

解决方案实践

基于 OBS 配置大数据存算分离集群

文档版本 1.0.0
发布日期 2024-03-29



版权所有 © 华为技术有限公司 2024。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



HUAWEI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暗示的声明或保证。

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。

安全声明

漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：

<https://www.huawei.com/cn/psirt/vul-response-process>

如企业客户须获取漏洞信息，请参见如下网址：

<https://securitybulletin.huawei.com/enterprise/cn/security-advisory>

目录

1 方案概述	1
2 资源和成本规划	3
3 实施步骤	7
3.1 准备工作	7
3.2 快速部署	10
3.3 开始使用	14
3.4 快速卸载	20
4 附录	22
5 修订记录	23

1 方案概述

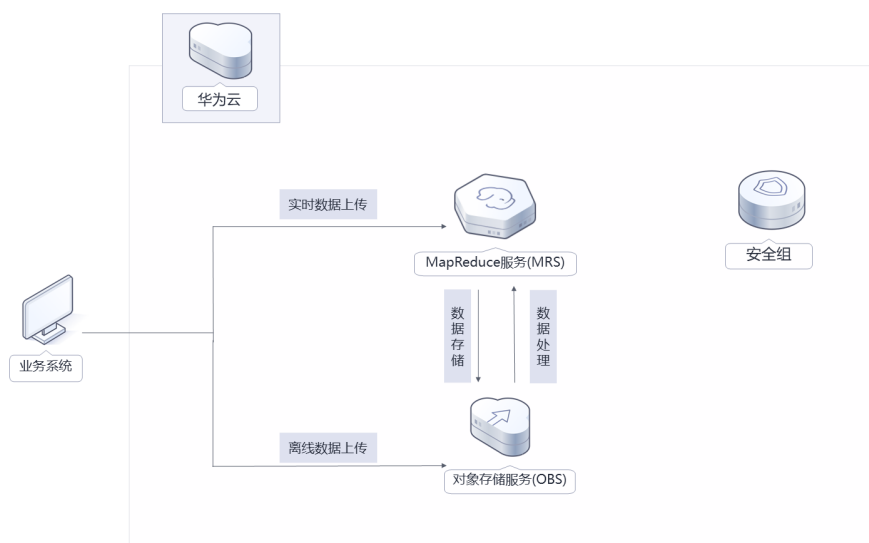
应用场景

该方案基于对象存储服务 OBS和MapReduce服务 MRS，帮助用户在华为云上快速搭建大数据存算分离集群。MRS是一个在华为云上部署和管理Hadoop系统的服务，通过为MRS集群绑定弹性云服务 ECS委托方式访问OBS，实现用户使用MRS集群作数据计算处理，而数据存储于OBS服务中。大数据在人们的生活中无处不在，在IoT、电子商务、金融、制造、医疗、能源和政府部门等行业均可以使用华为云MRS服务进行大数据处理。

方案架构

该方案基于对象存储服务 OBS和MapReduce服务 MRS，帮助用户在华为云上快速搭建大数据存算分离集群。架构如下图所示：

图 1-1 方案架构



该解决方案会部署如下资源：

- 在[对象存储服务 OBS](#)中创建一个并行文件系统，用于存储数据。
- 在[统一身份认证服务 IAM](#)创建一个委托，将账号Region内对象存储服务 OBS操作权限委托给弹性云服务器 ECS。
- 使用[MapReduce服务 MRS](#)的集群作为数据计算处理。

方案优势

- 计算存储分离
MapReduce服务 MRS集群作数据计算处理，而数据存储在对对象存储服务 OBS 中。
- 灵活弹性伸缩
MapReduce服务 MRS资源灵活配比，可以选择多种计算和存储资源进行组合，按需自动弹性伸缩，大大降低上云成本。
- 一键部署
一键轻松部署，即可完成对象存储服务 OBS并行文件系统、具有访问OBS权限的ECS委托的部署。

约束与限制

- 部署该解决方案之前，您需注册华为账号并开通华为云，完成实名认证，且账号不能处于欠费或冻结状态，如使用包周期部署确保余额充足。
- 如果选用IAM委托权限部署资源，请确保使用的华为云账号有IAM的足够权限，具体请参考[创建rf_admin_trust委托（可选）](#)；如果使用华为主账号或admin用户组下的IAM子账户可不选委托，将采用当前登录用户的权限进行部署。

2 资源和成本规划

该解决方案主要部署如下资源，不同产品的花费仅供参考，实际以收费账单为准，具体请参考华为云[官网价格](#)：

表 2-1 资源和成本规格(按需计费)

华为云服务	配置示例	每月预估花费
对象存储服务 OBS并行文件系统	<ul style="list-style-type: none">● 区域：华北-北京四● 产品类型：并行文件系统● 存储空间：数据存储（多AZ存储）0.1390元/GB/月● 请求费用：所有请求类型 0.0100元/万次● 默认存储类别：标准存储● 桶策略：私有● 流量费用：<ul style="list-style-type: none">● 内/公网流入流量（数据上传到 OBS）免费● 公网流出流量 / 00:00-08:00（闲时）0.2500元/GB● 公网流出流量 / 08:00-24:00（忙时）0.5000元/GB 费用包括存储空间、请求费用、流量费用三部分，具体请参考 OBS计费详情 。	费用包括存储空间、请求费用、流量费用三部分，详细请参考每月账单。

华为云服务	配置示例	每月预估花费
MapReduce服务 MRS	<ul style="list-style-type: none">● 按需计费：30.31元/小时● 区域：华北-北京四● 计费模式：按需计费● 集群版本：LTS集群● 集群类型：分析集群● Master节点规格：ac7.4xlarge.4 16核 64GB● Master系统盘：超高IO 480GB● Master数据盘：超高IO 600GB● Master节点数量：2个● 分析Core节点规格：ac7.4xlarge.4 16核 64GB● 分析Core系统盘：超高IO 480GB● 分析Core数据盘：超高IO 600GB● 分析Core节点数量：3个● 购买量：1	21819.60元
合计		21819.60元+OBS 费用

表 2-2 资源和成本规格(包年包月)

华为云服务	配置示例	每月预估花费
对象存储服务 OBS并行文件系统	<ul style="list-style-type: none"> ● 区域：华北-北京四 ● 产品类型：并行文件系统 ● 存储空间：数据存储（多AZ存储）0.1390元/GB/月 ● 请求费用：所有请求类型0.0100元/万次 ● 默认存储类别：标准存储 ● 桶策略：私有 ● 流量费用： <ul style="list-style-type: none"> ● 内/公网流入流量（数据上传到OBS）免费 ● 公网流出流量 / 00:00-08:00（闲时）0.2500元/GB ● 公网流出流量 / 08:00-24:00（忙时）0.5000元/GB <p>费用包括存储空间、请求费用、流量费用三部分，具体请参考OBS计费详情。</p>	费用包括存储空间、请求费用、流量费用三部分，详细请参考每月账单。
MapReduce服务 MRS	<ul style="list-style-type: none"> ● 区域：华北-北京四 ● 计费模式：包年包月 ● 集群版本：LTS集群 ● 集群类型：分析集群 ● Master节点规格：ac7.4xlarge.4 16核 64GB ● Master系统盘：超高IO 480GB ● Master数据盘：超高IO 600GB ● Master节点数量：2个 ● 分析Core节点规格：ac7.4xlarge.4 16核 64GB ● 分析Core系统盘：超高IO 480GB ● 分析Core数据盘：超高IO 600GB ● 分析Core节点数量：3个 ● 购买量：1 	16321.60元

华为云服务	配置示例	每月预估花费
合计		16321.60元+OBS费用

3 实施步骤

- 3.1 准备工作
- 3.2 快速部署
- 3.3 开始使用
- 3.4 快速卸载

3.1 准备工作

创建 rf_admin_trust 委托（可选）

步骤1 进入华为云官网，打开[控制台管理](#)界面，鼠标移动至个人账号处，打开“统一身份认证”菜单。

图 3-1 控制台管理界面



图 3-2 统一身份认证菜单



步骤2 进入“委托”菜单，搜索“rf_admin_trust”委托。

图 3-3 委托列表



- 如果委托存在，则不用执行接下来的创建委托的步骤
- 如果委托不存在时执行接下来的步骤创建委托

步骤3 单击步骤2界面中的“创建委托”按钮，在委托名称中输入“rf_admin_trust”，委托类型选择“云服务”，输入“RFS”，单击“下一步”。

图 3-4 创建委托



步骤4 在搜索框中输入“Tenant Administrator”权限，并勾选搜索结果，单击“下一步”。

图 3-5 选择策略



步骤5 选择“所有资源”，并单击“下一步“完成配置。”

图 3-6 设置授权范围



步骤6 “委托”列表中出现“rf_admin_trust”委托则创建成功。

图 3-7 委托列表



----结束

3.2 快速部署

本章节主要帮助用户快速部署该解决方案。

表 3-1 参数填写说明

参数名称	类型	是否可选	参数解释	默认值
agency_name	string	必填	委托名称，用于创建具有访问OBS权限的ECS委托。取值范围：1-64个字符的字符串。	空
obs_bucket_name	string	必填	对象存储服务 OBS 并行文件系统名称，命名格式为 {obs_bucket_name}-demo，用于存储数据，全局唯一。取值范围：1-56个字符，以小写字母或数字开头，仅支持小写字母、数字、-（中划线）。禁止以中划线（-）开头。	空

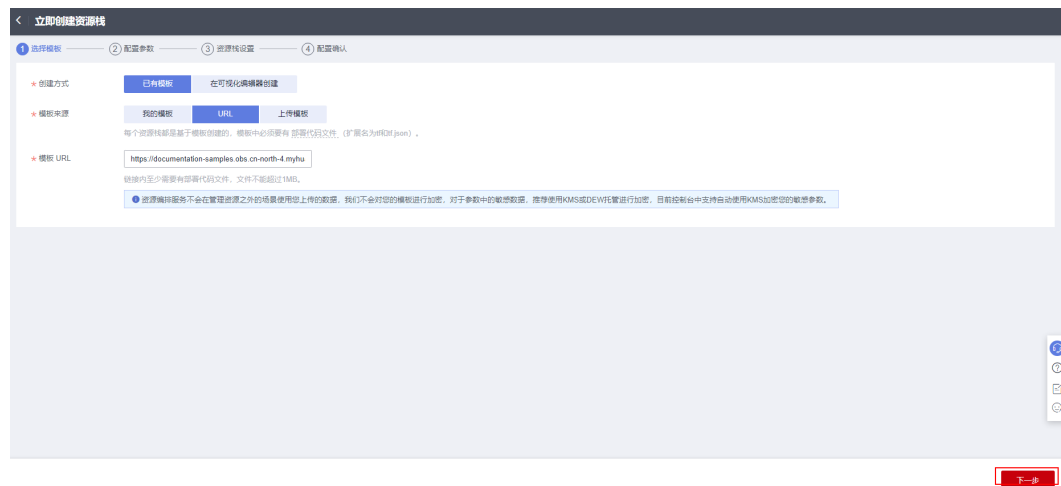
步骤1 登录[华为云解决方案实践](#)，选择“基于OBS配置大数据存算分离集群”解决方案。单击“一键部署”，跳转至解决方案创建堆栈界面。

图 3-8 解决方案实施库



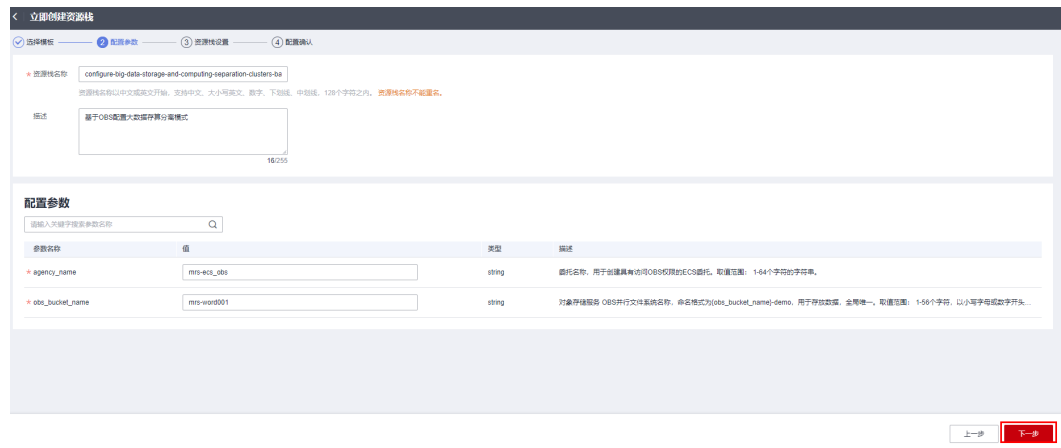
步骤2 在选择模板界面中，单击“下一步”。

图 3-9 选择模板



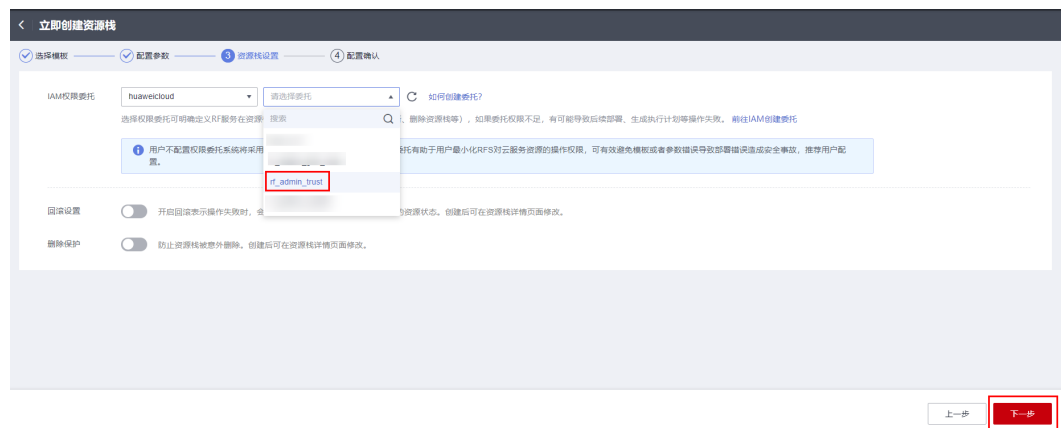
步骤3 在配置参数界面中，参考表3-1完成自定义参数填写，单击“下一步”。

图 3-10 配置参数



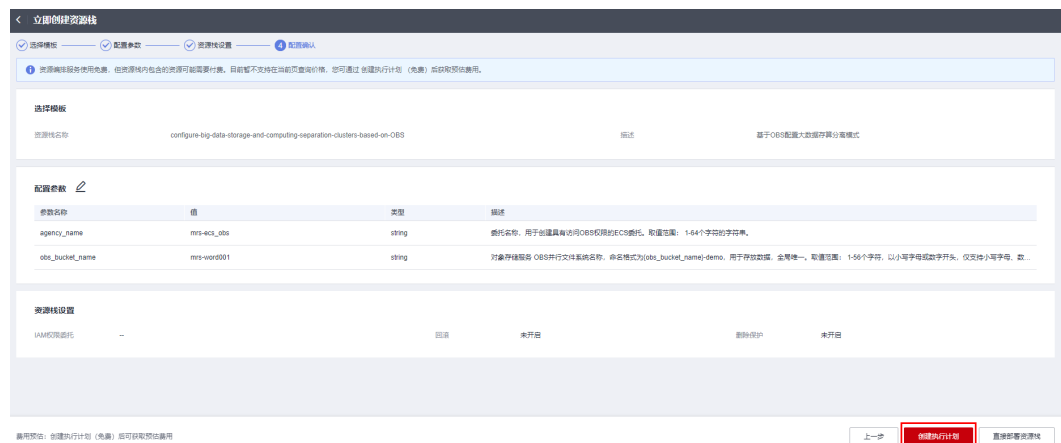
步骤4 （可选，如果使用华为主账号或admin用户组下的IAM子账户可不选委托）在资源设置界面中，在权限委托下拉框中选择“rf_admin_trust”委托，单击“下一步”。

图 3-11 资源栈设置



步骤5 在配置确认页面中，单击“创建执行计划”。

图 3-12 配置确认



步骤6 在弹出的创建执行计划框中，自定义填写执行计划名称，单击“确定”。

图 3-13 创建执行计划

创建执行计划

通过执行计划，可以预览您的资源变更信息。

* 执行计划名称: executionPlan_20240306_1632_4197

描述: 请输入对执行计划的描述

0/255

确定 取消

步骤7 待执行计划状态为“创建成功，待部署”后，单击“部署”，并且在弹出的执行计划确认框中单击“执行”。

图 3-14 执行计划

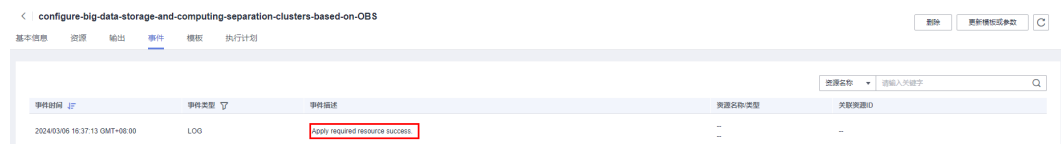
执行计划名称ID	状态	费用预估	创建时间	描述	操作
executionPlan_20240306_1632_4197	创建成功，待部署	查看费用预估	2024/03/06 16:34:36 GMT+08:00	--	部署

图 3-15 执行计划确认



步骤8 等待解决方案自动部署。部署成功后，单击“事件”，回显结果如下：

图 3-16 资源创建成功



----结束

3.3 开始使用

查看资源

步骤1 登录[统一身份认证服务 IAM](#)控制台，选择左侧列表栏中的“委托”，即可查看[3.2 快速部署](#)的委托。

图 3-17 查看创建的委托



步骤2 登录[对象存储服务 OBS](#)控制台，选择左侧列表的“并行文件系统”，即可查看[3.2 快速部署](#)的并行文件系统。

图 3-18 查看创建的并行文件系统



----结束

创建存算分离集群

说明

该示例帮助您创建MapReduce服务，在新建集群中配置现有委托实现存算分离，并开启Kerberos认证。

- MRS版本：3.2.0-LTS.1
- 集群类型：分析集群
- 有关MapReduce服务 MRS详细使用指导请参考[MRS产品文档](#)。

步骤1 登录MapReduce服务 MRS控制台，在导航栏选择“MRS集群 > 现有集群”，单击“购买集群”。

图 3-19 进入购买 MRS 集群页面



步骤2 在购买集群页面，选择“自定义购买”页签。

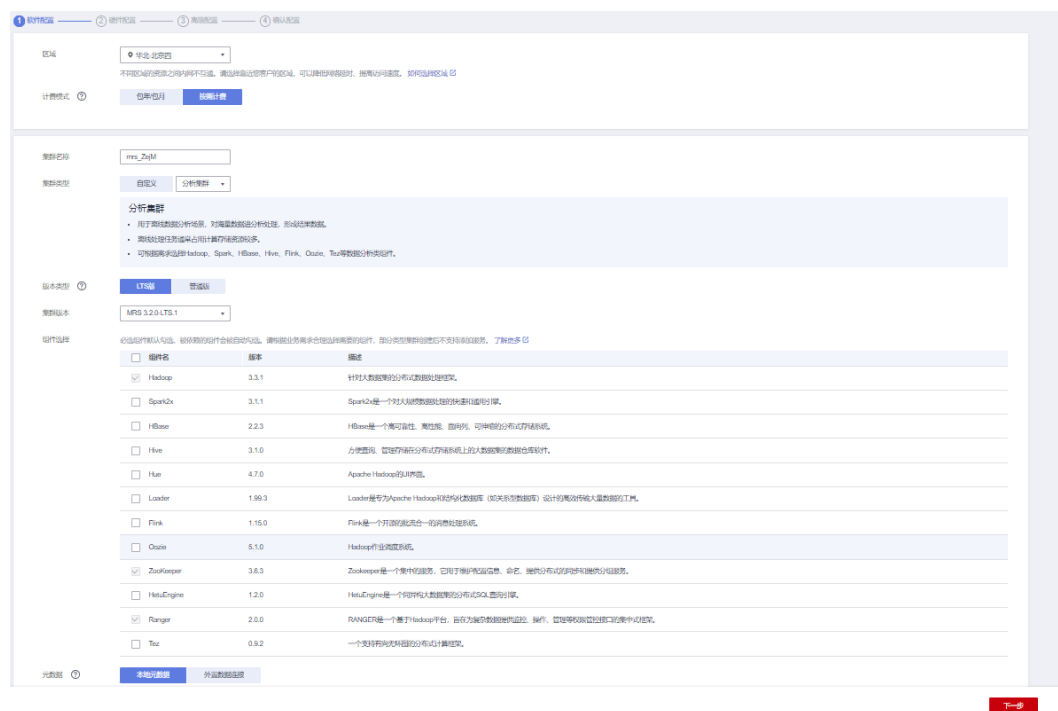
图 3-20 自定义购买



步骤3 在“自定义购买”页签，填写“硬件配置参数”。按照如下描述填写完成后单击“下一步”。

- 区域：请根据需要选择。
- 计费模式：选择“按需计费”。
- 集群名称：可以设置为系统默认名称，但为了区分和记忆，建议带上项目拼音缩写或者日期等。
- 集群版本：请选择集群版本。
- 集群类型：选择“分析集群”、“混合集群”、“流式集群”，根据业务需求勾选相应的组件。该示例选择“分析集群”。
- 元数据：选择“本地元数据”。

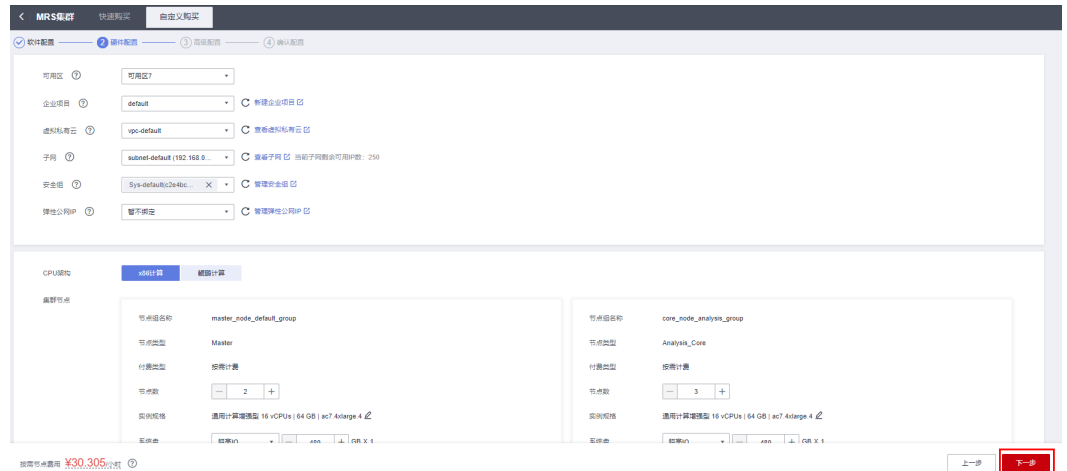
图 3-21 配置软件参数



步骤4 在“自定义购买”页签，填写“硬件配置参数”。按照如下描述填写完成后单击“下一步”。

- 可用区：默认即可。
- 虚拟私有云：默认即可。
- 子网：默认即可。
- 安全组：默认即可。
- 弹性公网IP：默认即可。
- 企业项目：默认即可。
- CPU架构：默认即可。MRS 3.x之前版本支持该参数。
- 集群节点：请根据自身需求选择节点规格和数量。

图 3-22 配置硬件参数



步骤5 在“自定义购买”页签，填写“高级配置参数”。按照如下描述填写完成后单击“下一步”。

- Kerberos认证：默认开启，请根据自身需要选择。
- 用户名：默认为“admin”，用于登录集群管理页面。
- 密码：设置admin用户密码。
- 确认密码：再次输入设置的admin用户密码。
- 登录方式：选择登录ECS节点的登录方式，本例选择密码方式。
- 用户名：默认为“root”，用于远程登录ECS机器。
- 密码：设置root用户密码。
- 确认密码：再次输入设置的root用户密码。
- 高级配置：勾选“现在配置”
- 委托：现有委托中选择图3-17

图 3-23 参数配置

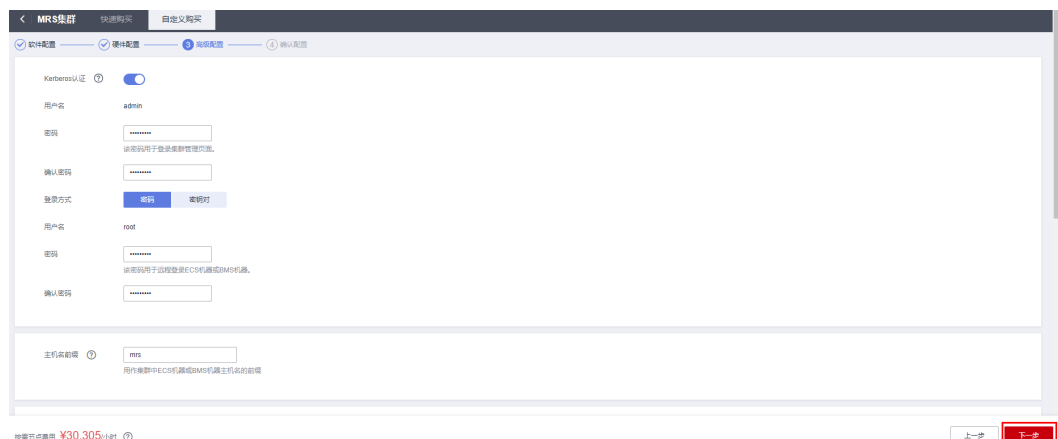
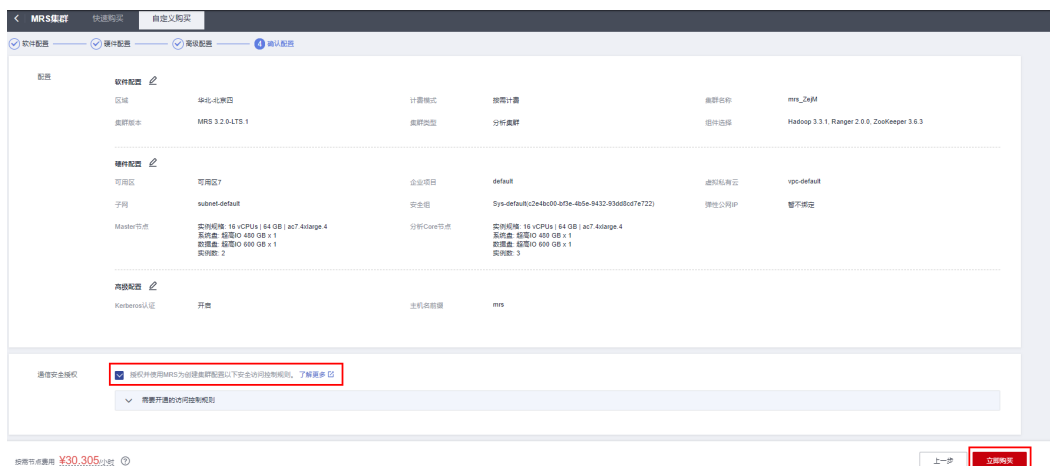


图 3-24 配置现有委托



步骤6 勾选通信安全授权，详细信息请参见[授权安全通信](#)。单击“立即购买”。单击“返回集群列表”，可以查看到集群创建的状态。集群创建需要时间，所创建集群的初始状态为“启动中”，创建成功后状态更新为“运行中”，请您耐心等待。

图 3-25 购买集群



----结束

为已有集群配置存算分离功能

步骤1 登录[MRS控制台](#)，在导航栏选择“集群列表 > 现有集群”。

图 3-26 登录 MRS 控制台



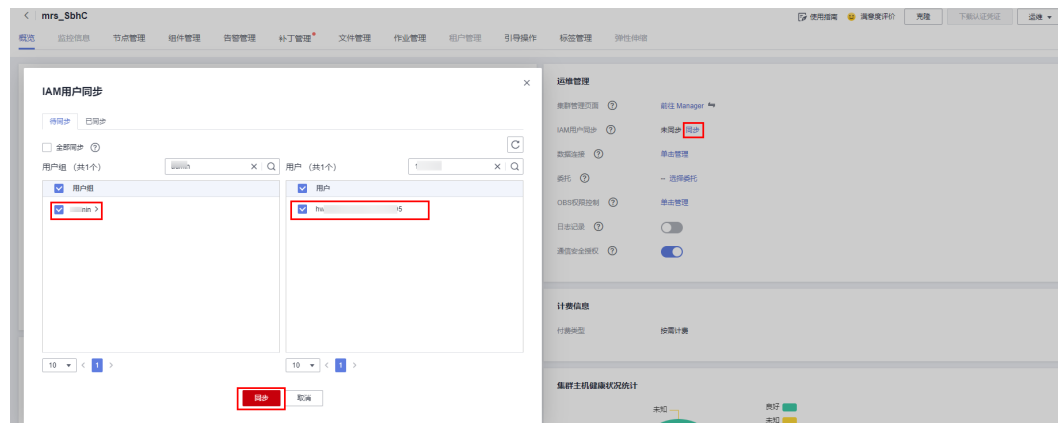
步骤2 单击集群名称，进入集群详情页面。

图 3-27 集群详情页面



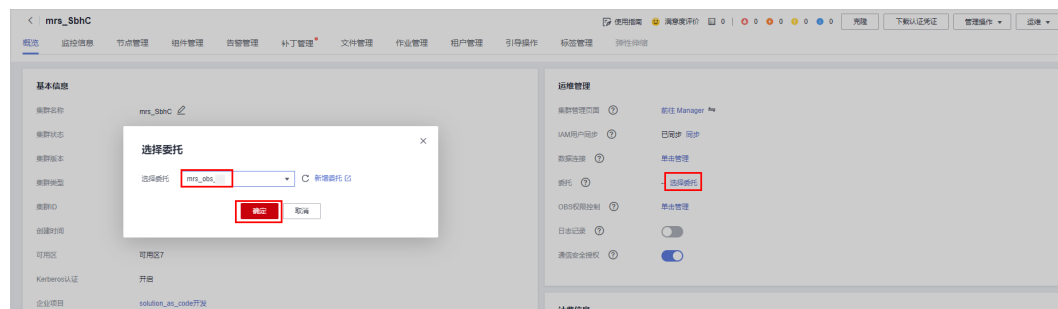
步骤3 在集群详情页的“概览”页签，单击“IAM用户同步”右侧的“同步”进行IAM用户同步，在对应用户组下选择要同步的用户，单击同步。

图 3-28 IAM 用户同步



步骤4 在集群详情页的“概览”页签，单击委托右侧的“选择委托”选择图3-17的委托并单击“确定”进行绑定，或单击“新建委托”进入IAM控制台进行创建后再在此处进行绑定。

图 3-29 配置委托



----结束

MRS 集群组件对接 OBS（基于委托）

📖 说明

该解决方案帮您仅搭建大数据存算分离集群，具体的业务使用可参考如下：

[Flink对接OBS文件系统](#)

[Flume对接OBS文件系统](#)

[HDFS客户端对接OBS文件系统](#)

[Hive对接OBS文件系统](#)

[MapReduce对接OBS文件系统](#)

[Spark2x对接OBS文件系统](#)

[Sqoop对接外部存储系统](#)

[Hudi对接OBS文件系统](#)

3.4 快速卸载

须知

快速卸载前请确认OBS桶中无文件，否则会导致删除失败。

步骤1 解决方案部署成功后，登录[资源编排服务 RFS](#)，进入“资源栈”，选择创建的资源栈名称，单击该方案堆栈后的“删除”。

图 3-30 一键卸载



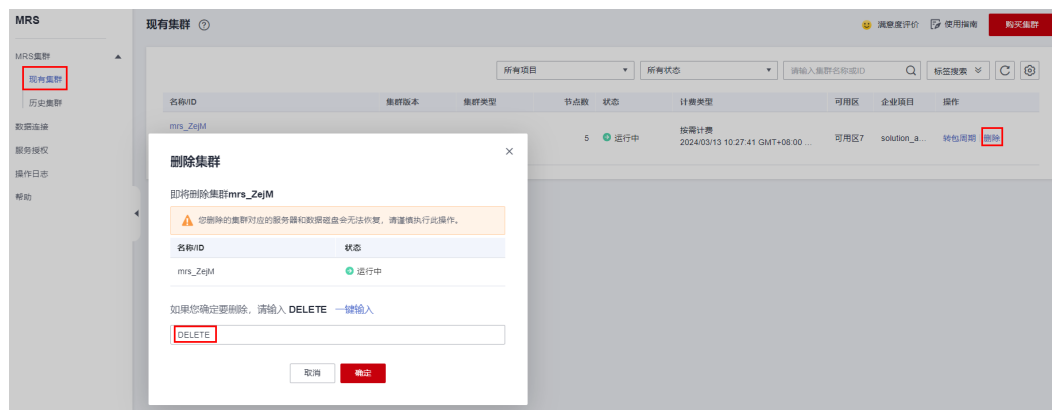
步骤2 在弹出的删除堆栈确认框中，输入Delete，单击“确定”，即可卸载解决方案。

图 3-31 删除堆栈确认



步骤3 登录MapReduce服务 MRS控制台，选择MRS集群-->现有集群，单击“删除”。

图 3-32 删除 MRS 集群



----结束

4 附录

名词解释

基本概念、云服务简介、专有名词解释

- **弹性云服务器 ECS**：是一种可随时自助获取、可弹性伸缩的云服务器，可帮助您打造可靠、安全、灵活、高效的应用环境，确保服务持久稳定运行，提升运维效率。
- **虚拟私有云 VPC**：为云服务器、云容器、云数据库等云上资源构建隔离、私密的虚拟网络环境。VPC丰富的功能帮助您灵活管理云上网络，包括创建子网、设置安全组和网络ACL、管理路由表、申请弹性公网IP和带宽等。
- **弹性公网IP EIP**：提供独立的公网IP资源，包括公网IP地址与公网出口带宽服务。
- **MapReduce服务 MRS**：提供Hudi、Doris、Spark、HBase、Flink、Clickhouse、Hadoop等开源大数据组件，支持湖仓一体、灵活的弹性控制能力。

5 修订记录

表 5-1 修订记录

发布日期	修订记录
2024-03-30	第一次正式发布。