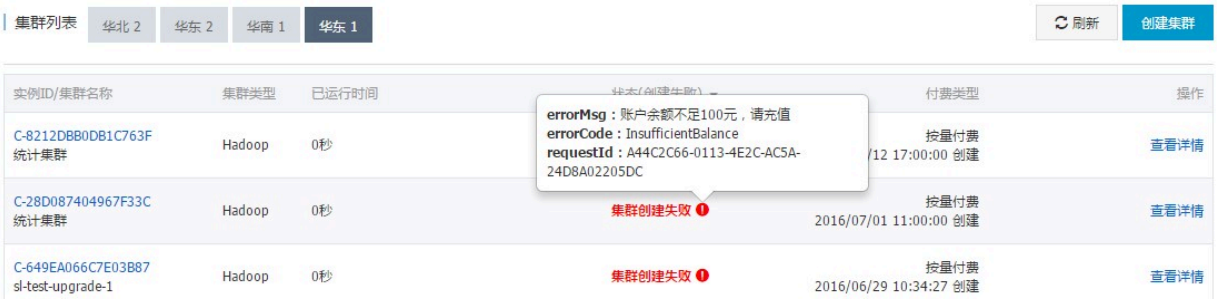
```
ssh emr-worker-1
```

3. 通过sudo命令可以获得root权限。

```
sudo vi /etc/hosts
```

创建失败

如果创建失败，在集群列表页上会显示集群创建失败，将鼠标移动到红色的感叹号上会看到失败原因。



创建失败的集群可以不用处理，对应的计算资源并没有真正的创建出来。这个集群会在停留 3 天以后自动隐藏。

2.2 集群列表与详情

您可以通过集群列表或集群详情了解您已创建的 E-MapReduce 集群信息。

集群列表

集群列表用来展示您所拥有的所有集群的基本信息。

集群列表展示的集群信息如下图所示：

列表栏各项说明如下：

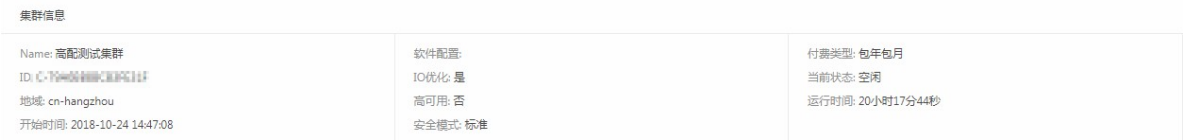
- 集群 ID/名称：集群的 ID 以及名称。在这里将鼠标移动到名称上，单击修改按钮可以对集群的名称进行修改。
- 集群类型：目前有 Hadoop 、Druid、Data Science、Kafka 四种类型。
- 状态：集群的状态，请参见集群状态。当集群出现异常的时候，比如创建失败，在右侧会有提示信息，鼠标悬停能够查看到详细的错误信息。您也可以通过列表上方的状态下拉菜单选择查看指定状态的集群。
- 创建时间：集群创建的时间。
- 运行时间：从创建开始到目前的运行时间。集群一旦被释放，计时终止。
- 付费类型：集群的付费类型。
- 操作：您可以对该集群执行的操作，包含以下操作：
  - 监控数据：监控 E-MapReduce 集群的 CPU 空闲率、内存容量、磁盘容量等多个监控项，帮助用户监测集群的运行状态。
  - 管理：进入集群与服务管理页面。
  - 详情：进入集群的详情页，查看集群建立以后的详细信息。
  - 更多
    - 扩容：集群扩容功能入口。
    - 释放：释放一个集群，请参见释放集群。
    - 重启：重启一个集群。

集群详情

集群详情用来展示用户集群的详细信息。

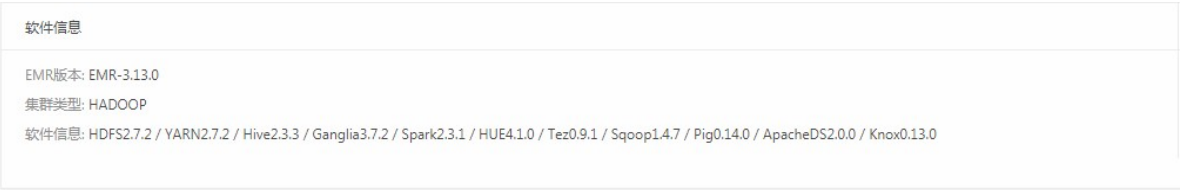
用户集群的详细信息包含以下四大部分：

- 集群信息



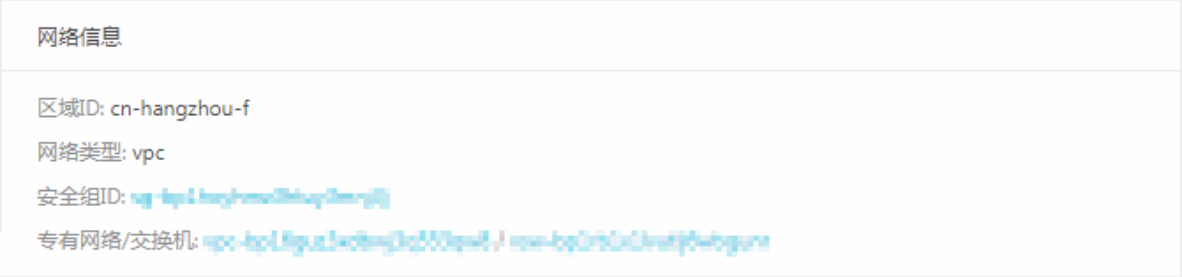
  - 名称：集群的名称
  - ID：集群的实例 ID 。
  - 地域：集群所在的 Region。
  - 开始时间：集群的创建时间。
  - 软件配置：软件的配置信息。
  - IO优化：是否开启了 IO 优化。
  - 高可用集群：是否开启了高可用集群。
  - 安全模式：集群中的软件以 Kerberos 安全模式启动。请参见Kerberos简介。
  - 付费类型：集群的付费类型。
  - 当前状态：请参见集群状态。
  - 运行时间：集群的运行时间。
  - 统一元数据：是否开启统一 Meta 数据库。
  - 引导操作：这里列出了所有配置的引导操作的名称、路径以及参数。
  - ECS 应用角色: 这个是当用户的程序在 EMR 计算节点上运行的时候，可以不填写阿里云的 AccessKey 来访问相关的云服务。例如 OSS，EMR 会自动的申请一个临时 AccessKey 来授权这次访问。而这个 AccessKey 的权限控制将由这个角色来控制。

- 软件信息



  - EMR 版本：使用的 E-MapReduce 的主版本。
  - 集群类型：选择的集群类型。
  - 软件信息：列出了用户安装的所有的应用程序及其版本，例如，HDFS 2.7.2、Hive 2.3.3、spark 2.3.1。

- **网络信息**



  - **区域 ID：集群所在的可用区，例如 cn-hangzhou-b，与 ECS 的一致。**
  - **网络类型：集群所在的网络。**
  - **安全组 ID：集群加入的安全组 ID**
  - **专有网络/交换机：用户集群所在的 VPC 与子网交换机的 ID。**

· 主机信息

- 主实例组（Master）：所有 Master 节点对应的配置，包含以下内容。
  - 主机数量：当前的节点数量和实际申请的节点数量。理论上这 2 个值一定是一样的，但是在创建过程中，当前节点会小于申请节点，直到创建完成。
  - CPU：单个节点的 CPU 的核数。
  - 内存：单个节点的内存的容量。
  - 数据盘配置：数据盘类型和单个节点的数据盘容量。
  - ECS ID：所购买的 ECS 的 ID
  - 状态：包含创建中、正常、扩容中和已释放。
  - 公网 ： Master 的公网 IP。
  - 内网 ： 机器的内网 IP，可以被集群中的所有节点访问到。
  - 创建时间：所购买的 ECS 的创建时间。

- 核心实例组（Core）：所有 Core 节点对应的配置，包含以下内容。
  - 主机数：当前的节点数量和实际申请的节点数量。
  - CPU：单个节点的 CPU 的核数。
  - 内存：单个节点的内存的容量。
  - 数据盘配置：数据盘类型和单个节点的数据盘容量。
  - ECS ID：所购买的 ECS 的ID
  - 状态：包含创建中，正常和扩容中。
  - 内网：机器的内网 IP，可以被集群中的所有节点访问到。
  - 创建时间：所购买的 ECS 的创建时间。

2.3 扩容集群

当您的集群资源（计算资源、存储资源）不足的时候，您可以将您的集群进行水平扩展。目前支持扩展您的 Core 节点和 Task 节点的数量，且使用的配置默认与您之前购买的 ECS 配置一致。

扩容入口

1. 在集群管理页上，找到对应的集群条目，单击详情，进入集群详情页面。
2. 单击右上角的资源变配，在下拉框中选择扩容。

扩容界面

如下图所示：

说明:

目前，只支持扩容，不支持缩容。

- 配置：当前实例组的配置。

- 付费类型：当前集群的付费类型。
- 当前 Core 台数：默认显示的是您当前的所有 Core 节点数量。
- 增加 Core 台数：输入您实际需要增加的量，右侧会显示扩容后的集群总费用，然后单击确定就会进行扩容。
- 交换机：当前集群的交换机。

扩容状态

如下图所示：

核心实例组

ECS ID	状态	公网	内网	创建时间
i-bp1htgltshh2o59kkxqz	正常		192.168.0.47	2018-10-25 10:36:59
i-bp1733wjev9dzvvp62n7	正常		192.168.0.48	2018-10-25 10:37:00
i-bp1dhriqana54rgvcfn2	扩容中		192.168.0.50	2018-10-25 10:53:07
i-bp1c0d1hnoaicxddc57t	扩容中		192.168.0.51	2018-10-25 10:53:09

您可以在集群基础信息页的核心实例组(CORE)的信息上看到集群的扩容情况，正在扩容的节点，其状态会显示为扩容中。当这台 ECS 的状态转为正常后，该 ECS 即已经加入该集群，并可正常提供服务了。

修改密码

扩容成功后用户可以 SSH 登录到扩容节点上修改 root 密码，具体步骤如下：

1. 使用如下命令 SSH 登录到 Master 主机。请在集群详情页的主机信息栏中获取集群 master 机器的公网 IP。

```
ssh root@ip.of.master
```

2. 切换到 hadoop 用户：

```
su hadoop
```

3. 登录到扩容节点，扩容节点的内网 IP 请在集群详情页的主机信息栏中获取：

```
ssh ip.of.worker
```

4. 使用如下命令修改 root 用户密码：

```
sudo passwd root
```

2.4 释放集群

在集群列表页面，您可以单击集群条目右侧的释放按钮对集群进行释放操作。

只有以下状态的按量付费的集群才可以被释放。

- 创建中
- 运行中
- 空闲中

普通释放

释放前会提示您再次确认，一旦确认释放，会发生以下的操作：

- 所有在集群上的作业都会被强制终止。
- 如果您选择了保存日志到 OSS，那么当前作业的日志会被保存到 OSS，所需时间取决于日志大小。日志的上传和作业的运行是并行的，作业生成日志的同时，就会进行日志的上传。所以最终作业停止时，需要上传的日志一般不会特别多，正常在几分钟内都会完成。
- 终止并释放所有的 ECS。这个过程取决于集群的大小，越小的集群会越快，正常都在几秒内完成，至多不会超过 5 分钟。最迟释放的 ECS 在等待释放时仍然计费。

> **警告：**
> 如果您想要节约成本而控制在整点前释放，请务必留出一定的释放时间保证确实在整点前释放。

强制释放

如果您不需要任何日志，只是想快速结束集群的运行，那么可以开启强制释放。释放过程就会跳过日志收集（如果有打开日志收集的话），直接进入 ECS 释放阶段。

释放失败的集群

由于系统错误等原因，集群有可能会在确认释放后，释放失败。您单击释放是一个异步的过程，可能等一会儿会发生集群释放失败的情况，但是不用担心，E-MapReduce 会启动后台保护，自动重试释放该集群，直到集群被成功释放为止。

2.5 集群续费管理

当您的包年包月集群服务即将到期的时候，您需要执行集群的续费操作，以继续您的 E-MapReduce 集群服务。集群续费包括 E-MapReduce 服务费的续费以及集群中 ECS 的续费。

续费流程

1. 登录阿里云 *E-MapReduce* 控制台。
2. 单击上方的集群管理页签，进入集群列表页面
3. 在需要续费的集群条目后单击详情进入集群详情页。

4. 单击右上角的费用管理，在下拉框中选择：

- 续费

  - 续费界面

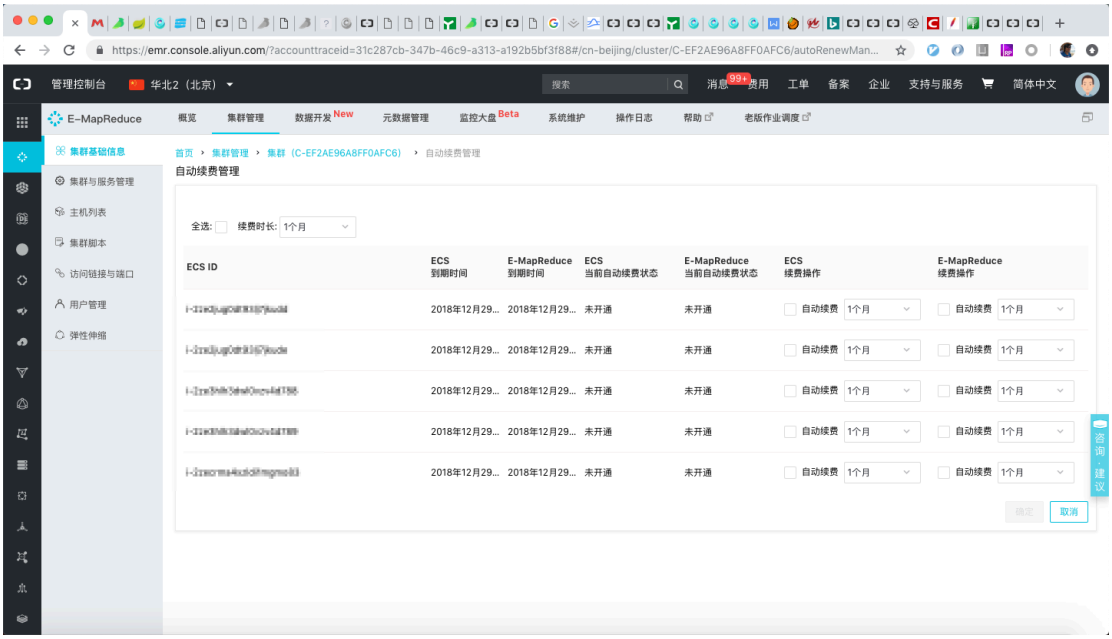
    如下图所示：

    - ECS 当前到期时间：这台 ECS 到期的时间。
    - EMR到期时间：E-MapReduce 服务到期的时间。
    - 数量：实例组的机器数量。
    - ECS列表：集群中机器的 ECS 实例 ID。
    - ECS 续费时长：需要对这个节点续费的时长，目前支持 1-9 个月、1 年、2 年和 3 年的续费时长。
    - EMR 续费时长：对应的这个节点的 E-MapReduce 服务费的时长，推荐和 ECS 的设置一致。
    - 续费价格：E-MapReduce 服务以及 ECS 节点对应的续费价格。

- 自动续费管理

  如果您购买的是包年包月集群服务，且要长期使用该项服务，您可以开通自动续费功能。开通自动续费功能后，系统将在资源到期前执行自动续费操作，无需您手动操作续费。以此保证不会因为资源未及时续费，而导致资源、数据被删除。

  - 续费界面



- ECS ID： 对应的节点的 ecsid
- ECS 到期时间： ECS 节点的到期时间
- EMR 到期时间：ECS 节点对应的EMR服务的到期时间
- ECS 当前自动续费状态： 开通或未开通，开通表示自动续费的。未开通，表示没有进行自动续费
- EMR 当前自动续费状态： 开通或未开通，开通表示自动续费的。未开通，表示没有进行自动续费
- ECS 续费操作：勾选上以后，并进行了保存。对应就会进入到自动续费的开通状态。按照后面设置的续费时长来进行续费，比如一个月那么就是每一个月进行一次续费；若不勾选，并进行保存。则取消自动续费的动态。
- EMR 续费操作： 勾选上以后，并进行了保存。对应就会进入到自动续费的开通状态。按照后面设置的续费时长来进行续费，比如一个月那么就是每一个月进行一次续费；若不勾选，并进行保存。则取消自动续费的动态。

说明:

- 如果您的实例将于下一天到期，需手动续费，无法设置自动续费。
- 如果您在自动续费扣款日前进行了手动续费，系统则按最新到期日期，重新计算下一次自动续费日期。关于计算方法，请参见 示例。
- 自动续费支持使用优惠券。

支付订单

> **注意:**
> 集群续费的费用为 ECS 续费价格与 E-MapReduce 服务产品价格的总和。当集群列表中存在未支付的订单时，您将无法执行集群的扩容和续费操作。

1. 单击确定后，会弹出下单成功提示的提示框（该提示信息可能会有较长的时间延迟，请耐心等待）。
2. 单击前往支付，跳转到订单支付页面。支付页面会显示您应付的总金额以及各订单的详情，其中一个为 E-MapReduce 产品费用订单，其它的为集群续费的 ECS 订单。
3. 单击确认支付，完成付款。
4. 完成支付之后，单击支付完成，返回集群列表页面。

此时，在集群列表页上续费成功集群的到期时间会更新为续费之后的到期时间，对应的 ECS 的到期时间更新目前存在一定的延迟，一般约 3-5 分钟之后会更新成续费之后的到期时间。

若您只是确认了续费订单但并未进行支付，您可在集群列表页面中找到该集群条目，在其右侧的操作栏中会出现前往支付和取消订单的按钮。您可单击前往支付，完成对应的订单支付和集群扩容流程；或单击取消订单，以取消本次续费操作。

2.6 SSH 登录集群

若您觉得在网页上的作业和执行计划无法满足您更加复杂的应用需求，您可以登录到 E-MapReduce 集群的主机上。找到集群的详情页，其中就有集群 master 机器的公网 IP 地址，您可以直接 SSH 登录到这台机器上，查看各种设置与状态。

集群环境变量

机器上已为您设置好相关的环境变量，其中包括以下常用的环境变量：

- JAVA_HOME
- HADOOP_HOME
- HADOOP_CONF_DIR
- HADOOP_OG_DIR
- YARN_LOG_DIR
- HIVE_HOME
- HIVE_CONF_DIR
- PIG_HOME
- PIG_CONF_DIR

您可以在脚本中直接引用这些变量，但请不要去修改这些变量的值，以免造成 E-MapReduce 的意外错误。

登录 Master 主机步骤

1. 使用如下命令 SSH 登录到 master 主机。请在集群详情页的主机信息栏中获取集群 master 机器的公网 IP。

```
ssh root@ip.of.master
```

2. 输入创建集群时设定的密码。

打通本地机器与集群 master 机器的 SSH 无密码登录

通常，您需要登录到集群上进行一些管理和操作。为便于登录集群 master 机器，您可打通与 master 机器的 SSH 无密码登录（集群 master 机器默认开通了公网 IP）。操作步骤如下：

1. 通过上面提到的 root + 密码的方式登录到 master 主机。
2. 切换到到 Hadoop 用户或者 HDFS 用户。

Linux 的 SSH 方式

1. 复制私钥到本地。

```
sz ~/.ssh/id_rsa
```

2. 回到您的本地机器，尝试重新登录 master 机器。

```
ssh -i 私钥存放路径/id_rsa hadoop@120.26.221.130
```

当然如果你只有这一个私钥，也可以直接放到你的 *~/.ssh/* 下，默认使用这个私钥，就不需要 `-i` 指定了。

Windows 的 SSH 方式

在 Windows 下你可以有多种方式来使用 SSH 免密码登录 master 机器。

- 方式一：使用 PuTTY

  1. 下载 *PuTTY*。
  2. 在同样的位置下载 PuTTYgen。
  3. 打开 PuTTYgen，并 Load 您的私钥。

注意：

请妥善保管这个私钥，保证该私钥的安全。若私钥不幸泄漏了，请立刻重新生成一个新的取代。

4. 使用默认的配置，并 Save private key。会保存出一个后缀为 ppk 的 PuTTY 使用的密钥文件。
5. 运行 PuTTY，并在配置页面选择 Session。
6. 输入您要连接的目标机器公网 IP 地址，要加上登录使用的用户名，类似 hadoop@MasterNodeIP。
7. 在配置页面，选择Connetion > SSH > Auth。
8. 选择之前生成好的 ppk 文件。
9. 最后单击 Open，就会自动登录到 master 节点了。

- 方式二：使用 Cygwin | MinGW

  这是在 Windows 上模拟 Linux 的非常方便的工具，使用起来也非常简单。

  如果采用这种方式，连接过程就可以参考上面的 Linux 的 SSH 方式了。

  推荐采用 MinGW 的方式，这个是最小巧的一种方式。如果官网打不开，可以下载 git 的客户端，默认带的 Git Bash 就可以满足。

查看 Hadoop、Spark、Ganglia 等系统的 Web UI

说明:

在进行本步骤前，请确认您已经完成了上面的*SSH* 无密码登录流程。

由于安全的缘故，E-MapReduce 集群的 Hadoop、Spark 和 Ganglia 等系统的 Web UI 监控系统的端口都没有对外开放。如果用户想要访问这些Web UI，需要建立一个 SSH 隧道，通过端口转发的方式来达到目的。有如下两种方式：

注意:

下面的操作是在您本地机器上完成的，不是集群内部机器。

- 方式一：端口动态转发
  - 密钥的方式

    创建一个 SSH 隧道，该隧道可打通您本地机器跟 E-MapReduce 集群的 master 机器的某个动态端口的连接：

    ```
    ssh -i /path/id_xxx -ND 8157 hadoop@masterNodeIP
    ```

  - 用户名密码的方式，需要输入密码：

    ```
    ssh -ND 8157 hadoop@masterNodeIP
    ```

    说明：

    8157 是您本地机器没有被使用过的任何一个端口，用户可以自定定义。

完成动态转发以后，您可以选择如下两种方式来查看：

- 推荐方式

  推荐使用 Chrome 浏览器，可以使用如下的方式来访问 Web UI：

  ```
  chrome --proxy-server="socks5://localhost:8157" --host-resolver-
  rules="MAP * 0.0.0.0 , EXCLUDE localhost" --user-data-dir=/tmp/
  ```

  若是 Windows 系统，这里的 *tmppath* 可以写成类似 *d:/tmppath*，若是 Linux 或者 OSX，可以直接写成 */tmp/*。

  在不同的操作系统中，Chrome 的默认安装路径不同，请参见下表：

操作系统	Chrome 位置
Mac OS X	/Applications/Google Chrome.app/Contents/MacOS/Google Chrome
Linux	/usr/bin/google-chrome

操作系统	Chrome 位置
Windows	C:\Program Files (x86)\Google\Chrome\Application\chrome.exe

- 插件方式

■ 使用 Chome 插件查看Web UI:

1. 首先安装 Chrome 的插件，*SwitchyOmega*。
2. 安装完成以后，单击插件，在弹出框中选择选项开始进行配置。
3. 单击新建情景模式，命名为 SSH tunnel，类型选择 PAC情景模式。
4. 配置以下内容到 PAC 脚本中：

```
function regExpMatch(url, pattern) {
  try { return new RegExp(pattern).test(url); } catch(ex) {
return false; }
}

function FindProxyForURL(url, host) {
    // Important: replace 172.31 below with the proper prefix
 for your VPC subnet

    if (shExpMatch(url, "*localhost*")) return "SOCKS5
localhost:8157";
    if (shExpMatch(url, "*emr-header*")) return "SOCKS5
localhost:8157";
    if (shExpMatch(url, "*emr-worker*")) return "SOCKS5
localhost:8157";

    return 'DIRECT';
}
```

5. 在左侧单击应用选项完成配置。
6. 运行一个命令行，在命令行选择一种方式运行以下的命令：

```
// 方式1，密钥的方式
ssh -i /path/id_xxx -ND 8157 hadoop@masterNodeIP
```

```
// 方式2，用户名密码的方式，需要输入密码
```

```
ssh -ND 8157 hadoop@masterNodeIP
```

7. 命令行跑起来以后，在 Chrome 上单击插件，切换到之前创建的 SSH tunnel 情景模式下。
8. 在浏览器地址栏输入机器 IP 地址 + 端口 即可访问。这个 IP 就是 SSH 命令行去连接的机器IP，一般是 Master 节点，例如 YARN :8088，HDFS 是：50070，等等。

这个方式下，正常网页浏览和集群 Web UI 访问互不干扰，可以同时使用。

■ 您本地机器跟 E-MapReduce 集群的 Master 主机的 SSH 通道打通后，要在浏览器中查看 Hadoop、Spark、Ganglia 的 Web UI，您还需要配置一个本地代理。操作步骤如下：

1. 假设您使用的是 Chrome 或者 Firefox 浏览器，请单击下载 *FoxyProxy Standard* 代理软件。
2. 安装完成并重启浏览器后，打开一个文本编辑器，编辑如下内容：

```
<?xml version="1.0" encoding="UTF-8"?>
<foxyproxy>
<proxies>
<proxy name="aliyun-emr-socks-proxy" id="2322596116" notes
="" fromSubscription="false" enabled="true" mode="manual"
selectedTabIndex="2" lastresort="false" animatedIcons="true
" includeInCycle="true" color="#0055E5" proxyDNS="true"
noInternalIPs="false" autoconfMode="pac" clearCacheBeforeUse
="false" disableCache="false" clearCookiesBeforeUse="false"
rejectCookies="false">
<matches>
<match enabled="true" name="120.*" pattern="http://120.*"
isRegEx="false" isBlackList="false" isMultiLine="false"
caseSensitive="false" fromSubscription="false" ></match>
</matches>
<manualconf host="localhost" port="8157" socksversion="
5" isSocks="true" username="" password="" domain="" ></
manualconf>
</proxy>
</proxies>
```

```
</foxyproxy>
```

其中：

- `Port 8157` 是您本地用来建立与集群 master 机器 SSH 连接的端口，这个需要跟您之前执行的在终端中执行的 SSH 命令中使用的端口匹配。
- `120.*` 这个匹配是用来匹配 master 主机的 IP 地址，请根据 master 的 IP 地址的情况来定。

3. 在浏览器中单击Foxyproxy按钮，选择 Options。
4. 选择 Import/Export。
5. 选择刚才您编辑的 xml 文件，单击 Open。
6. 在 Import FoxyProxy Setting 对话框中，单击 Add。
7. 单击浏览器中的 Foxyproxy 按钮，选择 Use Proxy aliyun-emr-socks-proxy for all URLs。
8. 在浏览器中输入 localhost:8088，就可以打开远端的 Hadoop 界面了。

· 方式二：本地端口转发

> **注意：**
> 这个方式的缺陷是只能看到最外层的界面，一旦要看详细的作业信息，就会出错。

- 密钥的方式：

```
ssh -i /path/id_rsa -N -L 8157:masterNodeIP:8088 hadoop@masterNode
IP
```

- 用户名密码的方式，需要输入密码：

```
ssh -N -L 8157:masterNodeIP:8088 hadoop@masterNodeIP
```

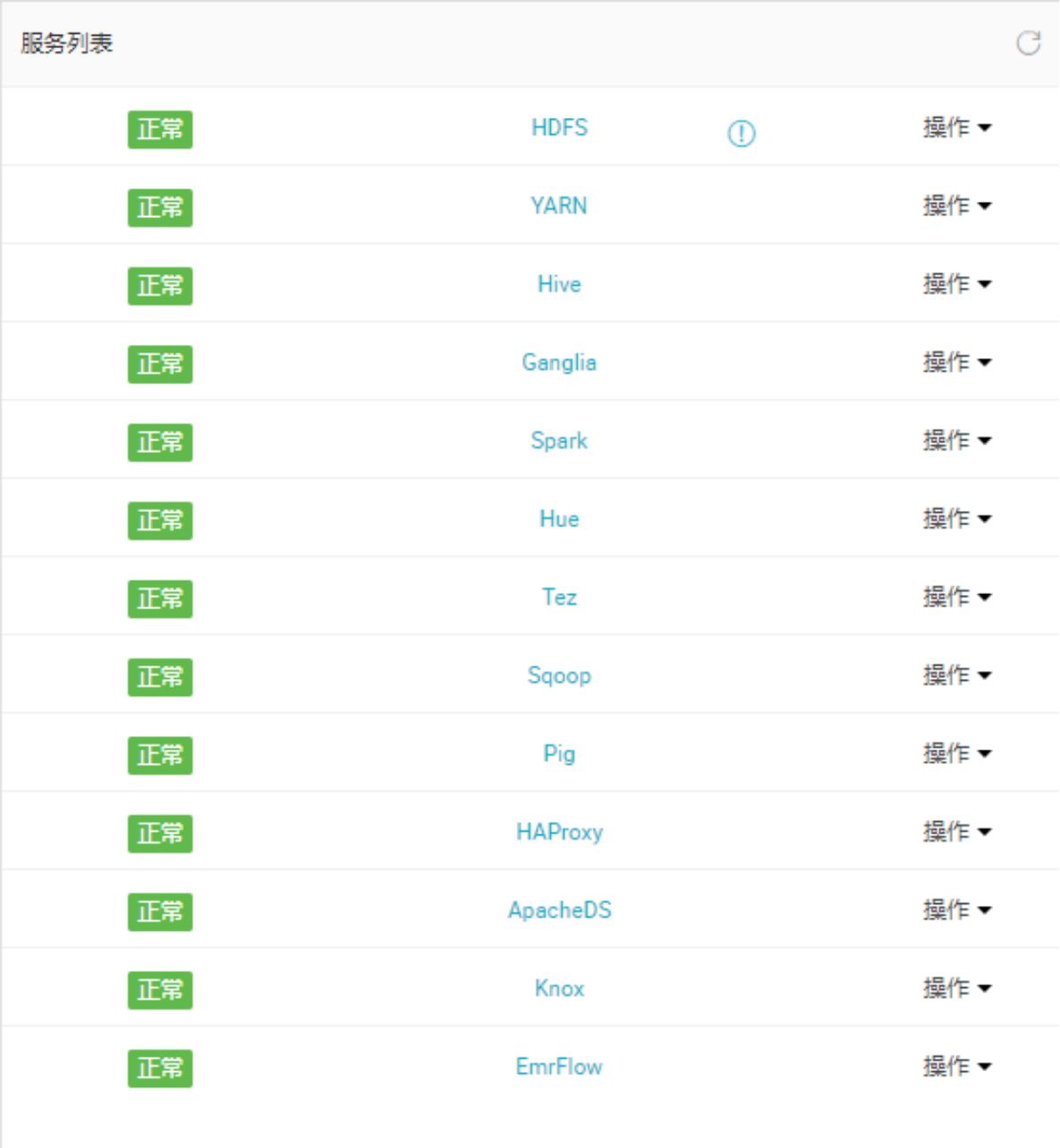
参数说明：

- `path`：私钥存放路径。
- `masterNodeIP`：要连接的 master节点 IP。
- `8088`：是 master 节点上 ResourceManager 的访问端口号。

2.7 服务列表

集群与服务管理页展示了 HDFS，YARN 等服务在集群节点上的运行状态。

服务列表展示的信息如下图所示。



创建集群时未勾选的服务，例如 Storm，将不会显示在服务列表中。

单击服务列表中的服务可查看对应服务的状态、部署拓扑、配置和配置修改历史。服务列表中服务的状态有正常和错误两种。如果某个节点上的服务状态是错误，您可以通过 master 节点跳转，登录到对应节点上查看服务进程情况。

2.8 集群脚本

集群，特别是包年包月集群，在使用过程中，可能会有新的第三方软件安装，修改集群运行环境的需求。集群脚本功能可以在集群创建好后批量选择节点，运行您指定的脚本，以实现个性化的需求。

集群脚本的作用

集群脚本类似引导操作，您可以在集群创建好后安装很多目前集群尚未支持的软件到您的集群上，例如：

- 使用 yum 安装已经提供的软件。
- 直接下载公网上的一些公开的软件。
- 读取 OSS 中您的自有数据。
- 安装并运行一个服务，例如 Flink 或者 Impala，但需要编写的脚本会复杂些。

强烈建议您先用单个节点进行集群脚本的测试，测试都正确以后再在整个集群上操作。

如何使用

1. 集群状态是空闲或者运行中的集群可以运行集群脚本，集群列表页面单击对应集群的详情按钮。
2. 左侧菜单单击集群脚本，即会进入该集群的集群脚本执行界面，右侧是已经执行过的集群脚本列表。
3. 单击右上角创建并执行，进入创建界面。
4. 填写创建界面上的配置项，选择执行的节点，点击执行，完成添加并执行操作。
5. 集群脚本列表可以看到新创建的集群脚本，点击刷新可以更新集群脚本的状态。
6. 单击查看详情可以看到脚本在各个节点上的运行情况，单击刷新可以更新脚本在各个节点上的运行状态。

只有空闲或者运行中的可用集群才能使用集群脚本功能。集群脚本适用于长期存在的集群，对按需创建的临时集群，应使用引导操作来完成集群初始化工作。

集群脚本会在您指定的节点上下载 OSS 上的脚本并运行，根据返回值是否为 0 判断执行成功还是失败。如果运行状态是失败，您可以登录到各个节点上查看运行日志，运行日志记录在每个节点的 `/var/log/cluster-scripts/clusterScriptId` 目录下。如果集群配置了 OSS 日志目录，运行日志也会上传到 `osslogpath/clusterId/ip/cluster-scripts/clusterScriptId` 目录下方便查看。

默认会使用 root 账户执行您指定的脚本，您可以在脚本中使用 `su hadoop` 切换到 Hadoop 账户。

集群脚本可能在部分节点上运行成功，部分节点上运行失败，例如节点重启导致的脚本运行失败。您可以在解决异常问题后，单独指定失败的节点再次运行。当集群扩容后，您也可以指定扩容的节点单独运行集群脚本。

1 个集群同一时间只能运行一个集群脚本，如果有正在运行的集群脚本，无法提交执行新的集群脚本。每个集群最多保留 10 个集群脚本记录，超过 10 个需要将之前的记录删除才能创建新的集群脚本。

脚本的例子

类似引导操作的脚本，您可以在脚本中指定从 OSS 下载需要的文件，下面的例子会将 *oss://yourbucket/myfile.tar.gz* 这个文件下载到本地，并解压到 */yourdir* 目录下：

```
#!/bin/bash
osscmd --id=<yourid> --key=<yourkey> --host=oss-cn-hangzhou-internal.
aliyuncs.com get oss://<yourbucket>/<myfile>.tar.gz ./<myfile>.tar.gz
mkdir -p /<yourdir>
tar -zxvf <myfile>.tar.gz -C /<yourdir>
```

osscmd 已预安装在节点上，可以直接调用来下载文件。

> **说明：**
> OSS 地址 host 有内网地址、外网地址和 VPC 网络地址之分。如果用经典网络，需要指定内网地址，杭州是 oss-cn-hangzhou-internal.aliyuncs.com。如果用 VPC 网络，要指定 VPC 内网可访问的域名，杭州是 vpc100-oss-cn-hangzhou.aliyuncs.com。

脚本也可以通过 yum 安装额外的系统软件包，下面的例子会安装 ld-linux.so.2：

```
#!/bin/bash
yum install -y ld-linux.so.2
```

2.9 访问链接与端口

本文介绍如何配置组件快捷入口以便您方便的访问集群上的开源组件。

访问链接与端口

当集群创建完成以后，我们会为您的集群创建默认的绑定几个域名，来方便您访问您的开源组件：

- YARN
- HDFS
- Ganglia
- Hue
- Spark

这些链接可以在集群管理中的访问链接与端口处找到。

> **说明:**
> 在 2.7.x 以上版本，或者是 3.5.x 以上版本，可以使用更加安全和方便的 UI 访问方式 KNOX，详细信息请参考Knox 使用说明。

设置安全组访问

当您初次使用组件快捷方式的时候，因为安全的原因，您需要遵循以下的步骤来打开您的安全组的访问。

1. 确认需要开放的安全组

   EMR 在创建的时候，会默认为您创建一个 EMR 专用的安全组。这个安全组默认的策略只开放了公网向内网方向的 22 端口。在集群的详情页面的网络信息处，找到集群对应所在的安全组名称和 ID。

2. 获取公网访问 IP

   为了安全的访问集群组件，我们推荐安全组在设置策略的时候，只针对当前的公网访问 IP 来开放。要获得当前的访问，请访问 http://ip.taobao.com/，在左下角会显示您当前的公网访问 IP 。

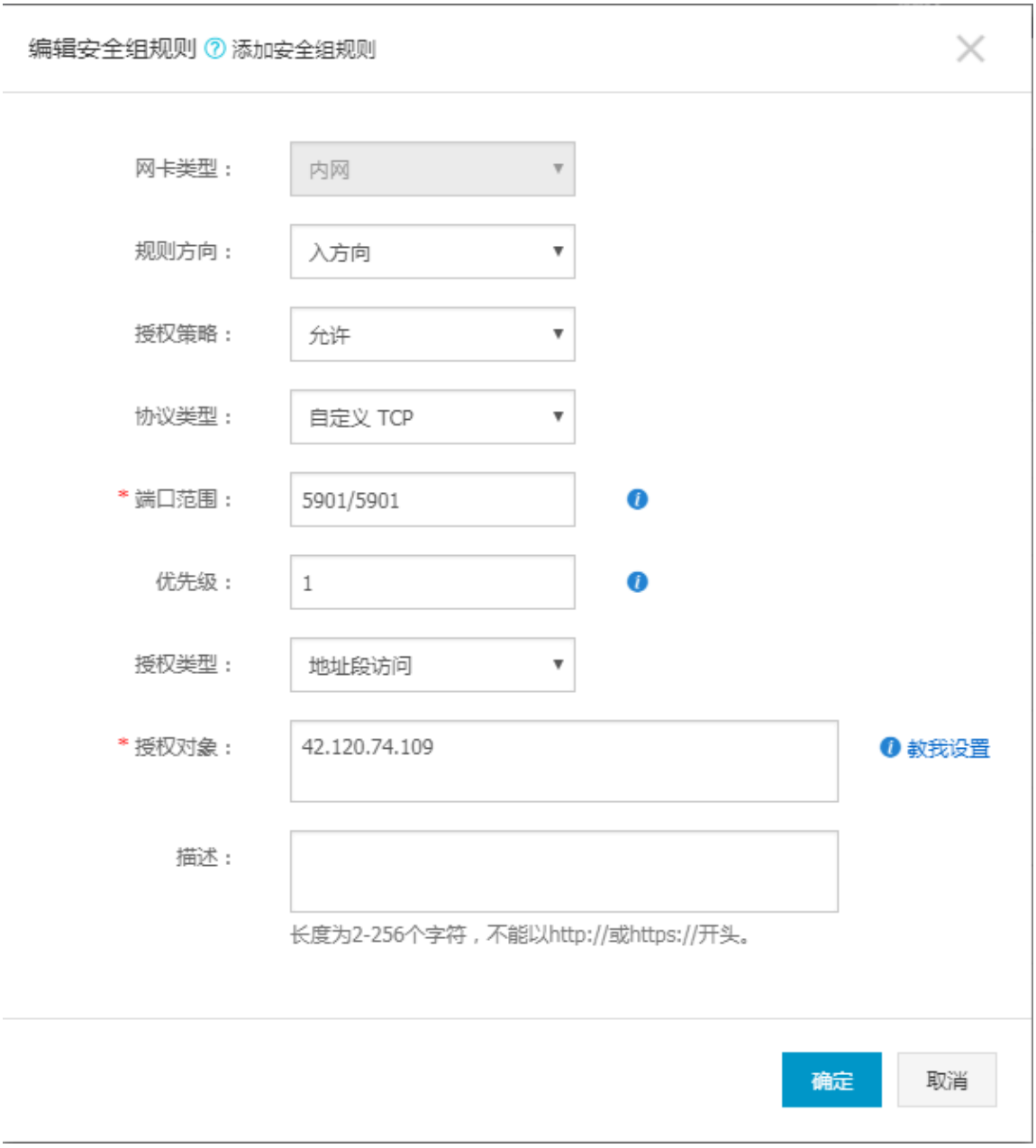
3. 修改安全组策略

a. 前往 ECS 控制台，在页面左侧导航栏中单击安全组，在安全组 ID 中输入刚才集群详情中的安全组 ID 进行查找。找到以后单击条目后面的配置规则。

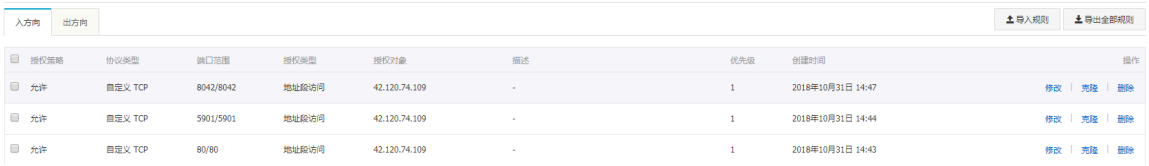
b. 如果集群是经典网络，那么配置公网入方向。

c. 如果集群是VPC，那么配置内网入方向。

d. 新增三条规则，如下图所示：


编辑安全组规则 添加安全组规则
网卡类型：
内网
规则方向：
入方向
授权策略：
允许
协议类型：
自定义 TCP
* 端口范围：
5901/5901
优先级：
1
授权类型：
地址段访问
* 授权对象：
42.120.74.109
教我设置
描述：
长度为2-256个字符，不能以http://或https://开头。
确定
取消

e. 全部完成以后，在策略列表中会看到新增的 3 条策略，如下图所示。



这个时候网络部分的设置就完成了。现在已经安全的开放了网络的访问路径。

设置开源组件访问

完成了网络的设置，这时候就可以设置访问权限了。默认的情况下，新集群的初次访问是没有设置用户名和密码的。所以无法通过 http 认证。您需要参照如下步骤设置您的访问权限。

1. 在对应集群 ID 右侧单击详情进入集群详情页面。

2. 单击左侧的访问链接与端口。
3. 单击右上方的设置用户名密码，会提示您设置访问用户名和密码。
4. 设置完成以后，再点击下方的链接，输入设置好的用户名和密码，就能正常的访问 WebUI 界面了。

说明:

- 只能存在一个用户名和密码，所以每次设置以后都将替换之前设置的用户名和密码。
- 目前只有 2.3 及以上的版本才支持此功能。

2.10 安全组

目前 E-MapReduce 创建集群的时候需要使用在 E-MapReduce 中创建的安全组。

创建 E-MapReduce 集群时，用户可以手动打开集群所在安全组的 22 端口（默认关闭，创建集群 > 基础配置 > 远程登录中打开），推荐您将 ECS 实例也按照功能划分，放于不同的用户安全组中例如，E-MapReduce 的安全组为 E-MapReduce 安全组，而您已有的安全组为用户-安全组，每个安全组按照不同的需要开启不同的访问控制。

将 E-MapReduce 集群加入现有安全组

说明:

- 经典网络类型的实例必须加入同一地域下经典网络类型的安全组。
- 专有网络类型的实例必须加入同一专有网络下的安全组。

1. 登录阿里云 *E-MapReduce* 控制台。
2. 单击上方的集群管理页签。
3. 在集群列表中找到需加入安全组的集群条目，单击其操作中的详情，即会进入集群详情页。
4. 在集群详情页上，找到网络信息下的所属安全组 ID，单击安全组的 ID。
5. 单击左侧的安全组内实例列表查看该安全组内所有 ECS 实例 ID/名称。
6. 前往阿里云 *ECS* 的管理控制台，在左侧导航栏中实例。
7. 在实例列表页面中，您会看到很多名字以 emr-xxx 开头的 ECS 实例，这些就是对应的 E-MapReduce 集群中的 ECS。找到需要加入安全组的实例，单击操作列下的管理。
8. 单击本实例安全组。
9. 单击加入安全组。
10. 选择需要加入的安全组。如果您需要加入到批量选择栏，选择安全组后单击加入多个安全组，将会显示一个选择栏，选中的安全组自动添加到选择栏中。

11.单击确定。

将现有集群加入 E-MapReduce 安全组

和上面的操作一样，先找到现有集群所在安全组，重复如上的操作，移入 E-MapReduce 的安全组即可。如果是一些零散的机器，也可以直接在 ECS 的控制台界面上选择机器，然后单击更多 > 网络和安全组 > 加入安全组批量将集群的实例移入 E-MapReduce 的安全组。

安全组的规则

一个 ECS 实例在多个不同的安全组的时候，安全组的规则是"或"的关系。举例来说就是，E-MapReduce 的安全组只开放了 22 端口，而用户-安全组开放了所有的端口。当 E-MapReduce 的集群加入用户-安全组以后，E-MapReduce 中的机器也会开放所有端口，所以在使用上请特别注意。

2.11 创建 Gateway

Gateway 是与 E-MapReduce 集群处于同一个内网中的 ECS 服务器，用户可以使用 Gateway 实现负载均衡和安全隔离，也可以通过 Gateway 向 E-MapReduce 集群提交作业。

您可以通过以下两种方式创建 Gateway：

- （推荐）通过阿里云 E-MapReduce 控制台创建。
- 手动搭建。

通过 E-MapReduce 控制台创建 Gateway

当前 E-MapReduce Gateway 仅支持 E-MapReduce Hadoop 类型的集群。在创建 Gateway 前，请确保您已经创建了 E-MapReduce Hadoop 类型集群。创建 Gateway，请按照如下步骤进行操作：

1. 登录阿里云 E-MapReduce 控制台。
2. 单击创建 Gateway 按钮。
3. 在创建 Gateway 页面中进行配置。
   - 集群名称：创建的 Gateway 的名称，长度限制为 1~64 个字符，只允许包含中文、字母、数字、连接号（-）、下划线（_）。
   - 密码/密钥对：
     - 登录密码：在文本框中输入登录 Gateway 的密码。长度限制为 8 - 30 个字符，且必须同时包含大写字母、小写字母、数字和特殊字符!@#$%^&*。
     - 密钥对：在下拉菜单中选择登录 Gateway 的密钥对名称。如果还未创建过密钥对，单击右侧的新建密钥对链接，进入 ECS 控制台创建。请妥善保管好密钥对所对应的私钥

文件（.pem 文件）。在 Gateway 创建成功后，该密钥对的公钥部分会自动绑定到 Gateway 所在的云服务器 ECS 上。在您通过 SSH 登录 Gateway 时，请输入该私钥文件中的私钥。

- 付费类型
  - 包年包月：一次性支付一段时间的费用，价格相对来说会比较便宜，特别是包三年的时候折扣会比较大。
  - 按量付费：根据实际使用的小时数来支付费用，每个小时计一次费用。
- 关联集群：为该集群创建 Gateway，即创建的 Gateway 可以向哪个集群提交作业。Gateway 将会自动配置与该集群一致的 Hadoop 环境。
- 可用区：关联集群所在的可用区（Zone）。
- 网络类型：关联集群的网络类型。
- VPC：关联集群所在地域下的 VPC。
- 交换机：选择在对应的 VPC 下的在对应可用区的交换机。
- 安全组名称:关联集群所属的安全组。
- 挂载公网：是否挂载弹性公网 IP。
- Gateway 实例：该区域内可选择的 ECS 实例规格。
- 系统盘配置：Gateway 节点使用的系统盘类型。系统盘有2种类型：SSD 云盘和高效云盘，根据不同机型和不同的 Region，系统盘显示类型会有不同。系统盘默认随着集群的释放而释放。
- 系统盘大小：最小为 40 GB，最大为 500 GB。默认值为 300 GB。
- 数据盘配置：Gateway 节点使用的数据盘类型。数据盘有两种类型：SSD 云盘和高效云盘，根据不同机型和不同的 Region，数据盘显示类型会有不同。数据盘默认随着集群的释放而释放。
- 数据盘大小：最小为 200 GB，最大为 4000 GB。默认值为 300 GB。
- 数量：数据盘的数量，最小设置为 1 台，最大设置为 10 台。

4. 单击创建完成 Gateway 的创建。

   新创建的 Gateway 会显示在集群列表中，创建完成后，状态列会变为空闲状态。

手动搭建 Gateway

- 网络环境

  首先要保证 Gateway 节点在 E-MapReduce 对应集群的安全组中，Gateway 节点可以顺利的访问 E-MapReduce 集群。设置节点的安全组请参考创建安全组。

- 软件环境
  - 系统环境：推荐使用 CentOS 7.2 及以上版本。
  - Java环境：安装 JDK 1.7 及以上版本，推荐使用 OpenJDK version 1.8.0 版本。
- 搭建步骤
  - E-MapReduce 2.7 及以上版本，3.2 及以上版本

    这些版本推荐直接使用 E-MapReduce 控制台来创建 Gateway。

    如果您选择手动搭建，请先创建一个脚本，脚本内容如下所示，然后在 Gataway 节点上执行。执行命令为：`sh deploy.sh <masteri_ip> master_password_file`。

    ■ *`deploy.sh`*：脚本名称，内容见下面代码。

    ■ *`masteri_ip`*：集群的master节点的IP，请确保可以访问。

    ■ *`master_password_file`*：保存 master 节点的密码文件，将 master 节点的密码直接写在文件内即可。

```
#!/usr/bin/bash
if [ $# != 2 ]
then
   echo "Usage: $0 master_ip master_password_file"
   exit 1;
fi
masterip=$1
masterpwdfile=$2
if ! type sshpass >/dev/null 2>&1; then
   yum install -y sshpass
fi
if ! type java >/dev/null 2>&1; then
   yum install -y java-1.8.0-openjdk
fi
mkdir -p /opt/apps
mkdir -p /etc/ecm
echo "Start to copy package from $masterip to local gateway(/opt/
apps)"
echo " -copying hadoop-2.7.2"
sshpass -f $masterpwdfile scp -r -o 'StrictHostKeyChecking no'
root@$masterip:/usr/lib/hadoop-current /opt/apps/
echo " -copying hive-2.0.1"
sshpass -f $masterpwdfile scp -r root@$masterip:/usr/lib/hive-
current /opt/apps/
echo " -copying spark-2.1.1"
sshpass -f $masterpwdfile scp -r root@$masterip:/usr/lib/spark-
current /opt/apps/
echo "Start to link /usr/lib/\${app}-current to /opt/apps/\${app}"
if [ -L /usr/lib/hadoop-current ]
then
   unlink /usr/lib/hadoop-current
fi
ln -s /opt/apps/hadoop-current  /usr/lib/hadoop-current
if [ -L /usr/lib/hive-current ]
then
   unlink /usr/lib/hive-current
fi
```

```
ln -s /opt/apps/hive-current  /usr/lib/hive-current
if [ -L /usr/lib/spark-current ]
then
   unlink /usr/lib/spark-current
fi
ln -s /opt/apps/spark-current /usr/lib/spark-current
echo "Start to copy conf from $masterip to local gateway(/etc/ecm
)"
sshpass -f $masterpwdfile scp -r root@$masterip:/etc/ecm/hadoop-
conf  /etc/ecm/hadoop-conf
sshpass -f $masterpwdfile scp -r root@$masterip:/etc/ecm/hive-conf
 /etc/ecm/hive-conf
sshpass -f $masterpwdfile scp -r root@$masterip:/etc/ecm/spark-
conf /etc/ecm/spark-conf
echo "Start to copy environment from $masterip to local gateway(/
etc/profile.d)"
sshpass -f $masterpwdfile scp root@$masterip:/etc/profile.d/hdfs.
sh /etc/profile.d/
sshpass -f $masterpwdfile scp root@$masterip:/etc/profile.d/yarn.
sh /etc/profile.d/
sshpass -f $masterpwdfile scp root@$masterip:/etc/profile.d/hive.
sh /etc/profile.d/
sshpass -f $masterpwdfile scp root@$masterip:/etc/profile.d/spark.
sh /etc/profile.d/
if [ -L /usr/lib/jvm/java ]
then
   unlink /usr/lib/jvm/java
fi
echo "" >>/etc/profile.d/hdfs.sh
echo export JAVA_HOME=/usr/lib/jvm/jre-1.8.0 >>/etc/profile.d/hdfs
.sh
echo "Start to copy host info from $masterip to local gateway(/etc
/hosts)"
sshpass -f $masterpwdfile scp root@$masterip:/etc/hosts /etc/
hosts_bak
cat /etc/hosts_bak | grep emr | grep cluster >>/etc/hosts
if ! id hadoop >& /dev/null
then
   useradd hadoop
fi
```

- **E-MapReduce 2.7 以下版本，3.2以下版本**

  创建一个脚本，脚本内容如下所示，然后在 Gataway 节点上执行。执行命令为：`sh deploy.sh <masteri_ip> master_password_file`。

  ■ *`deploy.sh`*：**脚本名称，内容见下面代码。**

  ■ *`masteri_ip`*：**集群的 master 节点的 IP，请确保可以访问。**

  ■ *`master_password_file`*：**保存 master 节点的密码文件，将 master 节点的密码直接写在文件内即可。**

```
!/usr/bin/bash
if [ $# != 2 ]
then
   echo "Usage: $0 master_ip master_password_file"
   exit 1;
fi
masterip=$1
```

```
masterpwdfile=$2
if ! type sshpass >/dev/null 2>&1; then
   yum install -y sshpass
fi
if ! type java >/dev/null 2>&1; then
   yum install -y java-1.8.0-openjdk
fi
mkdir -p /opt/apps
mkdir -p /etc/emr
echo "Start to copy package from $masterip to local gateway(/opt/
apps)"
echo " -copying hadoop-2.7.2"
sshpass -f $masterpwdfile scp -r -o 'StrictHostKeyChecking no'
root@$masterip:/usr/lib/hadoop-current /opt/apps/
echo " -copying hive-2.0.1"
sshpass -f $masterpwdfile scp -r root@$masterip:/usr/lib/hive-
current /opt/apps/
echo " -copying spark-2.1.1"
sshpass -f $masterpwdfile scp -r root@$masterip:/usr/lib/spark-
current /opt/apps/
echo "Start to link /usr/lib/\${app}-current to /opt/apps/\${app}"
if [ -L /usr/lib/hadoop-current ]
then
   unlink /usr/lib/hadoop-current
fi
ln -s /opt/apps/hadoop-current  /usr/lib/hadoop-current
if [ -L /usr/lib/hive-current ]
then
   unlink /usr/lib/hive-current
fi
ln -s /opt/apps/hive-current  /usr/lib/hive-current
if [ -L /usr/lib/spark-current ]
then
   unlink /usr/lib/spark-current
fi
ln -s /opt/apps/spark-current /usr/lib/spark-current
echo "Start to copy conf from $masterip to local gateway(/etc/emr
)"
sshpass -f $masterpwdfile scp -r root@$masterip:/etc/emr/hadoop-
conf  /etc/emr/hadoop-conf
sshpass -f $masterpwdfile scp -r root@$masterip:/etc/emr/hive-conf
 /etc/emr/hive-conf
sshpass -f $masterpwdfile scp -r root@$masterip:/etc/emr/spark-
conf /etc/emr/spark-conf
echo "Start to copy environment from $masterip to local gateway(/
etc/profile.d)"
sshpass -f $masterpwdfile scp root@$masterip:/etc/profile.d/hadoop
.sh /etc/profile.d/
if [ -L /usr/lib/jvm/java ]
then
   unlink /usr/lib/jvm/java
fi
ln -s /usr/lib/jvm/java-1.8.0-openjdk-1.8.0.131-3.b12.el7_3.x86_64
/jre /usr/lib/jvm/java
echo "Start to copy host info from $masterip to local gateway(/etc
/hosts)"
sshpass -f $masterpwdfile scp root@$masterip:/etc/hosts /etc/
hosts_bak
cat /etc/hosts_bak | grep emr | grep cluster >>/etc/hosts
if ! id hadoop >& /dev/null
then
   useradd hadoop
```

```
fi
```

- 测试
  - Hive

```
[hadoop@iZ23bc05hrvZ ~]$ hive
hive> show databases;
OK
default
Time taken: 1.124 seconds, Fetched: 1 row(s)
hive> create database school;
OK
Time taken: 0.362 seconds
hive>
```

  - 运行 Hadoop 作业

```
[hadoop@iZ23bc05hrvZ ~]$ hadoop  jar /usr/lib/hadoop-current/share
/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.0.jar pi 10 10
Number of Maps  = 10
Samples per Map = 10
Wrote input for Map #0
Wrote input for Map #1
Wrote input for Map #2
Wrote input for Map #3
Wrote input for Map #4
Wrote input for Map #5
Wrote input for Map #6
Wrote input for Map #7
Wrote input for Map #8
Wrote input for Map #9
  File Input Format Counters
      Bytes Read=1180
  File Output Format Counters
      Bytes Written=97
Job Finished in 29.798 seconds
Estimated value of Pi is 3.20000000000000000000
```

2.12 集群资源管理

E-MapReduce（以下简称 EMR） 集群资源管理主要应用于大集群多租户场景中。

EMR 集群资源管理可以帮助用户实现以下目标：

- 集群资源中不同部门或用户使用不同的资源队列，实现队列资源的隔离。
- 各队列具有一定的弹性，提高集群的使用效率。

说明:

- 目前 EMR 集群资源管理支持 *Capacity Scheduler* 和 *Fair Scheduler* 两种调度器。
- EMR 集群资源管理目前仅支持对 EMR Hadoop 类型的常驻集群进行管理。

开启资源队列

1. 登录阿里云 E-MapReduce 控制台，单击集群管理进入集群管理页面。
2. 单击对应集群右侧的管理。
3. 在左侧导航栏中单击集群资源管理。

4. 单击开启 YARN 资源队列

> **注意:**
>
> 开启集群资源队列后，YARN 组件配置中的 Capacity Scheduler 和 Fair Scheduler 配置区域将处于冻结状态，相关已有配置将会同步到集群资源管理页面中。如需继续在 YARN 组件配置中通过 XML 的方式设置集群资源，需在集群资源管理关闭 YARN 资源队列。

配置资源队列

- EMR 资源队列开启后，首先选择调度器类型，Fair Scheduler 或 Capacity Scheduler。
- 单击资源队列配置进入各队列配置页面。
- 单击设置调度器默认值设置的是调度器默认配置。

资源队列生效

配置完资源队列后，用户可以单击配置检查按钮，预览配置的 XML，确认无误后单击部署生效。

> **说明:**
>
> - 切换调度器类型（如从 Fair Scheduler 切换到 Capacity Scheduler），Fair Scheduler 中修改集群资源抢占相关参数，修改完成后需在集群与服务管理页面对 YARN 进行操作，重启 YARN all Component。
> - 新建资源队列等，设置资源队列的相关参数，直接点击部署生效即可让配置生效。
> - 配置生效过程可以在集群与服务管理页面，查看操作历史查看生效进度。

2.13 配置弹性伸缩

2.13.1 弹性伸缩概述

本文将介绍如何开启和关闭弹性伸缩功能。

在以下场景中，您可以通过 E-MapReduce 弹性伸缩功能节省成本，提高执行效率。

- 临时需要按照时间段添加计算节点，补充计算能力。
- 确保重要作业按时完成，按照某些集群指标扩充计算节点。

说明:

- 弹性伸缩功能仅能对 Task 节点进行扩容或缩容。
- 弹性伸缩仅支持 Hadoop 包年包月、按量付费集群。
- 如果能明确弹性扩容和缩容的时间，按照一定时间周期的伸缩资源，建议您选择按时间伸缩策略。
- 如果伸缩时间比较多，较难评估，但需要根据 YARN 的特定指标伸缩计算资源，建议您选择按负载伸缩策略。

开启弹性伸缩

1. 登录阿里云 E-MapReduce 控制台，单击前往集群列表进入集群列表页面。
2. 单击集群 ID 右侧的管理。
3. 在左侧导航栏中单击弹性伸缩。
4. 单击页面右上角的开启弹性伸缩按钮。

   如果当前账号是首次使用弹性伸缩功能，需对 E-MapReduce 授予弹性伸缩（ESS）的默认角色。

5. 在 ESS 的授权界面中单击确定即可。

关闭弹性伸缩

单击关闭弹性伸缩，当前已经通过弹性伸缩功能扩充的 Task 节点将会被全部释放，HDFS 存储的位于 Core 节点的数据不会受影响。

使用限制

EMR 弹性伸缩只能选择一种伸缩策略，按时间伸缩或按负载伸缩，两者只能选择一种。

如果切换伸缩策略，原伸缩规则会保留，但会处于失效状态，不会被触发执行；当前已经扩容的节点将会保留，除非缩容规则触发，否则不会被缩容。

2.13.2 按时间伸缩规则配置

如果 Hadoop 集群计算量在一定的周期内存在明显的波峰、波谷，您可以设置在每天、每周或每月的固定时间段扩出一定量的 Task 节点来补充计算能力，这样在保证作业完成的同时，可以节省您的成本。

由于弹性伸缩节点均为按量付费的购买方式，并且按量付费和包年包月的同等计算能力价格大概在 3：1 左右，所以需要根据您的弹性伸缩时间来设计包年包月计算能力和按量付费计算能力的比例。例如业务波峰时间段每天持续 8 小时，包年包月和按量付费的价格大致相同，当大于 8 小时时，包年包月比弹性伸缩的购买方式更优惠。

配置伸缩实例数

- 最大节点数：弹性伸缩的 Task 节点上限。一旦达到上限，即使满足弹性伸缩的规则，也不会继续进行弹性伸缩的动作。目前可设置的弹性伸缩最大上限为 1000。
- 最小节点数：弹性伸缩的 Task 节点下限。如果弹性伸缩规则中设置的增加或减少 Task 节点数小于此处的最小节点数，那么在首次执行时，集群会以最小节点数为准进行伸缩。

  例如，设置弹性扩容规则为每天零点动态添加 1 个节点，但最小节点数为 3。那么系统在第一天的零点时会添加 3 个节点，以满足最小节点数的要求。

配置伸缩规则

伸缩规则分为扩容规则和缩容规则。集群关闭弹性伸缩功能后，所有规则会被清空，再次开启弹性伸缩功能时，需要重新配置伸缩规则。

- 规则名称：在同一个集群中，伸缩规则名称（包括扩容规则和缩容规则）不允许重复。
- 规则执行周期：
  - 只执行一次：集群在指定的时间点执行一次弹性伸缩动作。
  - 重复执行：用户可以选择每天、每周或每月的某一特定时间点执行一次弹性伸缩动作。
- 重试过期时间：弹性伸缩在到达指定时间时可能由于各种原因不能执行，通过设置重试过期时间，系统会在该时间范围内每隔 30 秒一直检测可以执行伸缩的时机，直到在满足条件时执行伸缩。设置范围为 0 到 21600 秒。

  假设在指定时间段需要进行弹性伸缩动作A，如果有其他弹性伸缩动作B或正处在冷却期，则动作 A 无法执行。在您设置的重试过期时间内，每隔 30 秒会重试一次，尝试执行 A，一旦条件满足，集群会立刻执行弹性伸缩。
- 增加或减少 Task 节点数：规则被触发时，集群每次执行增加或减少的 Task 节点数量。
- 冷却时间：每次弹性伸缩动作执行完成，到可以再次进行弹性伸缩的时间间隔。在冷却时间内，不会发生弹性伸缩动作。

配置伸缩规格

弹性伸缩配置可以指定伸缩的节点的硬件规格。用户只能在开启弹性伸缩功能时配置，保存后不能更改。如出于特殊情况确实需要修改，可以关闭弹性伸缩功能后，再次开启。

- 选择 vCPU 和内存规格时，系统会根据您的选择自动匹配出满足条件的实例，显示在下面的备选实例列表中。您需要添加备选的实例到右侧列表中，以便集群按照已选的实例规格进行伸缩。
- 为避免由于 ECS 库存不足造成的弹性伸缩失败，您最多可以选择 3 种 ECS 实例。
- 无论是选择高效云盘还是 SSD 云盘，数据盘最小设置为 40 G。

2.13.3 按负载伸缩规则配置

在使用 E-MapReduce Hadoop 集群时，您如果无法准确的预估大数据计算的波峰、波谷，可以使用按负载伸缩的策略。

配置伸缩实例数

- 最大节点数：弹性伸缩的 Task 节点上限。一旦达到上限，即使满足弹性伸缩的规则，也不会继续进行弹性伸缩的动作。目前可设置的弹性伸缩最大上限为 1000。
- 最小节点数：弹性伸缩的 Task 节点下限。
  - 如果弹性伸缩规则中设置的增加后的 Task 节点总数小于最小节点数，那么在首次执行时，集群会以最小节点数为准，伸缩至最小节点数。例如，当前 Task 节点数为 0，设置弹性扩容规则为每天零点动态添加 1 个节点，但最小节点数为 3。那么系统在第一天的零点时会添加至 3 个节点，此后每天零点动态添加 1 个节点，以满足最小节点数的要求。
  - 如果弹性伸缩规则中设置减少后的 Task 节点总数小于最小节点数，那么集群不会进行弹性伸缩操作。

配置负载伸缩规则

伸缩规则分为扩容规则和缩容规则。集群关闭弹性伸缩功能后，所有规则会被清空，再次开启弹性伸缩功能时，需要重新配置伸缩规则。 切换伸缩策略，如从按负载伸缩切换到按时间伸缩时。原策略下的伸缩规则处于失效状态，不会被触发。已经弹出的节点会继续保留，不会被释放。

- 规则名称：在同一个集群中，伸缩规则名称（包括扩容规则和缩容规则）不允许重复。
- 集群负载指标：此处在 YARN 的负载指标，具体可以参见 *Hadoop* 官方文档。

E-MapReduce 弹性伸缩指标	YARN Metrics	说明
YARN.AvailableVCores	availableVirtualCores	The number of available virtual cores
YARN.PendingVCores	pendingVirtualCores	The number of pending virtual cores，EMR 补充采集
YARN.AllocatedVCores	allocatedVirtualCores	The number of allocated virtual cores
YARN.ReservedVCores	reservedVirtualCores	The number of reserved virtual cores
YARN.AvailableMemory	availableMB	The amount of memory available in MB
YARN.PendingMemory	pendingMB	The amount of memory pending in MB，EMR 补充采集
YARN.AllocatedMemory	allocatedMB	The amount of memory allocated in MB
YARN.ReservedMemory	reservedMB	The amount of memory reserved in MB
YARN.AppsRuning	appsRunning	The number of applications running

E-MapReduce 弹性伸缩指标	YARN Metrics	说明
YARN.AppsPending	appsPending	The number of applications pending
YARN.AppsKilled	appsKilled	The number of applications killed
YARN.AppsFailed	appsFailed	The number of applications failed
YARN.AppsCompleted	appsCompleted	The number of applications completed
YARN.AppsSubmitted	appsSubmitted	The number of applications submitted
YARN.AllocatedContainers	containersAllocated	The number of containers allocated
YARN.PendingContainers	containersPending	The number of containers pending
YARN.ReservedContainers	containersReserved	The number of containers reserved

- 统计周期与统计规则：用户选定的集群负载指标在一个统计周期内，按照选定的聚合维度（平均值、最大值、最小值），达到触发阈值为一次触发。
- 重复几次扩容：负载指标聚合后达到阈值触发的次数，达到该次数后触发集群弹性伸缩的动作。
- 扩容（台）：规则被触发时，集群每次执行增加或减少的 Task 节点数量。
- 冷却时间（秒）：每次弹性伸缩动作执行完成，到可以再次进行弹性伸缩的时间间隔。在冷却时间内，即使满足弹性伸缩条件也不会发生弹性伸缩动作。

配置伸缩规格

弹性伸缩配置可以指定伸缩的节点的硬件规格。用户只能在开启弹性伸缩功能时配置，保存后不能更改。如出于特殊情况确实需要修改，可以关闭弹性伸缩功能后，再次开启。

- 选择 vCPU 和内存规格时，系统会根据您的选择自动匹配出满足条件的实例，显示在下面的备选实例列表中。您需要添加备选的实例到右侧列表中，以便集群按照已选的实例规格进行伸缩。
- 为避免由于 ECS 库存不足造成的弹性伸缩失败，您最多可以选择 3 种 ECS 实例。
- 无论是选择高效云盘还是 SSD 云盘，数据盘最小设置为 40 G。

2.13.4 弹性伸缩抢占式实例

E-MapReduce 的抢占式实例适用于大数据作业执行成功与否没有强需求，但对计算资源价格非常敏感的场景。可以通过开通弹性伸缩功能，购买抢占式实例提升集群的计算资源。

开启弹性伸缩

开启弹性伸缩并设置伸缩规则可以参考以下步骤：

1. 登录阿里云 E-MapReduce 控制台。
2. 单击集群管理。
3. 单击需要添加抢占式实例集群 ID 右侧的管理。
4. 在左侧导航栏中单击弹性伸缩。
5. 如果集群没有开启过弹性伸缩，单击开启弹性伸缩。
6. 配置伸缩规则请参见按时间伸缩规则配置。
7. 在伸缩配置中，选择抢占式实例。

配置抢占式实例

> **说明：**
> 抢占式实例价格相对于普通按量付费实例有较大优惠，但阿里云会根据 供需资源或市场成交价的变化可能随时释放您的抢占式实例。

配置抢占式实例参考以下步骤：

1. 选择实例的 vCPU 和内存大小
2. 选择实例类型，EMR 会筛选出符合规格的所有实例类型，为保证您尽可能购买到抢占式实例，您可以选择 3 种实例规格。

3. 选中对应的实例规格后，设置单台实例规格的上限价格，单击确认。该实例类型会从可选实例到已选实例。如需改变出价，可再次在可选实例列表中选中实例规则，修改单台实例上限价格（按小时计算）。当您的出价高于当前市场价格时，您的实例就会运行。最终实例规格会按照市场价格计费。
4. 系统盘主要部署 OS 和 EMR 基础服务，无需您调整大小。可根据您的需要调整数据盘大小。
5. 实例出价上限+系统盘价格+数据盘价格为最终的配置价格。单击保存完成配置。

关于抢占式实例的更多信息，可以查看抢占式实例 FAQ。

2.13.5 弹性伸缩记录

弹性伸缩执行完成后，您可以单击弹性伸缩页面上方的弹性伸缩记录页签，查看弹性伸缩活动的执行记录，以及弹性伸缩活动执行完成后的节点数量等信息。

弹性伸缩的执行状态包括以下 4 类：

- 执行中：弹性伸缩活动正在执行。
- 成功：根据伸缩规则，所有弹性伸缩中的所有节点被加入或移出集群。
- 部分成功：根据伸缩规则，有部分节点成功被加入或移出集群，但是受磁盘配额管理或 ECS 库存的影响，部分节点执行失败。
- 全部失败：根据伸缩规则，没有一个节点被加入或移出集群。

2.14 配置专有网络

专有网络（Virtual Private Cloud，VPC）为用户创建一个隔离的网络环境，用户可以选择自有的 IP 地址范围、划分网络、配置路由表、网关等。

详见专有网络产品简介。另外通过高速通道 可以实现跨地域或跨用户的 VPC 内网互通、VPC 与物理 IDC 机房互通。

创建专有网络集群

E-MapReduce 在创建集群的时候可以选择网络类型，即经典网络/专有网络，若选择专有网络，需要如下额外操作：

- VPC：选择当前创建的 E-MapReduce 集群归属的 VPC，如果还没创建可以进入 VPC 控制台进行创建，一般一个账号最多创建 2 个 VPC 网络，超过 2 个需要提工单。
- 交换机：E-MapReduce 集群内的 ECS 实例通过交换机进行通信，如果还没创建可以进入 VPC 控制台，单击交换机页签进入交换机页面进行创建，因为交换机有可用区的属性，所以在 E-MapReduce 创建集群时选定了可用区后，创建的交换机也必须属于该可用区。

- 安全组名称：集群所属的安全组，经典网络的安全组不能在 VPC 中使用，VPC 的安全组只能在当前 VPC 中使用。这里只展示用户在 E-MapReduce 产品中创建的安全组。因为一些安全的原因目前尚不支持选择在 E-MapReduce 外创建的安全组。如果需要新建安全组，可以输入安全组的名字完成新建。

示例

不同 VPC 中的 EMR 集群通信（Hive 访问 HBase）

1. 创建集群。

   在 E-MapReduce 上面创建两个集群，Hive 集群 C1 处于 VPC1 中，HBase 集群 C2 处于 VPC2 中，两个集群都在杭州区域。

2. 配置高速通道。

   详细配置请参考同账号VPC互连，地域选择同地域即可。

3. SSH 登录 HBase 集群，通过 HBase Shell 创建表。

```
hbase(main):001:0> create 'testfromHbase','cf'
```

4. SSH 登录 Hive集群。

   a. 修改 hosts，增加如下一行：

```
$zk_ip emr-cluster //$zk_ip为Hbase集群的zk节点IP
```

   b. 通过 Hive Shell 访问 HBase。

```
hive> set hbase.zookeeper.quorum=172.16.126.111,172.16.126.112,172
.16.126.113;
hive> CREATE EXTERNAL TABLE IF NOT EXISTS testfromHive (rowkey
STRING, pageviews Int, bytes STRING) STORED BY 'org.apache.hadoop
.hive.hbase.HBaseStorageHandler' WITH SERDEPROPERTIES ('hbase.
columns.mapping' = ':key,cf:c1,cf:c2') TBLPROPERTIES ('hbase.table
.name' = 'testfromHbase');
```

   此时命令会卡住，然后会报 java.net.SocketTimeoutException 的异常，原因是 HBase 集群的 ECS 所在的安全组限制了相关端口的访问（E-MapReduce 创建的安全组默认只开放 22 端口）E-MapReduce，所以需要给 HBase 集群的安全组增加安全组规则开放端口给 Hive 集群，如下图所示：

内网入方向 | 内网出方向

授权策略	协议类型	端口范围	授权类型	授权对象	优先级	操作
允许	TCP	16000/16000	地址段访问	192.168.0.0/16	1	克隆 | 删除
允许	TCP	16020/16020	地址段访问	192.168.0.0/16	1	克隆 | 删除
允许	TCP	2181/2181	地址段访问	192.168.0.0/16	1	克隆 | 删除
允许	TCP	22/22	地址段访问	0.0.0.0/0	1	克隆 | 删除

自有 IDC 专线接入（访问 VPC 中 EMR 集群）

详见自行专线接入访问 VPC。

2.15 MetaService

E-MapReduce 环境下提供 MetaService 服务。基于此服务，您可以在 E-MapReduce 集群中以免 AK 的方式访问阿里云资源。

默认应用角色

默认地，您在创建集群时将需要向 E-MapReduce 服务授权一个应用角色（AliyunEmrEcsDefaultRole）。授权之后，您在 E-MapReduce 上的作业将可以无需显式输入 AK 来访问阿里云资源。AliyunEmrEcsDefaultRole 默认授予以下权限策略：

```
{
  "Version": "1",
  "Statement": [
    {
      "Action": [
        "oss:GetObject",
        "oss:ListObjects",
        "oss:PutObject",
        "oss:DeleteObject",
        "oss:ListBuckets",
        "oss:AbortMultipartUpload"
      ],
      "Resource": "*",
      "Effect": "Allow"
    }
  ]
}
```

所以默认情况下，基于 MetaService 的作业将只能访问 OSS 数据。如果您想基于 MetaService 访问其他阿里云资源，例如 LogService 等等，则需要给 AliyunEmrEcsDefaultRole 补充授予相应的权限。以上操作需要登录RAM 控制台完成。

> **注意:**
> 当前 MetaService 服务只支持 OSS，LogService 和 MNS 数据的免 AK 操作。请谨慎编辑，删除默认角色，否则会造成集群创建失败或者作业运行失败。

自定义应用角色

大多数情况下，您只需要使用默认应用角色或者修改默认应用角色即可。E-MapReduce 同时支持您使用自定义的应用角色。在创建集群时，您既可以使用默认应用角色，也可以选择自定义应用角色。如何创建角色并授权给服务，请参考 RAM 的相关文档。

访问 MetaService

MetaService 是一个 HTTP 服务，您可以直接访问这个 HTTP 服务来获取相关 Meta 信息：例如 `curl http://localhost:10011/cluster-region`可以获得当前集群所在 Region。

当前 MetaService 支持以下几类信息：

- Region： /cluster-region
- 角色名： /cluster-role-name
- AccessKeyId：/role-access-key-id
- AccessKeySecret：/role-access-key-secret
- SecurityToken：/role-security-token
- 网络类型：/cluster-network-type

使用MetaService

基于 MetaSerivce 服务，我们可以在作业中免 AK 地访问阿里云资源，这样可以带来两个优势：

- 降低 AK 泄漏的风险。基于 RAM 的使用方式，可以将安全风险降到最低。需要什么权限就给角色授予什么权限，做到权限最小化。
- 提高用户体验。尤其在交互式访问 OSS 资源时，可以避免写一长串的 OSS 路径。

下面示例几种使用方式：

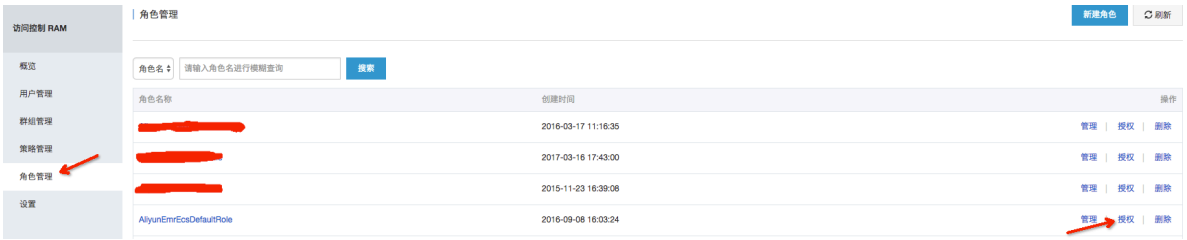
```
I. Hadoop 命令行查看 OSS 数据
    旧方式:  hadoop fs -ls oss://ZaH******As1s:Ba23N**************
sdaBj2@bucket.oss-cn-hangzhou-internal.aliyuncs.com/a/b/c
    新方式:  hadoop fs -ls oss://bucket/a/b/c
II. Hive建表
    旧方式:
        CREATE EXTERNAL TABLE test_table(id INT, name string)
        ROW FORMAT DELIMITED
        FIELDS TERMINATED BY '/t'
        LOCATION 'oss://ZaH******As1s:Ba23N**************sdaBj2@bucket
.oss-cn-hangzhou-internal.aliyuncs.com/a/b/c';
    新方式:
        CREATE EXTERNAL TABLE test_table(id INT, name string)
        ROW FORMAT DELIMITED
        FIELDS TERMINATED BY '/t'
        LOCATION 'oss://bucket/a/b/c';
III. Spark
    旧方式:  val data = sc.textFile("oss://ZaH******As1s:Ba23N
**************sdaBj2@bucket.oss-cn-hangzhou-internal.aliyuncs.com/a/b/
c")
    新方式:  val data = sc.textFile("oss://bucket/a/b/c")
```

支持MetaService的数据源

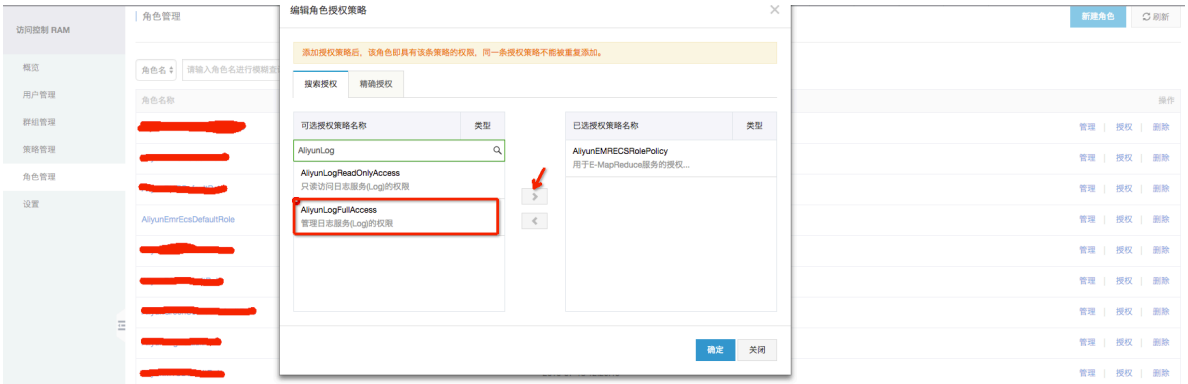
目前在 E-MapReduce 上支持 MetaService 的有 OSS，LogService 和 MNS。您可以在 E-MapReduce 集群上使用 E-MapReduce SDK 的接口免 AK 来读写上述数据源。需要强调的

是，MetaService 默认只有 OSS 的读写权限，如果您希望 MetaService 支持 LogService 或者 MNS，请前往 RAM 控制台修改 AliyunEmrEcsDefaultRole，增加 LogService 和 MNS 的权限。具体如何配置角色的权限，参考 RAM 的相关文档说明。下面将举例说明如何在给 AliyunEmrEcsDefaultRole 添加访问 LogService 的权限：

- 在 RAM 的角色管理中找到 AliyunEmrEcsDefaultRole，并单击授权。



- 搜索找到 AliyunLogFullAccess 权限，并添加。



- 添加完之后，我们就可以在 AliyunEmrEcsDefaultRole 的角色权限策略中看到已经添加的 AliyunLogFullAccess 权限策略。

这样，我们就可以在 E-MapReduce 集群中免 AK 访问 LogService 数据了。

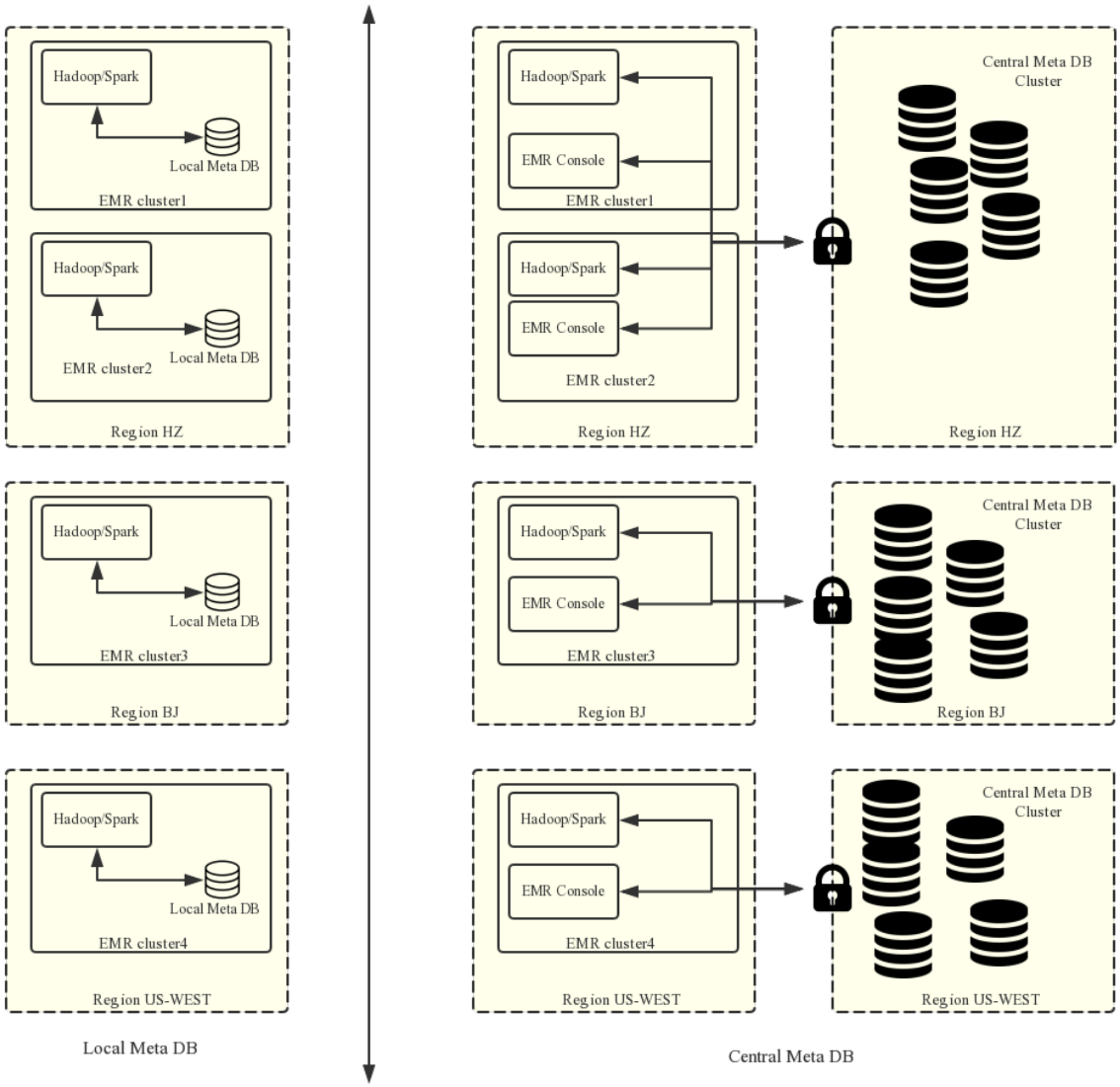
> **说明:**
> 这里演示了添加 AliyunLogFullAccess 权限策略，权限比较大，建议根据自己的实际需求，自定义一个权限策略，然后再授权给 AliyunEmrEcsDefaultRole。

2.16 元数据管理

2.16.1 Hive 元数据管理

从 E-MapReduce-2.4.0（以下简称 EMR） 版本开始，E-MapReduce 支持了统一元数据管理，在 E-MapReduce-2.4.0 版本之前，用户所有集群均采用的是集群本地的 mysql 数据库作为 Hive 元数据库，在 E-MapReduce-2.4.0 版本以及之后的版本中， E-MapReduce 会支持统一的高可靠的 Hive 元数据库。

介绍



有了统一的元数据管理之后，就可以实现：

- 提供持久化的元数据存储

  之前元数据都是在集群内部的 mysql 数据库，元数据会随着集群的释放而丢失，特别是 EMR 提供了灵活按量模式，集群可以按需创建用完就释放。如果用户需要保留现有的元数据信息，必须登录上集群手动将元数据信息导出。支持统一的元数据管理之后，不再存在该问题。

- 能更方便地实现计算存储分离

  EMR 上可以支持将数据存放在阿里云 OSS 中，在大数据量的情况下将数据存储在 OSS 上会大大降低使用的成本，EMR 集群主要用来作为计算资源，在计算完成之后机器可以随时释放，数据在 OSS 上，同时也不用再考虑元数据迁移的问题。

- 更方便地实现数据共享

  使用统一的元数据库，如果用户的所有数据都存放在 OSS 之上，则不需要做任何元数据的迁移和重建所有集群都是可以直接访问数据，这样每个 EMR 集群可以做不同的业务，但是可以很方便地实现数据的共享。

> **注意:**
>
> 在支持统一元数据之前，元数据是存储在每个集群本地环境的 Mysql 数据库中，所以元数据会随着集群的释放消亡。在支持统一元数据之后，释放集群不会清理元数据信息。所以，在任何时候删除 OSS 上的数据或者集群 HDFS 上的数据（包括释放集群操作）的时候，需要先确认该数据对应的元数据已经删除（即要 drop 掉数据对应的表和数据库）。否则元数据库中可能出现一些脏的元数据信息。

创建使用统一元数据的集群

- 页面方式

  创建集群时，在基础配置页面打开统一 meta 数据库开关。

- API 方式

  参考 API 文档中的 *CreateClusterV2* 定义，请指定此选项：useLocalMetaDb=false。

元数据基本操作

- 前提条件

  EMR 控制台统一元数据管理界面，当前只对使用统一元数据的集群生效，且必须有活跃中的统一元数据集群。

- 数据库操作

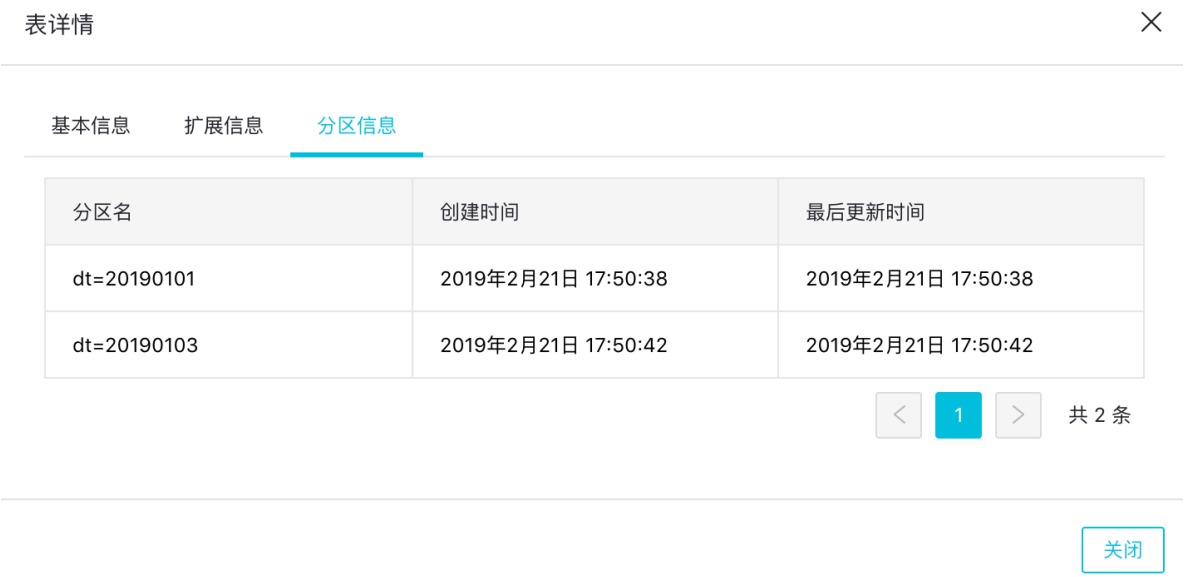
  可根据库名查找库，单击库名可查看当前库的所有表。

> **注意:**
>
> 删除数据库之前必须删除数据库下所有表。

- **表操作**

表创建：当前仅支持外部表创建和分区表创建。分隔符不仅支持普通的逗号、空格等字符，还支持四种特殊字符：TAB、^A、^B 和 ^C，或者自定义分隔符。

- **表详情**

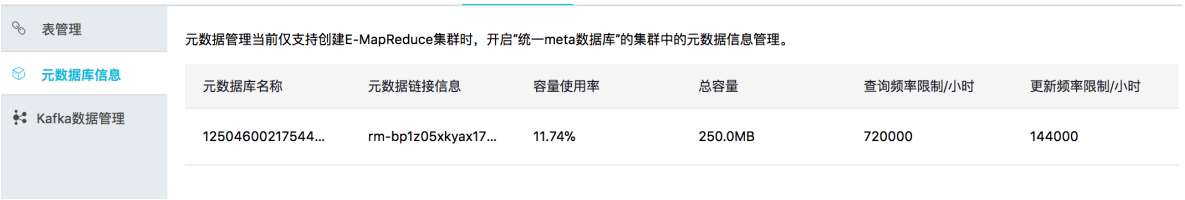


分区查看：对于分区表，可以在表详情 > 分区信息页面查看分区列表，最多显示 10000 个分区。

注意:

- **如果没有任何 EMR 集群，不支持进行对数据库和表的创建和删除操作。**
- **由于 HDFS 是每个集群内部文件系统，在没有进行特殊的网络环境设置的情况下，不同集群之间 的HDFS 无法相互访问的，所以 EMR 表管理功能对数据库和表的创建只支持基于 OSS 文件系统。**
- **数据库和表的 location 都不能选择整个 OSS bucket，需要选择到 OSS bucket 下面的目录。**

- 元数据库信息



可以在元数据库信息页查看当前 RDS 的用量和使用限制，如需修改，请提交工单联系我们。

迁移

- 从 EMR 统一元数据库迁出到用户自建 RDS

1. 购买 RDS 实例，并保证 RDS 可以和集群的 master 节点网络互通，建议跟E-MapReduce 的 ECS 在同一个安全组内，这样可以直接使用 RDS 的内网地址；

   说明：
   出于安全考虑，需要对 RDS 的 IP 白名单进行设置，设置只允许对应的 EMR 机器可以访问；

2. 在 RDS 中创建一个 Database，命名为 hivemeta，同时创建一个用户，把 hivemeta 的读写权限赋给这个用户；
3. 导出统一元数据库的内容（只导出数据，不用导表结构），为了保证数据的一致性，在执行这一步操作之前，最好将 hive 的 metastore 服务停掉，保证导出期间不会有新的 meta 数据变化。
4. 登录阿里云 E-MapReduce 控制台，单击上方集群管理页签进入集群管理页面。
5.
6. 单击对应集群 ID 进入集群与服务管理页面，在服务列表中依次单击Hive > 配置进入 Hive 管理页面，找到以下几个配置项的值（如果是不支持配置管理的老版本集群，在 $HIVE_CONF_DIR/hive-site.xml 中查找）：

```
javax.jdo.option.ConnectionUserName //对应数据库用户名；
javax.jdo.option.ConnectionPassword //对应数据库访问密码；
```

```
javax.jdo.option.ConnectionURL //对应数据库访问地址和库名;
```

7. 使用 *SSH* 方式登录到集群 master 节点，执行以下命令：

```
mysqldump -t hivemeta -h <统一元数据库地址> -u <统一元数据库用户名> -
p > /tmp/metastore.sql
```

密码就是上面提到的配置中的对应数据库访问密码。

8. 修改集群 master 节点（如果是 HA 集群则两个 master 都需要）上的*/usr/local/emr/emr-agent/run/meta_db_info.json*，将文件中的use_local_meta_db 设置为 false，meta 数据库信的链接地址、用户名和密码设置成 RDS 的信息。
9. 在 Hive 管理页面，把 meta 数据库信的链接地址、用户名和密码换成新 RDS 的信息；如果是不支持配置管理的老版本集群，修改 $HIVE_CONF_DIR/hive-site.xml 中的对应的配置为需要连接的数据库。
10. 在一台 master 节点上，把 *hive-site.xml*中的 meta 数据库链接地址、用户名和密码换成 RDS 的信息，然后执行 init schema：

```
cd /usr/lib/hive-current/bin
./schematool -initSchema -dbType mysql
```

11. 将之前导出来的 meta 数据导入 RDS。命令行登陆 mysql：

```
mysql -h {rds的url} -u {rds的用户名} -p
```

进入 mysql 的命令行之后，执行source /tmp/metastore.sql 正常情况可以完全导入，不会报错。

12. 在集群与服务管理，重启 Hive 所有组件，单击 RESTART ALL COMPONENTS。 验证 Hive功能是否正常，可以在 master 节点上，执行hive cli，查看数据是否与之前一致。

常见问题

- Wrong FS： *oss://yourbucket/xxx/xxx/xxx* 或 Wrong FS： *hdfs://yourhost:9000/xxx/xxx/xxx*

出现这个问题，是由于删除 OSS 上的表数据之前，没有删除数据表对应的元数据。导致表的 schema 还在，但实际的数据已不存在或已移动到别的路径。可以先修改表的 location 为一个存在的路径，然后再删除表：

```
alter table test set location 'oss://your_bucket/your_folder'
```

该操作可以直接在EMR控制台的交互式控制台中完成：

> 说明：
> *oss://your_bucket/your_folder*必须是一个存在的 OSS 路径。

- 删除 hive database 的时候出现：java.lang.IllegalArgumentException: java.net.UnknownHostException: xxxxxxx

  出现该问题的原因，是因为在之前的集群之上创建了 Hive 的数据库，并且数据库的位置是落在之前集群的 HDFS 之上，但是在集群释放的时候，没有清理掉对应的 hive database，导致新建集群之后，没法访问到之前已经释放集群的 HDFS 数据。所以如果是手动创建了在 HDFS 之上的数据库和表，在释放集群的时候请记得清理。

  解决方案：

  首先，通过命令行登录到集群 master 节点上，找到 hive meta db 的访问地址和用户名密码信息，在*$HIVE_CONF_DIR/hive-site.xml*中，找到对应属性。

```
javax.jdo.option.ConnectionUserName //对应数据库用户名;
javax.jdo.option.ConnectionPassword //对应数据库访问密码;
javax.jdo.option.ConnectionURL //对应数据库访问地址和库名;
```

```
  <property>
    <name>javax.jdo.option.ConnectionUserName</name>
    <value>${ConnectionUserName}</value>
    <description>Username to use against metastore database</description>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionPassword</name>
    <value>${ConnectionPassword}</value>
    <description>password to use against metastore database</description>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionURL</name>
    <value>jdbc:mysql://${DBConnectionURL}/${DBName}?createDatabaseIfNotExist=true&amp;amp;characterEncoding=UTF-8</value>
    <description>JDBC connect string for a JDBC metastore</description>
  </property>
```

  在集群 master 节点上登录 hive meta db:

```
mysql -h ${DBConnectionURL} -u ${ConnectionUserName} -p [回车]
[输入密码]${ConnectionPassword}
```

  登录上 hive meta db 之后，修改对应 hive database 的 location 为该 region 真实存在的 OSS 路径即可：

```
mysql> show databases;
+----------------------------------------------------+
| Database                                           |
+----------------------------------------------------+
| information_schema                                 |
| xxxxxxxxx77ac43c3bd0efae77e0bf1947d45fb4c896fb99 |
+----------------------------------------------------+

mysql> use xxxxxxxxx77ac43c3bd0efae77e0bf1947d45fb4c896fb99;

mysql> select * from dbs;
+-------+-----------------------+-----------------------------------+----------+------------+-------------+
| DB_ID | DESC                  | DB_LOCATION_URI                   | NAME     | OWNER_NAME | OWNER_TYPE |
+-------+-----------------------+-----------------------------------+----------+------------+-------------+
|     1 | Default Hive database | oss://mybucket/hive/warehouse     | default  | public     | ROLE        |
|     6 | NULL                  | hdfs://dirty-hostname/warehouse   | dirty_db | NULL       | USER        |
+-------+-----------------------+-----------------------------------+----------+------------+-------------+

mysql> update dbs set DB_LOCATION_URI = 'oss://your-bucket/your-db-folder' where DB_ID = 6;
```

2.16.2 Kafka 元数据管理

Kafka 元数据管理 适用于 E-MapReduce-3.17.0 以上版本。对于新创建的 Kafka 集群，后台会自动创建 Kafka 元数据。

当前 Kafka 元数据管理支持以下功能：

- 新增与删除 Topic
- 查看 Topic 详情
- 修改 Topic 分区数与配置

> 说明：
> 由于后台回调限制，Topic 列表与 Topic 信息的更新有一定延迟。

添加 Topic

1. 打开 Kafka 数据管理首页，单击右上角添加 Topic。
2. 指定 Topic 名，分区数等必填信息，以及对应的集群，在高级配置中可单击右上角+号增加配置项。
3. 创建完成后打开任务列表查看执行结果。

删除 Topic

您可以根据 Kafka 集群（只显示活跃的集群）和 Topic 名字筛选 Topic。对于已经释放的集群，Topic 列表只展示最后一次更新的详情。可以对活跃的 Topic 执行删除和更新配置，请在任务列表中查看执行结果。

> 注意：
> 内置的 Topic（如 __consumer_offsets）不支持删除与更新。

查看 Topic 详情

在 Topic 列表页单击对应 Topic 后的详情可查看该 Topic 的摘要信息，数据分布和消费组。

- 数据分布： 根据分区数显示各个分区的状态。
- 消费组： 当前 Topic 关联的所有消费组的概况，单击名称可查看消费组对各分区的 offset 与 lag。

修改 Topic 分区数与配置

- **批量调整分区**

  1. **打开 Kafka 数据管理首页，单击右上角批量调整分区。**
  2. **在弹出的操作框中，单击选择 Topic，在一个集群下选择一到多个 Topic，单击确定。**
  3. **分别指定调整后的分区数，单击确定。**
  4. **在任务列表查看执行结果，如果操作成功，稍后控制台上的 Topic 详情将发生变化。**

3 变更配置

3.1 升级节点配置

在实际使用中，我们可能会发现我们集群中的节点，尤其是 master 节点的 CPU 或者内存不够了。

> 说明:
>
> - 只有 EMR 包年包月集群才支持升级配置。
> - 本地盘实例（如 d1，i2 实例族）不能升级配置，只能增加节点个数。
> - 非本地盘实例只支持升级配置，不支持降低配置。

节点配置升级引导

1. 在集群管理页上，找到对应的集群条目，单击详情，进入集群详情页面。
2. 单击右上角的资源变配，在下拉框中选择配置升级。
3. 修改需要升级的节点配置。
4. 单击确认，等待一段时间生成订单。
5. 支付订单。
6. 返回到集群管理页面，刷新页面确认节点信息已经变为要升配的目标规格，例如，CPU：4 核内存：16 G。在机器组信息展示版块会显示如下提示：

7. 单击升级配置已完成，重启机器组生效。如下图所示：

8. 单击确认。

> 注意:

- 重启集群会重启集群的 ECS 实例，所以重启中的 ECS 实例上的大数据服务不可用，请务必确保不影响业务的情况下操作。
- 滚动重启
  - 勾选表示一个 ECS 实例重启完成且该实例上的大数据服务全部恢复后再启动下一个 ECS 实例。每个节点重启耗时约 5 分钟。
  - 不勾选表示同时重启所有 ECS 实例。
- 只重启变配节点
  - 变配节点，指已经完成过扩容磁盘或者升级配置操作的节点（如：CORE，MASTER 等）。
  - 勾选表示只重启变配节点，未变配的节点不会被重启。例如：只是对 CORE 组的节点做了升级配置，但未对 MASTER 升级配置操作，那么只会重启 CORE 组下的 ECS 实例，不会重启 MASTER 组下的ECS实例。
  - 不勾选表示所有节点都将重启。即：集群下的所有机器都会被重启。

9. 重启过程中，在对应的机器组（例如：CORE）提示机器组重启中，如下图所示：

10.待步骤 9 的提示消失后，升级配置全部完成并生效。可以登录集群查验。

11.如果只是升级了 CPU 而没有升级内存则忽略本步骤，升级配置结束；如果升级了 内存 ，需要修改集群服务配置，使得 YARN 可以使用新增的资源；如果 CPU 和内存都升级了，需要修改以上两个配置。配置步骤如下：

a. 在集群与服务管理页面，单击YARN。

b. 修改 CPU 配置

进入服务页面，单击配置，找到配置项 `yarn.nodemanager.resource.cpu-vcores` ，如果为计算密集型，建议调整为 ECS vCPU 的 1：1 比例，如果为混合型，可以调到 1：2 的比例内。例如计算节点为 32 vCore，且为计算密集型，需求大量正则匹配，则 `yarn`

`.nodemanager.resource.cpu-vcores=32`，如果计算节点为 32 vCore，计算为混合型，则 `yarn.nodemanager.resource.cpu-vcores` 可以调整到 32-64 之间。

c. 修改内存配置

进入服务页面，单击配置，找到配置项 `yarn.nodemanager.resource.memory-mb`。把个配置项的值修改为机器内存 * 0.8，单位为 MB。例如，当前新的配置下，内存是 32G，那么这里就配置为：`yarn.nodemanager.resource.memory-mb=26214`

d. 单击页面右上角的保存，保存配置项。

e. 单击页面右上角的操作，在下拉框中选择 CONFIGURE All Components。

f. 单击查看操作历史，待上一步 CONFIGURE YARN 的任务状态为成功之后继续操作。

g. 单击页面右上角的操作，在下拉框中选择 RESTART All Components。

h. 单击查看操作历史，待上一步的 Restart YARN 的任务状态为成功之后，新增的资源即可被 YARN 使用。

3.2 磁盘扩容

如果业务发展中发现数据存储空间不够了，EMR 3.11.0 及以上版本的集群可以在集群详情页对磁盘直接进行扩容来实现。

扩容入口

在集群管理页上，找到需要扩展的集群条目，单击详情，然后在详情页的右上角单击资源变配，在下拉框中选择磁盘扩容。

选择需要进行磁盘扩容的实例组。

> **注意：**
> 确保账号余额充足，扩容数据盘自动扣款，如果余额不够时，扩容数据盘流程会中断。

扩容界面

如下图所示：

- 配置：当前实例组数据盘的配置。
- 扩容至：扩容后数据盘大小。

说明:

- 目前，只支持扩容，不支持缩容。
- 磁盘扩容目前只支持数据盘扩容。

重启集群

1. 扩容结束后，在对应的机器组（比如：CORE）会提示“扩容磁盘已完成，机器组待重启”。如下图所示：

2. 单击扩容磁盘已完成，机器组待重启。如下图所示：

注意:

重启集群会重启集群的 ECS 实例，所以重启中的 ECS 实例上的大数据服务不可用，请务必确保不影响业务的情况下操作

- 滚动重启
  - 勾选表示一个 ECS 实例重启完成且该实例上的大数据服务全部恢复后再启动下一个 ECS 实例。每个节点重启耗时约 5 分钟
  - 不勾选表示同时重启所有 ECS 实例
- 只重启变配节点
  - 变配节点，指已经完成过扩容磁盘或者升级配置操作的节点（如：CORE，MASTER 等）
  - 勾选表示只重启变配节点，未变配的节点不会被重启。例如：只是对 CORE 组的节点做了磁盘扩容，但未对MASTER扩容磁盘和升级配 置操作，那么只会重启 CORE 组下的 ECS 实例，不会重启 MASTER 组下的 ECS 实例
  - 不勾选表示所有节点都将重启。即：集群下的所有机器都会被重启

3. 重启过程中，在对应的机器组（比如：CORE）会提示“机器组重启中”，如下图所示：

4. 待步骤 3 的提示消失后，扩容数据盘全部完成并生效。可以登录集群查验。

节点数据盘扩容

EMR 3.11.0 以下版本的集群可以通过在 ECS 控制台上对磁盘进行扩容。

1. 登录 EMR 控制台，在集群列表中找到待扩容的集群，单击详情进入集群详情页面。
2. 查看集群中的待扩容的 Core 节点的 `ECS Id`，类似 `i-bp1bsithym5hh9h93xxx`。默认的情况下，在扩容的时候，请将集群中所有的节点磁盘都进行统一的扩容，保证集群内部的节点的磁盘容量都是一致的。
3. 复制 `ECS Id`，并前往 ECS 控制台。 选择左侧的实例标签，然后在搜索中选择实例 ID 并输入 ECS ID。 注意要选择相同的地域。
4. 找到对应的 ECS 节点后，单击管理，进入实例详情页面，然后点击左侧的本实例磁盘标签。
5. 扩容数据磁盘。 由于目前无法指定多个磁盘批量扩容，所以需要对每一个磁盘重复如下的扩容操作。
6. 首先在 ECS 控制台，扩容所有磁盘，并重启节点。
7. 参考 *ECS 磁盘扩容说明* 进行磁盘扩容。

> **说明:**
> 这里 umount 操作失败的时候，需要先在集群上关闭 YARN 和 HDFS 服务。另外在 Disk1 操作的时候可能会碰到 ilogtail 写日志而无法 umount 的情况。需要通过 `sudo pgrep ilogtail | sudo xargs kill` 将 ilogtail 暂时杀掉。后续可以通过重启节点来恢复 ilogtail 服务。

8. 完成以后在节点通过 `df -h` 命令，能够看到所有的磁盘都完成了扩容。

节点系统盘扩容

系统盘扩容是一个比较复杂的操作，如无必要尽量避免操作。

1. 在 EMR 控制台上单击进入待扩容的集群详情页面。
2. 查看集群中的待扩容的 Master 节点的 `ECS Id`，类似 `i-bp1bsithym5hh9h93xxx`。默认的情况下，在扩容的时候，请将集群中所有的节点磁盘都进行统一的扩容，保证集群内部的节点的磁盘容量都是一致的。

3. 复制 `ECS Id`，并前往 ECS 控制台。 选择左侧的实例标签，然后在搜索中选择实例 ID 并输入 ECS ID。 注意要选择相同的地域。
4. 找到对应的 ECS 节点后，单击管理，进入实例详情页面，然后单击左侧的本实例磁盘标签。
5. 找到系统磁盘。系统磁盘只会有一块。
6. 参考 *ECS* 系统盘扩容说明 进行系统盘扩容。

> 注意:
>
> - 对于非 HA 集群，扩容系统盘会导致集群在扩容期间不可用。
> - 扩容完成以后，因为 ECS 会做一些磁盘的处理，可能会导致节点的 `/etc/hosts` 文件变化，需要在扩容完成以后修复。另外 SSH 的免登也会被破坏，但这个不会影响服务本身，可以手动修复。

3.3 切换支付类型

在 E-MapReduce 中有 2 种支付类型，按量（后付费）和包月（预付费），一般在刚开始进行产品尝试的时候，我们都希望是按量的，少量的费用的尝试。一旦确定以后我们再转换为包月的方式。

按量转包月

现在 EMR 支持先使用按量集群测试，然后在任何时间，在集群详情页面中，单击按量转包月按钮，将集群从按量付费模式切换为包月模式。整个集群的付费模式都将切换。

包月转按量

目前不支持从包月切换到按量模式。

3.4 移除异常节点

当组成 E-MapReduce 集群的 ECS 节点异常时，如果用户不需要该节点并移除该异常节点，可以使用移除异常节点功能。

移除异常节点的操作步骤为:

1. 登录阿里云 *E-MapReduce* 控制台。
2. 单击集群管理。
3. 单击需要移出异常节点的集群 ID
4. 在左侧导航栏中单击主机列表。

5. 查询找到需要移出的实例 ID，操作栏单击移除，只有 ECS 状态为停止，或已经释放的 ECS 才会被移除。
6. 确认要移除实例，单击确定。

3.5 多机器组

使用 E-MapReduce Hadoop 集群时，需要使用不同的实例组的机型。例如：

- 部分内存型实例节点（CPU: Mem = 1 vCore: 8 GiB）用于大数据离线处理，部分计算型实例（CPU: Mem=1 vCore: 2 GiB）用于模型训练。
- 不同的机器分配给不同的部门使用，部分是内存型实例（CPU: Mem=1 vCore: 8 GiB），部分是计算型实例（CPU:Mem=1 vCore: 2 GiB）。

EMR 可以通过 *Task* 节点设立多个机器组，在不同的机器组选择不同的配置达到以上目标。具体操作步骤如下：

1. 登录阿里云 *E-MapReduce* 控制台，单击集群概览区域的前往集群列表进入集群列表页面。
2. 单击对应集群右侧的管理。
3. 在左侧导航栏中单击集群基础信息。
4. 单击右上角资源变配，在下拉框中选择扩容。
5. 选择 Task（任务实例组），单击新增机器组。
6. 设置机器组名称，选择相关配置。

说明：

机器名在一个集群中是唯一的，不能重复。

注意:

EMR Hadoop 集群 Task 最多支持 10 个机器组。

4 第三方软件

4.1 引导操作

引导操作的作用是在集群启动 Hadoop 前执行您自定义的脚本，以便安装您需要的第三方软件或者修改集群运行环境。

引导操作的作用

通过引导操作，您可以完成很多目前 E-MapReduce 集群尚未支持的操作，例如：

- 使用 yum 安装已经提供的软件。
- 直接下载公网上的一些公开的软件。
- 读取 OSS 中您的自有数据。
- 安装并运行一个服务，例如 Flink 或者 Impala，但需要编写的脚本会复杂些。

> 说明：
> 强烈建议您先用按量付费的集群来进行引导操作的测试，测试都正确以后再创建包年包月的集群。

如何使用

1. 登录阿里云 E-MapReduce 控制台。
2. 在上方选择所在的地域（Region），所创建集群将会在对应的 Region 内。
3. 单击创建集群，即会进入创建集群的操作界面。
4. 在创建集群的基础配置页面的引导操作部分，单击添加，进入添加引导操作界面。
5. 填写添加引导操作界面上的配置项，完成添加。

> 说明：
> - 您最多可以添加 16 个引导操作，它们会按照您指定的顺序在集群初始化时执行。
> - 默认会使用 root 账户执行您指定的脚本，您可以在脚本中使用 `su hadoop` 切换到 Hadoop 账户。

引导操作可能会执行失败。为方便您的使用，引导操作失败并不会影响集群的创建。集群创建成功后，您可以在集群详情页集群信息栏内的引导/软件配置查看是否有异常发生。如果有异常，您可以登录到各个节点上查看运行日志，运行日志在 `/var/log/bootstrap-actions` 目录下。

引导操作类型

引导操作有两种，一种是自定义引导操作，另一种是运行条件引导操作。两者的主要区别是运行条件引导操作只会在满足条件的节点上运行您的指定操作。

- 自定义引导操作

  自定义引导操作需要指定引导操作名称和执行脚本在 OSS 中的位置，根据需要指定可选参数。集群初始化时各个节点会下载您指定的 OSS 脚本，直接执行或者附加上可选参数执行。

  您可以在脚本中指定需要从 OSS 下载的文件。例如，添加以下脚本，将 *oss://yourbucket/myfile.tar.gz* 文件下载到本地，并解压到 */yourdir* 目录下：

  ```
  #!/bin/bash
  osscmd --id=<yourid> --key=<yourkey> --host=oss-cn-hangzhou-internal
  .aliyuncs.com get oss://<yourbucket>/<myfile>.tar.gz ./<myfile>.tar.
  gz
  mkdir -p /<yourdir>
  tar -zxvf <myfile>.tar.gz -C /<yourdir>
  ```

  说明：

  osscmd 已预安装在节点上，可以直接调用来下载文件。

  注意：

  OSS 地址 host 有内网地址、外网地址和 VPC 网络地址之分。如果用经典网络，需要指定内网地址，杭州是 *oss-cn-hangzhou-internal.aliyuncs.com*。如果用 VPC 网络，要指定 VPC 内网可访问的域名，杭州是 *vpc100-oss-cn-hangzhou.aliyuncs.com*。

  引导操作也可以通过 yum 安装额外的系统软件包，下面的例子会安装 ld-linux.so.2：

  ```
  #!/bin/bash
  yum install -y ld-linux.so.2
  ```

- 运行条件引导操作

  运行条件引导操作的执行脚本是预定义的，您只需指定名称和可选参数。运行条件引导操作必须提供可选参数，可选参数需要包括运行条件和操作命令，以空格间隔。运行条件支持 `instance.isMaster=true/false`，指定只在 Master 或者在非 Master 节点上运行。在运行条件引导操作中添加以下可选参数，指定仅在 Master 节点创建目录：

  ```
  instance.isMaster=true mkdir -p /tmp/abc
  ```

  如果需要指定多个操作命令，您可以用分号 “；” 分割多个语句，例如：`instance.isMaster=true mkdir -p /tmp/abc;mkdir -p /tmp/def`。

4.2 软件配置

Hadoop、Hive、Pig 等软件含有大量的配置，当需要对其软件配置进行修改时，就可以使用软件配置功能来实现。目前软件配置操作只能在集群启动的时候执行一次。

例如，HDFS 服务器的服务线程数目 dfs.namenode.handler.count 默认是 10，假设要加大到 50；HDFS 的文件块的大小 dfs.blocksize 默认是 128 MB，假设系统都是小文件，想要改小到 64 MB。

操作步骤

1. 登录阿里云 E-MapReduce 控制台。
2. 在上方选择所在的地域（Region），所创建集群将会在对应的地域内。
3. 单击创建集群，即会进入创建集群的操作界面。
4. 在创建集群的软件配置这一步中可以看到所有包含的软件以及对应的版本。若想修改集群的配置，可以通过软件自定义配置（可选）框选择相应的 json 格式配置文件，对集群的默认参数进行覆盖或添加。json 文件的样例内容如下：

```
[
    {
        "ServiceName":"YARN",
        "FileName":"yarn-site",
        "ConfigKey":"yarn.nodemanager.resource.cpu-vcores",
        "ConfigValue":"8"
    },
    {
        "ServiceName":"YARN",
        "FileName":"yarn-site",
        "ConfigKey":"aaa",
        "ConfigValue":"bbb"
    }
]
```

说明：

- 实际传参的 FileName，需要去掉后缀。
- 服务名需全部大写。

ConfigKey 是配置项的名称，ConfigValue 为该配置项要设置的具体的值。

各个服务的配置文件如下所示：

- Hadoop

  Filename：

  - core-site.xml
  - log4j.properties
  - hdfs-site.xml
  - mapred-site.xml
  - yarn-site.xml
  - httpsfs-site.xml
  - capacity-scheduler.xml
  - hadoop-env.sh
  - httpfs-env.sh
  - mapred-env.sh
  - yarn-env.sh
- Pig

  Filename：

  - pig.properties
  - log4j.properties
- Hive

  Filename：

  - hive-env.sh
  - hive-site.xml
  - hive-exec-log4j.properties
  - hive-log4j.properties

设置好后，确认后单击下一步。

4.3 组件参数配置

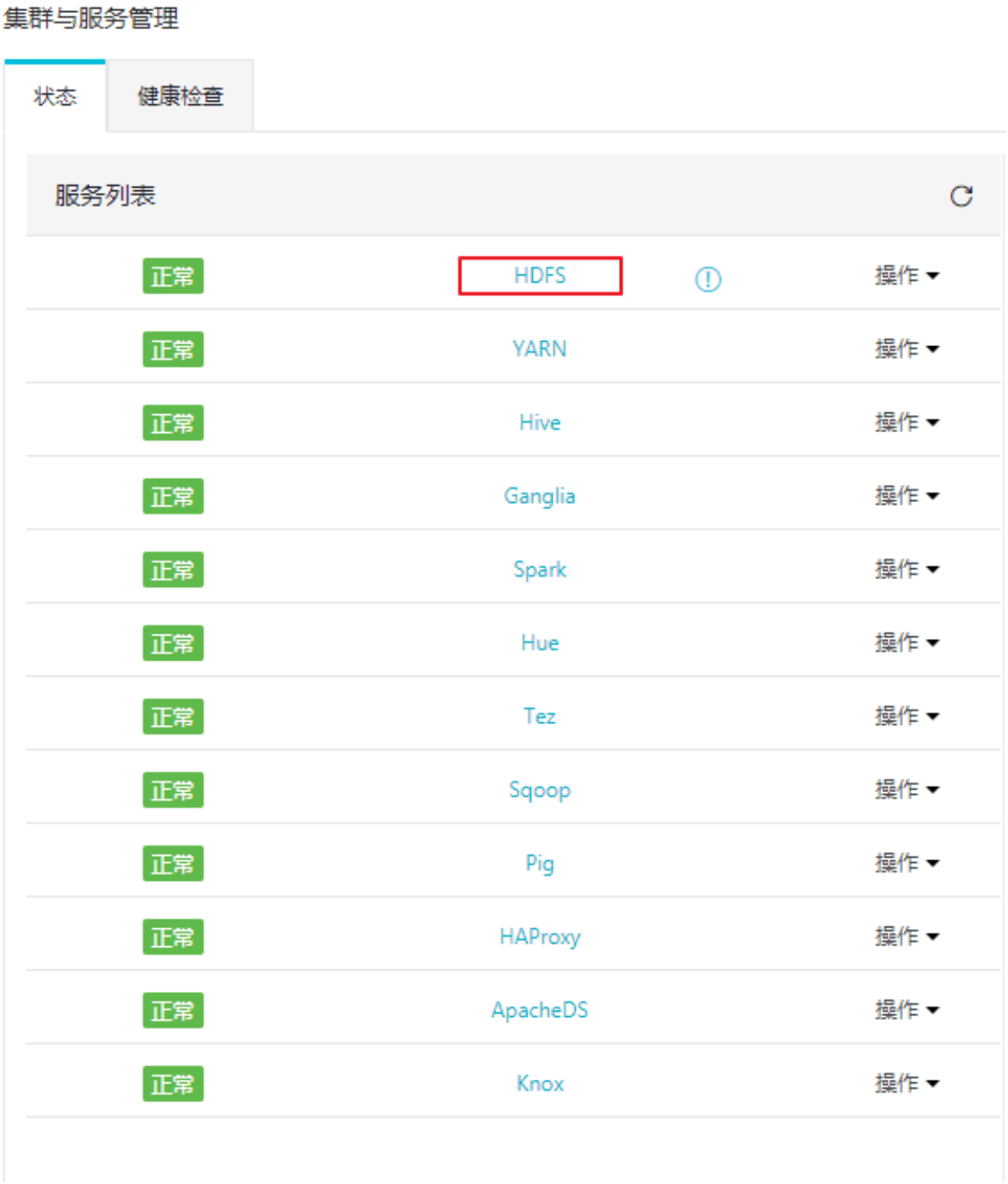
E-MapReduce 提供了 HDFS，YARN，Spark，Kafka，Druid 等组件的参数配置修改功能。

已有组件参数配置

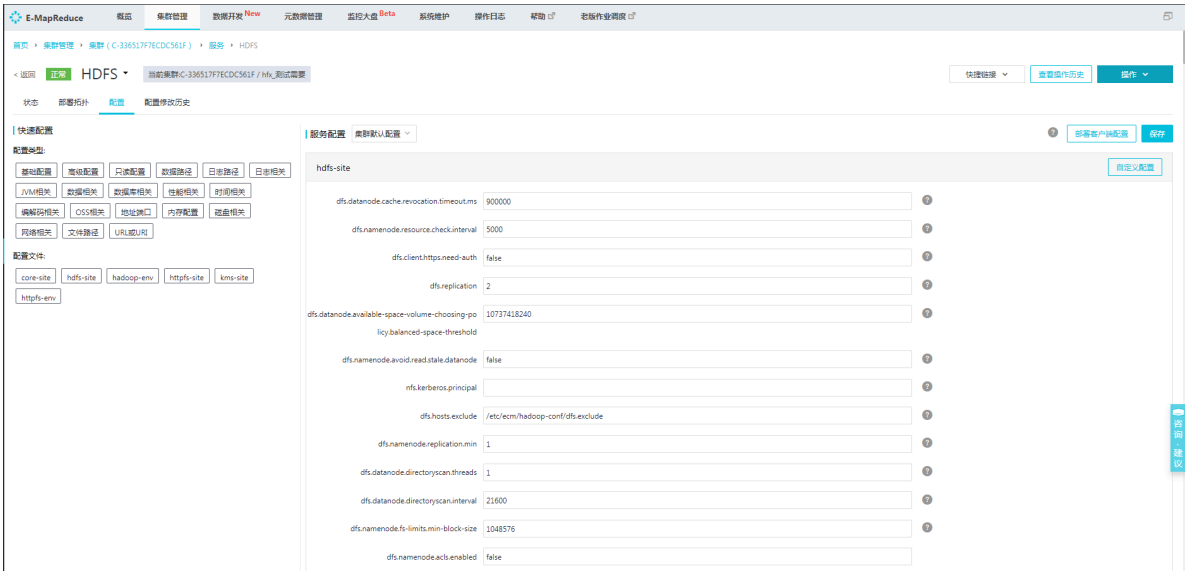
对已有组件参数的修改步骤如下：

1. 登录阿里云 E-MapReduce 控制台。

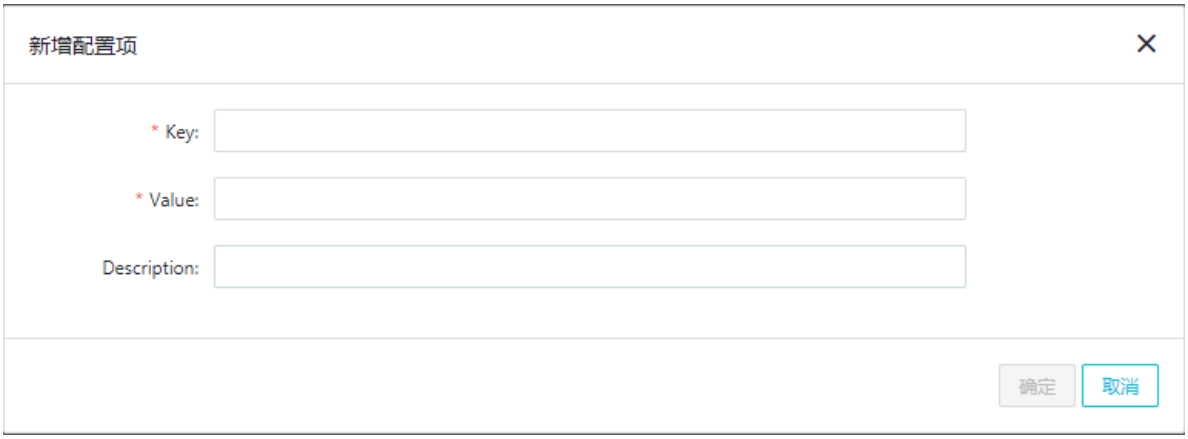
2. 单击上方的集群管理页签，进入集群列表页面。
3. 单击目标集群 ID 进入集群与服务管理页面。
4. 在服务列表中单击需要更新参数的组件，例如 HDFS。



5. 在 HDFS 服务页面中单击配置页签进入服务配置页面。



6. 您可以通过页面左侧的快速配置，选择参数的类型或配置文件。
7. 找到您要修改的参数，修改对应的值，如果没有该参数，请单击自定义配置新增配置项。



- Key：参数名称
- Value：参数值
- Description：参数描述

8. 单击保存。如果需要将修改的参数同步到与集群关联的Gateway，可以在确认修改对话框中勾选 Gateway 集群。

请参考本篇文档配置部署生效部分的操作，让您修改的参数生效。

配置部署生效

- 客户端类型配置
  1. 如果修改的参数类型为客户端类型配置，单击部署客户端配置弹出 Run Action 对话框。
  2. 执行范围，您可以选择参数同步的范围，可以将修改参数内同步到集群中的所有节点，也可以只同步选定节点。
  3. 填写执行原因，记录选择配置的原因。
  4. 单击确定，该客户端参数即可生效。
  5. 您可以单击查看操作历史查看执行状态和进度。
- 服务端类型配置
  1. 如果修改的参数类型为服务端类型配置，修改完成后，需要重启对应的服务。
  2. 单击部署客户端配置弹出 Run Action 对话框。
  3. 执行范围，您可以选择执行行操作的范围。
  4. 如果需要滚动重启，可以选择滚动重启服务。
  5. 填写执行原因，记录选择配置的原因。
  6. 单击确定，即开始滚动执行操作。
  7. 您可以单击查看操作历史查看执行状态和进度。

参数配置回滚

您可以对已经配置的参数进行回滚操作，具体操作步骤如下：

1. 登录阿里云 E-MapReduce 控制台。
2. 单击上方的集群管理页签，进入集群列表页面。

3. 单击目标集群ID进入集群与服务管理页面。
4. 在服务列表中单击需要更更新参数的组件，例如 HDFS。
5. 在 HDFS 服务页面中单击配置修改历史页签进入配置修改历史页面。
6. 在对应的记录的操作栏中单击回滚即可对参数进行回滚操作。

请参考本篇文档配置部署生效部分的操作，让回滚配置生效。