

Elastic MapReduce

Product Introduction

Product Documentation



Copyright Notice

©2013-2025 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



All trademarks associated with Tencent Cloud and its services are owned by the Tencent corporate group, including its parent, subsidiaries and affiliated companies, as the case may be. 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

Product Introduction

- Overview

- Strengths

- Architecture

- Features

- Use Cases

- Constraints and Limits

- Technical Support Scope

- Product release

 - Version Overview

 - Overview of Component Versions

Product Introduction

Overview

Last updated : 2023-12-26 17:39:56

Elastic MapReduce (EMR) is a secure, low-cost, and highly reliable open-source big data platform based on the cloud native technology and pan-Hadoop ecosystem. It provides diverse open-source big data components, such as Hive, Spark, HBase, Flink, StarRocks, Iceberg, and Alluxio, which are easy to deploy and manage, helping you efficiently build a cloud-based enterprise-grade data lake technology architecture.

Features

Completely derived from the open-source community, EMR enables you to seamlessly and smoothly migrate your existing big data clusters to Tencent Cloud. It is integrated with commonly used community components such as Hadoop, Hive, HBase, Spark, Presto, Flink, Sqoop, Hue, Iceberg, Druid, and StarRocks, fully meeting your various needs like online big data business, offline/online data warehousing, cloud-native data lake formation, and real-time stream computing.

EMR seamlessly incorporates the COS service, allowing you to migrate files stored in HDFS to COS with unlimited scalability, low storage costs, and high reliability for easy computing-storage separation. With COS, a cluster can be created whenever needed and terminated after the task is completed without the concerns over data loss. In addition, on-demand clusters can greatly reduce your big data processing costs.

EMR has five node types: master, core, task, router, and common. For the purposes of each type, see [Cluster Types](#). Currently, EMR supports multiple resource specifications, including Standard, Standard Network Optimized, MEM Optimized, High I/O, Computing, Computing Network Enhanced, and Big Data models. If you want to deploy a cluster on CPM, [contact us](#) for assistance.

Strengths

Last updated : 2023-12-26 17:40:05

Compared to self-created Hadoop clusters, Tencent Cloud EMR provides simpler, more stable, and more reliable Hadoop services.

Caution

In addition to the Hadoop cluster, Druid and ClickHouse big data clusters are also supported to provide more choices of big data architectures.

Flexibility

A secure and reliable Hadoop cluster can be created in just a few minutes to run mainstream open-source big data computing frameworks such as Hive, Spark, Presto, Impala, ClickHouse, Druid, and Flink, meeting your needs in scenarios such as **interactive BI, data warehousing, and real-time computation**.

Existing EMR clusters can be elastically and quickly scaled, and in-cloud computing resources can be scheduled in real time to respond to fast changes in your business data, reducing the high costs for reserving IT hardware.

Reliability

The master node is designed with disaster recovery in mind, and if it fails, a slave node will be started in seconds to ensure the availability of big data services.

A comprehensive monitoring system is in place, which can send SMS messages for exceptions in cluster components and tasks in a matter of seconds.

Hive metadata can be stored in MetaDB with a metadata reliability of 99.9996%.

Petabytes of high-persistence data stored in COS can be analyzed.

The recycle bin feature is enabled for clusters by default for you to restore devices that are deleted by mistake.

Security

The network policy for managed Hadoop clusters can be well planned through the convenient network isolation enabled by VPCs. Network ACLs and security groups can be created to filter traffic at the subnet and server levels, helping meet your network security needs in all aspects.

Tencent Cloud security reinforcement service provides an integrated security solution for EMR clusters, ranging from network protection and intrusion detection to vulnerability protection.

Kerberos authentication can be enabled for clusters to ensure secure access.

Ease of Use

Different clusters versions can be created to analyze the same data in COS in response to the actual business needs. Petabytes of data stored in data nodes or COS can be analyzed with the aid of out-of-the-box community components such as Hue and Oozie, eliminating your concerns over any knowledge migration costs.

A full-featured, intuitive, and easy-to-use monitoring system is provided to present nearly 1,000 cluster-level and component-level monitoring metrics on the monitoring overview page.

In-Cloud clusters consisting of multiple models are supported in a flexible manner, so that you can easily scale out or distribute configurations to heterogeneously configured clusters, enabling you to cope with business analysis challenges with higher-specced hardware.

Reduced Costs

EMR allows elastic scaling of your managed Hadoop cluster based on the business curve to reduce the high hardware costs.

It comes with a rich set of OPS tools which greatly improve the efficiency and enable you to focus on the business itself without having to worry about repeated construction of infrastructure for monitoring, security, and OPS.

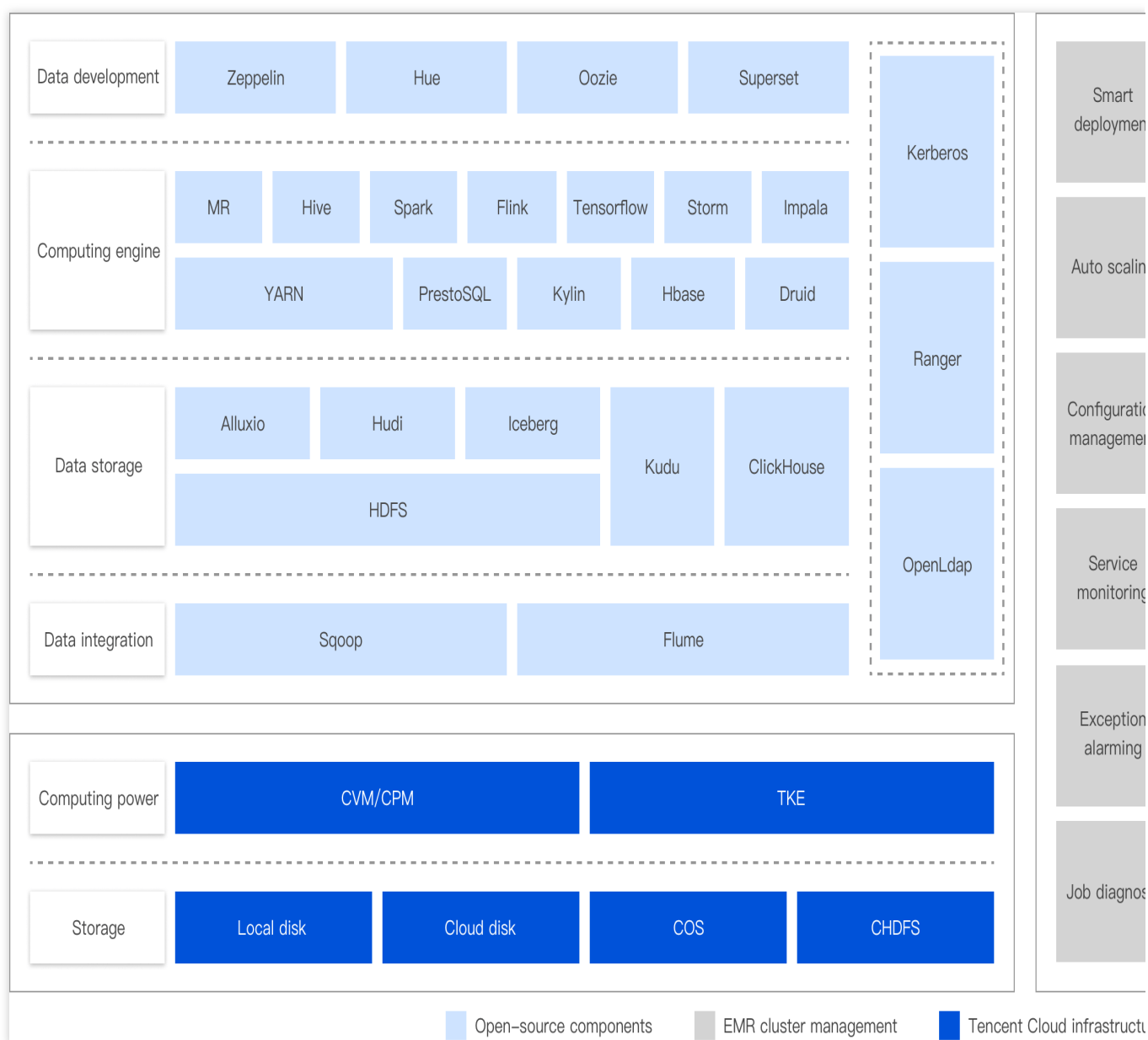
Warm and cold data can be stored on COS/CHDFS, effectively reducing the costs by **28%–50%**.

With unified Hive metadatabases and COS buckets, you can implement a cross-cluster architecture for analyzing the same dataset and create or terminate clusters as needed, reducing the cluster scaling costs.

Architecture

Last updated : 2025-01-24 09:49:24

The logical architecture of EMR is as shown below:



An EMR cluster consists mainly of three parts: open-source components, Tencent Cloud infrastructure, and cluster management.

Open-source components

EMR integrates dozens of cutting-edge open-source big data components from the Apache community, such as Hadoop, Hive, Spark, HBase, Presto, Flink, Alluxio, and Iceberg. For more information, see [Product Releases and Component Versions](#).

Based on optimized Iceberg, Alluxio and other open-source components, EMR offers Iceberg Z-Order algorithm, Alluxio transparent URI and other enhanced features.

Tencent Cloud infrastructure

EMR can be deployed based on multiple types of underlying computing resources, including CVM and CBM. It supports containerized deployment.

Data can be stored in a local disk, cloud disk, COS bucket, or CHDFS instance.

VPCs, network ACLs, and security groups provide EMR with a securely isolated network environment.

Cluster Management

EMR supports smart cloud deployment management, including fast creation, flexible scaling, and auto scaling.

EMR offers a great variety of convenient Ops tools, such as service configuration management, batch node management, and visual service Ops.

EMR provides complete cluster monitoring and diagnosis capabilities ranging from multidimensional metric monitoring, event, and inspection to alarming and log search.

Features

Last updated : 2023-12-26 17:51:47

Tencent Cloud Elastic MapReduce (EMR) is integrated with open-source frameworks and projects, such as Apache Hadoop, Apache Hive, Apache Spark, and Apache Storm. The cloud-based Hadoop services allow you to process vast amount of data securely, cost-efficiently, and elastically at scale. It has the following advantages:

Auto Scaling

Creating a cluster in minutes

You can create a secure, stable, cloud-managed Hadoop cluster in just a few minutes in the console.

Scaling in minutes

Your EMR cluster can be smoothly scaled up or down just a few minutes as your computing needs change.

API support

EMR clusters can be easily created, scaled, and terminated in programs through APIs.

Separation of Storage and Computation

Intra-cluster separation of storage and computation

At the cluster level, cloud-based Hadoop clusters can be planned in a manner where storage nodes and compute nodes are separated, so that you can scale the compute nodes as needed to lower the hardware costs.

COS-based separation of storage and computation

Massive amounts of data to be analyzed can be stored in COS. While the storage costs is reduced through COS, different versions of EMR clusters can be created to analyze the same data, which brings out extreme architectural flexibility.

OPS Support

Monitoring and multi-channel alarming

A comprehensive monitoring and OPS system is provided, which can detect exceptions in components such as Spark, Hive, and Presto and jobs within seconds after they occur to ensure the robust operations of big data clusters.

Technical support

In addition to comprehensive technical documentation, Tencent Cloud also has a technical service system where complete technical support is provided through various channels such as ticket.

Security

EMR uses security groups to control inbound and outbound traffic to your CVM instances. Components' web UIs can only be accessed through one specified instance assigned with public IP, and the access requests must be authenticated by username and password. In addition, the security group of this instance only allows SSH ports and proxy access ports.

Caution

Changing the project will cause the CVM instance to lose its security group.

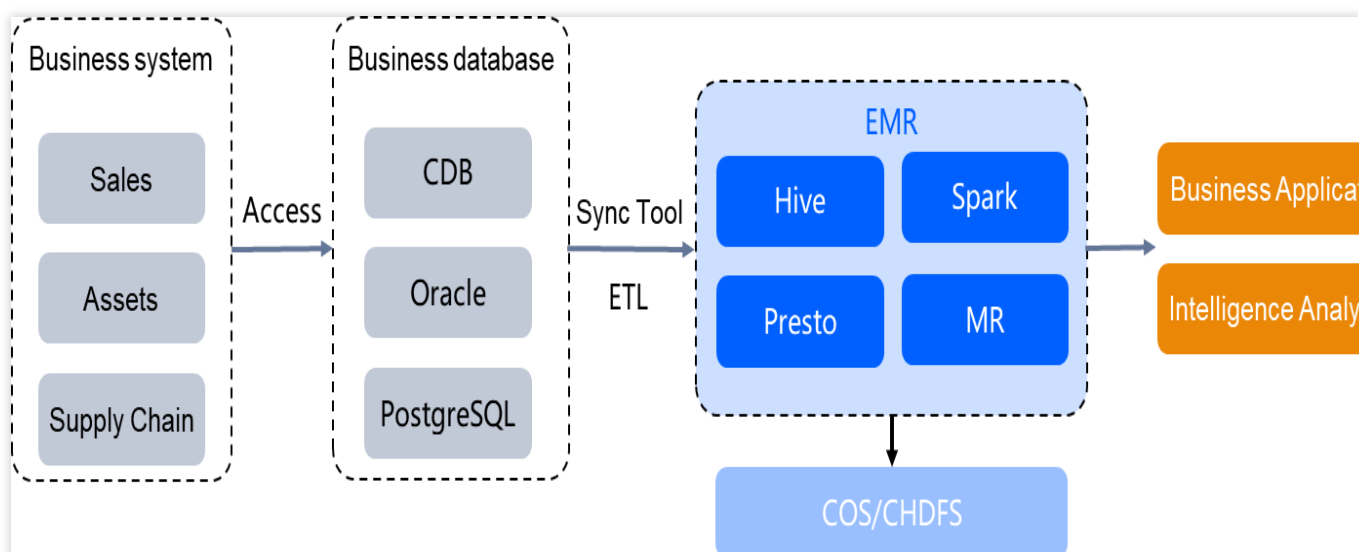
Use Cases

Last updated : 2025-01-03 15:07:59

Tencent Cloud EMR supports a wide range of open-source components and application scenarios. This article introduces the main application scenarios of EMR.

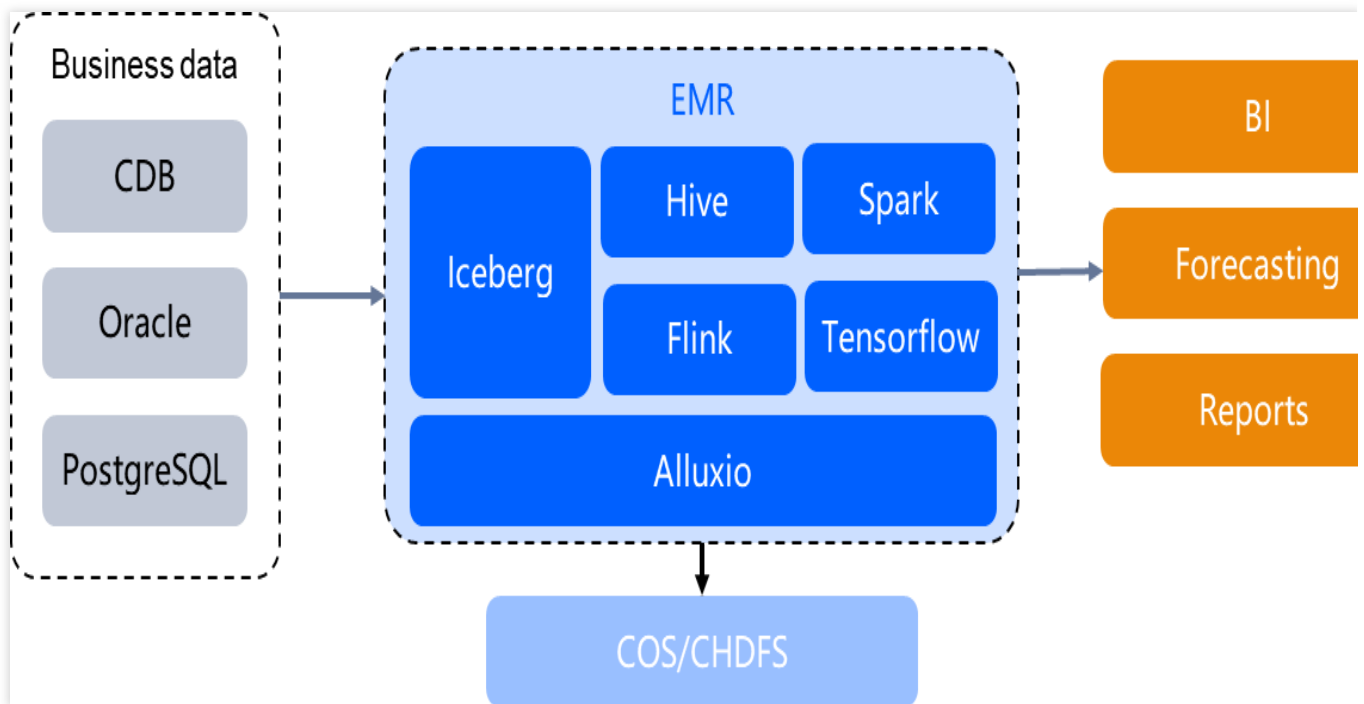
Enterprise Data Warehouse Construction

Summarizing and analyzing business data such as sales, assets, and the supply chain requires integrating different data sources and extracting data from various sources. With EMR's powerful petabyte-level data analysis capabilities and native support for Tencent Cloud COS and CHDFS storage, it provides a high-performance compute-storage separation solution for data warehouses, enabling the computation of massive data to uncover hidden business value for business decision-making.



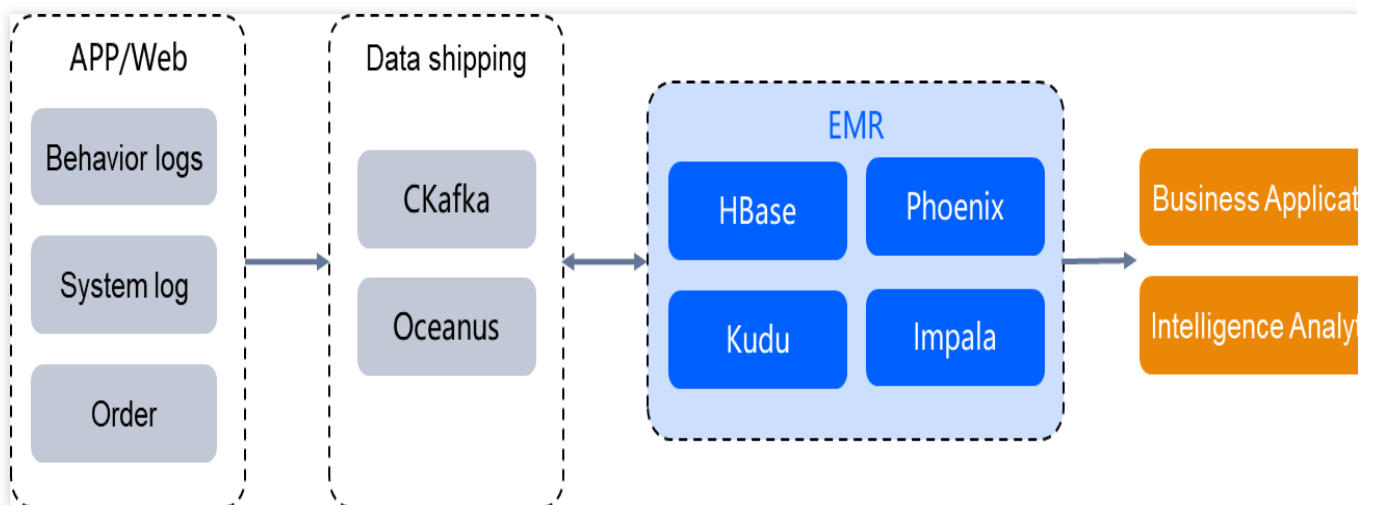
Corporate Data Lake Construction

In the domain of continuously accumulating full business data, enterprises need to store various types of data and adapt to multiple scenario data analysis tasks. EMR's data lake format and caching acceleration capabilities help build a data lake, fully utilizing resources and fitting scenarios such as offline computing, stream computing, interactive analysis, and machine learning. This grants customers higher data agility and lower data analysis costs.



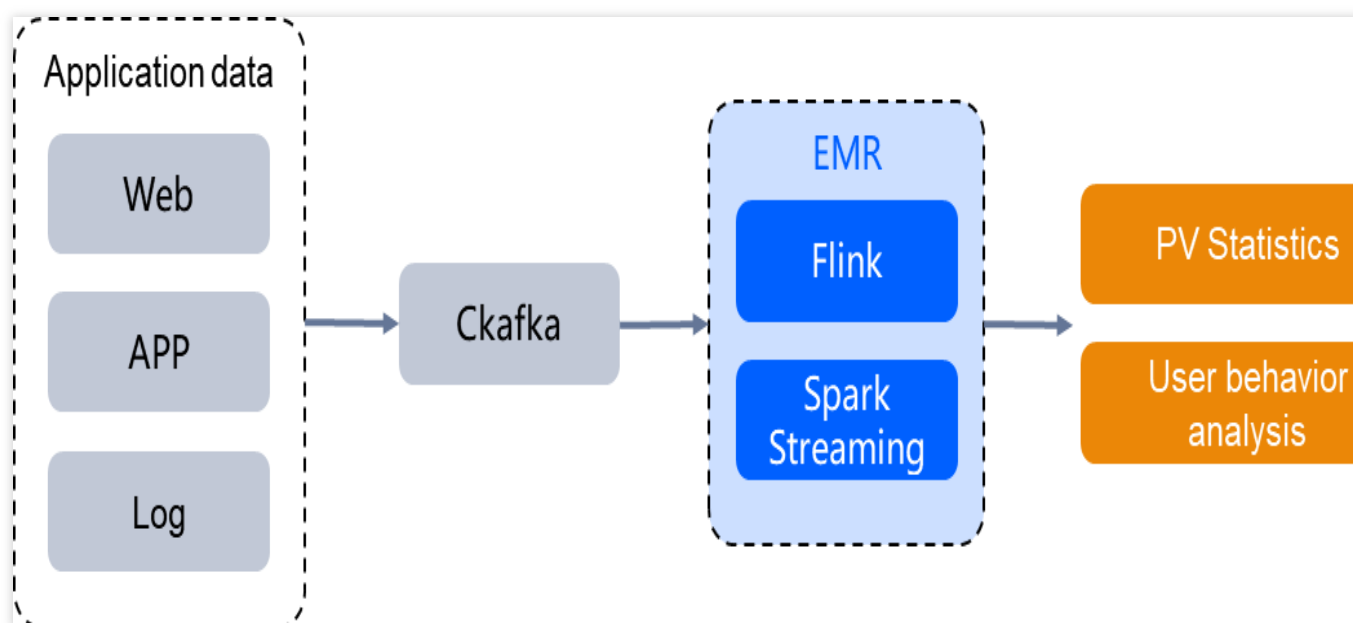
High Concurrency Online Data Query

Efficiently analyzing structured or semi-structured data such as user behavior, system logs, and orders requires collecting various business data from online websites, apps, and system user behaviors and logs. Through EMR's rich computing components, minute-level cluster construction, and horizontal scalability, it supports online business real-time queries, enhancing business response efficiency.



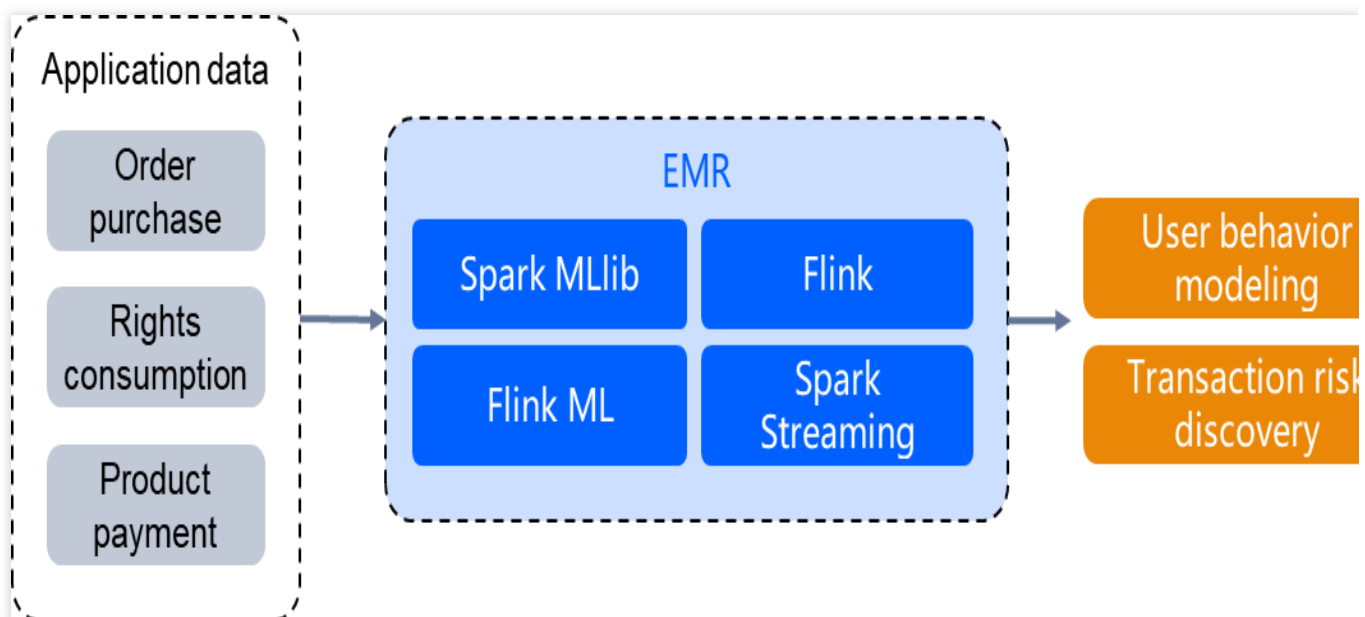
Real-time Stream Computing

In the domain of building real-time computing business, leveraging EMR cloud stream computing services enables real-time analysis construction within minutes. This facilitates real-time aggregation and analysis of user behavior data, helping improve user experience. By simultaneously setting up batch and stream processing systems, it realizes batch and stream integration, reduces resource investment, enhances data processing speed, timely analyzes business operation effectiveness, and quickly adjusts business strategies, supporting the better development of mainstream businesses.



Data Mining and Analysis

In scenarios requiring rapid computation and data mining capabilities, such as real-time risk control and real-time recommendation, EMR can provide efficient cloud stream computing services and data mining component support. This helps monitor abnormal transactions, quickly detect financial vulnerabilities, ensure funds security, and build user profiling models, enabling businesses to accurately and timely understand their user base.



Constraints and Limits

Last updated : 2023-12-27 09:47:39

Before using EMR, carefully read and understand the following use limits:

To ensure the cluster network security, new clusters will be placed in the same VPC. Do not change the VPC of an existing cluster or node; otherwise, the cluster network may fail.

When you create a cluster, EMR can help you create a new security group, or you can also manually select an existing EMR security group. Make sure that the selected security group has the necessary inbound/outbound rules for EMR. Do not delete or change the security group in use after creating the cluster; otherwise, cluster communication may fail, thus affecting the service.

Plan the storage space of nodes in advance based on your business needs, and promptly add storage nodes, as an insufficient storage space may cause data and node risks. Currently, core, task, and router nodes in the EMR cluster support mounting multiple cloud disks, and a node can have up to 15 cloud disks. Clusters in BM 2.0 and local disk models (IO and D series) don't support mounting multiple cloud disks.

When using EMR, do not perform operations in the CVM console, such as shutdown, restart, VPC switch, security group rule adjustment, so as to avoid cluster exceptions. OS reinstallation, instance termination, configuration adjustment, renewal, and billing mode change are also not recommended. You can perform necessary cluster maintenance operations in the EMR console.

Public IPs can increase the possibility of master nodes being attacked, so you need to manage and monitor the risks. EIPs (including the IPs on the secondary ENI) will be retained after the cluster is terminated, and the idle IPs will continue to incur fees. If you don't need to retain them, release them on the corresponding resource management page.

When you create a cluster, EMR provides component initialization parameters for general scenarios. Before you use components such as HDFS and HBase, we recommend you check whether the component parameters meet the needs of your business scenarios. To get the component initialization guide, contact us.

Keep the host login password of your EMR cluster secure. If you configure passwordless login for cross-node access, Tencent Cloud security services may detect vulnerability risks and prompt you.

Even if an exception occurs in your cluster, the cluster will continue to be billed. In this case, we recommend you promptly contact us for assistance. If we need to log in to your cluster for troubleshooting, we will gain your consent to request your account and password.

When you use or maintain your EMR cluster, some unexpected operations may render it unavailable or unstable, and you will receive a risk warning before performing such operations in the console. This document lists some of the prohibited and risky operations:

Prohibited Operations

Operation	Risk
Change the private IP of an EMR node in CVM	Node communication exception or cluster unavailability
Modify the security group of a CVM node while the cluster is running	Node communication exception or component service unavailability
Delete an existing process, application, or file from a node	Cluster/Component service unavailability
Delete or modify the `hosts` file in the `/etc` directory	Service exception, as the cluster cannot be associated with the service on the node
Delete or modify the HDFS metadata file `edit log`	HDFS cluster unavailability
Manually modify the data in the Hive metadatabase	Service exception caused by a Hive data parsing error
Delete a ZooKeeper data directory	Failure of dependent components

Risky Operations

Operation	Risk	Suggestion
Shut down or restart an EMR cluster node in CVM	Service unavailability	Check whether the operation is necessary and read the CVM operation limits
Mount a disk to an EMR node in the CVM console	Disk unavailability caused by EMR's failure to recognize and initialize the disk	Do this through core node scaling or under technical guidance
Unmount a disk from an EMR node in the CVM console	Data loss or cluster unavailability	Do this through core node scaling or under technical guidance
Modify the component configuration file parameters in CVM	Modified parameter overwrite after service restart	Modify the parameter configuration in the EMR console and seek technical guidance in special circumstances

Delete or modify the `resolv.conf` file in the `/etc` directory	Service exception, as the cluster cannot be associated with the service on the node	Check whether the operation is necessary and perform it under technical guidance
Modify the MetaDB password	EMR depends on the password configured in MetaDB, and after the password is changed, services such as Hive and Ranger will become unavailable	Modify the configuration in the EMR console under technical guidance
Modify the MetaDB floating IP	Hive and Ranger service unavailability, as EMR depends on the IP configured in MetaDB	Modify the configuration in the EMR console under technical guidance
Modify the MetaDB security group	Interruption of the communication between MetaDB and the cluster, or Hive and Ranger service unavailability	Perform the operation under technical guidance

Technical Support Scope

Last updated : 2023-12-27 09:49:04

Tencent Cloud Elastic MapReduce (EMR) cluster resources are owned by users. EMR offers semi-managed cloud services based on these resources. Users have full permissions for managing and operating purchased clusters, and are responsible for daily maintenance of the clusters. This document describes the technical support service for EMR.

Supported Services

Cluster purchase, creation, and termination

You can purchase and create a cluster, which involves software configuration, region and hardware configuration, and basic configuration. You can also terminate a cluster.

Cluster scale-out and scale-in

You can select different node types to complete the entire scale-out or scale-in process.

Cluster configuration changes

After a cluster is created, you can change the model configuration and expand the cloud disk involved.

Cluster service features

You can add optional components, enable or disable services, and use relevant management features.

Cluster alarms and monitoring

You can view the operating status of cluster nodes in the console, set event monitoring rules and inspection times, view the alarm history, and search for logs.

Automatic cluster scaling

You can enable or disable auto scaling. After auto scaling is enabled, you can choose custom scaling or managed scaling.

Assisted Services

Assists in identifying the defects or requirements of the open-source components of EMR, and solving these defects or requirements according to the product iteration schedule. This includes but is not limited to communicating with the open-source community and providing feasible solutions validated by the community and industry. However, due to the nature of open-source components, Tencent Cloud cannot promise to provide any solutions beyond the community's progress. For more information about open-source components, see [Product Releases and Component Versions](#).

Assists in identifying defects or requirements of Tencent Cloud basic products and resources that EMR depends on. Then, the responsible product team will solve the defects or requirements. The products and resources that EMR

depends on are listed below:

Computing products such as Cloud Virtual Machine (CVM) and Cloud Bare Metal (CBM).

Storage products such as local disks, Cloud Block Storage (CBS), and Cloud Object Storage (COS).

Network environment products such as Virtual Private Cloud (VPC), Network ACLs, and security groups.

Services Not Supported

EMR offers a range of reliable and useful Ops tools that cover service configuration management, batch node management, and visualized service Ops. However, EMR does not support Ops for specific clusters or components.

EMR does not support troubleshooting individual jobs when there are no obvious exceptions in the cluster components or confirmed product defects.

EMR does not support non-standard product services such as core node scaling and disk cleanup.

EMR does not support handling issues related to the development of user business applications.

EMR does not support handling issues related to third-party components installed by users.

EMR does not support handling issues related to unstable or unavailable clusters caused by unexpected user actions.

For more information, see [Constraints and Limits](#).

Support Methods

If you need technical support for EMR, please [submit a ticket](#).

Product release

Version Overview

Last updated : 2023-12-27 09:50:17

Product Release Overview

EMR consists of open-source applications in a series of big data ecosystems. It offers six [cluster types](#) for you to deploy as needed.

Product Release Number Format

1. EMR version numbers are in the format of `EMR va.b.c` as detailed below:

The meanings of `a` for different clusters are as follows:

For Hadoop clusters, `a` indicates the Hadoop versions supported by the current version. When `a` is `1` or `2`, Hadoop v2.x is supported; when `a` is `3`, Hadoop v3.x is supported.

For Druid clusters, `a` indicates the Druid versions supported by the current version. When `a` is `1`, Druid v0.17.x is supported.

For ClickHouse clusters, `a` indicates the ClickHouse versions supported by the current version. When `a` is `1`, ClickHouse v19.x and v20.x are supported.

For Kafka clusters, `a` indicates the Kafka versions supported by the current version. When `a` is `1`, Kafka v1.x is supported.

For Doris clusters, `a` indicates the Doris versions supported by the current version. When `a` is `1`, Doris v0.13x is supported.

For StarRocks clusters, `a` indicates the StarRocks versions supported by the current version. When `a` is `1`, StarRocks v2.x is supported.

`b` indicates that the version has new components or supports component version upgrade.

`c` indicates feature optimization.

Caution

The components and their versions bundled with each EMR version are fixed. Currently, neither selecting multiple versions of a component nor changing a component version in one EMR version is supported. For example, Hadoop v2.8.5 and Spark v3.2.1 are built into EMR v2.7.0.

Once a version of EMR is selected for cluster creation, the EMR and component version used by the cluster will not be automatically upgraded. For example, if EMR v2.7.0 is selected, then Hadoop will always be v2.8.5, and Spark will

always be v3.2.1. Even if EMR is upgraded to v2.8.0, Hadoop is upgraded to a higher version, and Spark is upgraded to v3.3.0 afterward, the previously created cluster will not be affected, and only new clusters will use the new versions. When you upgrade the cluster through data migration, for example, from EMR v2.6.0 to EMR v2.7.0, in order to avoid issues such as version incompatibility or environment changes, be sure to test the tasks to be migrated and ensure that they can work properly in the new software environment.

EMR v2.4.0 comes with Kona (based on OpenJDK8). We have developed and improved Kona based on the characteristics of cloud scenarios.

Overview of Component Versions

Last updated : 2023-12-27 09:49:54

Deprecated EMR Versions

We have deprecated some earlier EMR versions because their open-source components are too old to support new features of their communities. You can't create new clusters with deprecated versions, but you can still scale in or out existing clusters.

Deprecated EMR on CVM versions:

Hadoop cluster type: EMR v1.3.1, EMR v2.0.1, EMR v2.1.0, EMR v2.2.0, EMR v2.4.0, EMR v2.5.1, EMR v3.0.0, EMR v3.2.0, and EMR-TianQiong v1.0.0.

Druid cluster type: Druid v1.0.0.

For full-feature access and better stability, we recommend you create clusters with the latest stable version of each cluster type.

EMR on CVM Release Notes

EMR on CVM supports four cluster types: Hadoop, Druid, Kafka, and StarRocks. For Hadoop clusters, the Standard edition and JDK 11-Beta edition are available.

Hadoop Standard v2.x supports the following component versions:

Component	EMR v2.7.0	EMR v2.6.0	EMR v2.5.0	EMR v2.3.0
Release time	July 2022	July 2021	September 2020	May 2020
HDFS (required)	2.8.5	2.8.5	2.8.5	2.8.5
YARN (required)	2.8.5	2.8.5	2.8.5	2.8.5
ZooKeeper (required)	3.6.3	3.6.1	3.6.1	3.5.5
OpenLDAP (required)	2.4.44	2.4.44	-	-
Knox (required)	1.6.1	1.2.0	1.2.0	1.2.0
Tez	0.10.1	0.9.2	0.9.2	0.9.2
Hive	2.3.9	2.3.7	2.3.7	2.3.5

Spark	3.2.1	3.0.2	3.0.0	2.4.3
Livy	0.8.0	0.8.0	0.7.0	0.7.0
Kyuubi	1.4.1	1.4.1	-	-
Kylin	4.0.1	2.5.2	2.5.2	2.5.2
Presto	-	-	-	0.228
Trino (PrestoSQL)	385	332	332	-
Kudu	1.15.0	1.12.0	1.12.0	-
Impala	3.4.0	3.4.0	2.10.0	2.10.0
Storm	1.2.3	1.2.3	1.2.3	1.2.3
Flink	1.14.3	1.12.1	1.10.0	1.9.2
HBase	2.4.5	1.4.9	1.4.9	1.4.9
Phoenix (integrated in HBase)	5.1.2	4.14.0	4.13.0	4.13.0
Alluxio	2.8.0	2.5.0	2.3.0	1.8.1
Iceberg	0.13.0	0.11.0	-	-
Hudi	0.11.0	0.7.0	-	0.5.1
Hue	4.10.0	4.6.0	4.6.0	4.6.0
Oozie	5.2.1	5.1.0	5.1.0	5.1.0
Zeppelin	0.10.1	0.9.1	0.8.2	0.8.2
Superset	1.4.1	0.35.2	0.35.2	0.35.2
TensorFlowSpark	1.4.4	1.4.4	1.4.4	1.4.4
Jupyter (installed with TensorFlow)	4.6.3	4.6.3	4.6.3	4.6.3
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
Flume	1.9.0	1.9.0	1.9.0	1.9.0
Ranger	2.1.0	1.2.0	1.2.0	1.2.0
Kerberos (selectable only during cluster creation)	1.15.0	1.15.0	1.15.0	1.15.0

Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
GooseFS	1.2.0	-	-	-

Hadoop Standard v3.x supports the following component versions:

Component	EMR v3.5.0	EMR v3.4.0	EMR v3.3.0	EMR v3.2.1	EMR v3.1.0
Release time	October 2022	April 2022	September 2021	July 2021	December 2020
HDFS (required)	3.2.2	3.2.2	3.2.2	3.2.2	3.1.2
YARN (required)	3.2.2	3.2.2	3.2.2	3.2.2	3.1.2
ZooKeeper (required)	3.6.3	3.6.3	3.6.1	3.6.1	3.6.1
OpenLDAP (required)	2.4.44	2.4.44	2.4.44	2.4.44	-
Knox (required)	1.6.1	1.6.1	1.2.0	1.2.0	1.2.0
Tez	0.10.2	0.10.1	0.10.1	0.10.0	0.9.2
Hive	3.1.3	3.1.2	3.1.2	3.1.2	3.1.1
Spark	3.2.2	3.2.1	3.0.2	3.0.2	2.4.3
Livy	0.8.0	0.8.0	0.8.0	-	-
Kyuubi	1.6.0	1.4.1	1.1.0	-	-
Kylin	4.0.1	4.0.1	4.0.1	-	-
Presto	-	-	-	-	-
Trino (PrestoSQL)	389	372 (renamed Trino)	350	350	332
Impala	4.1.0	4.0.0	3.4.0	3.4.0	3.4.0
Kudu	1.16.0	1.15.0	1.15.0	1.13.0	1.13.0
HBase	2.4.5	2.4.5	2.3.5	2.3.3	2.3.3
Phoenix (integrated in HBase)	5.1.2	5.1.2	5.1.2	5.0.0	5.0.0
Flink	1.14.5	1.14.3	1.12.1	1.12.1	1.10.0

Hue	4.10.0	4.10.0	4.10.0	4.4.0	4.4.0
Oozie	5.2.1	5.1.0	5.1.0	5.1.0	5.1.0
Zeppelin	0.10.1	0.10.1	0.9.1	0.9.1	0.8.2
Superset	1.5.1	1.4.1	1.4.1	-	-
Alluxio	2.8.0	2.8.0	2.5.0	2.5.0	2.3.0
Iceberg	0.13.1	0.13.1	0.11.0	0.11.0	-
Hudi	0.12.0	0.11.0	0.8.0	-	-
Flume	1.10.0	1.9.0	1.9.0	1.9.0	1.9.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7	1.4.7
Ranger	2.3.0	2.1.0	2.1.0	2.1.0	2.0.0
Kerberos (selectable only during cluster creation)	1.15.1	1.15.1	1.15.1	1.51.1	1.15.1
Ganglia	3.7.2	3.7.2	3.7.2	-	-
Delta Lake	2.0.0	-	-	-	-
GooseFS	1.3.0	1.2.0	-	-	-

Hadoop JDK11-Beta supports the following component versions:

EMR v4.x is a beta version based on the JDK 11 environment, so all components are also based on the JDK 11 environment. Currently, you can only create Hadoop clusters with this version.

Component	EMR v4.0.0
Release time	March 2023
HDFS (required)	3.2.2
YARN (required)	3.2.2
ZooKeeper (required)	3.6.3
OpenLDAP (required)	2.4.44
Knox (required)	1.6.1
Tez	0.10.2

Hive	3.1.3
Spark	3.2.2
Livy	0.8.0
Kyuubi	1.6.0
Kylin	4.0.1
Presto	-
Trino (PrestoSQL)	389
Impala	4.1.0
Kudu	1.16.0
HBase	2.4.5
Phoenix (integrated in HBase)	5.1.2
Flink	1.14.5
Hue	4.10.0
Oozie	5.2.1
Zeppelin	0.10.1
Superset	1.5.1
Alluxio	2.8.0
Iceberg	0.13.1
Hudi	0.12.0
Flume	1.10.0
Sqoop	1.4.7
Ranger	2.3.0
Kerberos (selectable only during cluster creation)	1.15.1
Ganglia	3.7.2
Delta Lake	2.0.0

GooseFS	1.3.0
---------	-------

Druid clusters support the following component versions:

Component	Druid v1.1.0
Release time	August 2022
HDFS (required)	2.8.5
YARN (required)	2.8.5
Druid (required)	0.23.0
ZooKeeper (required)	3.6.3
Knox (required)	1.2.0
Superset	1.4.1
Ganglia	3.7.2

Kafka clusters support the following component versions:

Component	Kafka v2.0.0	Kafka v1.0.0
Release time	March 2023	May 2021
Kafka (required)	2.4.1	1.1.1
KafkaManager (required)	2.0.0.2	2.0.0.2
Knox (required)	1.2.0	1.2.0
ZooKeeper (required)	3.6.3	3.6.1

StarRocks clusters support the following component versions:

Component	StarRocks v1.4.0	StarRocks v1.3.0	StarRocks v1.2.0	StarRocks v1.1.0
Release time	April 2023	March 2023	November 2022	August 2022
StarRocks (required)	2.5.3	2.4.3	2.3.2	2.2.2
Knox (required)	1.2.0	1.2.0	1.2.0	1.2.0

EMR on TKE Release Notes

Component	EMR v4.0.0
Release time	May 2023
Spark	3.2.2
Kyuubi	1.6.0
ZooKeeper	3.6.3
OpenLDAP	2.4.44
Knox	1.6.1
Tez	0.10.2
Hive	3.1.3
Trino	389
Ranger	2.3.0
Hue	4.10.0
RSS	0.6.0