

Elastic MapReduce

EMR on TKE Operation Guide

Product Documentation



Copyright Notice

©2013-2024 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

Trademark Notice

 Tencent Cloud

All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

Contents

EMR on TKE Operation Guide

Introduction to EMR on TKE

Configuring Cluster

Permission Management

Role Authorization

Creating Cluster

Cluster Management

Adjusting the Number of Pods

Modifying Configuration

Service Management

Deployment Instructions

Adding Components

Restarting Services

WebUI Access

Role Management

Configuration Management

Configuration Update

Configuration Rollback

Monitoring and Ops

Monitoring Dashboard

Logs

Configuring Alarms

GC Analysis

Application Analysis

Spark Query

Hive Query

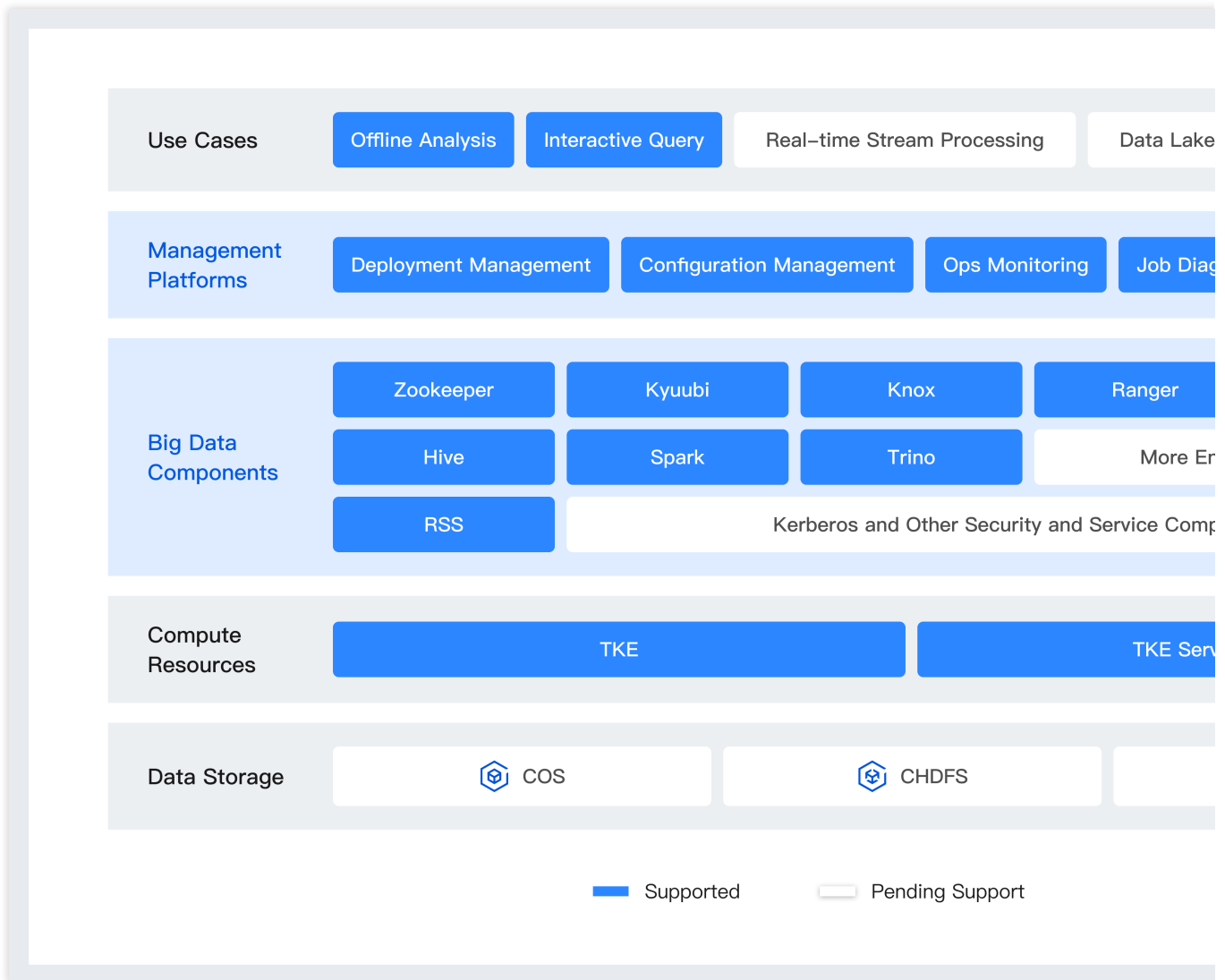
EMR on TKE Operation Guide

Introduction to EMR on TKE

Last updated : 2024-10-30 10:45:22

Tencent Cloud EMR on TKE is a big data service deployment platform based on containerized services and open-source big data engines, offering rapid deployment, flexible scaling, and efficient, secure services. Through features like application management in the console, users can focus more on business applications. The service engine includes big data components such as Spark, Hive, and Trino, allowing users to easily run, manage, and scale containerized applications.

Product Architecture



Description

Data storage: In the compute-storage separation scene applicable to EMR on TKE, multiple data storage products such as COS, CHDFS, and HDFS are provided for integration. Users can store data in these sources and perform processing and analysis using the EMR on TKE big data processing engine.

Computing resources: EMR on TKE supports deployment on Tencent Cloud TKE General Clusters and Serverless Clusters.

Big data components: EMR on TKE provides optional services including Hive, Spark, Trino, Zookeeper, Kyuubi, Knox, Ranger, Hue, and RSS.

Management platform: EMR on TKE provides a user-friendly interface through the EMR console for easy component deployment, configuration management, Ops monitoring, and exception alerts. It also offers advanced job analysis and diagnostics to help users gain insights into job costs.

Product Advantages

1. High resource utilization: EMR on TKE container services can automatically scale the number of cluster containers up or down based on preset policies, ensuring stable service operation while saving on resource costs. Flexible application resource configuration in offline scenes effectively improves resource utilization and optimizes costs.
2. Stability and reliability: EMR on TKE relies on the high-reliability features of TKE clusters, such as container self-check and self-healing. When a service pod node fails, the pod is automatically rebuilt, and the image is reloaded.
3. Simplified deployment: EMR on TKE can start a complete multi-service cluster in just a few minutes. Additionally, it allows users to easily and quickly adjust the number of service pods through console operations.
4. Granular security: EMR on TKE integrates with CAM to implement cluster access control. It also connects with COS using minimized storage permissions to achieve refined permission management in compute-storage separation scenes, ensuring the security of data access at the cluster usage level.

Configuring Cluster Permission Management Role Authorization

Last updated : 2024-10-30 15:50:01

When using the EMR service, users need to grant the service account the default system role `EMR_QCSRole`. Once the role is successfully granted, EMR can call related services (such as TKE and COS) to create clusters and save logs.

Note

When enabling EMR for the first time, you need to complete the role authorization process using the root account; otherwise, neither sub-accounts nor the root account can use EMR.

Role Authorization Process

1. When a user creates a cluster or creates an on-demand execution plan, if the `EMR_QCSRole` role authorization for the service account fails, the user will be redirected to a page notifying the permission limitations. Then click **Go to CAM** to proceed with role authorization.
2. Click **Agree to Authorize** to authorize the default role `EMR_QCSRole` to the EMR service account.
3. After authorization is completed, the user needs to refresh the EMR console or purchase page, after which normal operations can proceed. For more detailed information on `EMR_QCSRole` policies, you can log in to the [CAM Console](#). The permissions included in `EMR_QCSRole` can be found in [Collaborator/Sub-account Permissions](#).

Special Instructions for Service Role Authorization Related to EMR on TKE Clusters

When you create or use an EMR on TKE cluster, data needs to be directly written to or calculated in Cloud Object Storage (COS). To ensure data security, EMR should be granted temporary keys to read and write COS resources. Therefore, the relevant EMR service-related role `EMR_QCSLinkedRoleInApplicationDataAccess` should be authorized and bound to the `QcloudAccessForEMRLinkedRoleInApplicationDataAccess` preset policy.

1. When viewing the EMR on TKE cluster list, you need to check if the service-related role `EMR_QCSLinkedRoleInApplicationDataAccess` is bound to the EMR service.
2. If the EMR service-related role `EMR_QCSLinkedRoleInApplicationDataAccess` does not exist, authorization and binding need to be performed.

Note

If you need to specify cluster access permissions for the corresponding COS resources in a more refined manner, see [Custom Service Roles](#) for settings.

EMR on TKE Cluster Authentication Description

The permission settings for sub-accounts and collaborators are consistent with that of the EMR on CVM version. For details, see [Collaborator/Sub-account Permissions](#).

Tag authentication and API authentication settings are consistent with that of the EMR on CVM version. For details, see [Authentication Granularity Scheme](#).

Creating Cluster

Last updated : 2024-10-30 10:56:07

Overview

This document introduces the directions and configurations for creating an EMR on TKE cluster through the EMR console.

Directions

Log in to the [EMR Console](#), click **Create Cluster** on the EMR on TKE cluster list page, and complete the relevant configuration on the purchase page. When the cluster status in the list shows Running, it indicates that the cluster has been successfully created.

Configuration Item	Configuration Items Description
Cluster name	Set the cluster name to distinguish between different EMR clusters. The system generates a random name, which can be modified.
Region	The physical data center where the cluster is deployed. Each region represents an independent physical data center, and private networks of CVMs between different regions are not interconnected. Note: <ol style="list-style-type: none">1. Once the cluster is created, the region cannot be changed, so choose carefully.2. Only container clusters within this region are supported to provide resources for the cluster.
Container type	<ol style="list-style-type: none">1. The service role is deployed by using resources provided by the container layer, supporting both TKE General and TKE Serverless clusters.2. The computing resources are container deployment resources purchased synchronously with EMR, supporting resource association for EMR on TKE clusters. Note: Select a specific container cluster or computing resource to provide Pod resources for service deployment of the EMR on TKE cluster.
Cluster Network	Used for purchasing a db. It is necessary to ensure that the EMR cluster network is consistent with the container cluster network.
AZ and subnet	Used for purchasing a db. It is necessary to ensure that the EMR cluster network is consistent with the container cluster network.
Security group	The security group functions as a firewall to configure network access control for CVM. If

	<p>no security group is available, EMR will automatically create one for you. If there is an existing security group in use, you can select it directly. If the number of security groups has reached the upper limit and new ones cannot be created, you can delete some unused security groups. View the security groups currently in use.</p> <p>Create a security group: EMR helps users create a security group, enabling ports 22 and 30001, as well as the necessary private network IP ranges.</p> <p>Existing EMR security group: Select an already created EMR security group as the security group for the current instance, enabling ports 22 and 30001, as well as the necessary private network IP ranges.</p>
Billing mode	<p>Only supports pay-as-you-go mode. Pay-as-you-go: Pay based on usage duration. Account identity verification is required, and a 2-hour fee will be frozen at the time of activation (vouchers cannot be used as a freezing deposit). The frozen resource fee will be refunded upon termination.</p>
Product version	<p>The components and their versions bundled with different product versions vary.</p>
Deploy task	<p>Optional components that can be customized and combined based on your needs. Select at least one component service for deployment. Supported deployment types: default deployment and associated deployment.</p> <p>Associated deployment: Provides service sources for the current cluster, such as other cloud EMR clusters or DLC.</p> <p>Default deployment: Deploys the current service to the selected container resources, allowing customization of resource specifications and quantity (service role level deployment). Recommended for use.</p>
Associated deployment	<p>Select an EMR on CVM or EMR on TKE cluster that contains the current service.</p> <ol style="list-style-type: none"> 1. When you select an EMR on CVM cluster, the cluster needs to have the current service deployed. 2. When you select an EMR on TKE cluster, the current service in the cluster needs to be configured to provide external services. 3. MetaStore (HiveMetaStore) supports selecting DLC as the service source.
Default deployment	<p>Data volume directory: Provides data storage directories for containers. Currently, only mounting with HostPath or using the created PVC is supported. POD specifications set service role dimension deployment information as an optional setting.</p> <p>External access setting: Set whether the current role provides external services, with optional support for public network LB access and private network LB access.</p> <p>Node scheduling policy: Set the scheduling policy for the current role deployment, allowing Pods to be scheduled to nodes that match the desired Label according to the scheduling rules.</p>
Hive metadatabase	<p>If the Hive component is selected, Hive Metastore offers two storage options:</p> <p>The first is the default cluster option, where Hive metadata is stored in a separately purchased MetaDB for the cluster.</p>

	<p>The second option is to associate an external Hive Metastore, where you can choose to link to EMR-MetaDB or a self-built MySQL database, with metadata stored in the associated database, which will not be destroyed when the cluster is terminated. See Hive Metadata Management.</p> <p>Note: When you select one or more of the following components such as Hue, Ranger, Oozie, Druid, and Superset, the system will automatically purchase a MetaDB for storing metadata of components other than Hive.</p>
COS bucket	Used for storing logs, JAR packages, and other information.
Tag	You can add tags to cluster or node resources during creation to facilitate resource management. A maximum of 5 tags can be added, and tag keys should not be duplicated.
Set Password	Set the webUI password. The current password is only used to initially set up the service webUI access password.
Terms of agreement	Agree to the EMR Service Level Agreement and Refund Agreement .

Cluster Management

Adjusting the Number of Pods

Last updated : 2024-10-30 10:57:08

Overview

When the service role CPU or memory is insufficient, you can increase the CPU or memory size by changing the configuration. This document introduces the relevant operations for changing instance configurations through the [EMR on TKE console](#).

Note
Changing the configuration may cause running jobs to fail. Ensure that you perform this operation without impacting business operations.

During the configuration change process, resources may be preempted, causing pod reconstruction to fail and resulting in service unavailability. It is recommended to perform this operation during off-peak business hours.

Prerequisites

1. Pod resource configuration adjustments will be automatically billed item by item. Ensure sufficient account balance.
2. After the configuration of pay-as-you-go nodes is changed, the fees will be refrozen based on the latest configuration price. For more details, see [EMR on TKE Billing Instructions > Purchase Instructions](#).

Directions

1. Log in to the [EMR on TKE Console](#) and click the **Cluster ID/Name** in the cluster list to enter the cluster details page.
2. On the cluster details page, select **Cluster Services > A Certain Service**, and enter the role management page.
3. On the role management page, select the service role that requires configuration changes and click **Modify Configuration**.
4. On the configuration adjustment page, confirm the relevant information for the configuration change, select the target CPU and memory configurations, and make sure to carefully read the important notice before checking Agree to Change Information.
5. After the fees are confirmed, click **Confirm** to adjust the configuration.

Modifying Configuration

Last updated : 2024-10-30 10:59:06

Overview

When the service role CPU or memory is insufficient, you can increase the CPU or memory size by changing the configuration. This document introduces the relevant operations for changing instance configurations through the [EMR on TKE console](#).

Note

Changing the configuration may cause running jobs to fail. Ensure that you perform this operation without impacting business operations.

During the configuration change process, resources may be preempted, causing pod reconstruction to fail and resulting in service unavailability. It is recommended to perform this operation during off-peak business hours.

Prerequisites

1. Pod resource configuration adjustments will be automatically billed item by item. Ensure sufficient account balance.
2. After the configuration of pay-as-you-go nodes is changed, the fees will be refrozen based on the latest configuration price. For more details, see [EMR on TKE Billing Instructions > Purchase Instructions](#).

Directions

1. Log in to the [EMR on TKE Console](#) and click the **Cluster ID/Name** in the cluster list to enter the cluster details page.
2. On the cluster details page, select **Cluster Services > A Certain Service**, and enter the role management page.
3. On the role management page, select the service role that requires configuration changes and click **Modify Configuration**.
4. On the configuration adjustment page, confirm the relevant information for the configuration change, select the target CPU and memory configurations, and make sure to carefully read the important notice before checking Agree to Change Information.

Service Management

Deployment Instructions

Last updated : 2024-10-30 11:01:36

The deployment description is used to view the types of clusters supported for deployment in the EMR on TKE cluster, service role information, default Pod specifications, and data volume directories, providing users with a reference for adjusting service deployment resource configurations.

Deployment Information Description

Cluster type	Component	Role	Default Pod Specifications			Data Volume Directory
			CPU Limit (Cores)	Memory Limit	Pod Quantity Range	
TKE	hive	HiveMetaStore	1-100	2-100	1-100	New PVC; host path
		HiveServer2	1-100	2-100	1-100	New PVC; host path
	hue	Hue	2-100	4-100	1-100	New PVC; host path
	kyuubi	KyuubiServer	1-100	2-100	2-100	New PVC; host path
	ranger	Ranger	1-100	2-100	1-100	New PVC; host path
		RangerUsersync	1-100	2-100	1-100	New PVC; host path
	rss	Coordinator	1-100	2-100	2-100	New PVC; host path
		ShuffleServer	2-100	4-100	3-100	New PVC
	spark	SparkJobHistoryServer	1-16	1-64	1-2	New PVC; host path

	trino	TrinoCoordinator	1-100	2-100	1-2	New PVC; host path
		TrinoWorker	1-100	2-100	2-100	New PVC; host path
	zookeeper	QuorumPeerMain	1-100	2-100	3-100	New PVC
	knox	Knox	1-100	2-100	1-100	New PVC, Host Path
TKE Serverless	hive	HiveMetaStore	1-100	2-100	1-100	New 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		

Note

The default maximum value of the service role Pod in the cluster will be dynamically adjusted based on the available resources of the selected container cluster. The actual maximum value may be less than the maximum value given in the table.

Adding Components

Last updated : 2024-10-30 11:06:53

After the cluster is created, you can add other components supported by the selected product version through the Add Component operation. This document introduces how to add components through the console.

Note

Note Adding components only supports selecting components from the current EMR product version.

Directions

1. Log in to the [EMR on TKE Console](#) and click the **Cluster ID/Name** in the cluster list to enter the cluster details page.
2. Select **Cluster Information > Basic Configuration**, and then select **Add Components** from the service architecture diagram to add components that have not been installed in the cluster.
3. When the cluster lacks a metadatabase, a CloudDB instance storage unit should be purchased to serve as the metadata storage when you install the Hue or Ranger components.
4. The Hive component metadata storage provides two storage methods: the first is the default cluster option, where metadata is stored in MetaDB; the second is an association option, where the metadata can be stored in an associated EMR-MetaDB or an associated self-built MySQL database, and it will not be terminated when the cluster is terminated.
5. Whether to purchase MetaDB is consistent with the configuration during the new cluster purchase, and the Hive metastore selection is also consistent with the configuration during the new cluster purchase.
6. After the components are selected, click **Create** to proceed.

Restarting Services

Last updated : 2024-10-30 11:07:54

Overview

After you upgrade the service roles or modify the configuration in a cluster, it is necessary to restart the corresponding service for the configuration to take effect. This document introduces the directions for restarting services through the console.

Directions

1. Log in to the [EMR on TKE Console](#) and click the **Cluster ID/Name** in the cluster list to enter the cluster details page.
2. If you need to restart the entire component service, select **Cluster Service > Specific Service**, and then choose **Operations > Restart Service** in the right-hand sidebar of the page. When you select the service role as ALL, the entire service will be restarted.
3. If you need to restart a specific role type of the service, you can select **Cluster Service > Specific Service**, check the service role that needs to be restarted on the role management page, and click **Restart Service**.
4. After the information is confirmed, click **Confirm** to restart the service.

WebUI Access

Last updated : 2024-10-30 11:08:58

The Software WebUI Entry feature is provided by EMR to access the native UI of components. Through the Knox component, users can conveniently access the native UI of components via the public network.

Prerequisites

The cluster has installed the Knox component, and public network access for the Knox component is enabled.

Directions

1. Log in to the [EMR on TKE Console](#) and click the **Cluster ID/Name** in the cluster list to enter the cluster details page.
2. Select **Cluster Services > A certain service**, and then click **View Information > View WebUI** on the right side of the sidebar to access the WebUI.

Note

The access address requires authentication. The username is root, and the default password is the one entered during cluster creation.

During cluster creation, if the Knox component is not installed or public network access is not enabled for the Knox component, you will not be able to access the WebUI interface of related components via the native WebUI access address in the component management page.

Role Management

Last updated : 2024-10-30 11:09:43

Role management provides Ops features at both the service level and specific POD level, such as restart operations. POD health status monitoring is supported to understand the real-time status of role processes.

Directions

1. Log in to the [EMR on TKE Console](#) and click the **Cluster ID/Name** in the cluster list to enter the cluster details page.
2. On the cluster details page, select **Cluster Services** and click the corresponding service to enter the Role Management page.
3. The Role List displays information such as the current service role's POD health status, running/desired POD count, resource configuration, and latest restart time. The POD list displays information such as POD resource health status, POD status, POD IP, restart count, and latest restart time.

Description

When the POD status is Running or Succeeded, the POD health status is displayed as Normal; when the POD status is Pending, Failed, Unknown, or the corresponding service container is Not Ready, the POD health status is displayed as Abnormal.

Configuration Management

Configuration Update

Last updated : 2024-10-30 11:11:06

Overview

Configuration management supports the modification of key configuration parameters for commonly used open-source components such as HIVE and SPARK. You can modify the service configuration for different roles based on actual requirements. This document introduces the directions for configuring parameters through the console.

Note

When you delete custom configuration files in the console's configuration management, for security reasons, the client will not synchronize the deletion action.

Directions

Editing Configuration

1. Log in to the [EMR on TKE Console](#) and click the **Cluster ID/Name** in the cluster list to enter the cluster details page.
2. On the cluster details page, select **Cluster Services**, and click the corresponding service to enter the **Configuration Management page**.
3. After the configuration management page is entered, select the desired role as needed.
4. If you want to search for a specific configuration item or narrow down the search scope, you can use the filter on the left side.
5. Select the configuration file as needed, click **Edit Configuration** to enter edit mode, and proceed with adding, editing, or deleting configuration items as required.
Select the parameter to be modified and enter the new parameter value. If it is necessary, click **Restore** to recover to the original value or **Default Value** to reset to the system-recommended default value.
Some parameters support deletion. To delete the configuration, select **Delete > Confirm**.
If the desired parameter is not present in the file, you can click **Add Configuration Item** to open the add configuration item pop-up, and fill in the parameter name and value.
6. After the information is confirmed, click Save Configuration. Once the configuration is successfully applied, click Restart Service to complete the configuration modification.

Description

Note

If modifications are made to the service process configuration, a restart of the service is required after saving for the modifications to take effect.

If modifications are made to the client configuration, the modifications take effect immediately after saving, without restarting the service.

Adding Configuration File

1. Log in to the [EMR on TKE Console](#) and click the **Cluster ID/Name** in the cluster list to enter the cluster details page.
2. On the cluster details page, select **Cluster Services**, and click the corresponding service to enter the **Configuration Management page**.
3. If the desired configuration file is not available, click the **+** (Add Configuration File) on the right to enter the configuration file setup page and fill in the configuration details.
4. After you click **Save Configuration**, the parameters will be delivered and the configuration file name will be updated in the configuration file list.
5. Custom configuration files, once effective, support modification and deletion operations.

Configuration Rollback

Last updated : 2024-10-30 11:11:43

EMR supports adding, modifying, and deleting configuration items for various service parameters in the console, as well as rolling back configurations. This document introduces how to roll back service parameter configurations through the console.

Directions

1. Log in to the [EMR on TKE Console](#) and click the **Cluster ID/Name** in the cluster list to enter the cluster details page.
2. On the cluster details page, select **Cluster Services**, and click the corresponding service to enter the Configuration Management page.
3. On the Configuration Management page, select **Configuration History**. Click **Details** to view a comparison of parameter values before and after the changes. Click **Rollback** to recover the parameter configuration changes for that record. Select **Rollback > Confirm rollback**. After a successful rollback, the service is restarted, and the rollback will take effect shortly.

Description

Adding, modifying, and deleting configuration items support rollback, but adding configuration files and deleting the added configuration files do not support rollback.

Monitoring and Ops

Monitoring Dashboard

Last updated : 2024-10-30 11:18:41

Provides EMR on TKE monitoring policy configuration and alarm records features, and allows viewing of cluster service monitoring metrics and historical trends.

Monitoring Metric

EMR on TKE provides the following service monitoring metrics:

Service	Monitoring Metric Description
Hive	Hive Monitoring Metrics
Spark	Spark Monitoring Metrics
Trino	Trino Monitoring Metrics
Hue	-
Kyuubi	Kyuubi Monitoring Metrics
Ranger	Ranger Monitoring Metrics
Zookeeper	ZooKeeper Monitoring Metrics
Knox	-
cosranger	-

Logs

Last updated : 2024-10-30 11:21:08

Feature Introduction

The log search feature provides the ability to collect and search running logs of components, supporting keyword search for core service logs and node system logs of the current cluster, allowing quick access to key service logs directly from the console.

Directions

1. Log in to the [EMR Console](#) and click the corresponding **Cluster ID/Name** in the EMR on TKE cluster list to enter the cluster details page.
2. On the cluster details page, go to **Cluster Information > Service Architecture > Service Name** or select **Service Management > Service Name**, and then click **Logs**. You can filter by PodName and time range based on the current service to view the log content. Alternatively, on the cluster details page, go to **Basic Configuration > Service Architecture**, select **Service Name > Logs**, and choose PodName and Time Range to perform log searches.

Click **PodName** to navigate to the corresponding node status page. Click **Log Source** to navigate to the corresponding node monitoring metrics page.

2.1 **Keyword Instruction:** Supports full-text keyword search.

2.2 Supports special characters - . * > < = ! () { } / for search.

2.3 Supports phrase search, such as address=/ip:port.

3. When troubleshooting, you should often focus on the context around keywords in the logs. On the log search page, click **View Context** to enter the log context page, where you can select a time to view the log context within the specified time range.

Configuring Alarms

Last updated : 2024-10-30 11:21:50

For EMR on TKE monitoring and alarm configuration, see [EMR on CVM Alarm Configuration](#), and alarm records can be found in [Alert History](#).

GC Analysis

Last updated : 2024-10-30 11:22:40

Overview

Supports online GC analysis for Java processes by collecting, recording, and analyzing GC logs in real time, helping users identify whether process exceptions are caused by GC.

GC view: Filter by service, role, node, and time to view the related GC distribution and point information.

GC list: Filter GC log data as needed to view multidimensional GC information.

Directions

1. Log in to the [EMR console](#) and click the corresponding **Cluster ID/Name** in the cluster list to enter the cluster details page.
2. On the cluster details page, click **Cluster Monitoring**, and then select **JAVA Analysis > GC Online Analysis** to view the GC view and point information.
3. On the cluster details page, click **cluster information > Service Architecture > Service Card** or **Cluster Services > Service Name**, and then select GC Analysis to view the GC view and point information by selecting roles or nodes as needed.
4. Additionally, the GC query list information is provided, with some column headers supporting filtering or sorting operations.

The services and roles that support GC log collection are as follows:

Service	Role
Zookeeper	Zookeeper
Trino	Hmaster
	RegionServer
	HbaseThrift
Hive	HiveMetaStore
	HiveServer2
	HiveWebHcat
Knox	ldap

	gateway
Spark	SparkJobHistoryServer
Ranger	EmbeddedServer
	EnableUnixAuth
RSS	LivyServer
Kyuubi	KyuubiServer

Application Analysis

Spark Query

Last updated : 2024-10-30 11:23:37

Overview

Provides a quick overview of detailed metrics such as the user submitting Spark jobs, job status, and duration. It also supports historical job comparison, job insights, and task execution information at the job level.

Directions

1. Log in to the [EMR console](#), and click the corresponding **Cluster ID/Name** in the cluster list to enter the cluster details page.
2. On the cluster details page, select **Job Management > Spark** to view query management, access information on related Spark jobs, task details, application execution result insights, and application monitoring comparison. At the job level, it provides multi-dimensional filtering options such as user, application name, queue name, job type, duration, and related throughput resources.

Note

For the new features of task information, application insights, and application comparison for Spark-type applications, you need to check whether the Spark History version is compatible. Use the following command `curl https://localhost:10000/api/v1/applications" | json_pp` to check. If the returned data is not in a normal JSON format, then the Spark History version is not compatible, and you may [submit a ticket](#) to request to enable the relevant features.

Job queries collect ResourceManager data every 30 seconds. The impact of the collection operation on cluster operations is minimal and can be ignored.

3. In the job list, click **More > Application Insights** to view detailed insight items for the application, along with related insight rules, results, and suggestions.
4. In the job list, click **More > Task Information** to view the job's task list and the task's running logs.

Hive Query

Last updated : 2024-10-30 11:23:58

Overview

You can quickly view the status of submitted Hive queries through query management. The query list displays information such as the execution details and status of related queries and also helps quickly link to the execution jobs.

Directions

1. Log in to the [EMR console](#) and click the corresponding **Cluster ID/Name** in the cluster list to enter the cluster details page.
2. On the cluster details page, click **Job Management Directory**, and then select **Hive Name > Query** to filter and view relevant job query information.
3. Click Query Record to view detailed information, including the query statement, execution plan, execution overview, and profile.