

# CECSTACK 5.3.0

## 云大数据服务 CIK 用户指南

---

文档密级：公开

文档版本：01

发布日期：2025-01-23

### 【版权声明】

版权所有 © 中电云计算技术有限公司 2025。保留一切权利。

本文档的版权归中电云计算技术有限公司所有。非经中电云计算技术有限公司书面许可，任何人不得以包括通过程序或设备监视、复制、传播、展示、镜像、上载、下载、摘编等方式或以其他方式擅自使用本文档的任何内容。

### 【商标声明】

 中国电子云 和本文档所示其他中电云计算技术有限公司及/或其他关联公司的商标均为中电云计算技术有限公司及/或其关联公司所有。未经中电云计算技术有限公司及/或其关联公司书面许可，任何人不得以任何形式使用，也不得向他人表明您有权展示、使用或做其他处理。如您有宣传、展示等任何使用需要，您必须取得中电云计算技术有限公司及/或其关联公司事先书面授权。

本文档中出现的其他公司的商标或注册商标，由各自的所有人拥有。

### 【注意】

您购买的产品、服务或特性等应以中电云计算技术有限公司商业合同中的约定为准，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，中电云计算技术有限公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容可能会不定期进行更新。本文档仅作为使用指导，其中的陈述、信息或建议等均不构成任何明示或暗示的担保。

# 前言

## 概述

本文档主要介绍云大数据服务 Cealnsight on K8s（简称 CIK）的产品定位、产品优势、实例规格、使用流程等信息，以便读者全方位的了解云大数据服务 CIK 产品。

## 读者对象

本文档适用于以下读者：

- 维护工程师
- 技术支持工程师
- 系统管理员

## 本书约定

### 符号标志约定

本书采用各种醒目标志来表示在操作过程中应该特别注意的地方，这些标志的意义如下：

 警告	该标志后的注释需给予格外关注，不当的操作可能会对人身造成伤害。
 注意	提醒操作中应注意的事项，不当的操作可能会导致数据丢失或者设备损坏。 “注意”不涉及人身伤害。
 说明	对正文的重点信息进行必要的补充说明。 “说明”不是安全警示信息，不涉及人身、设备及环境伤害信息。
 提示	配置、操作、或使用产品的技巧、窍门。

## 修订记录

文档版本	发布时间	修订说明
01	2025-01-23	第一次正式发布。

# 目 录

<b>1 产品简介</b> .....	<b>1</b>
1.1 什么是云大数据服务 CIK .....	1
1.2 产品优势 .....	1
1.3 产品架构 .....	1
1.4 产品功能 .....	2
1.5 基本概念 .....	2
1.6 实例说明 .....	3
1.6.1 实例规格 .....	3
1.6.2 实例状态 .....	4
1.6.3 服务状态 .....	4
1.6.4 引擎版本 .....	5
<b>2 进入 CIK 实例列表页面</b> .....	<b>6</b>
2.1 简介 .....	6
2.2 前提条件 .....	6
2.3 操作步骤 .....	6
<b>3 集群实例管理</b> .....	<b>7</b>
3.1 创建实例 .....	7
3.2 查看实例列表 .....	8
3.3 删除实例 .....	9
3.4 变配实例 .....	9
<b>4 服务管理</b> .....	<b>11</b>
4.1 概述 .....	11
4.1.1 服务概览 .....	11
4.1.2 服务技术架构&基本概念介绍 .....	12
4.2 查看服务信息 .....	29
4.3 为服务绑定/解绑安全组 .....	29
4.3.1 为服务绑定安全组 .....	30
4.3.2 为服务解绑安全组 .....	30
4.4 为服务绑定/解绑弹性公网 IP .....	31
4.4.1 为服务绑定弹性公网 IP .....	31
4.4.2 为服务解绑弹性公网 IP .....	31
4.5 停止服务 .....	32

4.6 重启服务 .....	32
4.7 开启 HDFS 平衡器 .....	33
4.8 开启 HDFS 安全模式 .....	33
<b>5 服务配置 .....</b>	<b>34</b>
5.1 查看服务配置项 .....	34
5.2 编辑服务配置项 .....	34
5.3 下发服务配置项 .....	34
5.4 下载服务配置项 .....	35
<b>6 用户管理 .....</b>	<b>36</b>
6.1 用户管理 .....	36
6.1.1 新建用户 .....	36
6.1.2 查看用户列表 .....	36
6.1.3 手动同步用户信息 .....	37
6.1.4 删除用户 .....	37
6.1.5 关联用户组 .....	37
6.1.6 重置用户密码 .....	38
6.2 用户组管理 .....	38
6.2.1 新建用户组 .....	38
6.2.2 查看用户组列表 .....	38
6.2.3 同步用户组 .....	39
6.2.4 删除用户组 .....	39
6.2.5 关联用户 .....	40

# 1 产品简介

## 1.1 什么是云大数据服务 CIK

云大数据服务 Cealnsight on K8s（简称 CIK），基于云原生架构，融合分布式存储及计算引擎，满足海量数据的存储、计算及分析需求。同时，通过自主研发的云上大数据运维管理系统，实现平台的全链路监控、服务配置、运维管控一体化，极力打造简单可视的一站式开源大数据云平台。

## 1.2 产品优势

### 易于部署和管理

只需通过简单的操作，即可在几分钟内创建一个 CIK 集群，免去软硬件部署调试的复杂流程。此外，还提供了便捷的集群运营管理工具以及完善的集群监控和告警系统，以满足客户日常的集群运营管理需求。

### 弹性能力

具备良好的横向扩展能力，您可以根据业务需求弹性、灵活的增加节点，随时启停、灵活扩容，适应多变的业务场景，节省集群使用成本。

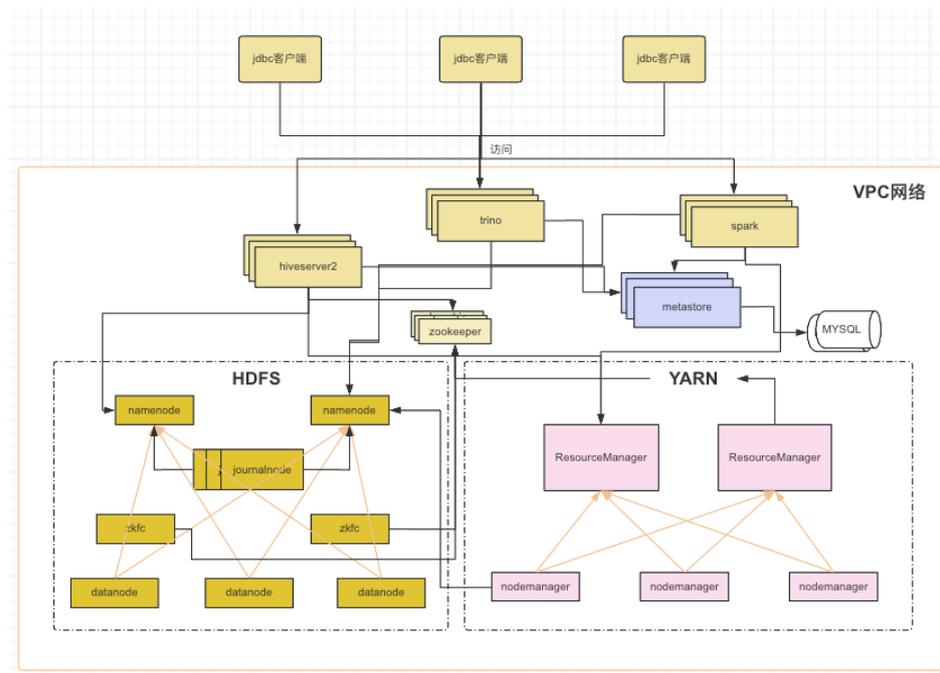
### 技术先进

支撑千台以上大规模国产服务器的集群调度，满足客户可持续业务发展；基于大数据组件的容器化托管，提供集群快速弹性扩缩能力，节点规模从 3 到千台级别，动态伸缩、普惠实用。

## 1.3 产品架构

当前版本仅提供基于 CCOS 底座的大数据集群，包含 ZooKeeper、HDFS、Yarn、Hive、Spark、Trino、Flink、OpenLDAP、Kerberos、Ranger 组件。

图1-1 云大数据服务 CIK 产品总体架构图



## 1.4 产品功能

### 实例管理

- 创建实例需要指定可用区进行创建，存储介质数量以及节点数量可选。
- 查看实例详情：通过控制台查看大数据服务实例列表以及详情信息。
- 释放实例：当您不再使用云大数据服务时，支持删除实例。

## 1.5 基本概念

### 实例

实例是云大数据服务 CIK 最小的集群管理单元，一个实例相当于一个独立运行的大数据集群，您可以在运营中心自行创建并管理云大数据服务的实例。

### 实例规格

云大数据服务 CIK 实例规格，包括 vCPU 核数、内存（GiB）、磁盘和节点个数信息，参见[实例规格](#)。

### 服务

一个可用于独立提供服务的单元，如 ZooKeeper 可以认为是一个服务，在云大数据服务中，多个服务构成一个集群。

## 子服务

服务内部可用于独立提供服务的单元，如 HDFS 组件中包含 NameNode、JournalNode、DataNode 等子服务。

## 本地盘

服务器的本地硬盘设备，NVMe SSD 磁盘类型（直通），可限制 quota 大小，适用于存储 I/O 性能、实时性等要求很高的业务场景。

## 裸盘

服务器的本地硬盘设备，SATA 磁盘类型（直通），适用于数据量大、存储可靠性高的业务场景。

# 1.6 实例说明

## 1.6.1 实例规格

实例规格由组件的规格组成，组件规格清单如下：

组件	子组件	规格族	vCPUs (核)	内存 (GiB)	节点数量	存储
ZooKeeper	ZooKeeper	通用型	1	2	3-3	本地磁盘 10~100GiB
HDFS	NameNode	通用型	4	16	2-2	本地磁盘 20~200GiB
HDFS	JournalNode	通用型	4	16	3-7	本地磁盘 20~200GiB
HDFS	DataNode	通用型	2	8	3-100	裸盘1~12个 单个裸盘10GiB
Yarn	ResourceManager	通用型	2	8	2-2	无
Yarn	NodeManager	通用型	2	8	1-100	裸盘1~2个 单个裸盘10GiB
Yarn	TimelineServer	通用型	1	2	1-1	无
Hive	MetaStore	通用型	4	16	1-3	无
Hive	Server2	通用型	4	16	0-3	无
Spark	HistoryServer	通用型	1	4	1-3	无
Spark	Client	通用型	1	4	1-3	无

组件	子组件	规格族	vCPUs (核)	内存 (GiB)	节点数量	存储
Trino	Coordinator	通用型	4	16	1-1	无
Trino	Worker	通用型	4	16	1-3	裸盘1~2个 单个裸盘10GiB
Flink	Historyserver	通用型	1	4	2	无
OpenLDAP	Ldapserver	通用型	1	4	2-2	裸盘1个
Ranger	Rangeradmin	通用型	1	4	1	无
Kerberos	Kdcserver	通用型	1	4	2	无

## 1.6.2 实例状态

实例状态	说明
创建中	正在创建实例。
变配中	正在变配实例
运行中	实例正常和可用。
删除中	正在删除实例。
已停止	实例已被停止，无法提供服务。
异常	实例不可用。
已释放实例	实例已经被删除。

## 1.6.3 服务状态

组件状态	说明
未知的	无法获取服务状态。
创建中	正在创建服务。
变配中	正在变配服务
运行中	服务正常和可用。
删除中	正在删除服务。
已停止	服务已被停止，无法提供服务。
异常	服务不可用。

不健康	服务可对外提供服务，但是服务中部分节点已出现异常。
-----	---------------------------

## 1.6.4 引擎版本

引擎版本	说明
CIK 1.0.0	包含的服务版本：ZooKeeper 3.6.3、HDFS 3.3.1、Yarn 3.3.1、Hive 3.1.2、Spark 3.2.1、Trino 414、Flink 1.15.2、OpenLDAP 2.5.17、Ranger 2.3.0、Kerberos 1.19.1。
ZooKeeper 3.6.3	云大数据服务 ZooKeeper 基于社区版 ZooKeeper 3.6.3内核优化并托管，兼容开源 ZooKeeper 协议和接口，并提供便捷的运维管理服务。
HDFS 3.3.1	云大数据服务 HDFS 基于社区版 Hadoop 3.3.1内核优化并托管，兼容开源 HDFS 协议和接口，并提供便捷的运维管理服务。
Yarn 3.3.1	云大数据服务 Yarn 基于社区版 Hadoop 3.3.1内核优化并托管，兼容开源 Yarn 协议和接口，并提供便捷的运维管理服务。
Hive 3.1.2	云大数据服务 Hive 基于社区版 Hive 3.1.2内核优化并托管，兼容开源 Hive 协议和接口，并提供便捷的运维管理服务。
Spark 3.2.1	云大数据服务 Spark 基于社区版 Spark 3.2.1内核优化并托管，兼容开源 Spark 协议和接口，并提供便捷的运维管理服务。
Trino 414	云大数据服务 Trino 基于社区版 Trino 414内核优化并托管，兼容开源 Trino 协议和接口，并提供便捷的运维管理服务。
Flink 1.15.2	云大数据服务 Flink 基于社区版 Flink 1.15.2内核优化并托管，兼容开源 Flink 协议和接口，并提供便捷的运维管理服务。
OpenLDAP 2.5.17	云大数据服务 OpenLDAP 基于社区版 OpenLDAP 2.5.17内核优化并托管，兼容开源 OpenLDAP 协议和接口，并提供便捷的运维管理服务。
Ranger 2.3.0	云大数据服务 Ranger 基于社区版 Ranger 2.3.0内核优化并托管，兼容开源 Ranger 协议和接口，并提供便捷的运维管理服务。
Kerberos 1.19.1	云大数据服务 Kerberos 基于社区版 Kerberos 1.19.1内核优化并托管，兼容开源 Kerberos 协议和接口，并提供便捷的运维管理服务。

# 2 进入 CIK 实例列表页面

## 2.1 简介

本节介绍如何在不同的平台进入云大数据服务 CIK 实例列表页面。

---

### 说明

产品界面因所在平台不同而有所差异，部分功能仅支持在 CECSTACK 平台使用，不支持在 CeaKE 平台使用，具体请以登录后的实际界面为准。

---

CeaKE 平台不支持以下功能：

- 绑定和解绑安全组
- 绑定和解绑弹性公网 IP
- 计费模式、专有网络、子网

## 2.2 前提条件

- 已获取系统访问地址。
- 已获取具备该云服务管理或操作权限的用户账号和密码。

## 2.3 操作步骤

- (1) 打开浏览器，在 Web 地址栏输入访问地址，进入登录页面。
- (2) 依次输入用户名、密码、验证码，单击“登录”，进入主页。
- (3) 单击“云服务”或“产品与服务”页签，选择“大数据”分类下的“云大数据服务 CIK”，进入实例列表页面。

---

### 说明

产品界面因所在平台不同而有所差异，具体请以登录后的实际界面为准。

---

# 3 集群实例管理

## 3.1 创建实例

### 简介

本章节介绍如何通过云大数据服务控制台创建实例。您可以根据业务需要选择适合的实例规格、存储空间等。

### 操作步骤

- (1) 在实例列表页面，单击“新建大数据实例”，进入实例创建页面。
- (2) 根据参数说明配置以下参数，完成后，单击“下一步”确认配置。

参数	说明
集群名称	系统会自动生成实例名称，支持自定义。 实例名称必须以中文或英文开头，最多包含64个字符，可以包含中文、英文、数字、点（.）、下划线（_）、中划线（-）。
描述	实例描述信息，最大为255个字符。
租户	实例所属的租户。
部门	实例所属的部门。
资源集	实例所属的资源集。
区域	实例所属的区域。 不同区域内的产品内网不互通，实例创建后不支持更换区域，请谨慎选择。
计费模式	<ul style="list-style-type: none"><li>包年包月：是一种预付费模式，按订单的购买周期计费，适用于可预估资源使用周期的场景，价格比按需计费模式更优惠。</li><li>按需计费：是一种先使用后付费的付费模式，按实际开通时长以小时为单位进行收费。</li></ul>  说明 该参数仅适用于CECSTACK平台。
可用区	实例所属的可用区。 可用区是指同一区域内，电力和网络互相隔离的物理区域，不同可用区之间确保故障隔离，即某个可用区发生故障不会影响到其他可用区。
CPU架构	实例所属主机的芯片架构，支持X86和ARM两种类型。
专有网络	实例所属的专有网络，可以实现不同业务之间的网络隔离。 可以单击“新建VPC”，跳转至VPC新建页面进行创建。  说明 该参数仅适用于CECSTACK平台。

参数	说明
子网	实例所属的子网。通过子网可以为实例提供与其他网络隔离且独享的网络资源，提高网络安全性。 可以单击“新建子网”，跳转至子网新建页面进行创建。  说明 该参数仅适用于CECSTACK平台。

(3) 进行容器实例配置，根据参数说明配置以下参数，完成后，单击“下一步”。

参数	说明
集群类型	集群类型，默认选择Hadoop集群。
集群版本	集群版本，默认为1.0.0。
必选服务	固定选择的服务，不可取消，目前必须服务包含OpenLDAP、ZooKeeper、HDFS、Yarn，单击服务右侧的“选择规格”功能，可选择规格、节点数量、磁盘大小和磁盘数量。
可选服务	可按需选择的服务，目前可选服务包含Ranger、Hive、Trino、Spark、Flink，单击服务右侧的“选择规格”功能，可选择规格、节点数量、磁盘大小和磁盘数量。权限服务Ranger只能在创建集群时开启，集群创建后不能删除。
Kerberos身份认证	启用Kerberos高安全认证服务后，集群中各组件均会通过Kerberos进行认证。Kerberos服务仅支持在创建集群时开启，完成集群创建后则不支持开启或关闭。

(4) 确认信息无误后，单击“确定新建”。

## 操作结果

创建成功后，显示于实例列表页面，状态由“创建中”变为“运行中”，表示实例状态正常。

## 后续操作

实例创建成功后，您可以在列表页面，对目标实例进行查看和管理。

## 3.2 查看实例列表

### 简介

您可以在实例列表页面创建实例，根据筛选条件查询指定的目标实例，并对目标实例进行删除操作。此外，您还可以修改实例名称、复制实例ID、定义显示列、刷新页面等。

### 操作步骤

在实例列表页面，可以进行以下操作：

任务	说明
创建实例	单击“新建大数据实例”，新建实例。

查询实例	选择筛选条件，输入关键字单击  模糊查询目标实例。
编辑实例名称	将鼠标置于实例名称上，单击  按钮，修改实例名称后单击“确定”。
复制实例ID	将鼠标置于实例名称上，单击  按钮，复制实例ID。
查看实例基本信息	单击实例ID链接，页面跳转至实例基本信息页面。
查看实例所属专有网络	单击专有网络ID链接，页面跳转至实例所属VPC页面。
删除实例	单击目标实例“操作”列的“删除”，删除实例。
定义显示列	单击  按钮，勾选实例列表显示列后，单击“确定”。
刷新页面	单击  按钮，刷新实例列表页面。

## 3.3 删除实例

### 简介

当您不再需要使用该实例时，可通过手动删除实例来释放资源。

### 限制与指导

当实例处于“运行中”、“已停止”、“异常”、“不健康”状态时，允许执行该操作。

### 操作步骤

- (1) 在实例列表页面，单击目标实例“操作”列的“删除”，弹出对话框。  
您也可以单击目标实例名称/ID，进入实例基本信息页面，单击页面右上角“删除”。
- (2) 确认信息无误后，单击“确定”。

### 操作结果

完成实例删除后，刷新实例列表页面，目标实例已从列表中删除。

## 3.4 变配实例

### 简介

当您需要修改实例的配置时，可通过变配功能进行修改，变配实例可以新增服务、删除服务以及修改服务。

### 限制与指导

- 首先实例必须处于“运行中”的状态，其次根据具体的服务状态有不同的限制：需要修改的服务必须处于“运行中”状态；需要创建的服务，它所依赖的服务必须处于“运行中”状态。
- 需要修改的服务，配置的资源必须大于当前，配置内容包含：规格、节点数量、磁盘大小和磁盘数量。

## 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击页面右上角“变配”，进入变更配置页面。
- (3) 确认当前配置，单击“下一步”。
- (4) 选择服务或取消服务、按需选择服务规格，单击“下一步”。
- (5) 确认已选服务实例的配置，单击“确认变更”。

## 操作结果

实例变配成功后，显示于实例列表页面，状态由“变配中”变为“运行中”，表示实例状态正常。

## 后续操作

实例变配成功后，您可以在实例列表页面，对目标实例进行查看和管理。

# 4 服务管理

## 4.1 概述

### 4.1.1 服务概览

云大数据服务 CIK 支持以下服务。

服务名称	简介
ZooKeeper	ZooKeeper是一个分布式、高可用性的协调服务。ZooKeeper提供分布式配置服务、同步服务和命名注册等功能。
HDFS	HDFS (Hadoop Distributed File System) 是一种Hadoop分布式文件系统, 具备高度容错性, 支持高吞吐量数据访问, 可以在处理海量数据 (TB或PB级别以上) 的同时最大可能的降低成本。HDFS适用于大规模数据的分布式读写, 特别是读多写少的场景。
Yarn	Yarn (Yet Another Resource Negotiator, 另一种资源协调者) 是一款用于部署、扩展和管理大型Yarn实例集群的大数据解决方案, 是Hadoop集群的资源管理系统, 可为上层应用提供统一的资源管理和调度。
Hive	Hive是基于Hadoop的一个数据仓库 (Data Warehouse, 简称数仓、DW)。可以将结构化的数据文件映射为一张数据库表, 并提供类SQL查询功能, 是用于存储、分析、报告的数据系统。
Spark	Spark是一款高性能大数据处理引擎, 基于内存进行迭代计算, 支持批处理、流处理、机器学习等场景, 并提供了一整套功能丰富的API, 支持Java、Scala、Python、SQL等编程语言。
Trino	Trino是一款高性能分布式查询引擎, Trino提供了一种简化、灵活的方法来分析PB级数据, 支持多种数据源联邦查询, 适用于交互式分析场景。
Flink	Flink是一个批处理和流处理结合的统一计算框架, 其核心是一个提供了数据分发以及并行化计算的流数据处理引擎。它的最大亮点是流处理, 是业界最顶级的开源流处理引擎。 适用于低时延数据处理 (Data Processing) 场景。高并发 pipeline 处理数据, 时延毫秒级, 且兼具可靠性, 支持 Java、Scala、Python、SQL 等编程语言。
OpenLDAP	OpenLDAP是一个开源的轻量级目录访问协议 (LDAP) 服务器, 它可以帮助组织和管理用户、组和其它目录对象。
Kerberos	Kerberos是一个网络身份验证协议, 它作为一种可信任的第三方认证服务, 通过对称加密的方式执行认证服务, 为客户端、服务端的应用程序提供强大的、严谨的认证服务。

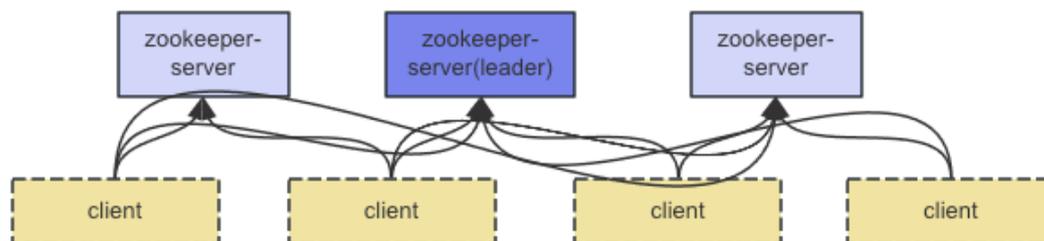
Ranger	Ranger是在Hadoop生态中启用、监控和管理全方位数据安全的一个框架，它可以以集中的方式为HDFS、YARN、Hive、Trino等组件提供统一的权限管理。同时Ranger还提供了一个易用的Web UI，不仅可以方便管理员操作权限策略，还可以对策略进行审计。
--------	--

## 4.1.2 服务技术架构&基本概念介绍

### 4.1.2.1 ZooKeeper

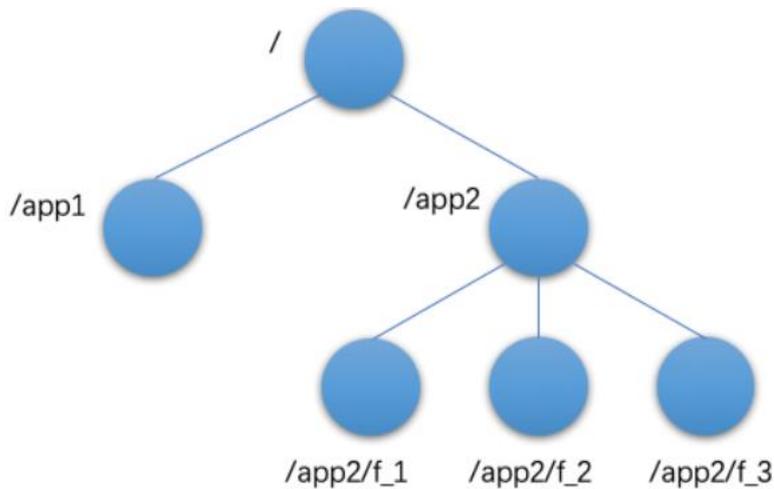
#### 技术架构

一个 ZooKeeper 集群需要由奇数个 ( $2N+1$ ) 节点构成，通过内部选举协议选出一个 Leader 节点，其余为 Follower 节点。写入数据时，由 Leader 节点负责统一协调写请求，至少 ( $N+1$ ) 个节点投票成功才能确定本次数据写入成功，因此至少有 ( $N+1$ ) 个存活的节点才能保证 ZooKeeper 整体服务可用。当 Leader 节点异常退出时，ZooKeeper 集群会重新发起选举，选出新的 Leader 节点，保证整体服务的高可用。



#### 数据组织

ZooKeeper 的数据组织方式与标准文件系统类似，组织成类似文件树的结构，在 ZooKeeper 中使用 znode (ZooKeeper node) 来描述文件，与标准文件系统不同的是，znode 并不区分目录或者文件的概念，每个 znode 都可以存储数据。



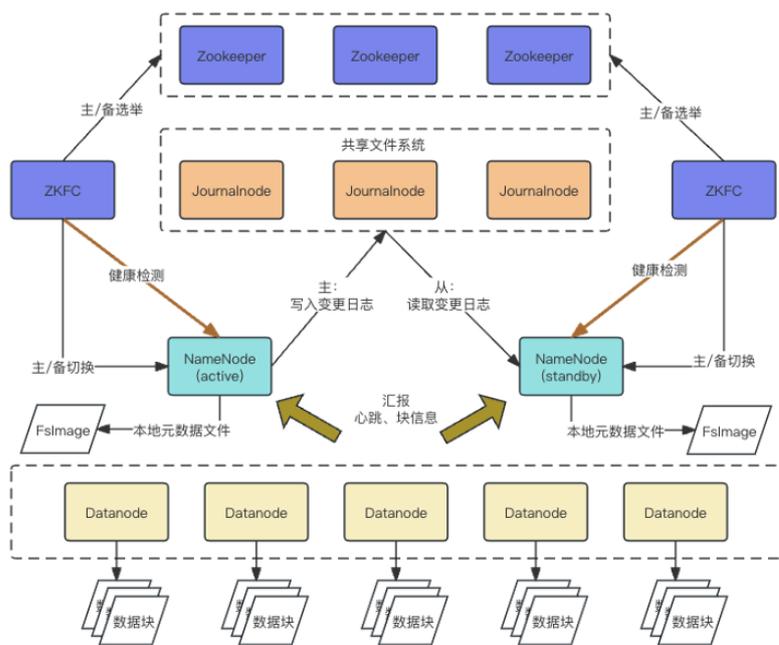
ZooKeeper 作为一个协调服务，znode 主要用来存储协调性数据，例如，服务状态信息和配置信息等，不应把 ZooKeeper 作为文件系统来存储大量数据。

## 4.1.2.2 HDFS

### 技术架构

HDFS 是经典的 Master 和 Slave 架构，每一个 HDFS 集群包括 2 个 NameNode 和多个 DataNode。NameNode 管理所有文件的元数据信息，并且负责与客户端交互。DataNode 负责管理存储在该节点上的文件。每一个上传到 HDFS 的文件都会被划分为一个或多个数据块，这些数据块根据 HDFS 集群的数据备份策略被分配到不同的 DataNode 上，位置信息交由 NameNode 统一管理。

图4-1 HDFS 高可用架构图



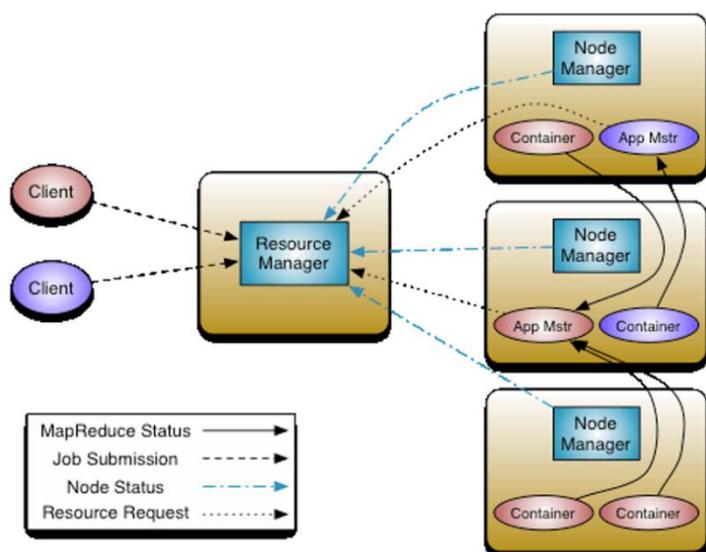
## 基础概念

名称	描述
NameNode	用于管理文件系统的命名空间、维护文件系统的目录结构树以及元数据信息，记录写入的每个数据块（Block）与其归属文件的对应关系。 此信息以命名空间镜像（FSImage）和编辑日志（EditsLog）两种形式持久化在本地磁盘中。
DataNode	DataNode是文件的实际存放位置。 DataNode会根据NameNode或Client的指令来存储或者提供数据块，并且定期的向NameNode汇报该DataNode存储的数据块信息。
JournalNode	在HDFS高可用的架构下，JournalNode主要作用是保证NameNode的高可用性，当某个节点出现故障时，其他节点可以从它的JournalNode中获取数据，继续提供服务，从而避免了系统单点故障的情况。 JournalNode的作用包含： <ul style="list-style-type: none"><li>• 存储数据编辑日志</li><li>• 数据同步和复制</li><li>• 数据恢复</li></ul>
Client	通过Client来访问文件系统，然后由Client与NameNode和DataNode进行通信。Client对外作为文件系统的接口。
Blocks	HDFS将文件拆分成128 MB大小的数据块进行存储，这些Block可能存储在不同的节点上。HDFS可以存储更大的单个文件，甚至超过任何一个磁盘所能容纳的大小。一个Block默认存储3个副本，以Block为粒度将副本存储在多个节点上。此方式不仅提高了数据的安全性，而且对于分布式作业可以更好地利用本地的数据进行计算，减少网络传输。
高可用	CIK中HDFS是高可用的，集群默认会启动2个NameNode，一个是Active NameNode，一个是Standby NameNode，Active NameNode 负责处理DataNode和Client的请求，StandByNameNode与Active NameNode拥有相同的最新的元数据，随时准备在Active NameNode异常时接管其服务，当Active NameNode异常时，Standby NameNode会感知并自动切换成Active状态来处理DataNode和Client的请求。 同时，在高可用的架构下，Standby NameNode会进行FSImage和EditsLog的合并，合并而成的新的FSImage会推送到Active NameNode上，并清理掉旧的FSImage和EditsLog。

### 4.1.2.3 Yarn

#### 技术架构

Hadoop Yarn 是基于 Master 与 Slave 主从架构的分布式集群资源管理系统，其中 RM（ResourceManager）是 Master 组件，负责整个集群的资源管理与任务调度，NM（NodeManager）是 Slave 组件，负责单个节点的任务管理与监控。



Hadoop Yarn 的高可用机制主要通过以下几个方面来确保系统的稳定性和可靠性：

- **ResourceManager High Availability (RM HA)**：RM HA 是通过在不同的节点上启动多个 ResourceManager 进程来实现的。其中一个 RM 被选为 Active，负责接收和处理客户端请求，而其他的 RM 处于 Standby 状态。当 Active RM 发生故障或需要升级时，Standby RM 会接管其职责成为 Active RM，从而避免单点故障。这种切换过程是无缝的，对于客户端和正在运行的任务来说是透明的。
- **ResourceManager Restart (RM 重启)**：为了保证集群升级或重启后应用能够自动恢复，YARN 使用了 RM 有状态重启机制。在 RM 启动时，它会从 ZooKeeper 等分布式存储中加载应用信息和当前状态，并恢复到之前的状态。这样，之前提交的应用可以继续运行，任务也可以继续执行，从而实现了应用的无感知重启。
- **NodeManager Restart (NM 重启)**：NM 有状态重启机制确保在节点升级或重启时，正在运行的任务不受影响。NM 会将运行时的任务 (Container) 信息和状态持续同步到本地存储 (如 LevelDB) 中。当 NM 重新启动时，它会重新加载任务状态，使得之前的任务可以继续执行，从而保证了任务的连续性。

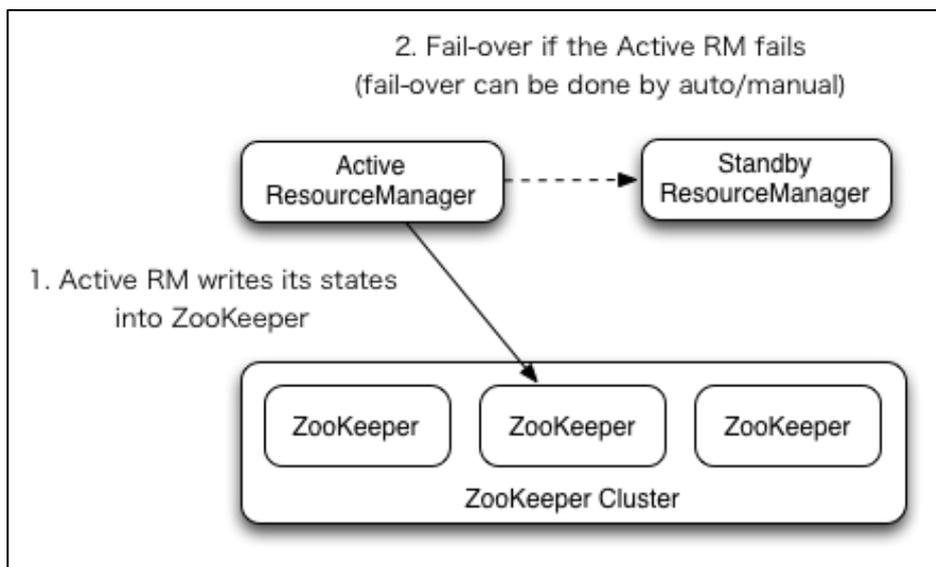
基于以上三大特性，通常情况下，可以很好做到当 RM 单点故障、RM 和 NM 升级、RM 和 NM 重启时，YARN 依然可以提供正常的服务且任务不受影响。

## 底层依赖

ResourceManager 的 HA 通过 Active/Standby 体系实现，其底层通过 ZooKeeper 集群来存储 RM 的状态信息、应用程序的状态。如果 Active 状态的 RM 遇到故障，会通过切换 Standby 状态的 RM 为 Active 来继续为集群提供正常服务。

配置属性	描述
hadoop.zk.address	ZK集群服务地址，用来存储内部状态和内部主节点选举。

图4-2 高可用架构图



## 基础概念

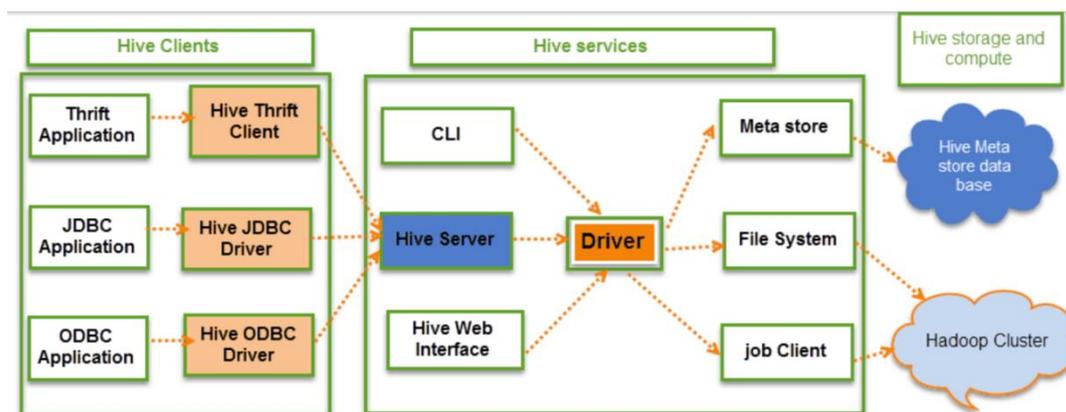
名称	描述
ResourceManager	<p>ResourceManager负责整个集群的资源管理和分配，是一个全局的资源管理系统。NodeManager以心跳的方式向ResourceManager汇报资源使用情况。</p> <p>ResourceManager由两个关键组件Scheduler和ApplicationsManager组成。</p>
NodeManager	<p>NodeManager是每个节点上的资源和任务管理器，它是这台机器的代理，负责该节点程序的运行，以及该节点资源的管理和监控。YARN集群每个节点都运行一个NodeManager。</p> <p>NodeManager定时向ResourceManager汇报本节点资源（CPU、内存）的使用情况和Container的运行状态。当ResourceManager宕机时NodeManager自动连接ResourceManager备用节点。</p> <p>NodeManager接收并处理来自ApplicationMaster的Container启动、停止等各种请求。</p>
ApplicationMaster	<p>用户提交的每个应用程序均包含一个ApplicationMaster，它可以运行在ResourceManager以外的机器上。负责与ResourceManager调度器协商以获取资源（Container）。</p>
Scheduler	<p>在容量和队列限制范围内负责为运行的容器分配资源。Scheduler是一个纯调度器（pure scheduler），只负责调度，它不会监视或跟踪应用程序的状态，也不负责重启失败任务，这些全都交给ApplicationMaster完成。Scheduler根据各个应用程序的资源需求进行资源分配。</p>
TimeLineServer	<p>不依赖于任何应用的Job History Server，负责收集作业的指标，以更通用的方式来检索YARN中当前运行的、以及历史运行的作业。</p>

## 4.1.2.4 Hive

### 技术架构

在 Hadoop 生态系统中，HDFS 用于存储数据，YARN 用于资源管理，MapReduce 用于数据处理，而 Hive 是构建在 Hadoop 之上的数据仓库，其包含以下几方面：

- 使用 HQL 作为数据查询接口。
- 使用 HDFS 作为数据存储，提供元数据管理服务，将存储在 HDFS 上的结构化数据映射成一张结构化的数据表，易于检索查询。
- 提供包括但不限于 MapReduce 在内的其他计算框架进行数据计算。
- 计算任务可调度到 Yarn 上进行执行。



### 基础概念

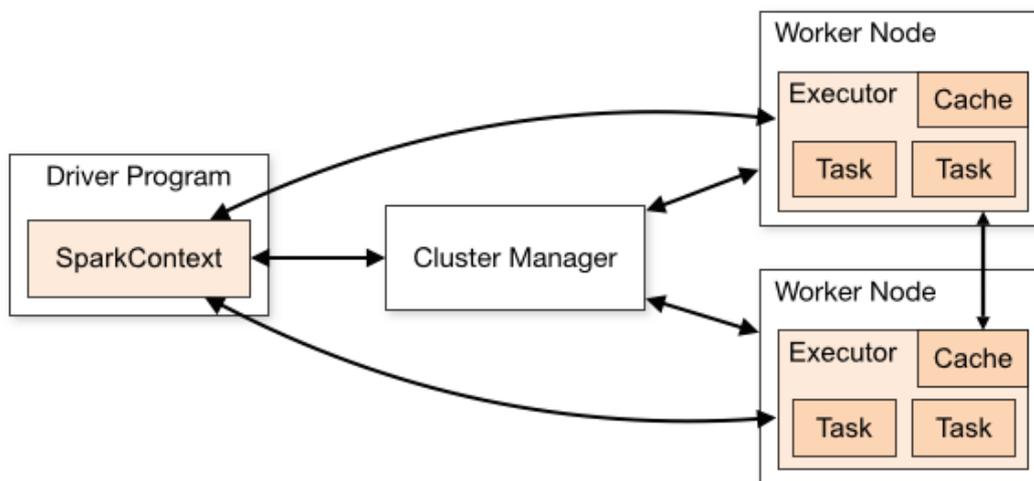
名称	描述
HiveServer2	HiveQL 查询服务器，可以配置为Thrift或者HTTP协议，接收来自JDBC客户端提交的SQL请求，支持多客户端并发以及身份验证。
Hive MetaStore	元数据管理模块，此模块被其他引擎所依赖，用于存储Database和Table等元信息。例如，Spark和Trino均依赖此模块作为其元数据管理。
Hive Client	Hive客户端，直接利用该客户端提交SQL作业，根据其设置运行引擎配置，可以将SQL转换成MR作业、Tez作业。

## 4.1.2.5 Spark

### 技术架构

Spark 整体架构主要由 Driver (SparkContext)、Worker (Executor)、集群资源管理等角色组成。SparkContext 可以连接到多种类型的集群资源管理器 (Spark 自己的独立集群管理器、Mesos、YARN 或 Kubernetes)，这些管理器在应用程序之间分配资源。当连接上资源管理器后，Spark 会向资源管理器申请资源并会在集群的节点上启动 Executors，接着应用程序代码 (由传递给

SparkContext 的 JAR 或 Python 文件定义) 将被分发至 Executors。最后, SparkContext 将任务 Task 发送给 Executor 执行。



## 基础概念

名称	描述
Application	基于Spark构建的用户程序。
Application jar	包含用户的Spark应用程序的jar。在某些情况下, 用户会希望创建一个包含其应用程序及其依赖项的“jar”, 用户的jar不应包括 Hadoop或Spark库, 这些库将在运行时添加。
Driver	运行应用程序的main()函数并创建SparkContext的进程。
Cluster manager	用于获取集群资源的外部服务(例如独立管理器Standalone manager、Mesos、YARN、Kubernetes等)。
Deploy mode	主要用于区分Driver在何处运行。在“cluster”模式下, Driver在集群内部启动运行。在“client”模式下, 提交者在集群外部启动运行Driver。
Worker node	可以在集群中运行应用程序代码的节点。
Executor	在Woker Node上为应用程序启动的进程, 用于运行任务并将数据保存在内存或磁盘存储中。每个应用程序都有自己的Executor。
Task	发送到Executor的工作单元。
Job	由多个任务组成的并行计算作业, 对应Spark RDD的Action操作。
Stage	每个Job作业都被划分为更小的任务集, 称为阶段Stage, Stage之间有依赖关系。
RDD	弹性分布式数据集, 是Spark中最基本的数据抽象, 它代表一个不可变、可分区、里面的元素可并行计算的集合。

## 4.1.2.6 Trino

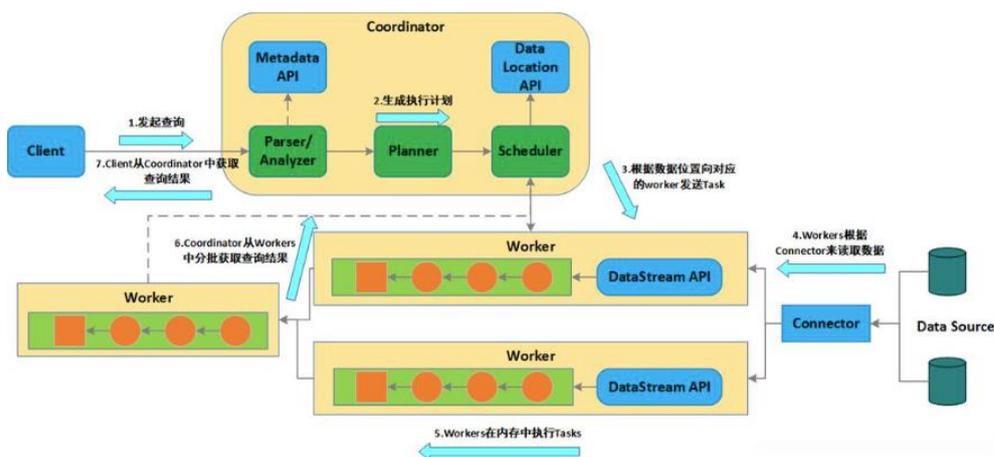
### 技术架构

Trino 是经典的 Master/Slave 架构，一个集群包含一个 Coordinator 节点和多个 Worker 节点。

Coordinator 负责工作：

- 接收用户查询请求，解析并生成执行计划，下发 Worker 节点执行。
- 监控 Worker 节点运行状态，各个 Worker 节点与 Coordinator 节点保持心跳连接，汇报节点状态。
- 维护 MetaStore 数据。

Worker 负责执行下发到任务，通过连接器读取外部存储系统到数据，进行处理，并将处理结果发送给 Coordinator 节点。



### 基础概念

名称	描述
Coordinator	负责解析语句、生成执行计划和管理Trino集群的服务器。
Worker	负责执行任务和处理数据的节点。
Trino Client	提交SQL作业终端。
Connector	Trino通过各种Connector来接入多种外部数据源。Trino提供了一套标准的SPI接口，用户可以使用这套接口开发自己的Connector，以便访问自定义的数据源。一个Catalog通常会绑定一种类型的Connector，在Catalog的Properties文件中设置。Trino内置了多种Connector。
数据模型	数据模型即数据的组织形式。Trino使用Catalog、Schema和Table三层结构来管理数据。
Catalog	一个Catalog可以包含多个Schema，物理上指向一个外部数据源，可以通过Connector访问该数据源。一次查询可以访问一个或多个Catalog。
Schema	相当于一个数据库实例，一个Schema包含多张数据表。
Table	数据表，与一般意义上的数据库表相同。
Statement	即输入的SQL语句。Trino支持符合ANSI标准的SQL语句，由字句、表达式和断言组成。

Query	即查询执行，当Trino接受SQL并执行时，会解析SQL并生成一个查询执行计划。查询由运行在多个Worker上且相互关联的Stage组成。
Stage	一个查询计划可以被分解为多个Stage，每个Stage代表了查询执行计划的一个逻辑部分或步骤。Stage之间存在依赖关系，不同Stage之间通过数据交换进行通信。
Task	Stage的具体执行单元。每个Stage由一个或多个Task组成，每个Task负责处理数据集的一部分，并与其他Task进行数据交换。
Split	Split是对数据源的逻辑划分，是实现数据并行处理的基本单元。每个Split对应数据源的一小部分数据，比如HDFS上的一个文件块或者表分区。Task会按照Split来进行数据处理，每一个Task会处理分配给它的一个或多个Split。
Operator	查询计划中最细粒度的操作单元，它们定义了对数据记录的具体操作，如过滤（Filter）、投影（Project）、聚合（Aggregation）等。 在执行过程中，多个Operator会按照优化后的执行计划组合成一个Operator树，这个树结构会分布在各个Task中执行，并通过网络通信进行跨Task的数据交换。
Driver	主要用于描述运算过程中的执行上下文，它包含了Operator实例的集合。

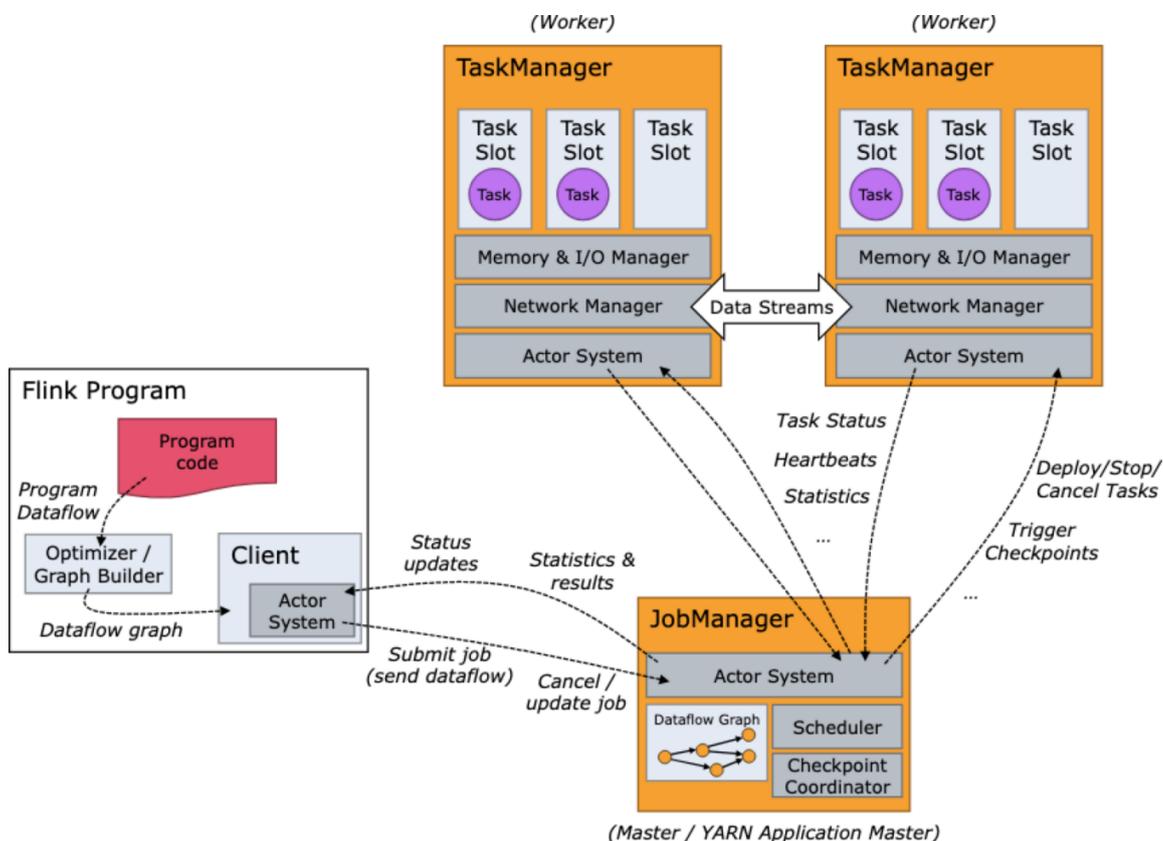
## 4.1.2.7 Flink

### 技术架构

Flink 运行时由两种类型的进程组成，包含一个 JobManager 和一个或者多个 TaskManager。

- JobManager
  - ResourceManager: 负责 Flink 集群中的资源提供、回收、分配——它管理 task slots，这是 Flink 集群中资源调度的单位（请参考 TaskManagers）。Flink 为不同的环境和资源提供者（例如 YARN、Kubernetes 和 standalone 部署）实现了对应的 ResourceManager。在 standalone 设置中，ResourceManager 只能分配可用 TaskManager 的 slots，而不能自行启动新的 TaskManager。
  - Dispatcher: 提供了一个 REST 接口，用来提交 Flink 应用程序执行，并为每个提交的作业启动一个新的 JobMaster。它还运行 Flink WebUI 用来提供作业执行信息。
  - JobMaster: JobMaster 负责管理单个 JobGraph 的执行。Flink 集群中可以同时运行多个作业，每个作业都有自己的 JobMaster。始终至少有一个 JobManager。高可用（HA）设置中可能有多个 JobManager，其中一个始终是 leader，其他的则是 standby。
- TaskManager
 

TaskManager（也称为 worker）执行作业流的 task，并且缓存和交换数据流。必须始终至少有一个 TaskManager。在 TaskManager 中资源调度的最小单位是 task slot。TaskManager 中 task slot 的数量表示并发处理 task 的数量。请注意一个 task slot 中可以执行多个算子。



## 基础概念

名称	描述
Application	基于Flink构建的用户程序。
Application jar	包含用户的Flink应用程序的jar。在某些情况下，用户会希望创建一个包含其应用程序及其依赖项的“jar”，用户的jar不应包括Hadoop或Flink库，这些库将在运行时添加。
JobManager	JobManager具有许多与协调Flink应用程序的分布式执行有关的职责：它决定何时调度下一个task（或一组task）、对完成的task或执行失败做出反应、协调checkpoint、并且协调从失败中恢复等等。
TaskManager	TaskManager（也称为worker）执行作业流的task，并且缓存和交换数据流。
Deploy mode	主要用于区分Driver在何处运行。在“Application”模式下，Client在集群内部启动运行。在“per-job”模式下，提交者在集群外部启动运行Client。
Worker node	可以在集群中运行应用程序代码的节点。
Task	对于分布式执行，Flink将算子的subtasks链接成tasks。
Operator Chains	经过优化后的多个task可以组成Operator Chains，提升资源的利用率。
Job	由多个任务组成的并行计算作业，对应Flink的Action操作。
StreamGraph	在运行时，Flink程序按照算子逻辑生成一个有向无环图DAG，这个图就是数据流图（StreamGraph）。

名称	描述
JobGraph	StreamGraph经过合并算子链优化后生成的就是JobGraph。提交给JobManager的数据结构。
ExecutionGraph	JobManager根据JobGraph生成ExecutionGraph。ExecutionGraph是JobGraph的并行化版本，是调度层最核心的数据结构。在JobGraph的基础上结合各算子的并行度生成的就是ExecutionGraph（ExecutionGraph与JobGraph最大的区别就是按照并行度将每个算子的任务拆分成了多个并行的子任务，并明确了任务间数据传输的方式）。
Physical Graph	JobManager根据ExecutionGraph对Job进行调度后，在各个TaskManager上部署Task后形成的“图”，并不是一个具体的数据结构。

## 4.1.2.8 OpenLDAP

### 技术架构

OpenLDAP 采取传统的 C/S 架构。通过配置服务器和客户端，实现账号的管理，并通过与第三方应用相结合，实现客户端所有账号均可通过服务端进行验证。

图4-3 C/S 架构图

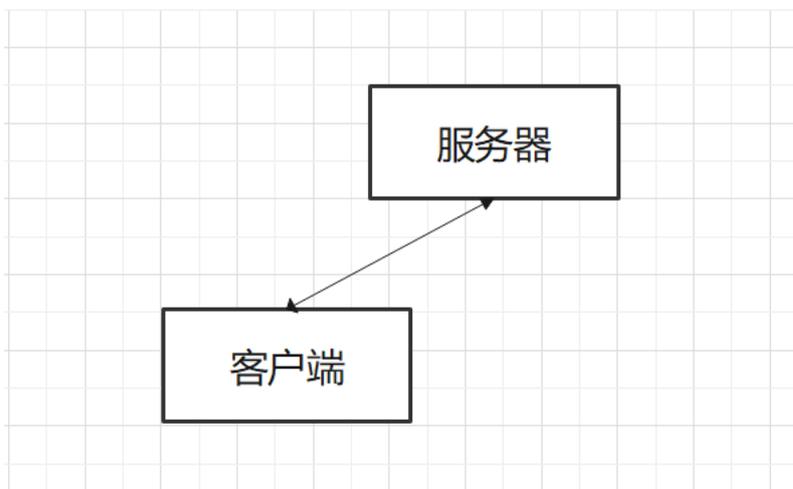
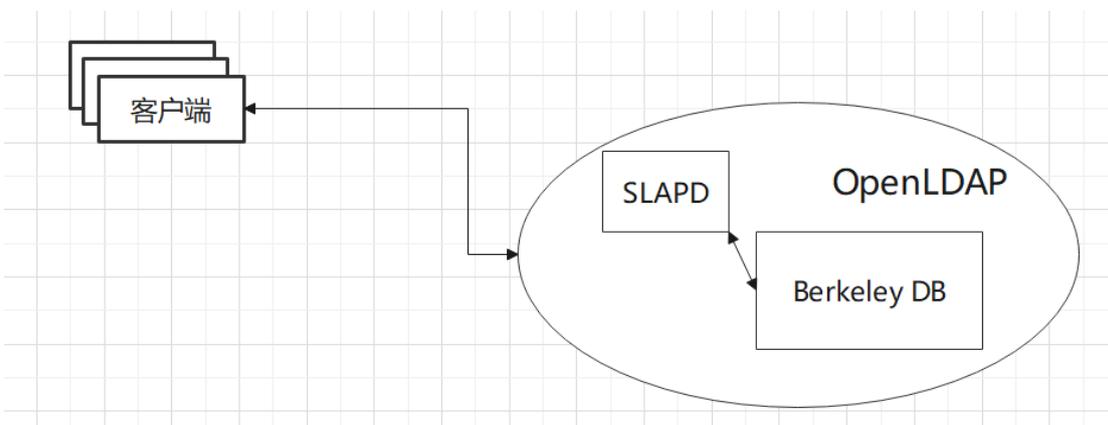


图4-4 认证工作模型



认证工作流程如下：

- (1) 客户端向 OpenLDAP 服务器发起验证请求。
- (2) 服务器接收用户请求后，并通过 slapd 进程向后端的数据库进行查询。
- (3) slapd 将查询的结果返回给客户端即可。如果有缓存机制，服务器端会先将查询的条目进行缓存，然后再发给客户端。

## 目录结构

LDAP 目录信息以树形结构进行存储，在树根一般定义国家（c=CN）或者域名（dc=com），其次往往定义一个或多个组织（organization, o）或组织单元（organization unit, ou）。一个组织单元可以包含员工、设备信息（计算机/打印机等）相关信息。例如 uid=babs, ou=People, dc=example, dc=com。

目前 OpenLDAP 目录架构分为两种：

- (1) 一种为互联网命名组织架构（如[图 4-5](#)）。
- (2) 另一种为企业级命名组织架构（如[图 4-6](#)）。

图4-5 互联网命名组织架构

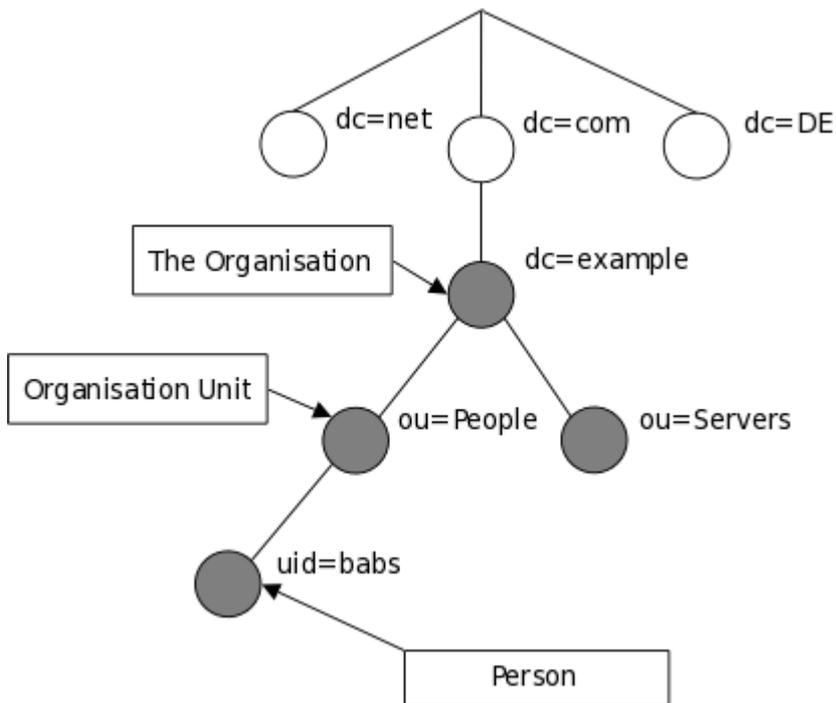
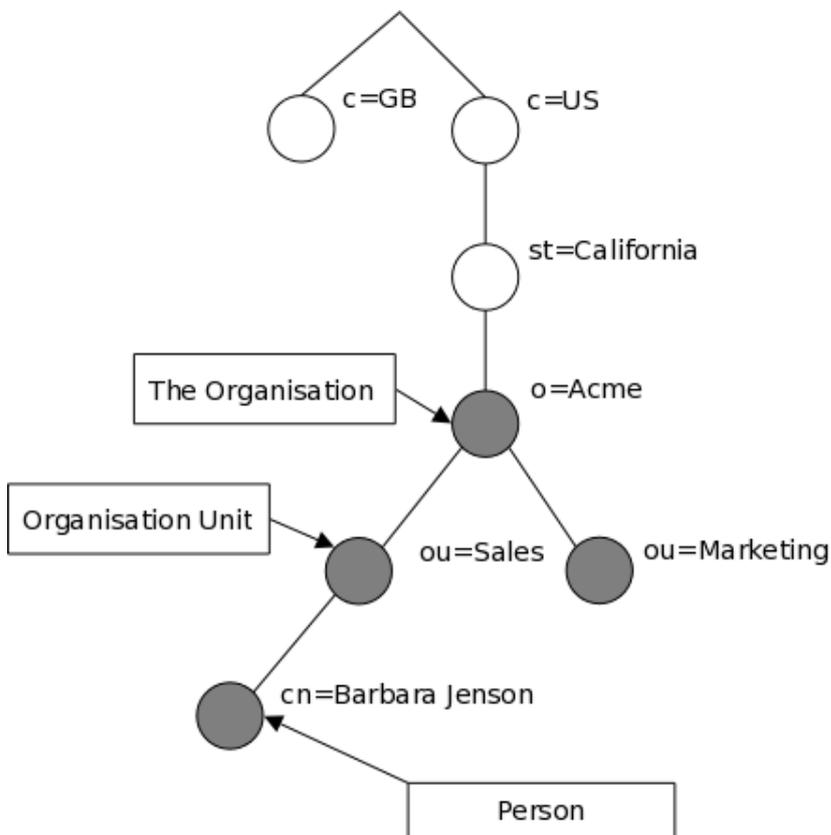


图4-6 企业级命名（传统命名）组织架构



## 基础概念

表4-1 基本概念

名称	描述
目录树 (Directory Tree)	在一个目录服务系统中，整个目录信息集可以表示为一个目录信息树，树中的每个节点是一个条目。
条目 (Entry)	每个条目就是一条记录，每个条目有自己的唯一可区别的名称 (DN)。
对象类 (ObjectClass)	与某个实体类型对应的一组属性，对象类是可以继承的，这样父类的必须属性也会被继承下来。
属性 (Attribute)	描述条目的某个方面的信息，一个属性由一个属性类型和一个或多个属性值组成，属性有必须属性和非必须属性。

表4-2 常见名称说明

名称	英文全称	描述
dc	Domain Component	域名的部分，其格式是将完整的域名分成几部分，如域名为 example.com 变成 dc=example, dc=com。
uid	User Id	用户ID，如“tom”。

ou	Organization Unit	组织单位，类似于Linux文件系统中的子目录，它是一个容器对象，组织单位可以包含其他各种对象（包括其他组织单元），如“market”。
cn	Common Name	公共名称，如“Thomas Johansson”。
sn	Surname	姓，如“Johansson”。
dn	Distinguished Name	惟一辨别名，类似于Linux文件系统中的绝对路径，每个对象都有一个惟一的名称，如“uid= tom, ou=market, dc=example, dc=com”，在一个目录树中DN总是惟一的。
rdn	Relative Dn	相对辨别名，类似于文件系统中的相对路径，它是与目录树结构无关的部分，如“uid=tom”或“cn=Thomas Johansson”。
c	Country	国家，如“CN”或“US”等。
o	Organization	组织名，如“Example, Inc.”。

表4-3 常用 objectClass

名称	含义
olcGlobal	全局配置文件类型，主要是cn=config.ldif的配置项。
top	顶层的对象。
organization	组织，比如公司名称，顶层的对象。
organizationalUnit	一个目录节点，通常是group，或者部门这样的含义。
inetOrgPerson	用户节点类型，person类型，叶子节点。
groupOfNames	分组的group类型，标记一个group节点。
olcModuleList	配置模块的对象。

## 4.1.2.9 Kerberos

### 技术架构

图4-7 架构图

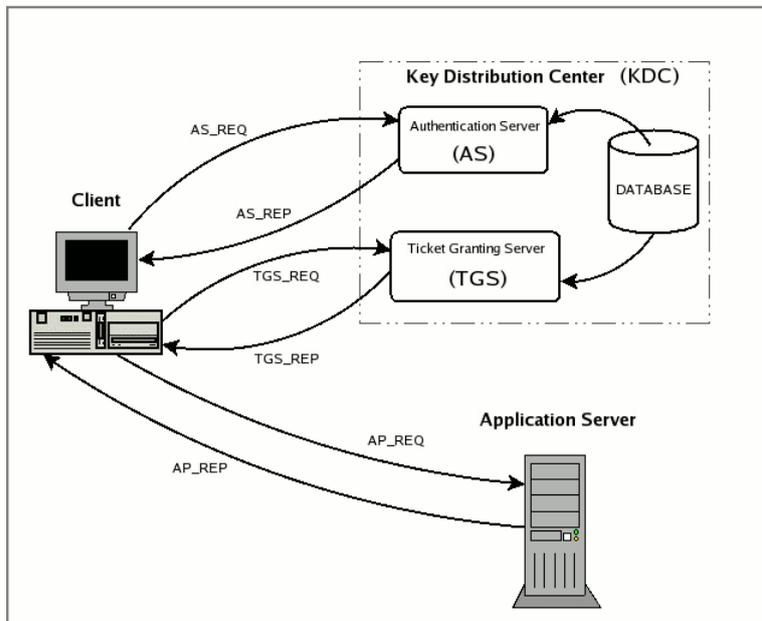


表4-4 架构图参数说明

名称	说明
AS_REQ	初始化一个用户（kinit）时发出的用户认证请求，这个请求是发给KDC中的Authentication Server（AS）。
AS_REP	AS回复给client的信息，其中包括TGT（用TGS secret key加密过的）and the session key（用发请求的用户的secret key加密过的）。
TGS_REQ	client为了一个service ticket向Ticket Granting Server（TGS）的信息，其中包括上一条信息中获得的TGT（用TGS secret key加密过的），一个客户端产生的一个authenticator（用session key加密过的）。
TGS_REP	TGS回复给TGS_REQ的信息，其中包括service ticket（用appservice的secret key加密过），和一个TGS产生的service session key（之前AS产生的一条session key给加密过的）。
AP_REQ	客户端发给appserver的一条访问服务请求，其中包括service ticket和一个authenticator（使用TGS产生的service session key加密过的）。
AP_REP	appserver回复给client的一条信息，证明该appserver确实是客户端希望访问的server。但是，这条信息也不一定总是要回复的。比如当客户端和appserver需要相互认证的时候，客户端向appserver发出请求，这个时候就不需要回复信息。

## 基础概念

名称	描述
KDC	密钥分发中心，负责管理发放票据，记录授权。
Realm	Kerberos管理领域的标识。
principal	Kerberos下的用户可以称为Principal，当每添加一个用户或服务的时候都需要向kdc添加一条principal，principal的形式为：主名称/实例名@领域名。
主名称	主名称可以是用户名或服务名，表示是用于提供各种网络服务（如hdfs、yam、hive）的主体。
实例名	实例名简单理解为主机名。
keytab文件	存储了用户的加密密码，常用这种方式认证。

### 4.1.2.10 Ranger

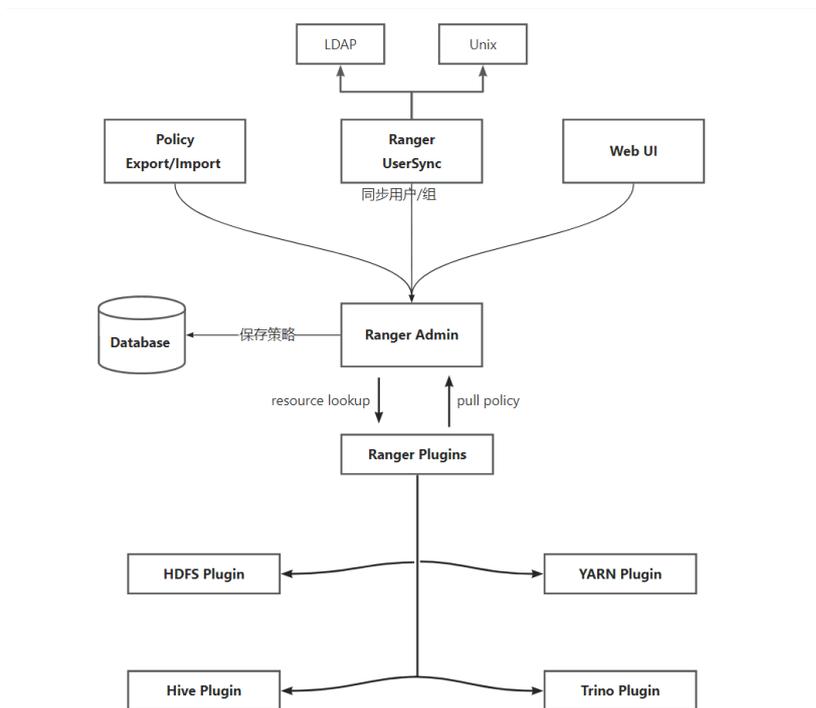
#### 技术架构

Ranger 的架构中主要由三个组件组成：

- **Ranger Admin:** 您可以创建和更新安全访问策略，这些策略被存储在数据库中。各个组件的 Plugin 定期对这些策略进行轮询。
- **Ranger Plugins:** Plugin 嵌入在各个集群组件的进程里，是一个轻量级的 Java 程序。例如，Ranger 对 Hive 的组件，就被嵌入在 Hiveserver2 里。这些 Plugin 从 Ranger Admin 服务端拉取策略，并把它们存储在本地文件中。当接收到来自组件的用户请求时，对应组件的 Plugin 会拦截该请求，并根据安全策略对其进行评估。
- **Ranger UserSync:** Ranger 提供了一个用户同步工具。您可以从 Unix 或者 LDAP 中拉取用户和用户组的信息。这些用户和用户组的信息被存储在 Ranger Admin 的数据库中，可以在定义策略时使用。

从架构图中可以看到，Ranger 的核心服务是 Ranger Admin（后端是一个 EmbeddedServer Java 进程），无论是 Web UI 还是 UserSync，亦或是 Ranger Plugins 都是直接跟 Ranger Admin 进行通信，而不会直接跟其元数据库进行通信，因为 Ranger Admin 提供了丰富的 RESTful 的接口，满足了其周边的所有数据和操作的需求。

在我们的架构中，比较核心的 Ranger Admin 支持多服务，多个服务通过负载方式向其他应用或者用户提供服务，来保证服务的高可用。所用到的数据库为 MySQL，MySQL 也以集群方式提供，从而保证了数据存储的安全可用。



## 基础概念

名称	描述
Admin	Ranger安全管理服务。
Web UI	提供给管理员的用户交互的管理页面。
EmbeddedServer	Admin的后端进程服务名。
UserSync	一个用户同步的服务，可将Unix和LDAP的用户同步到Ranger Admin。
Database	Ranger的元数据库，用于保存Ranger正常运行的元数据，例如用户、组、角色、权限策略等。
Ranger Plugin	Ranger提供的插件，可以与其他组件集成，从而使用Ranger Admin的策略来控制组件的权限，Ranger本身提供多种插件，可满足Hadoop生态中多数使用场景的需求。
Service Name	一个服务命名，在Ranger中表示相同引擎或组件的一个策略集的命名空间，通过Service Name可以实现一个Ranger Admin管理相同的引擎或者组件的多个集群。
User	用户，Ranger Admin中的用户，也是权限策略的用户名。
Group	用户组，Ranger Admin中的组名，也是引擎或组件的组名，一个组名必须对应到用户才会生效。
Role	角色，Ranger Admin中的角色，一个Role中可以包含多个User或者Group。
Security Zone	安全区域，Ranger Admin中提供的多个资源和权限策略隔离管理的一种方式。
Audit	审计，Ranger中关于访问、系统、账户、插件、权限策略、用户同步的审计信息，其数据存储在Elasticsearch或者Solr中。
Ranger Policy	权限策略，对资源、用户、权限控制的一个集合。
Allow Conditions	Policy允许的条件集合。

Exclude from Allow Conditions:	Policy允许的条件集合中排除的权限条件集合。
Deny Conditions	Policy拒绝的条件集合。
Exclude from Deny Conditions	Policy拒绝的条件集合中排除的权限条件集合。
Access	访问控制类型的策略。
Masking	数据脱敏类型的策略。
Row Level Filter	行级过滤类型的策略。

## 4.2 查看服务信息

### 简介

完成创建云大数据服务 CIK 实例后，您可以查看该实例的服务信息。

### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“服务”，进入服务列表页面，可以进行以下操作。

任务	说明
查看服务信息	列表查看对应服务的信息，包括状态、节点运行数量/总数量、总容量规格等。
停止服务	单击目标服务操作列的“停止”，停止当前服务。
重启服务	单击目标服务操作列的“重启”，重新启动当前已停止的服务。
查看服务监控信息	单击目标服务操作列的“监控信息”，跳转至云监控页面，可查看服务的监控指标信息，新建和查看告警规则（参见《云监控 CMS 用户指南》）。
查看服务详情	单击服务名称，进入服务详情页面。 <ul style="list-style-type: none"> <li>• 基本信息：例如服务 ID、状态、版本、规格、节点配置信息、连接信息等。</li> <li>• 部署拓扑：例如节点名称、类型、状态、IP 等。</li> <li>• 资源队列：例如队列名称、容量等，支持新建队列和刷新。</li> </ul>

## 4.3 为服务绑定/解绑安全组



说明

该功能仅适用于 CECSTACK 平台。

## 4.3.1 为服务绑定安全组

### 简介

实例创建完成后，支持为服务绑定安全组。安全组类似防火墙功能，是一个逻辑上的分组，用于设置网络访问控制。

### 限制与指导

- 当前版本仅 Kerberos、Ranger、HDFS、Yarn、Hive、Trino 服务支持绑定安全组。
- 确保安全组中至少放行 CIK 实例所在子网的网段，并设置为最高优先级，且开放 8088 端口，否则会导致实例出现异常。

### 操作步骤

- (1) 在实例列表页面，单击目标实例名称/ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“服务”，进入服务列表页面。
- (3) 单击服务名称，进入服务基本信息页面。
- (4) 在“连接信息”区域，单击安全组名称右侧“绑定”，弹出对话框。
- (5) 在下拉框中选择要绑定的安全组。  
若无可用安全组，单击“新建安全组”。完成新建后，单击“”刷新下拉框，即可选择已创建的安全组。
- (6) 单击“确定”。

## 4.3.2 为服务解绑安全组

### 简介

当服务不再需要绑定安全组时，可以通过该功能进行解绑。

### 前提条件

服务已绑定安全组。

### 限制与指导

当前版本仅 Kerberos、Ranger、HDFS、Yarn、Hive、Trino 服务支持解绑安全组。

### 操作步骤

- (1) 在实例列表页面，单击目标实例名称/ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“服务”，进入服务列表页面。
- (3) 单击服务名称，进入服务基本信息页面。
- (4) 在“连接信息”区域，单击安全组名称右侧“解绑”，弹出对话框。
- (5) 确认信息无误后，单击“确定”。

## 4.4 为服务绑定/解绑弹性公网IP



该功能仅适用于 CECSTACK 平台。

---

### 4.4.1 为服务绑定弹性公网 IP

#### 简介

绑定弹性公网 IP 后，即可通过外网 IP 地址访问该 CIK 实例的指定服务。

#### 前提条件

- 在绑定弹性公网 IP 前，需要先为服务绑定安全组。
- 绑定后，请第一时间确认该弹性公网 IP 对应安全组中已配置相关访问规则，否则可能无法外部地址无法访问该服务。

#### 限制与指导

当前版本仅 Kerberos、Ranger、HDFS、Yarn、Hive、Trino 服务支持绑定弹性公网 IP。

#### 操作步骤

- (1) 在实例列表页面，单击目标实例名称/ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“服务”，进入服务列表页面。
- (3) 单击服务名称，进入服务基本信息页面。
- (4) 在“连接信息”区域，单击外网地址右侧“绑定”，弹出对话框。
- (5) 在下拉框中选择要绑定的弹性公网 IP。  
若无可用弹性公网 IP，单击“新建弹性公网 IP”。完成新建后，单击“”刷新下拉框，即可选择已创建的弹性公网 IP。
- (6) 单击“确定”。

### 4.4.2 为服务解绑弹性公网 IP

#### 简介

当不再使用当前绑定的弹性公网 IP 或需要更换弹性公网 IP 时，可以进行解绑。

#### 前提条件

已[为服务绑定弹性公网 IP](#)。

#### 操作步骤

- (1) 在实例列表页面，单击目标实例名称/ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“服务”，进入服务列表页面。

- (3) 单击服务名称，进入服务基本信息页面。
- (4) 在“连接信息”区域，单击外网地址右侧“解绑”，弹出对话框。
- (5) 单击“确定”。

## 4.5 停止服务

### 简介

当您需要停止正在使用中的服务时，可通过停止服务功能暂停服务。

### 限制与指导

实例必须处于“运行中”状态，且依赖当前服务的其他服务必须处于“停止中”状态。

### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“服务”，进入服务页面。
- (3) 单击目标服务“操作”列的“停止”，弹出对话框。
- (4) 单击“确定”。

### 操作结果

服务停止成功后，显示于服务列表页面，状态由“停止中”变为“已停止”，表示服务停止成功。

## 4.6 重启服务

### 简介

当服务配置项信息更改需要重启服务才能生效，或需要重启已停止的服务时，可使用重启服务功能。

### 限制与指导

- 实例必须处于“运行中”状态。
- 当前服务必须处于“运行中”、“已停止”和“异常”状态中的一种。
- 当前服务依赖的其他服务处于“运行中”状态。
- 若当前服务依赖 HDFS 服务，则 HDFS 服务必须处于非安全模式。

### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“服务”，进入服务页面。
- (3) 单击目标服务“操作”列的“重启”，弹出对话框。
- (4) 单击“确定”。

### 操作结果

服务重启成功后，显示于服务列表页面，状态由“重启中”变为“运行中”，表示服务重启成功。

## 4.7 开启HDFS平衡器

### 简介

当集群水位分布不均衡时，需要开启平衡器以平衡水位，使高水位的节点数据迁移至低水位的节点上。

### 操作步骤

- (1) 在实例列表页面，单击目标实例名称/ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“服务”，进入服务列表页面。
- (3) 单击服务名称“HDFS”，进入 HDFS 服务基本信息页面。
- (4) 单击页面右上角“更多 > balancer 平衡”，弹出对话框。
- (5) 设置平衡器阈值，单击“确定”。

### 操作结果

HDFS 服务开始进行数据平衡，详情页面显示“平衡器阈值”为设置值。

## 4.8 开启HDFS安全模式

### 简介

开启安全模式后，集群将处于只读不写的状态，数据无法写入和修改。在集群维护期间，开启该功能可以防止元数据被修改或者丢失。

### 操作步骤

- (1) 在实例列表页面，单击目标实例名称/ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“服务”，进入服务列表页面。
- (3) 单击服务名称“HDFS”，进入 HDFS 服务基本信息页面。
- (4) 在基本信息区域，将“安全模式”右侧开关置于  状态以开启该功能，弹出对话框。
- (5) 单击“确定”。

# 5 服务配置

## 5.1 查看服务配置项

### 简介

在服务配置页面，您可以查看服务不同版本的配置项。

### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“配置”，进入实例服务配置页面。
- (3) 页面左侧显示服务名称和状态，右侧显示所选服务的配置信息。
- (4) 单击“版本”输入框，下拉框中显示所有已创建的配置项版本，可按照版本查看配置项信息。

## 5.2 编辑服务配置项

### 简介

实例创建完成后，您可以通过配置页面修改或新增服务的配置项信息。

### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“配置”，进入实例配置页面。
- (3) 单击页面左侧目标服务名称，右侧显示该服务的配置信息。
- (4) 单击“版本”输入框，选择并切换至“Editable”版本，单击“编辑”。
- (5) 修改已有配置项或新增配置项后，单击“保存”。

### 后续操作

配置项信息修改后，只有再次下发配置并重启该服务后，才能使配置生效。

## 5.3 下发服务配置项

### 简介

若您修改了服务的配置项信息，则需要再次下发配置并重启该服务后，才能使配置生效。下发配置后，系统将自动生成一个新版本的配置项信息，可以为服务下发不同版本的配置项信息。

### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“配置”，进入实例配置页面。

- (3) 单击页面左侧目标服务名称，右侧显示该服务的配置信息。
- (4) 单击“版本”输入框，选择并切换至“Editable”或“未激活”版本，单击“下发配置”。
- (5) 在弹出的对话框中输入“备注”信息，单击“确定”。

## 后续操作

[重启服务](#)

## 5.4 下载服务配置项

### 简介

服务配置项信息支持以.xml 文件的形式下载至本地，作为配置文件供客户端使用。

### 限制与指导

当前版本仅 HDFS、Yarn、Hive、Spark、Flink 服务支持下载客户端配置文件，且该服务必须处于“运行中”状态。

### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“配置”，进入实例配置页面。
- (3) 单击页面左侧目标服务名称，右侧显示该服务的配置信息。
- (4) 单击“下载客户端配置”，选择“内网”或“外网”。
- (5) 单击“确定”。

### 操作结果

当前服务的配置项信息以.xml 文件形式下载至本地。

# 6 用户管理

## 6.1 用户管理

### 6.1.1 新建用户

#### 简介

完成创建 CIK 实例后，您可以通过用户管理页面为实例创建用户并设置密码。

#### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“用户”，默认进入用户列表页面。
- (3) 单击“新建用户”，弹出对话框。
- (4) 输入用户名、密码、确认密码、描述信息。
- (5) 单击“确定”。

#### 操作结果

刷新页面，可在下方列表查看到该用户信息。

### 6.1.2 查看用户列表

#### 简介

您可以在用户列表页面新建用户，根据筛选条件查询指定的用户，并对目标用户进行删除、同步、关联用户组、重置密码操作。

#### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“用户”，默认进入用户列表页面。
- (3) 在用户列表页面，可以进行以下操作。

任务	说明
新建用户	单击“新建用户”。
查询用户	选择筛选条件，输入关键字单击  模糊查询目标用户。
同步用户	若用户“同步状态”显示失败，则单击该用户“操作”列的“同步”，进行手动同步。
删除用户	单击目标用户“操作”列的“删除”。
关联用户组	单击目标用户“操作”列的“关联用户组”，将用户与用户组相关联。
重置密码	单击目标用户“操作”列的“重置密码”。

任务	说明
刷新页面	单击  按钮，刷新用户列表页面。

### 6.1.3 手动同步用户信息

#### 简介

完成创建用户后，若该用户“同步状态”显示失败，则需要通过“同步”功能将用户信息手动同步至大数据安全服务中。

#### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“用户”，默认进入用户列表页面。
- (3) 单击目标用户的操作栏“同步”，弹出对话框。
- (4) 单击“确定”。

### 6.1.4 删除用户

#### 简介

您可以删除不再需要的用户信息。

#### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“用户”，默认进入用户列表页面。
- (3) 单击目标用户的操作栏“删除”，弹出对话框。
- (4) 确认信息无误后，单击“确定”。

### 6.1.5 关联用户组

#### 简介

完成用户创建后，可以将指定用户与某个用户组进行关联，以对用户进行分组管理。

#### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“用户”，默认进入用户列表页面。
- (3) 单击目标用户的操作栏“更多 > 关联用户组”，弹出对话框。
- (4) 在左侧未关联用户组列表中选择用户组，单击  移入右侧已关联用户组。
- (5) 单击“确定”。

## 6.1.6 重置用户密码

### 简介

若忘记密码，可以手动对指定用户密码进行重置。

### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“用户”，默认进入用户列表页面。
- (3) 单击目标用户的操作栏“更多 > 重置密码”，弹出对话框。
- (4) 输入新密码和确认密码。
- (5) 单击“确定”。

## 6.2 用户组管理

### 6.2.1 新建用户组

#### 简介

完成创建用户组后，可以将用户与用户组相关联，以实现对用户的分组管理。

#### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“用户”，默认进入用户列表页面。
- (3) 单击“用户组”页签，切换到用户组内容。
- (4) 在用户组列表页面，单击“新建用户组”，弹出对话框。
- (5) 填写用户组名称、描述信息。
- (6) 确认信息无误后，单击“确定”。

#### 操作结果

完成新建用户后，刷新用户列表页面，即可看到用户信息。

### 6.2.2 查看用户组列表

#### 简介

您可以在用户组列表页面新建用户组，根据筛选条件查询指定的用户组，并对目标用户组进行删除、同步、关联用户组操作。

#### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“用户”，默认进入用户列表页面。
- (3) 单击“用户组”页签，切换到用户组内容。

(4) 用户组内容页面，显示用户组列表。

在用户组列表页面，可以进行以下操作：

任务	说明
新建用户组	单击“新建用户组”，新建用户组。
查询用户组	选择筛选条件，输入关键字单击  模糊查询目标用户组。
同步用户组	单击同步失败的目标用户组“操作”列的“同步”，对用户组进行同步。
删除用户组	单击目标用户组“操作”列的“删除”，删除用户组。
关联用户	单击目标用户组“操作”列的“关联用户”，关联用户信息。
刷新页面	单击  按钮，刷新用户组列表页面。

## 6.2.3 同步用户组

### 简介

完成创建用户组后，若用户组“同步状态”列显示失败，则需要通过“同步”功能，将用户组信息手动同步至大数据安全组件中。

### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“用户”，默认进入用户列表页面。
- (3) 单击“用户组”页签，切换到用户组内容。
- (4) 在用户组列表页面，单击目标用户组的操作栏“同步”按钮。

## 6.2.4 删除用户组

### 简介

您可以删除不再需要的用户组信息。

### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“用户”，默认进入用户列表页面。
- (3) 单击“用户组”页签，切换到用户组内容。
- (4) 在用户组列表页面，单击目标用户组的操作栏“删除”。
- (5) 确认信息无误后，单击“确定”。

## 6.2.5 关联用户

### 简介

完成创建用户组后，可以将其与用户进行关联。

### 操作步骤

- (1) 在实例列表页面，单击目标实例 ID，进入实例基本信息页面。
- (2) 单击左侧导航栏“用户”，默认进入用户列表页面。
- (3) 单击“用户组”页签，切换到用户组内容。
- (4) 在用户组列表页面，单击目标用户组的操作栏“更多”，选择“关联用户”，弹出对话框。
- (5) 在对话框中，选择用户，可以多选。
- (6) 确认信息无误后，单击“确定”。