Tencent Cloud

Elastic MapReduce

EMR on TKE Operation Guide

Product Documentation

Contents

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
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.
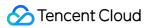
Directions

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: 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	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 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. 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. 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.
Billing mode	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.
Product version	The components and their versions bundled with different product versions vary.
Deploy task	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. Associated deployment: Provides service sources for the current cluster, such as other cloud EMR clusters or DLC. 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.
Associated deployment	Select an EMR on CVM or EMR on TKE cluster that contains the current service. 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	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. External access setting: Set whether the current role provides external services, with optional support for public network LB access and private network LB access. 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.
Hive metadatabase	If the Hive component is selected, Hive Metastore offers two storage options: The first is the default cluster option, where Hive metadata is stored in a separately purchased MetaDB for the cluster.

	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. 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.
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.