CDH集群

硬件规划说明书

创建日期：2022-07-10

**文档修改记录**

|  |  |  |  |
| --- | --- | --- | --- |
| **修改日期** | **修改人** | **修改说明** | **版本号** |
| 2022/7/10 | 谢宗铭 |  |  |
|  |  |  |  |
|  |  |  |  |

目录

1 引言 5

1.1 编写目的 5

2 建设目标 6

2.1 项目总体建设目标 6

2.2 项目建设规划 6

3 集群架构设计 7

3.1 整体体系架构 7

3.2 产品规格说明 9

3.3 CDH 6.3开源组件清单 9

3.4 物理架构设计 11

3.5 硬件配置 12

4 基础软件安装设计 14

5 集群配置规划 15

5.1 集群运维及管理节点 15

5.2 集群主节点 16

5.3 边缘节点 17

5.4 数据节点 17

5.5 磁盘规划建议 18

5.6 数据库用户及schema设计 19

5.7 高可用支持 20

# 引言

## 编写目的

本文档是针对利得新建CDH大数据集群提出的硬件规划设计方案。本文档描述了平台建设的目标与功能，介绍了系统关键问题的解决方案，明确各子系统的划分、及其相互关系，明确外围需要集成的系统的集成方案。

本文档便于各开发小组协调工作，使得开发人员对系统的设计都进行周密思考、全盘权衡，为后续概要、详细设计奠定基础。

# 建设目标

## 项目总体建设目标

基于CDH的Hadoop技术，构建新一代大数据平台；平台的技术应该在立足业务发展需要的前提下，引入主流、前瞻性的技术，满足利得打造下一代大数据平台的战略需要。

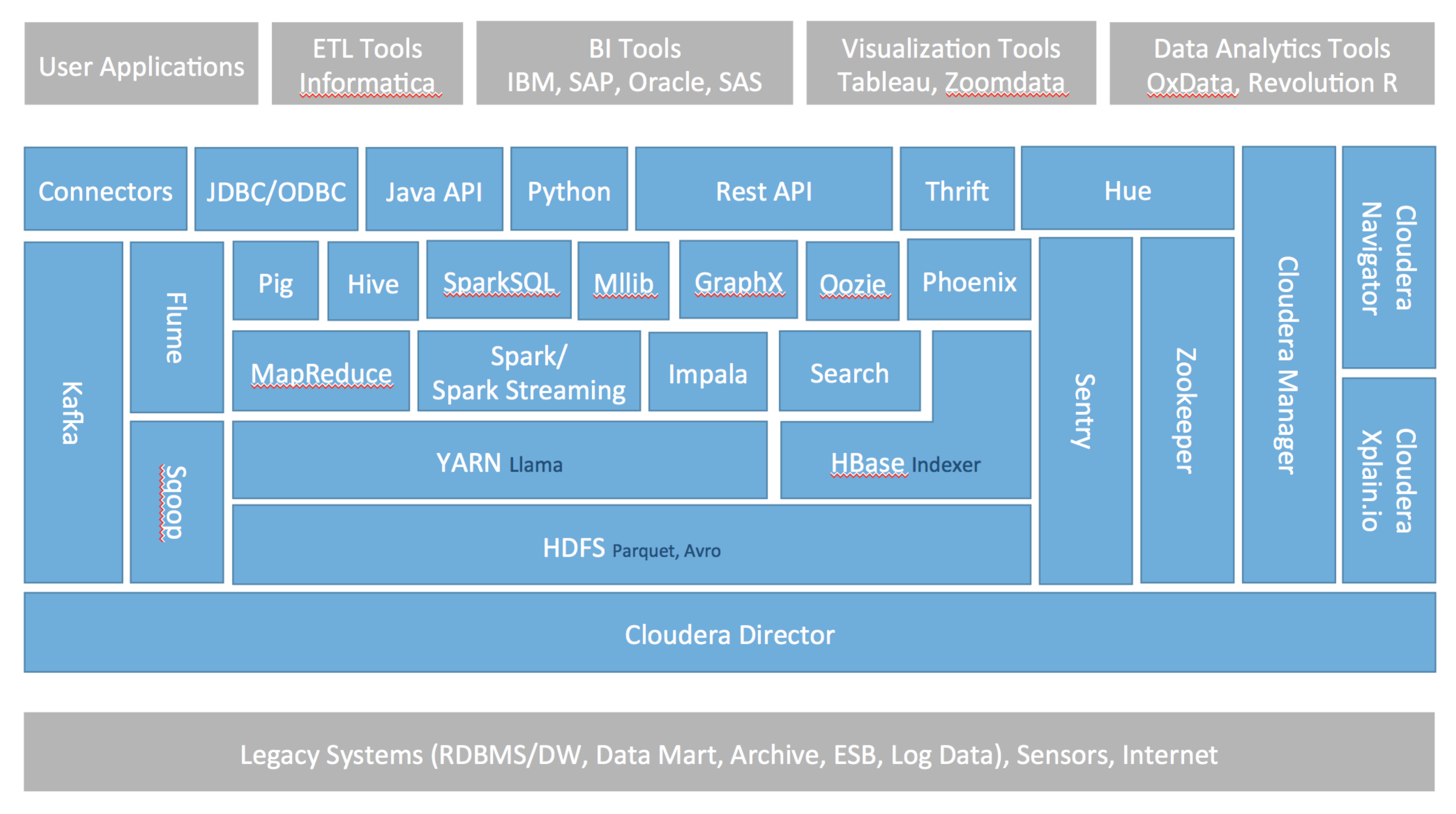
平台除了提供数据的采集、转换、入库、查询、数据共享能力，以及相应的管理和监控功能，还应该具有开放性，易于第三方的深度机器学习工具的集成。

## 项目建设规划

项目的初步建设规划为建设部署一套设计容量为70TB的CDH大数据集群。

# 集群架构设计

## 整体体系架构



Cloudera的软件体系结构中包含了以下模块：系统部署和管理，数据存储，资源管理，处理引擎，安全，数据管理，工具库以及访问接口。关键组件的角色信息如下表所示：

|  |  |  |  |
| --- | --- | --- | --- |
| **模块** | **组件** | **管理角色** | **工作角色** |
| 系统部署和管理 | Cloudera Manager | Cloudera Manager Server | Cloudera Manager Agent |
| Host Monitor |  |
| Service Monitor |  |
| Reports Manager |  |
| Alert Publisher |  |
| Event Server |  |
| 数据存储 | HDFS | NameNode | DataNode |
| Secondary NameNode |  |
| JournalNode |  |
| FailoberController |  |
| Kudu | Kudu Master | Tablet Server |
| HBase | HBase Master | RegionServer |
| 资源管理 | YARN | ResourceManager | NodeManager |
| Job HistoryServer |  |
| 处理引擎 | Spark | History Server |  |
| Impala | Impala Catalog Server | Impala Daemon |
| Impala StateStore |  |
| Search |  | Solr Server |
| 安全、数据管理 | Sentry | Sentry Server |  |
| Cloudera Navigator | Navigator KeyTrustee |  |
| Navigator Metadata Server |  |
| Navigator Audit Server |  |
| 工具库 | Hive | Hive Metastore |  |
| Hive Server2 |  |

## 产品规格说明

|  |  |
| --- | --- |
| 支持OS (仅64-bit) | RHEL/CentOS/OL with RHCK kernel 7.6, 7.5, 7.4, 7.3, 7.2 |
| 支持的JDK | JDK 8 |
| 支持的开发打包组件 | Apache Maven |
| 支持的云平台 | Amazon EC2, SoftLayer, Microsoft Azure, Google Cloud Engine |
| 支持的数据库  （用于repository存储） | MariaDB 5.x、Mysql 5.x、Oracle 12.2、9.2 (default for RHEL/CentOS/OEL 7)  <https://www.cloudera.com/documentation/enterprise/6/release-notes/topics/rg_database_requirements.html> |

## CDH 6.3开源组件清单：

|  |  |  |
| --- | --- | --- |
| 组件 | 描述 | 版本 |
| Apache Avro | 一个序列化系统，用于通过网络存储和传输数据 | v1.8.2 |
| Apache Flume | 用于收集、聚合日志和事件数据的分布式框架，同时将数据实时接入HDFS或HBase | v1.9.0 |
| Apache Hadoop | 可靠的、可伸缩的分布式存储和计算框架 | v3.0.0 |
| Apache HBase | 可伸缩性的实时读写访问的、基于KV形式的记录和表的存储 | v2.1.2 |
| HBase Indexer | Apache许可的模块，能够实时的为HBase数据添加索引 | v1.5.1 |
| Apache Hive | 具备类SQL接口的元数据存储库，加上为BI应用提供的连接ODBC/JDBC驱动 | v2.1.1 |
| Hue | Hadoop基于浏览器的UI | v4.3.0 |
| Apache Impala | 针对HDFS或HBase数据的即席交互式的SQL查询引擎 | v3.2 |
| Apache Kafka | 基于发布/订阅模式的分布式消息队列，提供稳定、高可用、高吞吐的消息服务 | v2.1.0 |
| Kite SDK | Apache许可的库、工具盒示例的集合，用于简化Hadoop应用程序开发 | v1.0.0 |
| Apache Oozie | Hadoop作业及任务协调工作流引擎 | v5.1.0 |
| Apache Parquet | Apache许可的列式文件格式 | v1.9.0 |
| Apache Pig | 处理Hadoop平台数据的高级数据流语言 | v0.17.0 |
| Apache Kudu | 高性能存储查询引擎 | v1.9.0 |
| Apache Solr | 自由文本、模糊匹配和分面搜索引擎 | v7.4.0 |
| Apache Sentry | 为Impala、Hive、Search和HDFS模块提供细粒度的、基于角色的授权 | v2.1.0 |
| Apache Spark | 快速、通用的数据处理引擎，支持循环数据流、内存计算、机器学习和图计算 | v2.4.0 |
| Apache Sqoop | Hadoop平台与关系型数据库之间整合数据的数据传输引擎 | v1.4.7 |
| Apache Zookeeper | 高可靠的分布式协调服务 | v3.4.5 |

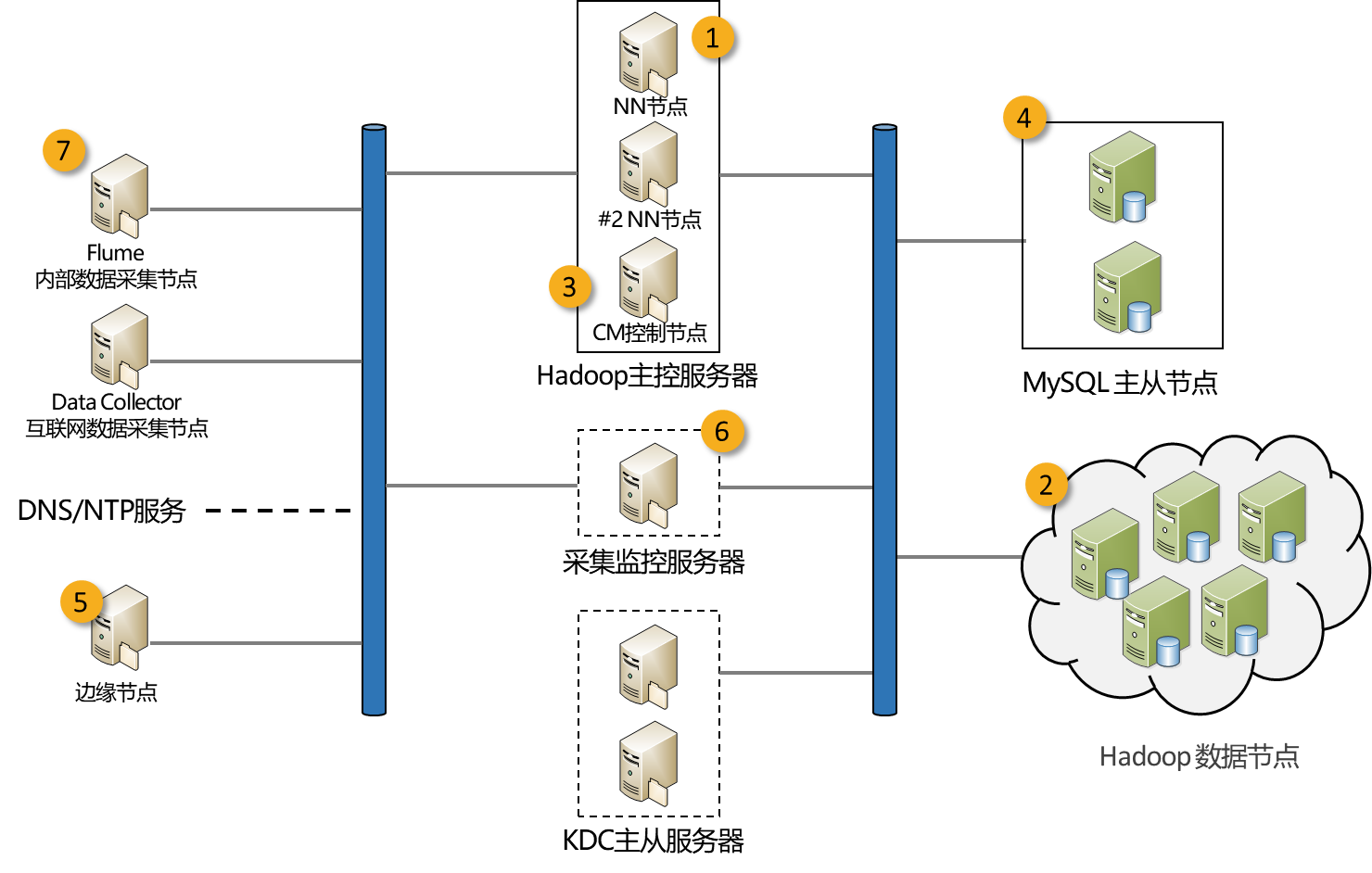
## 物理架构设计

本节提供利得大数据集群服务器规划和配置信息：

* 管理控制节点（3台）：提供集群的管理控制功能。
* 数据节点（7台）：提供集群的存储和计算。
* 数据采集节点（3台）：提供数据交换服务。
* 边缘节点（2台）：提供集群访问网关。
* KDC节点（2台）：提供集群账户统一身份认证和管理。

由于数据采集监控节点、边缘节点会随着数据采集和计算任务增加而增加节点，因此建议数据采集监控节点、边缘节点、KDC安全节点可以使用虚拟机。

服务器拓扑图如下图所示：



## 生产环境硬件配置

**服务器配置**

我们建议生产环境采用如下服务器配置：

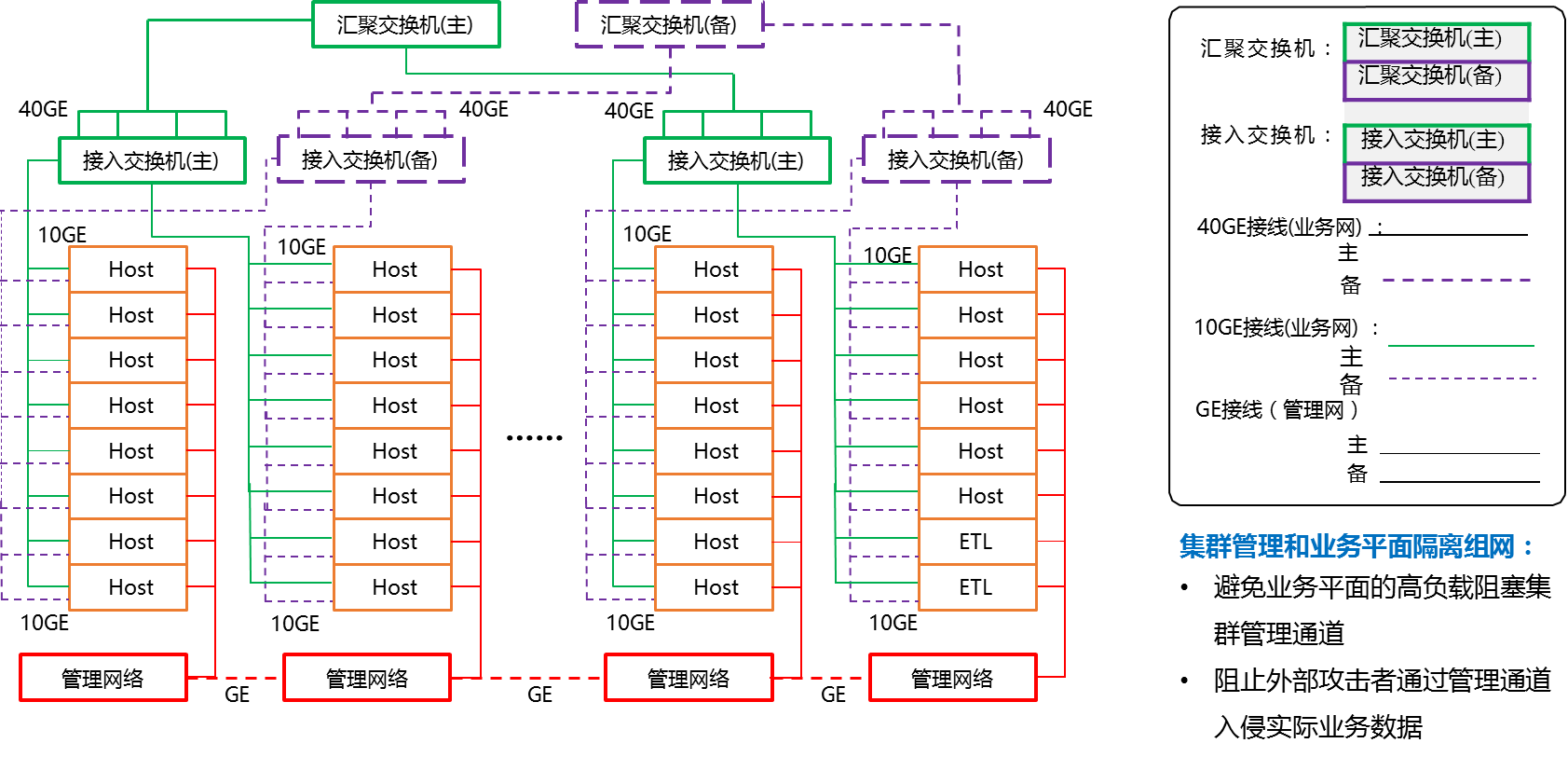
|  |  |  |  |
| --- | --- | --- | --- |
| **设备类型/型号** | **角色** | **硬件配置** | **数量** |
| X86机架服务器 | 管理节点 | 2路16核CPU；256GB内存； 2块900GB SAS硬盘 RAID 1，4块1T SAS硬盘 RAID 1； 2个万兆网口，1个千兆管理网口 | 3 |
| X86机架服务器 | 数据节点 | 2路16核CPU；512GB内存； 2块900GB SAS硬盘 RAID 1，  4块1T SAS硬盘 单盘RAID 0，  10块4TB SAS硬盘 单盘RAID 0； 2个万兆网口，1个千兆管理网口 | 7 |
| X86机架服务器或VM虚拟机  （可以虚拟机） | 数据采集(交换)监控节点 | 2路8核CPU；64GB内存； 2\*500GB SAS硬盘 RAID 1； 2个万兆网口，1个千兆管理网口 | 3 |
| X86机架服务器或VM虚拟机  （建议虚拟机） | 边缘节点 | 1路8核CPU；16GB内存； 2\*100GB SAS硬盘 RAID 1； 2个万兆网口，1个千兆管理网口 | 2 |
| X86机架服务器或VM虚拟机  （建议虚拟机） | KDC安全节点 | 1路4核CPU；16GB内存； 2块100GB SAS硬盘 RAID1； 2个万兆网口，1个千兆管理网口 | 2 |

注：

1. 由于数据采集监控节点、边缘节点会随着数据采集和计算任务增加而增加节点，因此建议数据采集监控节点、边缘节点、KDC安全节点都使用虚拟机。
2. 1T SAS硬盘用于存储主节点的元数据信息和Kudu数据，可提升整体性能。

**网络配置**

集群建议下面的网络结构。



* + 聚合层交换与核心交换机之间采用全网状互联的方式，以实现较高的冗余性。
  + 汇聚交换机之间采用多链路互联，并通过多个10GE以太网络链路，上联核心层交换机。
  + 接入层交换机以2个10GE口上联聚合层交换机，接入层交换机采用二层或者三层交换设备，支持VLAN等链路层协议，支持各种用户隔离和策略控制功能，如PVLAN、ACL、CAR等技术，为不同的用户提供高级别的安全隔离和网络QoS服务。

交换机配置：

|  |  |  |
| --- | --- | --- |
| **名称** | **配置** | **数量** |
| 接入交换机 | 10GE可堆叠交换机 | 2台 |
| 汇聚交换机 | 40GE可堆叠交换机 | 2台 |
| 管理交换机 | 24口1GE交换机 | 2台 |

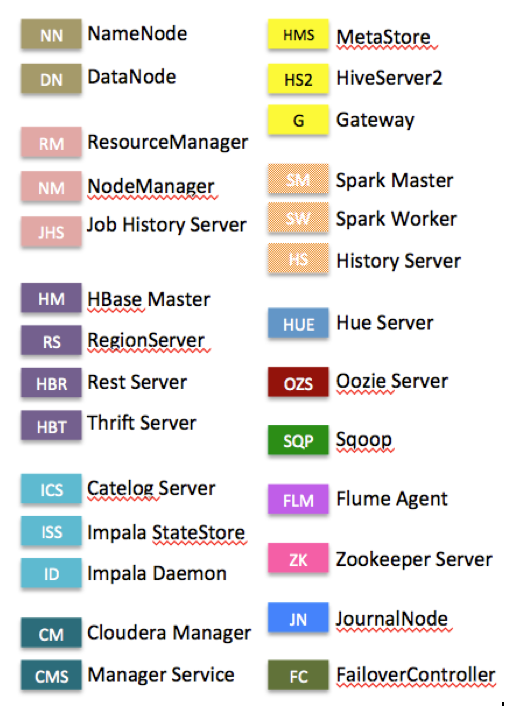
# 基础软件安装设计

|  |  |  |
| --- | --- | --- |
| **软件名称** | **类型** | **版本** |
| 操作系统 | Redhat/Centos | 7.6（最小化安装即可） |
| 数据库 | MySQL | 5.7 |
| JDK | Oracle JDK | 8 |
| CDH | Express Edition | 6.3.2 |
| Cloudera Manager | Express Edition | 6.3.2 |

# 集群配置规划

本集群采用CDH6.3.x版本进行部署。

一个中等规模的集群，集群的节点数一般在10到200左右，通常的数据存储可以规划到几百TB至几PB，适用于一个中型企业的数据平台，或者大型企业的业务部门数据平台。节点的复用程度可以降低，可以按照管理节点、主节点、工作节点、工具（边界）节点来划分。



## 集群运维及管理节点

集群管理控制节点上安装Cloudera Manager，以及Cloudera Management Service服务。

* 使用1台服务器（管理控制节点中的一台）作为CM管理节点
* 考虑到CM的可靠性，使用3台管理节点中的一台作为CM管理服务的冷备（不启动服务，只是在管理节点的服务器完全无法启动和修复时，进行手动切换）

|  |
| --- |
| 监控及管理节点安装角色 |
| Cloudera Manager Server |
| Cloudera Management Services |
| Hue Server |
| Oozie Server |

## 集群主节点

主节点上安装有个CDH服务的管理节点以及HA的组件，可以如下方式部署角色：

|  |  |  |  |
| --- | --- | --- | --- |
| 服务 | 主节点1 | 主节点2 | 主节点3 |
| HDFS | NameNode | NameNode |  |
| FailoverController | FailoverController |  |
| JournalNode | JournalNode | JournalNode |
| YARN | Resource Manager | Resource Manager |  |
|  |  | Job History Server |
| Zookeeper | Zookeeper Server | Zookeeper Server | Zookeeper Server |
| HBase | HBase Master | HBase Master | HBase Master |
| Impala | Impala StateStore | Impala Category Server |  |
| Hive |  | Hive Metastore | Hive Metastore |
|  | hiveserver2 | hiveserver2 |
| Spark |  |  | History Server |
| CM |  |  | CM及CMS冷备 |

## 边缘节点

边缘节点可以部署以下一些角色：

|  |
| --- |
| 边缘节点 |
| Flume Agent |
| Sqoop Client |
| Kafka broker |

## 数据节点

数据节点的角色部署（待定）：

|  |
| --- |
| 数据节点 |
| HDFS DataNode |
| YARN NodeManager |
| Impala Daemon |
| HBase RegionServer |
| Gateway |

## 磁盘规划建议

|  |  |  |
| --- | --- | --- |
| **角色节点** | **RAID推荐** | **磁盘要求** |
| 所有节点操作系统 | RAID10 | 根目录空间不小于50GB |
| 所有节点 | RAID10 | /var/log 空间不小于50GB |
| 所有节点 | RAID10 | /opt/cloudera空间不小于20GB |
| Host Monitor （A standalone service for Cloudera Manager） | RAID1 | /var/lib/cloudera-host-monitor不小于15GB （目录可调整，大小不可调整）；可以单独部署 |
| Service Monitor（A standalone service for Cloudera Manager） | RAID1 | /var/lib/cloudera-service-monitor不小于15GB （目录可调整，大小不可调整）；可以单独部署 |
| Event Server（A standalone service for Cloudera Manager） | RAID1 | /var/lib/cloudera-scm-eventserver，可以通过调整存储的最大事件数来控制存储空间占用 |
| Reports Manager（A standalone service for Cloudera Manager） | RAID1 | /var/lib/cloudera-scm-headlamp，目录空间不确定，取决于fsimage的大小和HDFS中文件或者目录的路径深度 |
| Navigator Audit Server（A standalone service for Cloudera Manager） | RAID1 | /var/lib不小于10GB，可以通过调整保存的审计日志天数来调整存储 |
| Navigator Metadata Server（A standalone service for Cloudera Manager） | RAID1 | /var/lib不小于10GB |
| NameNode (Active/Standby) | JBOD | 例如：/hadoop/dfs/nn不小于128GB |
| JournalNode | JBOD | 例如：/hadoop/dfs/jn不小于64GB |
| Zookeeper | JBOD/RAID1 | 例如：/var/lib/zookeeper不小于10GB |
| DataNode | JBOD | 增加noatime的挂载选项 |

## 数据库用户及schema设计

大数据平台的Repository数据库采用Oracle 12.2或MySQL 5.7版本，其依赖的服务所需的数据库用户及schema设计如下：

|  |  |  |  |
| --- | --- | --- | --- |
| **schema 名称** | **对应表空间名称** | **对应表空间大小** | **数据文件是否自行增长** |
| hive | TB\_HIVE\_METASTORE | 20GB | 是，每次增加500MB |
| amon | TB\_AMON | 20GB | 是，每次增加500MB |
| rman | TB\_RMAN | 5GB | 是，每次增加500MB |
| sentry | TB\_SENTRY | 5GB | 是，每次增加500MB |
| cm | TB\_CM | 10GB | 是，每次增加500MB |
| oozie | TB\_OOZIE | 5GB | 是，每次增加500MB |

## 高可用支持

|  |  |  |  |
| --- | --- | --- | --- |
| 服务 | 支持方式 | 描述 |  |
| HDFS | 内建的HA功能 |  |  |
| YARN | 内建的HA功能 |  |  |
| Zookeeper | 内建的HA功能 |  |  |
| HBase | 内建的HA功能 |  |  |
| Hive Metastore | 内建的HA功能 |  |  |
| HiveServer2 | 通过外部的Load Balancer支持 | 配置Load Balancer的地址和端口即可；Load Balancer只需要配置HiveServer2的10000端口 |  |
| Impala | 通过外部的Load Balancer支持 | 配置Load Balancer的地址即可；Load Balancer可以分别配置Impala Daemon的WebUI，JDBC，Thrift的端口 |  |
| Oozie | 通过外部的Load Balancer支持 | 配置Load Balancer的地址和端口即可；Load Balancer需要配置Oozie Server的11000端口 |  |
| Hue Server | 内建的HA支持 | 配置Hue Server和相应的Load Balancer即可 |  |