

弹性 MapReduce

EMR on TKE 操作指南



腾讯云

【 版权声明 】

©2013–2024 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分內容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【 商标声明 】

及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【 服务声明 】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。

您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

【 联系我们 】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或95716。

文档目录

EMR on TKE 操作指南

EMR on TKE 简介

配置集群

管理权限

角色授权

创建集群

管理集群

调整 Pod 数量

变更配置

任务中心

重建

管理服务

部署说明

用户管理

新增组件

重启服务

WebUI 访问

角色管理

配置管理

配置更新

配置回滚

监控运维

监控大盘

日志

配置告警

GC 分析

应用分析

Spark 查询

Hive 查询

EMR on TKE 操作指南

EMR on TKE 简介

最近更新时间：2024-05-31 08:59:32

腾讯云 EMR on TKE 基于容器化服务、开源大数据引擎，提供可快速部署、灵活伸缩、高效安全的大数据服务部署平台。通过控制台的应用管理等功能优势使用户聚焦于业务应用。服务引擎覆盖 Spark、Hive、Trino 等大数据组件，方便用户轻松的运行、管理和扩展容器化应用程序。

产品架构



说明

- 数据存储：在 EMR on TKE 适用的存算分离场景中，提供了 COS、CHDFS、HDFS 等多种数据存储产品对接，用户可以将数据存储在以上数据源中，通过 EMR on TKE 的大数据处理引擎进行处理和分析。
- 算力资源：EMR on TKE 支持基于腾讯云 TKE 标准集群和 Serverless 集群部署外，也提供 EMR 计算资源同步购买和使用。
- 大数据组件：EMR on TKE 提供 Hive、Spark、Trino、Zookeeper、Kyuubi、Knox、Ranger、Hue、RSS 等服务可选。
- 管控平台：EMR on TKE 通过 EMR 控制台提供用户简单便捷组件部署、配置管理、运维监控、异常告警等的基础运维管理，同时也提供了高阶的作业分析诊断功能帮助用户分析作业洞察成本。

产品优势

1. 高资源利用率：EMR on TKE 容器服务可根据场景需要按照预设策略自动扩展或缩减集群容器数量，既保障服务稳定运行又可节省资源开销。在离线场景的灵活的应用资源配置有效提升资源利用率，优化成本。
2. 稳定可靠：EMR on TKE 依赖 TKE 集群的容器自检及自愈等高可靠性特性，在服务 Pod 节点出现故障时自动重建 Pod 拉起镜像。
3. 简化部署：EMR on TKE 可以在几分钟内启动一个完整的 EMR on TKE 的多服务集群。同时通过控制台的服务操作简单快速的变更服务 Pod 数量。
4. 安全精细：EMR on TKE 联动 CAM 实现集群权限访问控制，同时通过最小化存储权限对接 COS 实现存算分离场景下的精细化权限管理，保护集群使用级数据访问的安全性。

配置集群
管理权限

角色授权

最近更新时间：2023-12-14 14:53:12

使用弹性 MapReduce 服务时，用户需要为服务账号授予系统默认角色 EMR_QCSRole。当该角色授予成功后，弹性 MapReduce 才能调用相关服务（TKE、COS 等）创建集群和保存日志等。

注意

首次开通弹性 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弹性 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，若为服务账号授予 EMR_QCSRole 角色失败，会有如下提示。然后单击前往访问管理，进行角色授权。



2. 单击同意授权，将默认角色 EMR_QCSRole 授予弹性 MapReduce 的服务账号。



3. 授权完成后，用户需刷新弹性 MapReduce 的控制台或购买页，刷新后即可正常操作。更多 EMR_QCSRole 相关的详细策略信息，可登录访问管理控制台查看。EMR_QCSRole 包含的权限信息请参见协作者/子账号权限。

EMR on TKE 集群相关服务角色授权特别说明

在创建或使用 EMR on TKE 集群时，需要对 COS（对象存储）进行数据的直接写入或计算，为保证数据安全需要授予 EMR 通过临时密钥方式对 COS 资源进行读写，需要授予 EMR 服务相关角色 EMR_QCSLinkedRoleInApplicationDataAccess 且绑定 QcloudAccessForEMRLinkedRoleInApplicationDataAccess 预设策略。

1. 当进行 EMR on TKE 集群列表时，需要检查是否存在绑定 EMR 服务相关角色 EMR_QCSLinkedRoleInApplicationDataAccess。



2. 如果不存在 EMR 服务相关角色 EMR_QCSLinkedRoleInApplicationDataAccess 则需要授权绑定。

服务授权

同意赋予 [弹性MapReduce](#) 权限后，将创建服务预设角色并授予 [弹性MapReduce](#) 相关权限

角色名称 EMR_QCSLinkedRoleInApplicationDataAccess

角色类型 服务角色

角色描述 当前角色为 [弹性MapReduce](#) 服务角色，该角色将在已关联策略的权限范围内访问您的其他云服务资源。

授权策略 预设策略 QcloudAccessForEMRLinkedRoleInApplicationDataAccess ⓘ

同意授权

取消

⚠ 注意

若需要精细化指定集群访问对应的 COS 资源权限，详情请参见 [自定义服务角色](#) 设置。

EMR onTKE 集群鉴权说明

- 子账号与协作者权限设置与 EMR on CVM 版本一致，详情请参见 [协作者/子账号权限](#)。
- 标签鉴权和接口鉴权设置与 EMR on CVM 版本一致，详情请参见 [鉴权粒度方案](#)。

创建集群

最近更新时间：2024-05-31 08:59:32

操作场景

本文为您介绍通过 EMR 控制台创建一个 EMR on TKE 集群的操作步骤和相关配置。

操作步骤

登录 [EMR 控制台](#)，在 EMR on TKE 集群列表页单击**创建集群**，在购买页面完成相关配置；当集群列表中集群状态显示为运行中时，表示集群创建成功。

配置项	配置项说明
集群名称	通过设置集群名称，来区分不同的 EMR 集群，系统随机生成支持修改。
地域	<p>集群所部署的物理数据中心，每个地域（region）都指一个独立的物理数据中心，不同地域间的云服务器内网不互通。</p> <p>注意：</p> <ol style="list-style-type: none"> 1. 集群创建后，无法更改地域，请谨慎选择。 2. 仅支持该地域下的容器集群为集群提供资源。
容器类型	<ol style="list-style-type: none"> 1. 服务角色由底层容器提供资源进行部署，支持 TKE 标准集群和 TKE Serverless 集群。 2. 计算资源为 EMR 同步购买的容器部署资源，支持 EMR on TKE 集群可关联使用资源。 <p>注意：选择具体某个容器集群或计算资源为该 EMR on TKE 集群提供 Pod 资源进行服务部署。</p>
集群网络	用于购买 db 使用，保持网络与容器集群网络一致。
可用区及子网	用于购买 db 使用，保持网络与容器集群网络一致。
安全组	<p>安全组具有防火墙功能，用于设置云服务器 CVM 的网络访问控制。如果没有安全组，EMR 会自动帮您新建一个安全组。若已经有在使用的安全组可以直接选择使用。若安全组数量已达到上限无法新建，可删除部分不再使用的安全组。查看已在使用的安全组。</p> <p>创建安全组：EMR 帮助用户创建一个安全组，开启22和30001端口及必要的内网通信网段。</p> <p>已有 EMR 安全组：选择已创建的 EMR 安全组作为当前实例的安全组，开启22和30001端口及必要的内网通信网段。</p>
计费模式	仅支持按量计费模式。按量计费：按照使用时长付费，需对账户进行实名认证，在开通时需冻结2小时的费用（代金券不可用作冻结凭证），销毁时退还冻结资源费用。
产品版本	不同产品版本上捆绑的组件和组件的版本不同。
部署服务	<p>非必选组件，根据自身需求组合搭配自定义部署，最少选择一个组件服务部署；部署形式支持：默认部署和关联部署。</p> <p>关联部署：是提供当前集群的服务来源（其他云上 EMR 集群或 DLC）。</p> <p>默认部署：将当前服务部署到已选择的容器资源中，可自定义部署资源规格和数量（服务角色维度部署），推荐使用。</p>
关联部署	<p>选择 EMR on CVM EMR on TKE 集群中含有当前服务的集群。</p> <ol style="list-style-type: none"> 1. 选择 EMR on CVM 集群时，该集群必须部署当前服务。 2. 选择 EMR on TKE 集群时，该集群中当前服务必选设置可对外提供服务。 3. MetaStore（HiveMetaStore）支持选择DLC服务来源。
默认部署	<p>数据卷目录：为容器提供数据存储目录，目前仅支持使用主机路径（HostPath 挂载方式）或使用新建 PVCPOD 规格设置服务角色维度部署信息可选项。</p> <p>外部访问设置：设置当前角色是否开启对外提供服务，支持公网 LB 访问和内网 LB 访问可选项。</p> <p>节点调度策略：设置当前角色部署调度策略，可根据调度规则，将 Pod 调度到符合预期的 Label 的节点中。</p>
Hive 元数据库	<p>如果选择了 Hive 组件，Hive 元数据库提供了两种存储方式：</p> <p>第一种集群默认，Hive 元数据存储于集群独立购买的 MetaDB。</p> <p>第二种是关联外部 Hive 元数据库，可选择关联 EMR-MetaDB 或自建 MySQL 数据库，元数据将存储于关联的数据库中，不随集群销毁而销毁。可参考 Hive 元数据管理。</p> <p>注意：当选择 Hue、Ranger、Oozie、Druid、Superset 一个或多个组件时系统会自动购买一个 MetaDB 用于除 Hive 外的组件元数据存储。</p>
cos 存储桶	用于存储日志，jar 包等信息。
标签	您在创建时对集群或节点资源添加标签，以便于管理集群和节点资源，最多可绑定5条，标签键不可重复。
设置密码	设置 WebUI 密码，当前密码仅用于初始设置服务 WebUI 访问密码。

协议条款	同意《弹性 MapReduce 服务等级协议》和《退费协议》。
------	---------------------------------

管理集群

调整 Pod 数量

最近更新时间：2023-09-12 17:23:12

操作场景

集群中服务使用的资源不足或过剩时，可通过调整服务角色部署的 POD 数量改变集群负载能力。

说明

当前扩容 POD 规格为默认新建集群时选择 POD 规格，若需要调整服务资源配置，需在 [变更配置](#) 进行设置。

操作步骤

1. 登录 [EMR on TKE 控制台](#)，在集群列表中单击**集群 ID/名称**进入集群详情页。
2. 在集群详情页选择**集群服务 > 某服务**，并进入角色管理页。
3. 在角色管理页选择需要扩容/缩容的服务角色并单击**调整 POD 数量**。
4. 根据实际需求调整 POD 数量，单击**确定**即可开始调整。

变更配置

最近更新时间：2023-09-12 17:23:13

操作场景

服务角色 CPU 或内存大小不足时，可以通过变更配置提升服务角色 CPU 或内存大小。本文为您介绍通过 [EMR on TKE 控制台](#) 变更实例配置的相关操作。

⚠ 注意

- 变更配置会导致正在运行的作业运行失败，请确保在不影响业务的情况下操作。
- 变更配置过程中，资源可能被抢占使 POD 重建失败，导致服务不可用，建议在业务低谷期操作。

前提条件

1. POD 资源配置调整将自动逐条扣费，请保持账户余额充足。
2. 按量计费的节点变配后，将重新按最新配置的单价进行费用冻结，详情请参见 [EMR on TKE计费说明 > 购买说明](#)。

操作步骤

1. 登录 [EMR on TKE 控制台](#)，在集群列表中单击**集群 ID/名称**进入集群详情页。
2. 在集群详情页选择**集群服务 > 某服务**，并进入角色管理页。
3. 在角色管理页选择需要变配的服务角色并单击**变更配置**。
4. 在调整配置页，确认相关变配信息，选择目标 CPU、内存配置，并且需认真阅读重要提示并勾选同意变更信息。
5. 费用确认无误后，单击**确定**即可调整配置。

任务中心

最近更新时间：2023-11-24 10:47:01

EMR on TKE 版本集群任务详情信息与 EMR on CVM 版本一致，详情请参见 [任务中心](#)。

重建

最近更新时间：2023-11-24 10:47:01

操作场景

重建是指销毁当前所有角色维度所部署的 POD 资源，并重新购买与原有配置规格一致的 POD 资源，主要适用场景是当配置文件发生“增加或减少”时，可利用重建功能使其配置生效。

⚠ 注意：

- 重启：仅适用于配置项参数发生变化，使其生效。
- 重建：重建存在失败情况，重建失败可能当值服务角色的负载无法拉起，从而导致服务不可用，请谨慎操作，请确保在不影响业务的情况下操作。

操作步骤

1. 登录 [EMR on TKE 控制台](#)，在集群列表中单击**集群 ID/名称**进入集群详情页。
2. 在集群详情页选择**集群服务 > 服务(如：选择 Hive)**，并进入角色管理页。
3. 在角色管理页选择需要重建的服务角色并单击**重建**。
4. 在重建配置页，确认相关变配信息，选择重建方式（支持：滚动重建和快速重建），并且需认真阅读重要提示并勾选同意重建信息。
5. 单击**确定**即可重建 POD。
6. POD 列表中，POD 状态点击查看状态详情，可查看当前 POD 最近一小时且最多10条状态数据。

管理服务 部署说明

最近更新时间：2023-09-12 17:23:13

部署说明用于查看 EMR on TKE 集群支持部署的集群类型、服务角色信息、Pod 默认规格信息、以及数据卷目录，便于用户参考调整服务部署资源配置。

部署信息说明

集群类型	组件	角色	Pod 默认规格			数据卷目录
			Cpu 限制 (核)	内存限制 (GB)	Pod 数量范围	
TKE	hive	HiveMetaStore	1-100	2-100	1-100	新建 PVC, 主机路径
		HiveServer2	1-100	2-100	1-100	新建 PVC, 主机路径
	hue	Hue	2-100	4-100	1-100	新建 PVC, 主机路径
	kyuubi	KyuubiServer	1-100	2-100	2-100	新建 PVC, 主机路径
	ranger	Ranger	1-100	2-100	1-100	新建 PVC, 主机路径
		RangerUsersync	1-100	2-100	1-100	新建 PVC, 主机路径
	rss	Coordinator	1-100	2-100	2-100	新建 PVC, 主机路径
		ShuffleServer	2-100	4-100	3-100	新建 PVC
	spark	SparkJobHistoryServer	1-16	1-64	1-2	新建 PVC, 主机路径
	trino	TrinoCoordinator	1-100	2-100	1-2	新建 PVC, 主机路径
		TrinoWorker	1-100	2-100	2-100	新建 PVC, 主机路径
	zookeeper	QuorumPeerMain	1-100	2-100	3-100	新建 PVC
	knox	Knox	1-100	2-100	1-100	新建 PVC, 主机路径
TKE Serverless	hive	HiveMetaStore	1-100	2-100	1-100	新建 PVC
		HiveServer2	1-100	2-100	1-100	
	hue	Hue	2-100	4-100	1-100	
	kyuubi	KyuubiServer	1-100	2-100	2-100	
	ranger	Ranger	1-100	2-100	1-100	
		RangerUsersync	1-100	2-100	1-100	
	rss	Coordinator	1-100	2-100	2-100	

		ShuffleServer	2-100	4-100	3-100
	spark	SparkJobHistoryServer	1-16	1-64	1-2
	trino	TrinoCoordinator	1-100	2-100	1-2
		TrinoWorker	1-100	2-100	2-100
	zookeeper	QuorumPeerMain	1-100	2-100	3-100
	knox	Knox	1-100	2-100	1-100

注意

集群中服务角色 Pod 默认规格的最大值，将结合所选容器集群可用资源动态调整，实际最大值可能小于表中给定的最大值。

用户管理

最近更新时间：2023-11-24 10:47:01

在 EMR on TKE 版本下，集群部署 OpenLdap 服务时支持用户管理，用户管理相关操作和查看可参考 EMR on CVM 版本的 [用户管理](#)。

新增组件

最近更新时间：2023-09-12 17:23:13

集群创建后，可以通过新增组件操作增加部署当前选定产品版本中支持的其它组件，本文为您介绍通过控制台新增组件的操作。

注意

新增组件仅支持选择当前 EMR 产品版本的组件。

操作步骤

1. 登录 [EMR on TKE 控制台](#)，在集群列表中单击**集群 ID/名称**进入集群详情页。
2. 选择**集群信息 > 基础配置**，在服务架构图中选择**新增组件**，新增集群中未安装的组件。
3. 当集群无元数据库，在安装 Hue、Ranger 组件时，需要购买一个云数据库实例存储单元为元数据存储地。
4. 其中 Hive 组件元数据存储提供了两种存储方式：第一种集群默认，元数据存储于 MetaDB；第二种是关联，关联 EMR-MetaDB 或关联自建 MySQL 数据库，元数据将存储于关联的数据库中，不随集群销毁而销毁。
5. 是否购买 MetaDB 与新购集群时一致，Hive 元数据库选择也与新购集群时一致。
6. 勾选组件后，单击**创建**即可。

重启服务

最近更新时间：2023-09-12 17:23:13

操作场景

集群中服务角色升级更新或修改配置后，需要重启对应的服务使配置生效。本文为您介绍通过控制台重启服务的操作。

操作步骤

1. 登录 [EMR on TKE 控制台](#)，在集群列表中单击**集群 ID/名称**进入集群详情页。
2. 若您需要重启整个组件服务，可以选择**集群服务>某服务**，在该页面导航栏右侧选择**操作 > 重启服务**。当您选择服务角色为 ALL 时，将重启整个服务。



3. 若您需要重启该服务的某个角色类型，可以选择**集群服务 > 某服务**，在角色管理页勾选需要重启的服务角色并单击**重启服务**。
4. 确认无误后，单击**确定**，即可重启服务。

WebUI 访问

最近更新时间：2023-09-12 17:23:13

软件 WebUI 入口功能是 EMR 提供的组件原生 UI 访问能力，通过 Knox 组件即可在公网快捷访问组件原生 UI。

前提条件

集群安装了 Knox 组件，且 Knox 组件开启了公网访问。

操作步骤

1. 登录 [EMR on TKE 控制台](#)，在集群列表中单击**集群 ID/名称**进入集群详情页。
2. 根据需要选择**集群服务 > 某服务**，在该页面导航栏右侧选择**查看信息 > 查看WebUI**即可访问。

说明

- 访问地址需要进行身份验证，用户名为 root，默认密码为创建集群时输入的密码。
- 集群创建时，如果没有安装 Knox 组件，或 Knox 组件未开启公网访问，将不能通过组件管理页面的原生 WebUI 访问地址进入相关组件的 WebUI 界面。

角色管理

最近更新时间：2023-12-14 14:53:12

角色管理提供了角色服务级别和具体的 POD 维度的运维操作功能，如重启操作。POD 健康状态支持监控，以便于了解角色进程实时状态。

操作步骤

1. 登录 [EMR on TKE 控制台](#)，在集群列表中单击**集群 ID/名称**进入集群详情页。
2. 在集群详情页中选择**集群服务**，单击对应服务进入角色管理页。
3. 角色列表展示当前服务角色的 POD 健康状态、运行/期望 POD 数、资源配置、最近重启时间等信息。POD 列表展示 POD 资源的健康状态、POD 状态、POD IP、重启次数、最近重启时间等信息。

① 说明

POD 状态为 Running 和 Succeeded 时，POD 健康状态显示为正常；POD 状态为 Pending、Failed、Unknown，或服务对应的 Container not ready 时，POD 健康状态显示为异常。

配置管理

配置更新

最近更新时间：2024-02-27 14:30:32

操作场景

配置管理支持 HIVE、SPARK 等常用开源组件的关键配置参数的修改，可以根据实际需要选择不同角色对服务的配置进行修改。本文为您介绍通过控制台配置参数的操作。

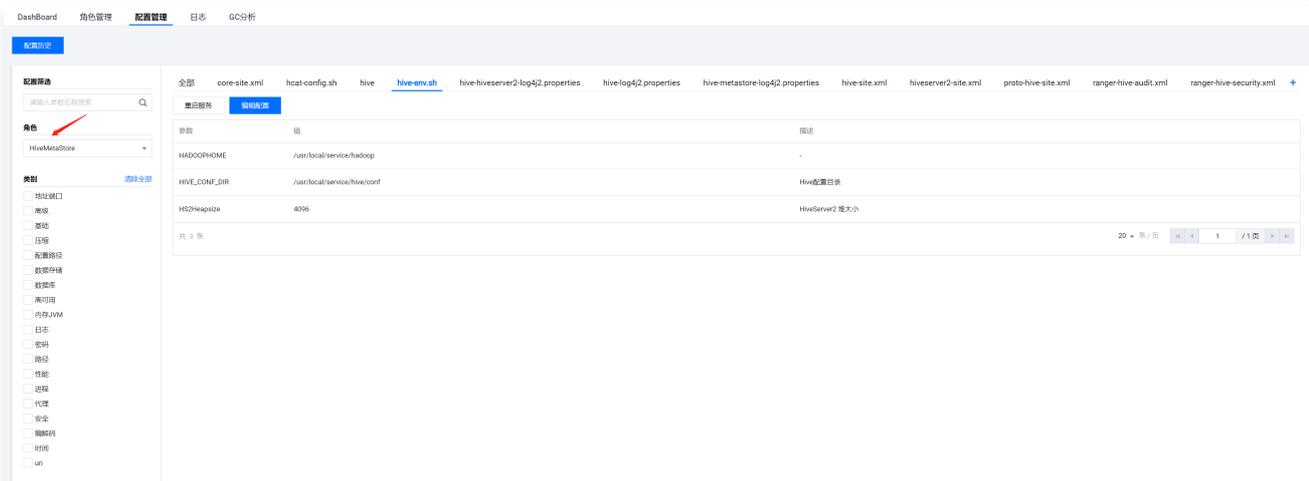
注意

在控制台配置管理中删除自定义配置文件，基于安全性考量，客户端不会同步删除动作。

操作步骤

编辑配置项

1. 登录 [EMR on TKE 控制台](#)，在集群列表中单击**集群 ID/名称**进入集群详情页。
2. 在集群详情页中选择**集群服务**，单击对应服务进入**配置管理**页。
3. 进入配置管理页后，根据需要选择角色。



4. 若您想搜索某个配置项或者缩小配置项查找范围，可通过左侧筛选器进行过滤。
5. 根据需要选择配置文件，单击**编辑配置**进入编辑状态，根据需要进行新增、编辑、删除配置项操作。
 - 选择需要修改的参数填入新的参数值，如有需要可单击**复原**恢复为原始值，也可单击**默认值**恢复到系统推荐默认值。
 - 部分参数支持删除操作，如需删除该配置，选择**删除** > **确认**即可。
 - 若该文件下没有您想要配置参数，可单击**新增配置项**打开新增配置项弹框页面，填写新增配置项的参数名、参数值。
6. 确认无误后单击**保存配置**，配置下发成功后单击**重启服务**，修改配置项设置完成。

说明

- 若修改的是服务进程配置，保存后需要重启服务才可配置生效。
- 若修改的是客户端配置，保存后无需重启服务即可生效。

新增配置文件

1. 登录 [EMR on TKE 控制台](#)，在集群列表中单击**集群 ID/名称**进入集群详情页。
2. 在集群详情页中选择**集群服务**，单击对应服务进入**配置管理**页。
3. 若没有您想要配置的文件，可单击右侧“+”（新增配置文件）进入配置文件配置页面，填写配置文件。
4. 单击**保存配置**后，参数下发并更新配置文件列表中的配置文件名。
5. 自定义配置文件下发生效后支持修改和删除操作。

配置回滚

最近更新时间：2023-09-12 17:23:13

弹性 MapReduce 支持在控制台对各服务参数新增、修改、删除配置项等操作进行配置回滚，本文为您介绍如何通过控制台回滚各服务参数配置。

操作步骤

1. 登录 [EMR on TKE 控制台](#)，在集群列表中单击**集群 ID/名称**进入集群详情页。
2. 在集群详情页中选择**集群服务**，单击对应服务进入配置管理页。
3. 在配置管理页选择**配置历史**，单击**详情**可以看到配置变更前后的参数值对比，单击**回滚**可对该条记录的参数配置变更回滚。选择**回滚 > 确认回滚**，回滚成功重启服务，稍作等待回滚即可生效。

① 说明

新增、修改、删除配置项支持回滚，新增配置文件和删除新增配置文件不支持回滚。

监控运维

监控大盘

最近更新时间：2023-09-12 17:23:13

提供 EMR on TKE 监控策略配置、告警历史记录功能，同时可查看集群服务监控指标及历史趋势。

监控指标

EMR on TKE 提供服务监控指标如下：

服务	监控指标说明
Hive	Hive 监控指标
Spark	Spark 监控指标
Trino	Trino 监控指标
Hue	-
Kyuubi	Kyuubi 监控指标
RSS	-
Ranger	Ranger 监控指标
Zookeeper	Zookeeper 监控指标
Knox	-

日志

最近更新時間：2023-09-12 17:23:14

功能介绍

日志搜索功能提供组件的运行日志采集和搜索功能，支持当前集群核心服务日志和节点系统日志进行关键词搜索，在控制台即可快速查看服务关键日志。

操作步骤

1. 登录 [EMR 控制台](#)，在 EMR on TKE 集群列表中单击对应的 **集群 ID/名称** 进入集群详情页。
2. 在集群详情页 **集群信息 > 服务架构 > 服务名称** 或者选择 **服务管理 > 服务名称**，单击日志可根据当前服务选择 PodName 和时间范围条件过滤，查看日志内容。
 或者在集群详情页 **基础配置 > 服务架构**，选择 **服务名称 > 日志**，选择 PodName 和时间范围，即可进行日志搜索。

The screenshot displays the HIVE log search interface. At the top, there are tabs for Dashboard, 角色管理, 配置管理, 日志 (selected), and Java分析. Below the tabs, there are search filters: Time (1小时, 24小时, 近7天, 2023-06-05 16:14:05 ~ 2023-06-06 16:14:05), Keyword (请输入日志内容关键字), PodName (请选择PodName), and Log Level (最低日志级别: WARN). A blue '查询' button is present. The main area shows a table of log entries under the '全部' tab. The table has columns: 时间, PodName, 级别, 信息, and 操作. Two log entries are visible, both with a 'WARN' level and a message indicating a failure to download a jar file: '[main] fs.TemrfsHadoopFileSystemAdapter: init temrfs impl failed, we will retry again, retryInfo: 4/5 java.io.IOException: download jar failed, localJarPath: /data1/emr/hdfs/tmp/temrfs/hadoop-temrfs-3.1.0-8.1.4-1.0.6.jar...'. There are '查看上下文' and '更多' links for each entry.

单击 **PodName**，可跳转到对应节点状态页面。单击 **日志源**，可跳转到对应节点监控指标展示页。

2.1 **关键字说明**：支持关键字全文检索。

2.2 支持特殊字符 - . * > < = ! () { } / 检索。

2.3 支持短语检索，例如：address=/ip:port。

3. 在排查问题的时候，经常需要关注关键词的上下文日志，在日志搜索页单击 **查看上下文**，进入日志上下文页面，支持选择时间查看时间范围内的日志上下文信息。

日志上下文

角色 **HiveServer2** PodName **emr-4zjq66fk-hive-hiveserver2-0**

日志路径 **/data1/emr/hive/logs/hadoop-hiveserver2**

时间 **2023-06-06 07:14:56 ~ 2023-06-06 08:16:30** [定位至最新一条](#)

```
1 2023-06-06 07:14:56.250 WARN [main] fs.TemrfsHadoopFileSystemAdapter: init temrfs
2 java.io.IOException: download jar failed, localJarPath: /data1/emr/hdfs/tmp/temrfs
3   at com.qcloud.emr.fs.TemrfsHadoopFileSystemJarLoader.downloadJarPath(TemrfsHad
4   at com.qcloud.emr.fs.TemrfsHadoopFileSystemJarLoader.getAlreadyLoadedClassInfo
5   at com.qcloud.emr.fs.TemrfsHadoopFileSystemJarLoader.init(TemrfsHadoopFileSyst
6   at com.qcloud.emr.fs.TemrfsHadoopFileSystemAdapter.initJarLoadWithRetry(Temrfs
7   at com.qcloud.emr.fs.TemrfsHadoopFileSystemAdapter.initialize(TemrfsHadoopFile
8   at org.apache.hadoop.fs.FileSystem.createFileSystem(FileSystem.java:3355) ~[ha
9   at org.apache.hadoop.fs.FileSystem.access$200(FileSystem.java:124) ~[hadoop-co
10  at org.apache.hadoop.fs.FileSystem$Cache.getInternal(FileSystem.java:3404) ~[h
11  at org.apache.hadoop.fs.FileSystem$Cache.get(FileSystem.java:3372) ~[hadoop-co
12  at org.apache.hadoop.fs.FileSystem.get(FileSystem.java:484) ~[hadoop-common-3.
13  at org.apache.hadoop.fs.Path.getFileSystem(Path.java:365) ~[hadoop-common-3.2.
14  at org.apache.hadoop.hive.ql.exec.Utilities.ensurePathIsWritable(Utilities.jav
15  at org.apache.hadoop.hive.ql.session.SessionState.createRootHDFSDir(SessionSta
16  at org.apache.hadoop.hive.ql.session.SessionState.createSessionDirs(SessionSta
17  at org.apache.hadoop.hive.ql.session.SessionState.start(SessionState.java:627)
18  at org.apache.hadoop.hive.ql.session.SessionState.start(SessionState.java:586)
19  at org.apache.hive.service.cli.CLIService.applyAuthorizationConfigPolicy(CLISer
20  at org.apache.hive.service.cli.CLIService.init(CLIService.java:115) ~[hive-ser
21  at org.apache.hive.service.CompositeService.init(CompositeService.java:59) ~[h
22  at org.apache.hive.service.server.HiveServer2.init(HiveServer2.java:227) ~[hiv
23  at org.apache.hive.service.server.HiveServer2.startHiveServer2(HiveServer2.jav
24  at org.apache.hive.service.server.HiveServer2.access$1600(HiveServer2.java:137
25  at org.apache.hive.service.server.HiveServer2$StartOptionExecutor.execute(Hive
26  at org.apache.hive.service.server.HiveServer2.main(HiveServer2.java:1155) ~[hi
27  at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method) ~[?:1.8.0_322]
28  at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:6
```

时间 ↓	PodName
2023-06-06 08:11:11.355	emr-4zjq66fk-hive-hiveser...
2023-06-06 08:09:00.282	emr-4zjq66fk-hive-hiveser...

配置告警

最近更新时间：2023-09-12 17:23:14

EMR on TKE 监控告警配置，参考 [EMR on CVM 告警配置](#)，告警历史请参见 [告警历史](#)。

GC 分析

最近更新时间：2023-09-12 17:23:14

操作场景

支持 Java 进程的 GC 在线分析，通过实时采集、记录、分析 GC 日志，帮助用户排查是否因 GC 导致的进程异常。

- GC 视图：筛选服务、角色、节点及时间，查看相关 GC 分布情况及点位信息。
- GC 列表：可根据需要筛选过滤 GC 日志数据，查看 GC 记录的多维度信息。

操作步骤

1. 登录 [EMR 控制台](#)，在集群列表中单击对应的**集群 ID/名称**进入集群详情页。
2. 在集群详情页中单击**集群监控**，然后选择 **JAVA 分析 > GC 在线分析**，即可查看 GC 视图及点位信息。
3. 单击集群详情页**集群信息 > 服务架构 > 服务卡片**或者**集群服务 > 服务名称**，然后选择 GC 分析即可按需选择角色、节点查看 GC 视图及点位信息。
4. 同时提供 GC 查询列表信息，部分列头字段支持筛选或排序等操作。

支持 GC 日志采集的服务及角色如下：

服务	角色
Zookeeper	Zookeeper
Trino	Hmaster
	RegionServer
	HbaseThrift
Hive	HiveMetaStore
	HiveServer2
	HiveWebHcat
Knox	Idap
	gateway
Spark	SparkJobHistoryServer
Ranger	EmbeddedServer
	EnableUnixAuth
RSS	LivyServer
Kyuubi	KyuubiServer

应用分析

Spark 查询

最近更新时间：2023-10-26 14:35:31

操作场景

提供快速查看 Spark 作业的提交用户、状态、持续时间等多项明细指标，并支持作业级历史任务对比、作业洞察、任务执行信息等。

操作步骤

1. 登录 [EMR 控制台](#)，在集群列表中单击对应的**集群 ID/名称**进入集群详情页。
2. 在集群详情页选择**作业管理 > Spark**，即可查看查询管理，查询相关 Spark 作业信息、任务信息查看、应用执行结果洞察及应用监控对比等。作业级别提供用户、应用名、队列名、作业类型、持续时间及相关吞吐资源等多维信息筛查。

The screenshot displays the 'SPARK' job management page. On the left is a navigation sidebar with options like '集群信息', '集群服务', '作业管理', and '监控大盘'. The main area shows a table of jobs with the following data:

应用ID	洞察结果	用户	应用名	持续时间	状态	操作
spark-81236172c4514e4585cc5...	OK	hadoop	Hive on Spark (sessio...	54s527ms	completed	详情 更多
spark-fde10038a6df4869997bf4...	OK	hadoop	Hive on Spark (sessio...	1h32min	completed	详情 更多
spark-3ba71846bfad431dbe1a6...	OK	hadoop	Hive on Spark (sessio...	47s63ms	completed	详情 更多

注意

- 其中 Spark 类型应用的任务信息、应用洞察、应用对比新功能需 check Spark History 版本是否符合，check 命令如下：
`curl "http://localhost:10000/api/v1/applications" | json_pp`，返回数据为非正常的 json 格式数据则 SparkHistory 版本不符合，可 [提交工单](#) 申请开启相关功能。
- 作业查询将每30s采集一次 ResourceManager 数据，采集操作对集群业务影响微小可忽略。

3. 在作业列表中单击**更多 > 应用洞察**，查看应用的详细洞察项及相关的洞察规则、结果、建议。

应用洞察
✕

🔄

洞察项	等级 ▾	规则	结果/建议
CPU资源浪费	一般	空闲的CPU资源 占 Executor申请的CPU资源 超过...	✔
ExecutorGC	中等	Executor的GC时间占比超过20%	✔
Memory资源浪费	一般	空闲的内存资源 占 Executor申请的内存资源 超过 ...	✔
ScheduleOverhead	严重	花费过多的时间用于调度Task	✔
慢Task	中等	Task处理时间大于所有Task平均处理时间的2倍	✔
数据倾斜	严重	Task处理的数据大于所有Task平均处理数据的2倍	✔
调度倾斜	严重	Task调度倾斜, 某些Executor执行了过多的Tasks	✔

4. 在作业列表中单击更多 > 任务信息，查看作业的任务列表及任务的运行日志。

任务信息
✕

🔍
🔄

Id	HostPort	IsActive	RddBlocks	TotalCores	MemoryUsed	操作
1	<div style="width: 20px; height: 10px; background-color: #ccc; display: inline-block;"></div>	true	--	1	--	--
2	<div style="width: 20px; height: 10px; background-color: #ccc; display: inline-block;"></div>	true	--	1	--	--
driver	<div style="width: 20px; height: 10px; background-color: #ccc; display: inline-block;"></div>	true	--	--	--	--

共 3 条
10 条 / 页

⏪
⏩
1
/ 1 页
⏪
⏩

Hive 查询

最近更新时间: 2023-09-12 17:23:14

操作场景

提交后的 Hive 查询可通过查询管理快速查看查询的运行状况。查询列表展示了相关查询的执行信息、执行状态等信息，同时可帮助快速关联查询得到执行作业。

操作步骤

1. 登录 [EMR 控制台](#)，在集群列表中单击对应的**集群 ID/名称**进入集群详情页。
2. 在集群详情页中单击**作业管理目录**，然后选择 **Hive 名称 > 查询**，即可筛选、查看相关作业查询信息。

The screenshot shows the 'HIVE' query management interface. On the left is a navigation menu with options like '集群信息', '集群服务', '作业管理', and 'HIVE'. The main area displays a '查询列表' (Query List) with a search bar and filters. The table below shows several queries with their execution details.

执行语句	执行时长	执行状态	开始时间	用户	操作
insert into cos_hs2_Native_TEXTFILE_spa...	31ms	FINISHED	2023-06-14 03:09:32	hadoop	详情
use hivetest_hs2	0ms	FINISHED	2023-06-14 03:09:04	hadoop	详情
select * from cos_hs2_crontrib_serde_spar...	0ms	FINISHED	2023-06-12 11:12:58	hadoop	详情
create table cos_hs2_crontrib_serde_spar...	0ms	FINISHED	2023-06-12 11:12:45	hadoop	详情
insert into cos_hs2_crontrib_serde_spar ...	6ms	ERROR	2023-06-12 11:12:45	hadoop	详情

3. 单击查询记录操作可查看查询语句、查询计划、执行总览、Profile 等详细信息。

The screenshot shows the 'Profile' tab of a query's details. It contains performance logs for compile-time and execution-time. The 'Compile-time metadata operations' section shows a table of call names and their durations.

Call Name	Time (ms)
isCompatibleWith_(Configuration,)	0
flushCache_()	1

Below this, there are sections for 'Compile-Time Perf-Logger' and 'Execution-Time Perf-Logger', each with a table structure for 'Compile-time Call Name' and 'Time (ms)'.