

# 弹性 MapReduce

产品简介

产品文档



腾讯云

**【版权声明】**

©2013-2023 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传播全部或部分本文档内容。

**【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。

**【服务声明】**

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。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模式的承诺或保证。

---

## 文档目录

### 产品简介

产品概述

产品优势

产品架构

产品功能

应用场景

集群类型

约束与限制

产品发行版及组件版本预览表

技术支持范围

产品发行版

组件版本概览

版本概述

# 产品简介

## 产品概述

最近更新时间：2023-12-26 17:39:47

弹性 MapReduce (EMR) 是基云原生技术和泛 Hadoop 生态开源技术的安全、低成本、高可靠的开源大数据平台。提供易于部署及管理的 Hive、Spark、HBase、Flink、StarRocks、Iceberg、Alluxio 等开源大数据组件，帮助客户高效构建云端企业级数据湖技术架构。

## 功能特性

弹性 MapReduce 的组件完全源于开源社区，您可以将现有的大数据集群无缝平滑迁移至腾讯云上。弹性 MapReduce 产品中集成了社区中常见的热门组件，包括但不限于 Hadoop、Hive、Hbase、Spark、Presto、Flink、Sqoop、Hue、Iceberg、Druid、Starrocks 等，可以满足您对大数据在线业务、离线/在线数据仓库、云原生数据湖构建、实时流式计算等全方位场景的需求。

弹性 MapReduce 无缝集成了腾讯云对象存储（COS）服务，您可将原本存储于 HDFS 中的文件放置在可无限扩展、存储成本低且高可靠的 COS 中，实现计算存储分离。依托于 COS，您可以在需要的时候创建集群，并在任务完成后销毁集群。与此同时，您无需担心数据的丢失。按需创建的集群，可以大幅度降低您的大数据处理成本。

弹性 MapReduce 定义了5种节点类型：Master 节点、Core 节点、Task 节点、Router 节点和 Common 节点。各类型节点作用，请参见 [节点类型说明](#)。

弹性 MapReduce 目前支持多种资源规格，您可以采用标准型、标准网络优化型、内存型、高 IO 型、计算型、计算网络增强型及大数据机型实例作为计算资源。若您需要在裸金属云服务器上部署集群，可[联系我们](#)。

# 产品优势

最近更新时间：2023-12-26 17:40:15

与自建 Hadoop 集群相比，弹性 MapReduce 能提供更方便、更安全、更可靠的云端 Hadoop 服务。

## 注意

除提供 Hadoop 集群类型外，还支持 Druid 和 ClickHouse 大数据集群，提供更丰富的大数据架构。。

## 灵活

只需几分钟即可获得一个安全可靠的 Hadoop 集群，以运行 Hive、Spark、Presto、Impala、ClickHouse、Druid、Flink 等主流开源大数据计算框架，覆盖用户**交互式 BI、数仓场景、实时计算**等场景的需求。

提供对现有弹性 MapReduce 集群进行快速弹性伸缩的能力，实时调配云端计算资源以应对业务数据的快速波动，节省高昂的预留 IT 硬件成本。

## 可靠

Master 节点容灾设计，备节点秒级拉起，保障大数据服务可用性。

完善的监控体系建设，您可以通过短信渠道秒级感知集群组件及任务的运行异常状况。

支持将 Hive 元数据存放于 MetaDB，元数据可靠性达99.9996%。

支持分析存放于 COS 的高存储耐久性的 PB 级数据。

集群默认开启回收站功能，提供误删除设备的找回机制。

## 安全

可通过便捷的 VPC 网络安全隔离手段规划托管 Hadoop 集群网络策略，支持网络 ACL 和安全组，可从子网和节点维度筛选流量，全方位满足网络安全需求。

腾讯云品质的安全加固服务为 EMR 集群提供一体化的安全服务，涵盖网络防护、入侵检测、漏洞防护等。

提供集群级别的 Kerberos 认证，保障集群访问安全。

## 易用

可以响应业务需求创建不同版本的集群分析 COS 上的同一份数据。

可以借助开箱即用的 Hue、Oozie 等社区组件随心分析位于数据节点或 COS 上的 PB 级数据，无需担心产生任何知识迁移成本。

近千项集群级、组件级监控指标，搭配监控概览页面，提供丰富且清晰易用的监控系统。

灵活支撑云端多机型集群，实现对异构配置集群在扩容、配置下发等场景下的轻松应对，以更优硬件配置应对业务分析挑战。

## 节约成本

通过 EMR 服务，可以按业务曲线随心伸缩托管 Hadoop 集群，缩减高昂的硬件成本。

丰富的运维工具支持，大幅提升运维工作效率，让工程师更专注于业务本身的商业价值，摆脱重复搭建监控、安全、运维工具等基础设施。

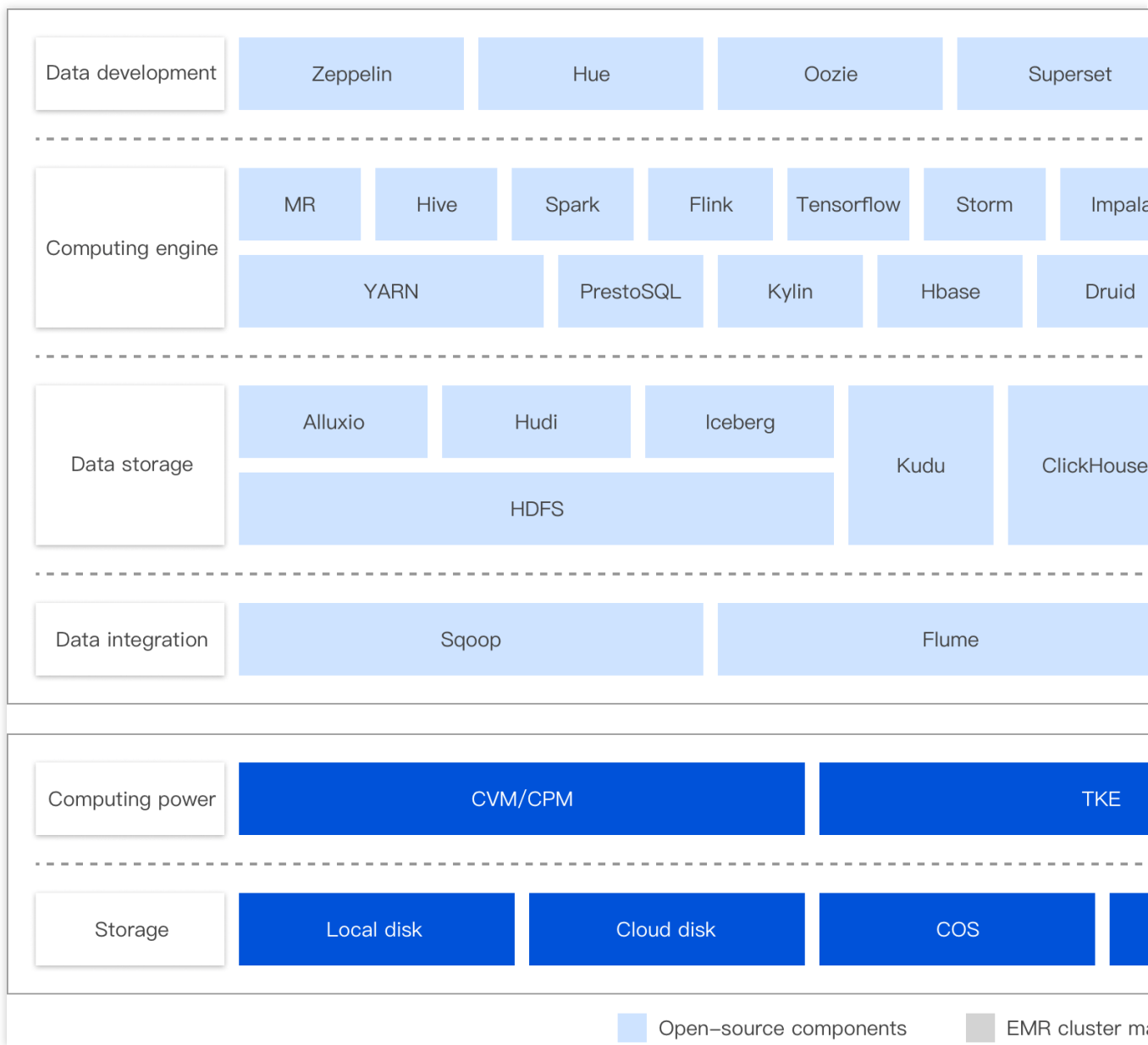
支持温冷数据的对象存储 COS/CHDFS 存储，成本有效降低\*\*28% - 50%\*\*。

结合统一 Hive 元数据库以及统一对象存储，实现跨集群的同数据集分析架构，集群按需创建或销毁，节省集群柔性成本。

# 产品架构

最近更新时间：2023-12-26 17:40:34

腾讯云弹性 MapReduce 产品逻辑架构如图所示：



EMR 主要由开源组件、腾讯云基础设施和集群管理三部分构成：

## 开源组件

集成 Apache 社区 Hadoop、Hive、Spark、Hbase、Presto、Flink、Alluxio、Iceberg 等几十种丰富、前沿的开源大数据组件（详见 [产品发行版及组件版本预览表](#)）。

基于 Iceberg、Alluxio 等开源组件优化，提供 Iceberg Z-Order 算法、Alluxio 透明 URI 等性能及功能增强。

## 腾讯云基础设施

可基于云服务器（CVM）、裸金属云服务器等多种底层计算资源部署，支持容器化部署。

数据可存储在本地盘、云硬盘，或腾讯云对象存储（COS）、云 HDFS 服务上。

私有网络（VPC）、网络 ACL、安全组为 EMR 提供安全隔离的网络环境。

#### 集群管理

快速创建、灵活扩缩、自动伸缩的云端智能部署管理。

服务配置管理、批量节点管理、服务运维可视化操作等丰富便捷的运维工具。

多维度指标监控、事件、巡检、告警、日志搜索等完善的集群监控诊断能力。



# 产品功能

最近更新时间：2023-12-26 17:52:14

弹性 MapReduce 结合云技术和 Hadoop、Hive、Spark、Storm 等社区开源技术，为您提供安全、低成本、高可靠、可弹性伸缩的云端 Hadoop 服务。其主要功能体现在以下方面：

## 弹性伸缩

### 分钟级集群创建

通过控制台数分钟就可创建一个安全、稳定的云端托管 Hadoop 集群。

### 分钟级集群扩缩容

仅需数分钟即可对现有 EMR 集群进行平滑扩缩容，以适应互联网业务需求的快速变化。

### API 支持

支持通过 API 方式便捷的在程序中创建、扩/缩容、销毁 EMR 集群。

## 存储计算分离

### 集群内存存储计算分离

集群内支持按照存储节点、计算节点的模式来规划云端 Hadoop 集群，以支持客户对计算节点的随意伸缩来降低硬件成本。

### 基于 COS 的存储计算分离

支持把待分析海量数据存放于 COS，在通过 COS 规模化效应降低存储成本的同时，您还可以创建不同 EMR 版本分析同一份数据，这将为您带来极度的架构灵活性。

## 运维支撑

### 监控与多渠道告警

提供完善的监控运维体系，对包含 Spark、Hive、Presto 等在内的组件异常和任务异常的秒级感知，以保障大数据集群的稳健运行。

### 技术服务支持

在提供完善技术文档之外，还支持包含工单等渠道在内的技术服务体系，为客户提供完备的技术支持。

## 安全

---

EMR 创建的 CVM 子机同时会创建安全组来限制外网访问。各组件 Web UI 均通过其中一台有外网 IP 的子机进行访问，并且通过用户名和密码进行验证，有外网 IP 的子机安全组只开放 SSH 端口和代理访问端口。

#### 注意

CVM 子机如果更换项目会导致 CVM 安全组丢失。

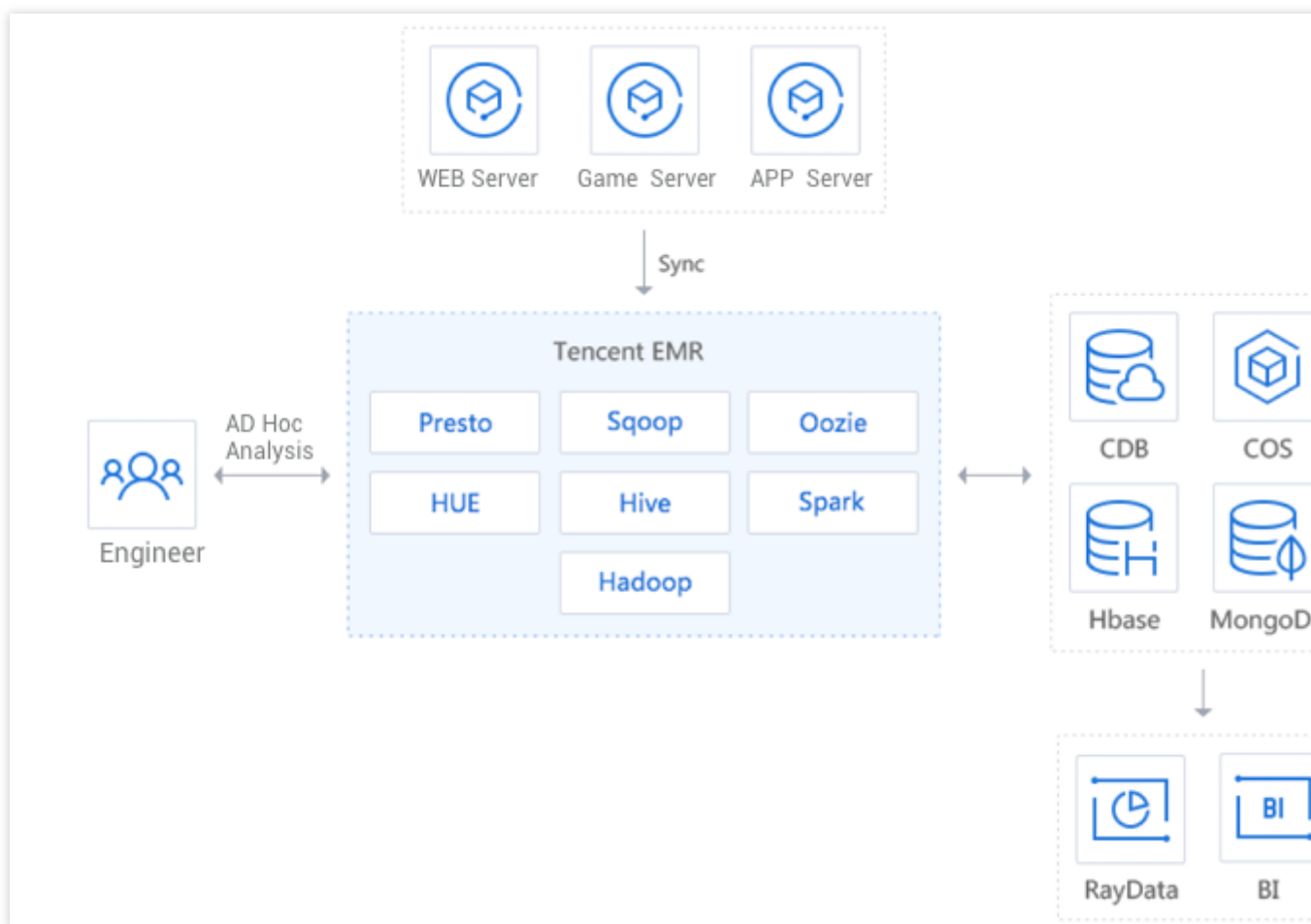
# 应用场景

最近更新时间：2023-12-26 17:55:10

弹性 MapReduce (EMR) 集群应用场景很多, Hadoop 和 Spark 能够支持的场景 EMR 都可以支持, 因为 EMR 本质就是 Hadoop 和 Spark 的集群服务。下文为 EMR 应用的经典场景。

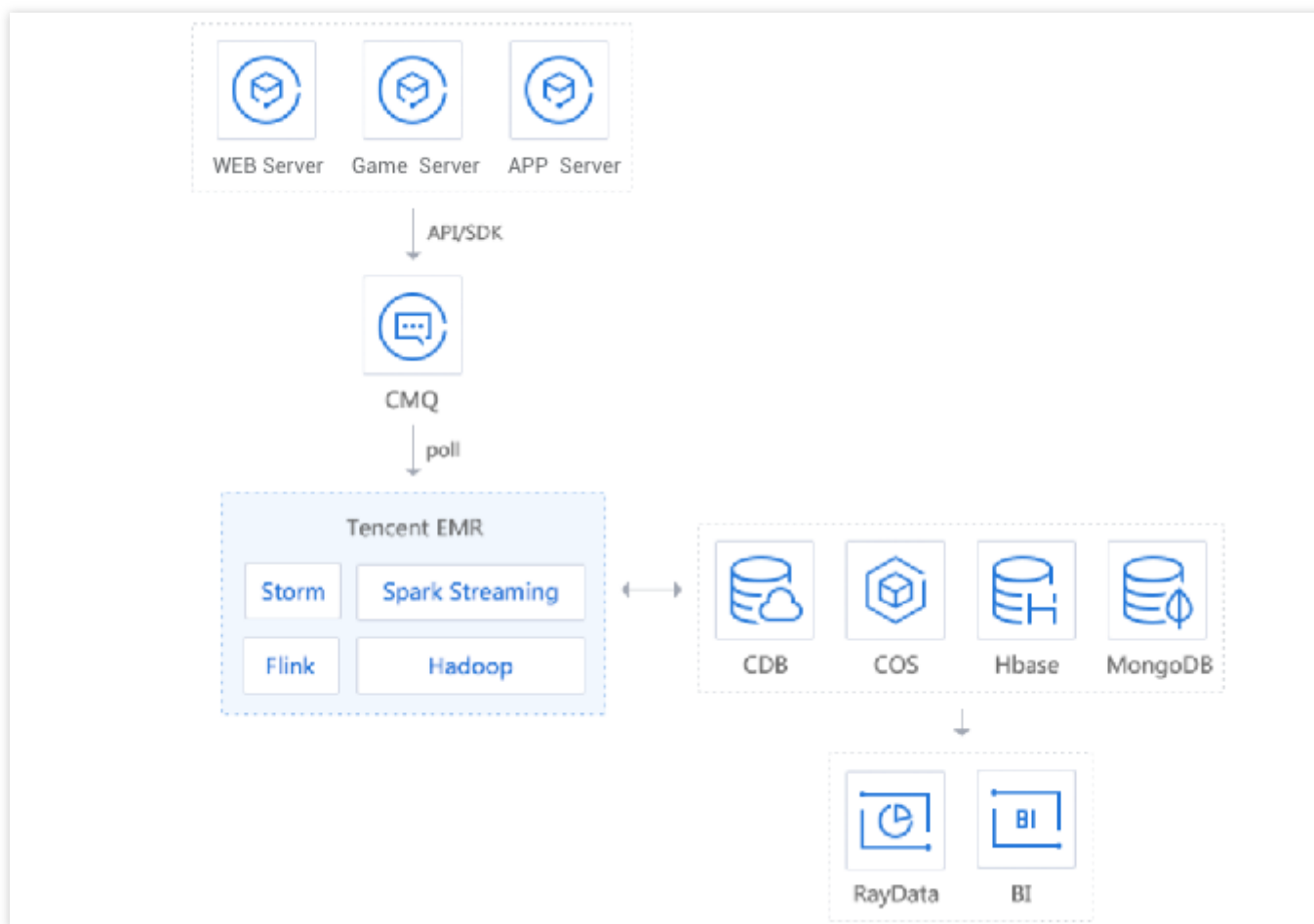
## 离线数据分析

把游戏、Web 应用、手机 App 等业务服务器上的海量日志同步到 EMR 的数据节点或 COS 后, 可借助于 Hue 等工具, 使用 Hive、Spark、Presto 等主流计算框架快速获取数据洞察力。可使用 Sqoop 等工具加载分散于各 TencentDB 或其他存储引擎的数据, 并把分析后的数据同步到 TencentDB, 为 RayData 这样的数据可视化产品提供数据支撑。



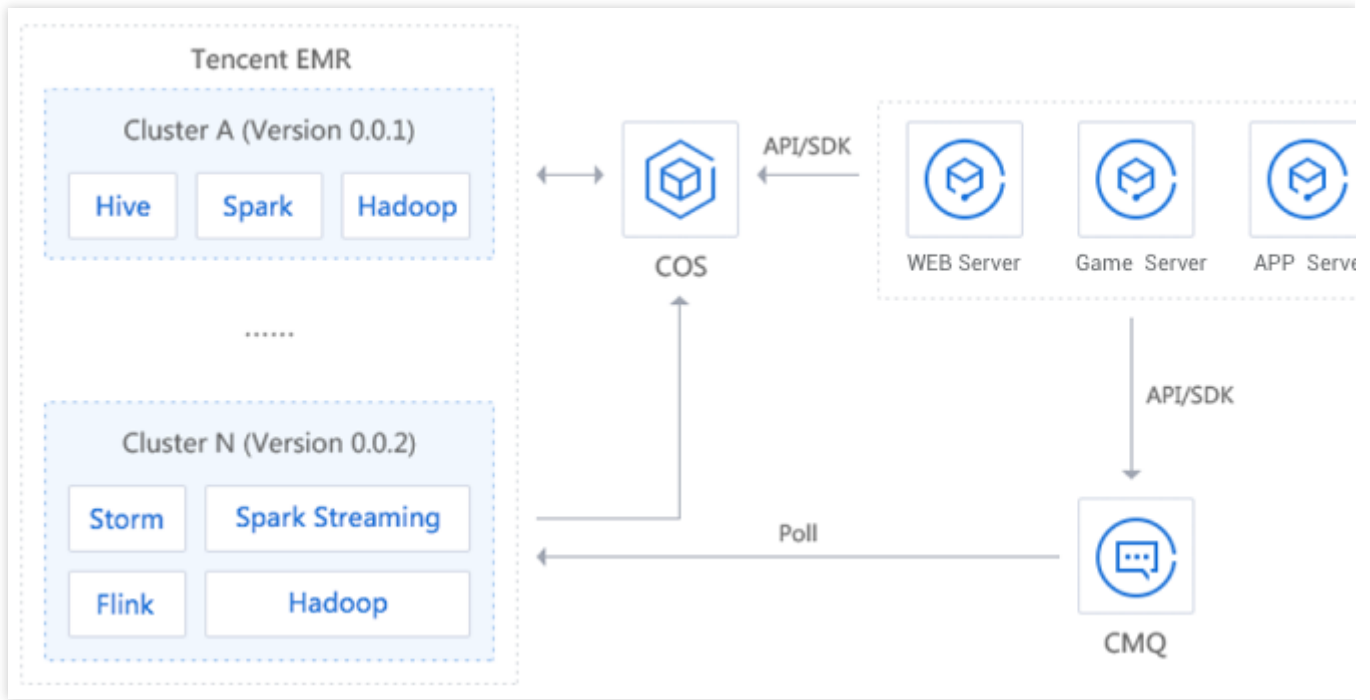
## 流式数据处理

在程序/工具中通过 API、SDK 把位于业务服务器上实时产生的数据 Push 到 CMQ 消息中间件之后，可在 EMR 产品中选择合适的流式数据处理引擎来分析数据，以实现对业务变动的实时告警；还可以把分析结果实时同步到 TencentDB 等存储引擎，以便于通过 RayData 等数据可视化产品对业务状态进行实时可视化检测。



## 分析 COS 数据

可通过 EMR 产品快速分析存储于 COS 上的海量数据，以实现彻底的存储计算分离。通过这样的设计，可充分利用 COS 提供的丰富数据同步工具，同时还可以让多个不同版本 Hadoop 集群分析同一份数据，以满足数据一致性及历史原因导致的多版本 Hadoop 集群共存的问题。



# 集群类型

最近更新时间：2023-12-27 09:47:16

EMR 支持6种集群类型及相应的应用场景，并定义了5种节点类型，不同集群类型及应用场景支持的节点类型及部署节点数、部署服务不同；可根据业务选择不同的集群类型及应用场景创建集群。

## 说明

ClickHouse、Doris、Kafka 集群类型未默认开放，如需要可 [联系工单](#) 开通。

## 集群类型说明

### Hadoop 集群

应用场景	描述	节点部署说明
默认场景	基于开源 Hadoop 及其周边生态组件，提供了海量数据存储、离线/实时数据分析、流式数据计算、机器学习等场景的大数据解决方案。	<p><b>Master 节点</b>：为管理节点，保证集群的调度正常进行；主要部署 NameNode、ResourceManager、HMaster 等进程；非 HA 模式下数量为1，HA 模式下数量为2。注意：部署组件中含 Kudu 时，集群仅支持 HA 模式，Master 节点数量为3。</p> <p><b>Core 节点</b>：为计算及存储节点，您在 HDFS 中的数据全部存储于 Core 节点中，因此为了保证数据安全，扩容 Core 节点后不允许缩容；主要部署 DataNode、NodeManager、RegionServer 等进程。非 HA 模式下数量<math>\geq 2</math>，HA 模式下数量<math>\geq 3</math>。</p> <p><b>Task 节点</b>：为纯计算节点，不存储数据，被计算的数据来自 Core 节点及 COS 中，因此 Task 节点往往被作为弹性节点，可随时扩容和缩容；主要部署 NodeManager、PrestoWork 等进程；可随时更改 Task 节点数，实现集群弹性伸缩，最小值为0。</p> <p><b>Common 节点</b>：为 HA 集群 Master 节点提供数据共享同步以及高可用容错服务；主要部署分布式协调器组件，如 ZooKeeper、JournalNode 等节点。非HA模式数量为0，HA 模式下数量<math>\geq 3</math>。</p> <p><b>Router 节点</b>：用以分担 Master 节点的负载或者作为集群的任务提交机，可以随时扩容和缩容；主要部署 Hadoop 软件包，可选择部署 Hive、Hue、Spark 等软件和进程；可随时更改 Router 节点数，最小值为0。</p>
ZooKeeper	适用于大规模集群建立分布式、高可用性的协调服务。	<b>Common 节点</b> ：主要部署分布式协调器组件 ZooKeeper，部署节点个数必须是单数，最少3个 Common 节点，仅支持高可用（HA）。
HBase	适用于存储海量非结构化数据	<b>Master 节点</b> ：为管理节点，保证集群的调度正常进行；主要

	<p>或半结构化数据，提供高可靠性、高性能、面向列和可伸缩的，实时数据读写的分布式存储系统。</p>	<p>部署 NameNode、ResourceManager、HMaster 等进程；非 HA 模式下数量为1，HA 模式下数量为2。</p> <p>Core 节点：为计算及存储节点，您在 HDFS 中的数据全部存储于 Core 节点中，因此为了保证数据安全，扩容 Core 节点后不允许缩容；主要部署 DataNode、NodeManager、RegionServer 等进程。非 HA 模式下数量<math>\geq 2</math>，HA 模式下数量<math>\geq 3</math>。</p> <p>Task 节点：为纯计算节点，不存储数据，被计算的数据来自 Core 节点及 COS 中，因此 Task 节点往往被作为弹性节点，可随时扩容和缩容；主要部署 NodeManager 等进程；可随时更改 Task 节点数，实现集群弹性伸缩，最小值为0。</p> <p>Common 节点：为 HA 集群 Master 节点提供数据共享同步以及高可用容错服务；主要部署分布式协调器组件，如 ZooKeeper、JournalNode 等节点。非 HA 模式数量为0，HA 模式下数量<math>\geq 3</math>。</p> <p>Router 节点：用以分担 Master 节点的负载或者作为集群的任务提交机，可以随时扩容和缩容；可随时更改 Router 节点数，最小值为0。</p>
<p>Presto</p>	<p>提供开源的分布式 SQL 查询引擎，适用于交互式分析查询，支持对海量数据进行快速查询分析。</p>	<p>Master 节点：为管理节点，保证集群的调度正常进行；主要部署 NameNode、ResourceManager 等进程；非 HA 模式下数量为1，HA 模式下数量为2。</p> <p>Core 节点：为计算及存储节点，您在 HDFS 中的数据全部存储于 Core 节点中，因此为了保证数据安全，扩容 Core 节点后不允许缩容；主要部署 DataNode、NodeManager 等进程。非 HA 模式下数量<math>\geq 2</math>，HA 模式下数量<math>\geq 3</math>。</p> <p>Task 节点：为纯计算节点，不存储数据，被计算的数据来自 Core 节点及 COS 中，因此 Task 节点往往被作为弹性节点，可随时扩容和缩容；主要部署 NodeManager、PrestoWork 等进程；可随时更改 Task 节点数，实现集群弹性伸缩，最小值为0。</p> <p>Common 节点：为 HA 集群 Master 节点提供数据共享同步以及高可用容错服务；主要部署分布式协调器组件，如 ZooKeeper、JournalNode 等节点。非 HA 模式数量为0，HA 模式下数量<math>\geq 3</math>。</p> <p>Router 节点：用以分担 Master 节点的负载或者作为集群的任务提交机，可以随时扩容和缩容；可随时更改 Router 节点数，最小值为0。</p>
<p>Kudu</p>	<p>提供分布式可扩展性的列式存储管理器，支持随机读写和 OLAP 分析对更新较快的数据进行处理。</p>	<p>Master 节点：为管理节点，保证集群的调度正常进行；主要部署 NameNode、ResourceManager 等进程；非 HA 模式下数量为1，HA 模式下数量为2。</p> <p>Core 节点：为计算及存储节点，您在 HDFS 中的数据全部存储于 Core 节点中，因此为了保证数据安全，扩容 Core 节点后不允许缩容；非 HA 模式下数量<math>\geq 2</math>，HA 模式下数量<math>\geq 3</math>。</p>

		<p><b>Task 节点：</b>为纯计算节点，不存储数据，被计算的数据来自 Core 节点及 COS 中，因此 Task 节点往往被作为弹性节点，可随时扩容和缩容；可随时更改 Task 节点数，实现集群弹性伸缩，最小值为0。</p> <p><b>Common 节点：</b>为 HA 集群 Master 节点提供数据共享同步以及高可用容错服务；主要部署分布式协调器组件，如 ZooKeeper、JournalNode 等节点，非HA模式数量为0，HA 模式下数量≥3。</p> <p><b>Router 节点：</b>用以分担 Master 节点的负载或者作为集群的任务提交机，可以随时扩容和缩容；可随时更改 Router 节点数，最小值为0。</p>
--	--	--

## Druid 集群

应用场景	描述	节点部署说明
默认场景	支持高性能实时分析，提供了大数据查询毫秒级延迟，支持多种数据摄入方式，适用于大数据实时查询场景。	<p><b>Master 节点：</b>为管理节点，保证集群的调度正常进行；主要部署 NameNode、ResourceManager等进程；非 HA 模式下数量为1，HA 模式下数量为2。</p> <p><b>Core 节点：</b>为计算及存储节点，您在 HDFS 中的数据全部存储于 Core 节点中，因此为了保证数据安全，扩容 Core 节点后不允许缩容；主要部署 DataNode、NodeManager等进程，非 HA 模式下数量≥2，HA 模式下数量≥3。</p> <p><b>Task 节点：</b>为纯计算节点，不存储数据，被计算的数据来自 Core 节点及 COS 中，因此 Task 节点往往被作为弹性节点，可随时扩容和缩容；主要部署 NodeManager等进程；可随时更改 Task 节点数，实现集群弹性伸缩，最小值为0。</p> <p><b>Common 节点：</b>为 HA 集群 Master 节点提供数据共享同步以及高可用容错服务；主要部署分布式协调器组件，如 ZooKeeper、JournalNode 等节点，非 HA 模式数量为0，HA 模式下数量≥3。</p> <p><b>Router 节点：</b>用以分担 Master 节点的负载或者作为集群的任务提交机，可以随时扩容和缩容；可随时更改 Router 节点数，最小值为0。</p>

## ClickHouse 集群

应用	描述	节点部署说明



场景		
默认场景	提供列式数据库管理系统，适用于大宽表实时分析、实时 BI 报表分析、用户行为分析等高性能数仓分析业务场景。	<p><b>Core 节点：</b>为计算及存储节点；主要部署 ClickHouseServer 进程。</p> <p><b>Common 节点：</b>为 HA 集群 Master 节点提供数据共享同步以及高可用容错服务；主要部署分布式协调器组件 ZooKeeper 节点，非 HA 模式数量为0，HA 模式下数量≥3。</p>

## Doris 集群

应用场景	描述	节点部署说明
默认场景	提供 MPP 分析型数据库产品，对于 PB 数量级、结构化数据可以做到亚秒级查询响应，使用上兼容 MySQL 协议，语法是标准的 SQL。适用于固定历史报表分析、实时数据分析、交互式数据分析等场景。	<p><b>Master 节点：</b>为 Frontend 模块，同时提供 Web UI 的功能；部署 FE Follower、Broker 等进程，非 HA 模式下数量≥1，HA 模式下数量≥3。</p> <p><b>Core 节点：</b>为 Backend 模块，主要提供数据存储功能；部署 BE、Broker 等进程，部署数量≥3。</p> <p><b>Router 节点：</b>部署 Frontend 模块，实现读写高可用；可选择部署 FE Observer、Broker 等进程，可扩容增加 Router 节点，不支持缩容。</p>

## Kafka 集群

应用场景	描述	节点部署说明
默认场景	提供一个分布式、分区的、多副本的、多订阅者，基于 ZooKeeper 协调的消息处理系统，主要适用于异步处理，消息通讯以及流式数据接收和分发场景。	<p><b>Core 节点：</b>提供一个分布式、分区的、多副本的、多订阅者，基于 ZooKeeper 协调的消息处理系统，主要适用于异步处理，消息通讯以及流式数据接收和分发场景。</p> <p><b>Common 节点：</b>为 HA 集群 Core 节点提供数据共享同步以及高可用容错服务，非 HA 模式数量为0，HA</p>

模式下数量≥3。

## StarRocks 集群

应用场景	描述	节点部署说明
默认场景	StarRocks 采用了全面向量化技术，支持极速统一的OLAP分析数据库，适用多维分析，实时分析，高并发等场景等多种数据分析场景。	<p><b>Master 节点：</b>为 Frontend 模块，同时提供 Web UI 的功能；部署 FE Follower、Broker 等进程，非 HA 模式下数量≥1，HA 模式下数量≥3。</p> <p><b>Core 节点：</b>为 Backend 模块，主要提供数据存储功能；部署 BE、Broker 等进程，部署数量≥3。</p> <p><b>Router 节点：</b>部署 Frontend 模块，实现读写高可用；可选择部署 FE Observer、Broker 等进程，可扩容增加 Router 节点，不支持缩容。</p>

# 约束与限制

最近更新时间：2023-12-27 09:48:17

在使用腾讯云 EMR 服务之前，请您详细阅读并了解以下使用限制：

新建集群为保证集群网络安全，集群将放置在同一 VPC 中，请勿随意变更已有集群或节点的 VPC，避免造成集群网络不互通。

在创建集群时，EMR 可以帮您创建一个新的安全组，您也可以手动选择已有 EMR 安全组。请仔细确认手动选择的安全组具备 EMR 必要的出入站规则，并在集群创建后不要随意删除或更改正在使用的安全组，避免造成集群通信异常，影响服务。

请根据业务需要提前规划节点的存储空间，并及时扩充存储节点，避免因存储空间不足造成数据及节点运行风险。目前 EMR 集群 Core、Task、Router 三类节点支持挂载多云盘，单个节点总多云盘累计不超过15块，黑石2.0机型和本地盘机型（IO 系列和 D 系列）集群暂不支持多云盘。

使用 EMR 服务时，请您尽可能避免在 CVM 控制台进行操作，如关机、重启、切换私有网络、调整安全组规则等，以免造成集群异常；OS 重装、实例销毁/变配、续费、变更计费类型等操作也会受限，您可以在 EMR 控制台进行集群维护的必要操作。

公网 IP 会在一定程度上增加 Master 节点被网络攻击的可能性，请您管理和监控相关风险。弹性公网 IP（含辅助网卡上的 IP）会在集群销毁后继续保留，闲置 IP 会产生费用，如不需保留，请到对应资源管理器页面进行释放。

在创建集群时，EMR 提供了满足通用场景的组件初始化参数，在使用组件服务前，建议您检查 HDFS/HBase 等组件参数以确保匹配您的业务场景。如需相关组件初始化指南，可以联系技术支持人员获取。

请您妥善保管 EMR 集群的主机登录密码。在您配置节点间访问免密登录后，腾讯云安全可能会检测到漏洞风险并对您进行提示。

集群出现异常状态时，会仍然持续计费，建议及时联系我方技术支持人员处理。如需登录集群排查的故障，技术支持人员会在征得您同意后请您提供账号密码。

在您使用及维护 EMR 集群时，一些非预期的操作可能会导致集群不可用或不稳定，您在控制台执行部分操作前会有相应的风险提示，本文也为您列举了一些禁止及高危操作：

## 禁止操作

操作	操作风险
在 CVM 中修改 EMR 节点内网 IP	节点通信异常、集群不可用
在集群运行中修改 CVM 节点的安全组	节点通信异常、组件服务不可用
删除节点上已有进程/应用程序/文件	集群/组件服务不可用
删除或者修改/etc 目录下的 hosts 文件	集群关联不到节点上的服务，导致服务异常

删除或者修改 HDFS 元数据文件 edit log	导致 HDFS 集群不可用
手动修改 Hive 元数据库的数据	Hive 数据解析错误，服务异常
删除 ZooKeeper 相关数据目录	相关依赖组件无法运行

## 高危操作

操作	操作风险	建议
在 CVM 中对 EMR 集群的节点进行关机、重启	重启、关机导致服务不可用	确认操作必要性，并详细阅读 CVM 相关操作限制
在 CVM 控制台对 EMR 节点挂载磁盘	EMR 无法识别和初始化，导致磁盘不可用	建议通过 Core 节点扩缩实现，或在技术人员指导下进行
在 CVM 控制台对 EMR 节点卸载磁盘	会导致数据丢失或集群不可用	建议通过 Core 节点扩缩实现，或在技术人员指导下进行
直接在 CVM 上修改组件配置文件的参数	服务重启后，导致修改的参数被覆盖	在 EMR 控制台上修改参数配置，特殊情况请在技术支持人员指导下进行
删除或者修改/etc 目录下的 resolv.conf 文件	集群关联不到节点上的服务，导致服务异常	确认操作必要性，并在技术指导下进行
修改 MetaDB 密码	EMR 依赖 MetaDB 中配置的密码，修改后导致 Hive/Ranger 等服务不可用	在 EMR 控制台同步修改配置，并在技术人员指导下进行
修改 MetaDB 浮动 IP	EMR 依赖 MetaDB 中配置的 IP，修改后导致 Hive/Ranger 等服务不可用	在 EMR 控制台同步修改配置，并在技术人员指导下进行
修改 MetaDB 安全组	导致 MetaDB 与集群通信受阻，Hive/Ranger 等服务不可用	在技术人员指导下进行

# 产品发行版及组件版本预览表

最近更新时间：2023-12-27 09:48:50

## EMR 已停止购买版本说明

部分历史 EMR 发行版因开源组件版本较低，无法体验社区新特性，现已下线。下线的 EMR 发行版不再支持新集群创建，已购的集群仍可正常使用扩缩容等服务。

Hadoop 集群类型已下线版本：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、EMR-TIANQIONG-V1.0.0。

Duird 集群类型已下线版本：DUIRD- V1.0.0。

推荐您使用各集群类型最新稳定发行版进行集群创建，以获得更多的特性及更好的稳定性。

## EMR 标准版本更新记录

标准版本，当前支持 Hadoop 集群、Druid 集群、ClickHouse 集群、Kafka 集群、Doris 集群、StarRocks 集群。

Hadoop2.x 标准版本支持组件产品版本

组件名称	EMR-V 2.7.0	EMR-V 2.6.0	EMR-V 2.5.0	EMR-V 2.3.0
发布时间	2022.07	2021.07	2020.09	2020.05
hdfs (必选组件)	2.8.5	2.8.5	2.8.5	2.8.5
yarn (必选组件)	2.8.5	2.8.5	2.8.5	2.8.5
zookeeper (必选组件)	3.6.3	3.6.1	3.6.1	3.5.5
openldap (必选组件)	2.4.44	2.4.44	-	-
knox (必选组件)	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 (集成在 hbase 中)	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 (伴随 tensorflow 安装)	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 (仅创建时支持选择)	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	-	-	-

## Hadoop3.x 标准版本支持组件产品版本

组件名称	EMR-	EMR-V3.4.0	EMR-	EMR-V	EMR-V
------	------	------------	------	-------	-------

	V3.5.0		V3.3.0	3.2.1	3.1.0
发布时间	2022.10	2022.04	2021.09	2021.07	2020.12
hdfs (必选)	3.2.2	3.2.2	3.2.2	3.2.2	3.1.2
yarn (必选)	3.2.2	3.2.2	3.2.2	3.2.2	3.1.2
zookeeper (必选组件)	3.6.3	3.6.3	3.6.1	3.6.1	3.6.1
openldap (必选组件)	2.4.44	2.4.44	2.4.44	2.4.44	-
knox (必选组件)	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 (改名 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 (集成在 hbase 中)	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	0.10.1	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 (仅创建时支持选择)	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	-	-
deltalake	2.0.0	-	-	-	-
goosefs	1.3.0	1.2.0	-	-	-

## Druid 集群已支持组件产品版本

组件名称	DRUID-V1.1.0
发布时间	2022.08
hdfs (必选组件)	2.8.5
yarn (必选组件)	2.8.5
druid (必选组件)	0.23.0
zookeeper (必选组件)	3.6.3
knox (必选组件)	1.2.0
superset	1.4.1
ganglia	3.7.2

## ClickHouse 集群已支持组件产品版本

组件名称	CLICKHOUSE-V1.2.0	CLICKHOUSE-V 1.1.0	CLICKHOUSE-V 1.0.0
发布时间	2020.09	2020.05	2020.04
clickhouse (必选组件)	20.7.2.30	20.3.10.75	19.16.12.49
zookeeper (必选组件)	3.4.9	3.4.9	3.4.9



superset	0.35.2	0.35.2	-
----------	--------	--------	---

## Kafka 集群已支持组件产品版本

组件名称	KAFKA-V 1.0.0
发布时间	2021.05
Kafka (必选组件)	1.1.1
kafkamanager (必选组件)	2.0.0.2
knox (必选组件)	1.2.0
zookeeper (必选组件)	3.6.1

## Doris 集群已支持组件产品版本

组件名称	DORIS-V 1.2.0	DORIS-V 1.1.0	DORIS-V 1.0.0
发布时间	2022.01	2021.09	2021.05
doris (必选组件)	0.15.0	0.14.0	0.13.0
knox (必选组件)	1.2.0	1.2.0	1.2.0

## StarRocks 集群已支持组件产品版本

组件名称	STARROCKS-V1.1.0	STARROCKS-V 1.0.0
发布时间	2022.08	2022.03
starrocks (必选组件)	2.2.2	2.0.0
knox (必选组件)	1.2.0	1.2.0

# 技术支持范围

最近更新时间：2023-12-27 09:49:14

腾讯云弹性 MapReduce（EMR）集群资源归属于用户，EMR提供基于该资源的半托管云服务能力，用户对集群拥有完全的管理操作权限，集群日常运维管理由用户负责。为了更好的支持使用，现将 EMR 产品技术支持服务规范进行相应的说明。

## 支持的服务

### 集群购买创建及销毁流程

支持客户成功完成软件配置、区域与硬件配置、基础配置等购买创建全部流程，支持销毁。

### 集群扩容和缩容流程

支持客户选择不同的节点类型，完成扩容及缩容全部流程。

### 集群配置变更流程

支持创建成功后客户完成选择变更机型配置，单独选择云数据盘扩容变更配置流程。

### 集群服务功能

支持在可选组件范围内添加新增组件功能，支持启停服务功能及相关管理功能。

### 集群告警监控功能

支持客户在控制台查看集群节点运行状态，支持设置监控事件规则和巡检时间，查看告警历史，搜索日志功能。

### 集群自动伸缩功能

支持开启或关闭自动伸缩，开启后客户可选择自定义伸缩或托管伸缩。

## 协助支持的服务

协助排查 EMR 产品开源组件的缺陷或需求，将根据产品规划迭代排期解决。包括但不限于与开源社区积极沟通、提供社区及业界验证明确可行的解决方案等，但鉴于开源组件性质，腾讯云无法承诺超越社区进度的解决方案，开源组件详情请参见 [产品发行版及组件版本预览表](#)。

协助排查 EMR 产品依赖的其他腾讯云基础产品的缺陷或需求，流转对应产品团队负责解决，依赖的其他产品如下：

云服务器（CVM）、裸金属云服务器等底层计算产品。

本地盘、云硬盘，或腾讯云对象存储（COS）等存储产品。

私有网络（VPC）、网络 ACL、安全组等网络环境。

## 不支持的服务

EMR 提供服务配置管理、批量节点管理、服务运维可视化操作等丰富便捷的运维工具，并保证工具功能的可靠性和可用性，不提供具体集群和组件管理的运维操作。

在集群组件无明显异常或明确产品缺陷的情况下，不负责单个作业的问题排查。

不支持非产品标准能力的服务，例如 Core 节点缩容、磁盘容量清理等。。

不支持客户业务应用开发相关问题处理。

不支持用户自行安装的第三方组件问题处理。

不支持用户进行非产品预期操作导致集群不稳定或不可用问题处理，详情请参见 [约束与限制](#)。

## 支持方式

如果您在使用弹性 MapReduce（EMR）过程中需要相关技术支持，可 [提交工单](#) 联系客服。

# 产品发行版

## 组件版本概览

最近更新时间：2023-12-27 09:50:06

### EMR 已停止购买版本说明

部分历史 EMR 发行版因开源组件版本较低，无法体验社区新特性，现已下线。下线的 EMR 发行版不再支持新集群创建，已购的集群仍可正常使用扩缩容等服务。

EMR on CVM 已下线版本：

Hadoop 集群类型：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、EMR-TIANQIONG-V1.0.0。

Duird 集群类型已下线版本：DUIRD-V1.0.0。

推荐您使用各集群类型最新稳定发行版进行集群创建，以获得更多的特性及更好的稳定性。

### EMR on CVM 发行版本更新记录

EMR on CVM 支持 Hadoop、Druid、Kafka、StarRocks 四种集群类型，Hadoop 集群类型当前提供标准版本和 Jdk11-beta 版本两种选择。

Hadoop 集群标准版本 Hadoop2.x 支持组件版本

组件名称	EMR-V 2.7.0	EMR-V 2.6.0	EMR-V 2.5.0	EMR-V 2.3.0
发布时间	2022.07	2021.07	2020.09	2020.05
hdfs (必选组件)	2.8.5	2.8.5	2.8.5	2.8.5
yarn (必选组件)	2.8.5	2.8.5	2.8.5	2.8.5
zookeeper (必选组件)	3.6.3	3.6.1	3.6.1	3.5.5
openldap (必选组件)	2.4.44	2.4.44	-	-
knox (必选组件)	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 (集成在 hbase 中)	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 (伴随 tensorflow 安装)	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 (仅创建时支持选择)	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 集群标准版本 Hadoop3.x支持组件版本

组件名称	EMR-V3.5.0	EMR-V3.4.0	EMR-V3.3.0	EMR-V3.2.1	EMR-V3.1.0
发布时间	2022.10	2022.04	2021.09	2021.07	2020.12
hdfs (必选)	3.2.2	3.2.2	3.2.2	3.2.2	3.1.2
yarn (必选)	3.2.2	3.2.2	3.2.2	3.2.2	3.1.2
zookeeper (必选组件)	3.6.3	3.6.3	3.6.1	3.6.1	3.6.1
openldap (必选组件)	2.4.44	2.4.44	2.4.44	2.4.44	-
knox (必选组件)	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 (改名 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 (集成在 hbase 中)	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 (仅创建时支持选择)	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	-	-
deltalake	2.0.0	-	-	-	-
goosefs	1.3.0	1.2.0	-	-	-

#### Hadoop 集群JDK11-Beta 版本支持组件版本

EMR-V4.x 为基于 JDK11 环境的 Beta 版本，各组件均基于 JDK11 环境运行，当前仅支持 Hadoop 集群类型。

组件名称	EMR-V4.0.0
发布时间	2023.3
hdfs (必选)	3.2.2
yarn (必选)	3.2.2
zookeeper (必选组件)	3.6.3
openldap (必选组件)	2.4.44
knox (必选组件)	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 (集成在 hbase 中)	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 (仅创建时支持选择)	1.15.1
ganglia	3.7.2
deltalake	2.0.0
goosefs	1.3.0

Druid 集群已支持组件产品版本

组件名称	DRUID-V1.1.0
------	--------------



发布时间	2022.08
hdfs (必选组件)	2.8.5
yarn (必选组件)	2.8.5
druid (必选组件)	0.23.0
zookeeper (必选组件)	3.6.3
knox (必选组件)	1.2.0
superset	1.4.1
ganglia	3.7.2

#### Kafka 集群已支持组件产品版本

组件名称	KAFKA- V2.0.0	KAFKA-V 1.0.0
发布时间	2023.03	2021.05
Kafka (必选组件)	2.4.1	1.1.1
kafkamanager (必选组件)	2.0.0.2	2.0.0.2
knox (必选组件)	1.2.0	1.2.0
zookeeper (必选组件)	3.6.3	3.6.1

#### StarRocks 集群已支持组件产品版本

组件名称	STARROCKS-V1.4.0	STARROCKS-V1.3.0	STARROCKS-V1.2.0	STARROCKS-V1.1.0
发布时间	2023.04	2023.03	2022.11	2022.08
starrocks (必选组件)	2.5.3	2.4.3	2.3.2	2.2.2
knox (必选组件)	1.2.0	1.2.0	1.2.0	1.2.0

## EMR on TKE 发行版本更新记录

组件名称	EMR-V4.0.0
------	------------

发布时间	2023.5
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

# 版本概述

最近更新时间：2023-12-27 09:50:27

## 产品发行版本介绍

腾讯云弹性MapReduce（以下简称：EMR）由一系列大数据生态的开源应用程序组成，提供六种 [集群类型](#) 供客户选择部署。

## 产品发行版本号格式

1. 弹性 MapReduce 采用 EMR-Va.b.c 格式的版本号，详细说明如下：

a. 在不同集群中代表的版本含义如下：

Hadoop 集群中 a 代表当前版本支持的 Hadoop 版本，a 等于1或2为支持 Hadoop 为 2.x 版本，a 等于3表示为支持 Hadoop 3.x 版本。

Druid 集群中 a 代表当前版本支持的 Druid 版本，a 等于1为支持 Druid 为 0.17.x 版本。

ClickHouse 集群中 a 代表支持的 ClickHouse 的版本，a 等于1为支持 ClickHouse 为 19.x 和 20.x 版本。

Kafka 集群中 a 代表当前版本支持 Kafka 的版本，a 等于1为支持 Kafka 1.x。

Doris 集群中 a 代表当前版本支持 Doris 的版本，a 等于1为支持 Doris 0.13x。

StarRocks 集群中 a 代表当前版本支持 StarRocks 的版本，a 等于1为支持 StarRocks 2.x。

b. 代表版本中新增组件或支持组件版本升级。

c. 代表功能优化。

### 注意

每一个版本上捆绑的组件和组件的版本都是固定的。目前还不支持组件的多个不同版本的选择，也不支持用户自行更改组件的版本。例如在 EMR-V2.7.0 版本中内置的是 Hadoop 2.8.5、Spark 3.2.1 等。

一旦选择了 EMR 某个版本创建集群，该集群使用的 EMR 版本和组件版本不会自动升级，例如选 EMR-V2.7.0 版本，那么 Hadoop 就一直保持在 2.8.5 版本，Spark 就一直保持在 3.2.1 版本。后续如果版本升级到了 EMR-V2.8.0 版本，Hadoop 到了更高版本，Spark 到了 3.3.0 版本，也不会影响已创建的集群。只有新的集群才会使用新的镜像。

当您通过数据迁移的方式升级集群版本时（例如，从 EMR-V2.6.0 版本升级到 EMR-V2.7.0 版本），为防止一些升级不兼容、环境变化等问题的出现，请务必测试需要迁移的任务，以确保在新的软件环境中可以正常运行。

EMR-V2.4.0 版本安装 kona（基于 OpenJDK8），基于云场景的支撑及特性，我们在 kona 进行了开发及优化。