

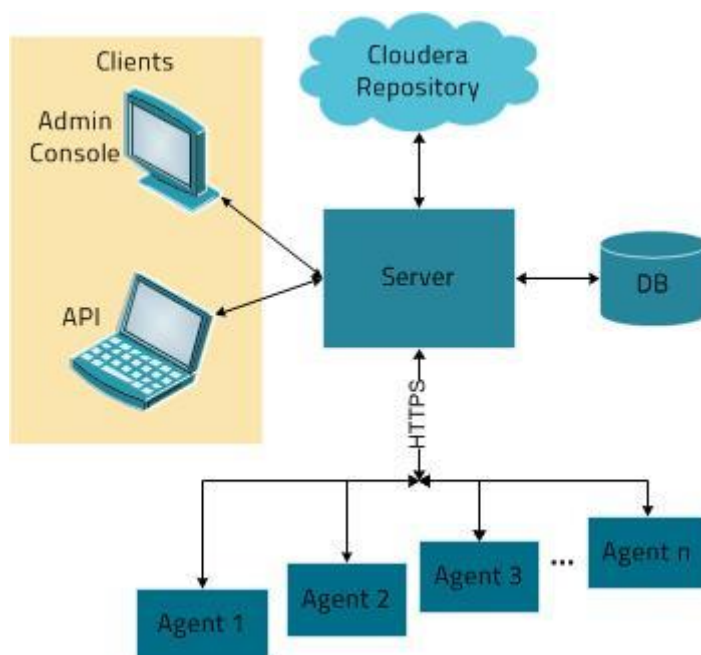
Cloudera Manager 产品介绍

1. 介绍

Cloudera Manager 是 CDH 市场领先的管理平台。作为业界第一的端到端 Apache Hadoop 的管理应用，Cloudera Manager 对 CDH 的每个部件都提供了细粒度的可视化和控制，从而设立了企业部署的标准。通过 Cloudera Manager，运维人员得以提高集群的性能，提升服务质量，提高合规性并降低管理成本。

Cloudera Manager 设计的目的是为了使得对于企业数据中心的的管理变得简单和直观。通过 Cloudera Manager，可以方便地部署，并且集中式的操作完整的大数据软件栈。该应用软件会自动化安装过程，从而减少了部署集群的时间。通过 Cloudera Manager 可以提供一个集群范围内的节点实时运行状态视图。同时，还提供了一个中央控制台，可以用于配置集群。不仅如此，Cloudera Manager 通过包含一系列的报道和诊断工具，可以帮助您优化集群性能，并且提高利用率。Cloudera Manager 能够为您提供以下的功能：

- 自动化 Hadoop 安装过程，大幅缩短部署时间；
- 提供实时的集群概况，例如节点、服务的运行状况；
- 提供了集中的中央控制台对集群的配置进行更改。
- 包含全面的报告和诊断工具，帮助优化性能和利用率。



Cloudera Manager 的架构如上图所示，主要由如下几部分组成：

- 服务端/Server: Cloudera Manager 的核心。主要用于管理 web server 和应用逻辑。它用于安装软件，配置，开始和停止服务，以及管理服务运行的集群。
- 代理/agent: 安装在每台主机上。它负责启动和停止的进程，部署配置，触发安装和监控主机。
- 数据库/Database: 存储配置和监控信息。通常可以在一个或多个数据库服务器上运行的多个逻辑数据库。例如，所述的 Cloudera 管理器服务和监视后台程序使用不同的逻辑数据库。
- Cloudera Repository: 用于软件分发由 Cloudera 软件仓库。
- 客户端/Clients: 提供了一个与 Server 交互的接口：
 - 管理平台/Admin Console: 提供一个管理员管理集群和 Cloudera Manage 的基于网页的交互界面。
 - API: 为开发者提供了创造自定义 Cloudera Manager 程序的 API。

2. 功能

此部分内容在免费版本 Cloudera Express 产品以及商业版本中可用。

2.1. 管理

2.1.1. 参数配置

2.1.1.1. 可视化参数配置界面

cloudera manager

[主页](#)
[群集](#)
[主机](#)
[诊断](#)
[审核](#)
[图表](#)
[脚本](#)
[备份](#)
[管理](#)

Cluster 1

HDFS

状态

实例

配置

命令

审核

文件浏览器

图表

配置

命令

审核

文件浏览器

图表

搜索

搜索

角色组

历史记录和观察

备注

保存更改

下方有 1 个验证警告。

类别	属性	值	说明
<div> <div>基础范围</div> <div> <div>Balancer Default Group</div> <div>Datatalde Default Group</div> <div>Datatalde Group 1</div> <div>Datatalde Group 2</div> <div>Datatalde Group 3</div> <div>Fallover Controller Default Group</div> <div>Gateway Default Group</div> <div>HTTPFS Default Group</div> <div>JournalNode Default Group</div> <div>NFS Gateway Default Group</div> <div>NameNode Default Group</div> <div>SecondaryNameNode Default Group</div> </div> </div>	<div>ZooKeeper 服务</div> <div>KMS Service</div> <div> <div>HDFS 块大小</div> <div>dfs.block.size, dfs.blocksize</div> </div> <div> <div>默认值 Umask</div> <div>dfs.umaskmode, fs.permissions umask-mode</div> </div> <div> <div>启用 WebHDfs</div> <div>dfs.webhdfs.enabled</div> </div> <div> <div>检查 HDfs 权限</div> <div>dfs.permissions</div> </div> <div> <div>压缩编码解码器</div> <div>io.compression.codecs</div> </div>	<div> <div> <div> <div>● ZooKeeper</div> <div>● none</div> <div>重选为空时默认值</div> </div> </div> <div>默认值为空，单选以圆点。</div> <div>128 字节</div> <div>默认值</div> <div>022</div> <div>默认值</div> <div> <div> <div> <div>默认值</div> <div>默认值</div> </div> </div> <div>默认值</div> </div></div>	<div>此 HDfs 服务实例依赖的 ZooKeeper 服务的名称</div> <div>The Key Management Server used by HDfs. This must be set in order to use encryption for data at rest</div> <div>以字节计算的新建 HDfs 文件默认块大小。请注意该值也用作 HBase 区域服务器 HLog 块大小。</div> <div>将文件和目录创建的默认值设为 umask，指定为十进制数（开头数字为 0）</div> <div>启用 WebHDfs 界面</div> <div>如果失败，根据检查将关闭 HDfs 中的文件。</div> <div>可用于作业或 map 压缩中的压缩编码解码器的逗号分隔列表。</div>

2.1.1.2. 过期设置报警

cloudera manager 主页 群集 主机 诊断

主页 状态 所有运行状况问题 所有配置问题 6

Cluster 1 (CDH 5.3.2, Parcel)

主机		
HBase		C
HDFS	✖ 1	C
Hive		C
Hue	✖ 1	C
Impala		C
Key-Value Stor...		C
Oozie		C
Solr		C
Spark		C
Sqoop 2		C
YARN (MR2 Inc...		C
ZooKeeper		

Cloudera Management Service

Cloudera Mana...	✖ 4	
------------------	-----	--

配置过期需要重启

客户端配置过期

2.1.1.3. 智能参数验证及优化

cloudera manager 主页 群集 主机 诊断 审核 图表 备份 管理

主页 状态 所有运行状况问题 所有配置问题 5 所有最新命令 1

所有配置问题

▼ 其他

- Cloudera Management Service: [Service Monitor 的 Java 堆栈大小 \(字节\)](#)
推荐的堆大小为 256.0 MiB 字节，大于配置 27.0 MiB。
- Cloudera Management Service: [Service Monitor 的最大非 Java 内存](#)
建议的非 Java 内存大小为 1.5 GiB，大于配置 98.0 MiB。
- Cloudera Management Service: [Host Monitor 的 Java 堆栈大小 \(字节\)](#)
推荐的堆大小为 256.0 MiB 字节，大于配置 27.0 MiB。
- Cloudera Management Service: [Host Monitor 的最大非 Java 内存](#)
建议的非 Java 内存大小为 1.5 GiB，大于配置 98.0 MiB。

▼ Cluster 1

- HDFS: [Namenode 的 Java 堆栈大小 \(字节\)](#)
Java Heap Size of Namenode in Bytes is recommended to be at least 1GB for every million HDFS blocks. Suggested minimum value: 1073741824

2.1.2. 管理员权限管理

对系统管理员、数据库管理员及其他管理员必须授予不同级别的管理权限。

2.2. 监控

Cloudera Manager 提供了许多监控功能，用于监测群集（主机，服务守护进程）健康、组件性能以及集群中运行的作业的性能和资源需求。具体包括以下监控功能：

- 服务监控

查看服务和角色实例级别健康检查的结果，并通过图表显示，有助于诊断问题。如果健康检查发现某个组件的状态需要特别关注甚至已经出现问题，系统会对管理员应该采取的行动提出建议。同时，系统管理员还可以查看上的服务或角色上操作的执行历史，也可以查看配置更改的审计日志。

- 主机监控

监控群集内所有主机的有关信息，包括：哪些主机上线或下线，主机上目前消耗的内存，主机上运行的角色实例分配，不同的机架上的主机分配等等。汇总视图中显示了所有主机群集，并且可以进一步查看单个主机丰富的细节，

包括显示主机关键指标的直观图表。

- 行为监控

Cloudera Manager 提供了列表以及图表的方式来查看集群上进行的活动，不仅显示当前正在执行的任务行为，还可以通过仪表盘查看历史活动。同时提供了各个作业所使用资源的许多统计，系统管理员可以通过比较相似任务的不同性能数据以及比较查看同一任务中不同执行的性能数据来诊断性能问题或行为问题。

- 事件活动

监控界面可以查看事件，并使它们用于报警和搜索，使得系统管理员可以深入了解发生集群范围内所有相关事件的历史记录。系统管理员可以通过时间范围、服务、主机、关键字等字段信息过滤事件。

- 报警

通过配置 Cloudera Manager 可以对指定的事件产生警报。系统管理员可以针对关键事件配置其报警阈值、启用或禁用报警等，并通过电子邮件或者通过 SNMP 的事件得到制定的警报通知。系统也可以暂时抑制报警事件，此限制可以基于个人角色、服务、主机、甚至整个集群配置，使得进行系统维护/故障排除时不会产生过多的警报流量。

- 审计事件

Cloudera Manager 记录了有关服务、角色和主机的生命周期的事件，如创建角色或服务、修改角色或服务配置、退役主机和运行 Cloudera Manager 管理服务命令等等。系统管理员可以通过管理员终端查看，界面提供了按时间范围、服务、主机、关键字等字段信息来过滤审计事件条目。

- 可视化的时间序列数据图表

系统管理员可以通过搜索度量数据，系统将根据指定规则创建数据，组（方面）数据的图表，并把这些图表保存到用户自定义的仪表板。

- 日志

介绍如何访问日志中的各种考虑到你正在查看当前上下文的方式。例如，监控服务时，你可以轻松地点击一个链接，查看相关的特定服务的日志条目，通过相同的用户界面。当查看关于用户的活动信息，您可以方便地查看了作业运行时所用的作业的主机上发生的相关日志条目。

- 报告

Cloudera Manager 可以将收集到的历史监控数据统计生成报表，比如按目录查看集群作业活动的用户、按组或作业 ID 查看有关用户的磁盘利用率，用

户组的历史信息等。这些报告可以根据选定的时间段（每小时，每天，每周，等等）汇总数据，并可以导出为 XLS 或 CSV 文件。同时系统管理员还可以管理包括搜索和配额等 HDFS 目录设置。

2.3. 批量部署

Hadoop 本身是一个分布式的系统，因此在安装时，需要对每一个节点进行组件的安装，并且由于是开软软件，其安装过程相对比较复杂，Hadoop 每个组件都需要做很多的配置工作，如果手动去完成这个工作，工作量将会非常可观。

针对这种情况，CDH 提供了 Cloudera Manager 来自动化安装部署 Hadoop。大大缩短了 Hadoop 的安装时间，同时也简化了安装 Hadoop 的过程。

CDH 的安装针对不同的客户环境，提供了多种不同的安装方式：

1. 如果用户的集群能够访问到 Cloudera 的网站，直接可以进行完全自动化的安装。
2. 如果用户的集群不能够访问公网，则只需要在集群内部建立一个软件仓库，然后再进行自动化的集群部署

其中上述提到的自动化安装的过程如下：

1. 安装环境准备，下载 Cloudera Manager 以及 CDH 的安装文件，安装 JDK，yum 等基本软件。
2. 挑选一台节点，安装 Cloudera Manager，用户只需要启动安装脚本即可，通常情况下 2 分钟就能够完成。
3. Cloudera Manger 是一个 Web 应用，提供了基于浏览器的界面，用户可以通过浏览器可视化的进行 CDH 的安装部署。
4. 通过 Cloudera Manager 界面，添加其他需要安装的节点，选择要安装的 Hadoop 组件，以及每个节点承担的角色，选择安装，Cloudera Manager 会自动的将需要安装的软件分发到对应的节点，并完成安装。
5. 同时 Cloudera Manager 会提供一个默认的 CDH 配置，并且提供了配置的管理界面，供用户在后续进行定制化的配置。
6. 在安装过程中，如果任意一台机器安装出现问题，Cloudera Manager 都会进行汇报，并提供日志，供用户分析，解决问题。

7. 当所有节点的软件都安装完成之后，Cloudera Manager 会启动所有的服务。

从上述的安装过程可以看出，CDH 的安装主要体现两个特点，批量化以及自动化。只需要在其中一个节点完成，其他节点都可以进行批量化的自动安装。

2.4. 集成

2.4.1. 安全集成

通常在一个企业内部，都会对验证系统作集中的部署，例如使用 AD，LDAP 等验证服务作为集中式的验证服务器，提供对所有业务系统的验证工作。为了方便 Hadoop 大数据平台与原有验证系统的集成，CM 提供了丰富的集成功能，只需要在界面，进行简单的配置，即可完成。

2.4.2. ClouderaManager API

Cloudera 产品具有开放的特性，这种开放性的其中一个体现就是 CM 提供了丰富的 API，供客户调用，基本上所有在界面上提供的功能，通过 API 都可以完成同样的工作。Cloudera Manager API 支持的功能包括配置和服务生命周期管理、服务健康信息和指标，并允许配置 Cloudera Manager 本身。API 复用 Cloudera Manager 管理控制台（Cloudera Manager Admin Console）相同的主机和端口，无需额外的操作流程或参数配置。API 支持 HTTP 基本身份验证（HTTP Basic Authentication），接受与 Cloudera Manager 管理控制台相同的用户和凭据。并且这些 API 都是标准的 REST API，使用 HTTP 标准的 CRUD 所有的请求和响应都可以用 JSON（JavaScript Object Notation）来表示，基本上所有的业务系统都应该支持，通过这些 API，能够方便的将 CM，集成到企业原有的集中管理系统。

下面是一些可以通过 Cloudera Manager API 做的操作：

- 通过编程部署整个 Hadoop 集群。Cloudera Manager 支持 HDFS、MapReduce、Yarn、Zookeeper、HBase、Hive、Oozie、Hue、Flume、Impala、Solr、Sqoop、Spark 和 Accumulo 等；
- 配置各种 Hadoop 服务并验证；
- 开展服务和角色的管理行为，如启动、停止、重新启动、故障转移等，同时提供了更先进的工作流程，如设立高可用性和退役；

- 通过具有智能化服务的健康检查和指标来监控服务和主机；
- 监控用户的工作和其他集群活动；
- 检索基于时间序列的度量数据；
- 搜索 Hadoop 系统内事件；
- 管理 ClouderaManager 自身；
- 将 Hadoop 集群的整个部署描述下载为一个 JSON 文件；
- 等等

此外，基于企业版本产品，，该 API 还允许一些高级功能：

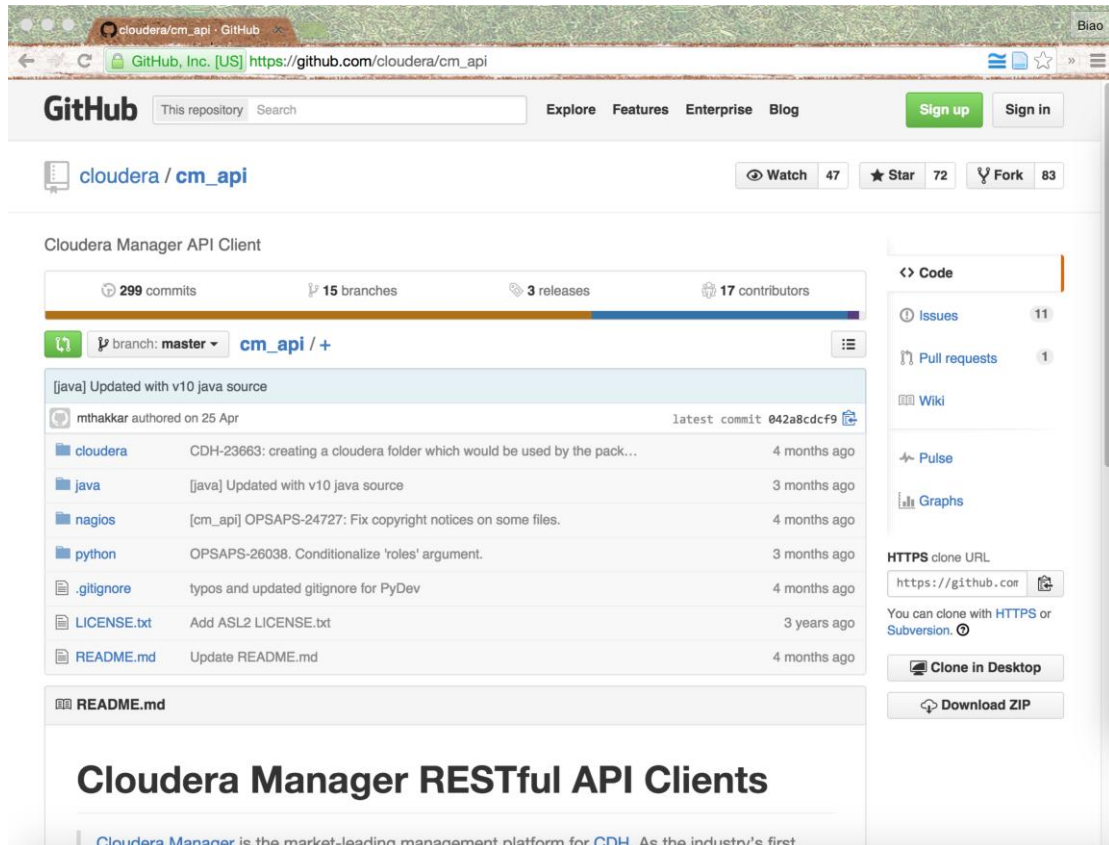
- 执行滚动重启和滚动升级；
- 审计用户在 Hadoop 上的活动和访问；
- 为 HDFS 和 Hive 执行备份和跨数据中心复制；
- 检索每个用户的 HDFS 使用情况报告和每个用户的 MapReduce 的资源使用情况报告。

例如，下面的 HTTP 请求示例可以得到当前所有集群的基本信息：

```
$ curl -u admin:admin 'http://localhost:7180/api/v1/clusters'
{
  "items" : [ {
    "name" : "Cluster 1 - CDH4",
    "version" : "CDH4"
  }, {
    "name" : "Cluster 2 - CDH3",
    "version" : "CDH3"
  } ]
}
```

2.4.2.1. Cloudera Manager 客户端实现

用户可以通过 Java 或 Python 轻松地构建一个 Cloudera 的管理客户端应用，或者直接使用 HTTP 协议发送 Rest 风格请求。Cloudera 在 github 上开源了 Cloudera Manger 的客户端 Rest API 项目，包含所有源代码以及例子，并采用 Apache 协议授权。



这些客户端的代码使用 Cloudera 的管理器进行各种交互式管理，比如：

- 管理多个集群
- 启动和停止全部或单个服务或角色
- 升级群集上运行的服务
- 查系统中的任何活动资源使用的时间序列数据对
- 读取系统中的所有进程日志以及输出和错误
- 编程配置各个部署的参数
- 收集诊断数据的调试问题，以帮助调试问题
- 运行分布式命令来管理自动故障转移（fail over）、主机退役等
- 查看所偶遇发生在系统中的事件和警报
- 从系统中添加和删除用户

比如，如下的示例演示了在系统显示终端打印出集群、服务以及角色资源：

```

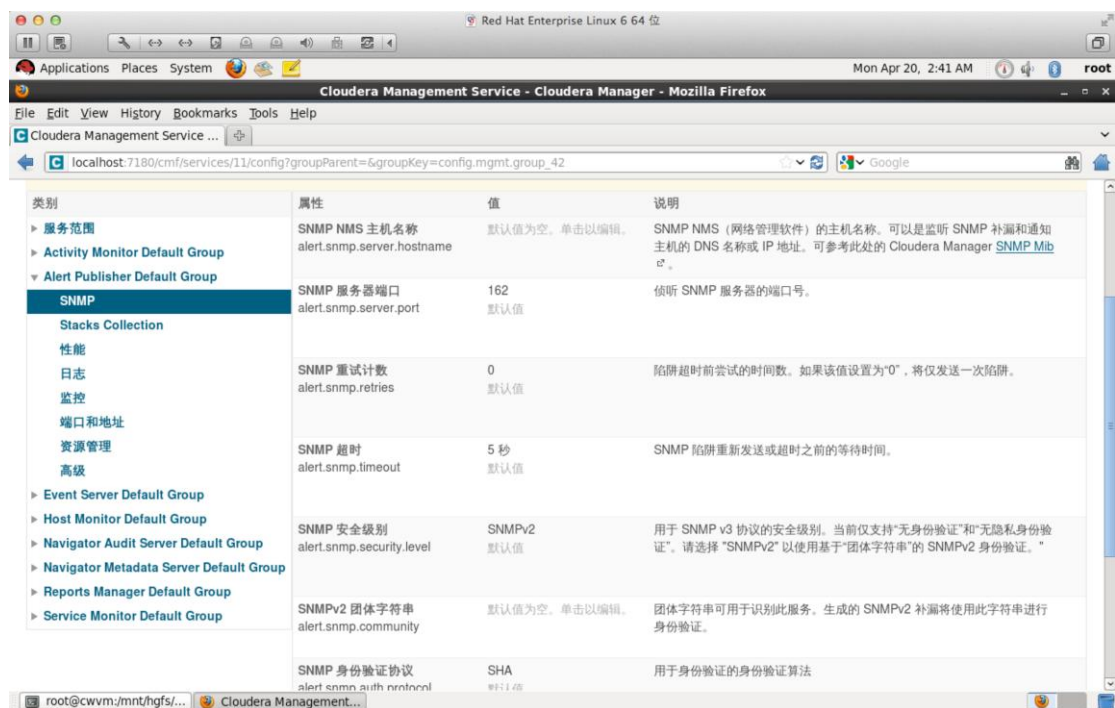
ClustersResourceV2 clustersResource = v2.getClustersResource();
for (ApiCluster cluster : clustersResource.readClusters(DataView.FULL)) {
    System.out.println(cluster.getName());
    ServicesResourceV2 servicesResource = clustersResource.getServicesResource(cluster.getName());
    for (ApiService service : servicesResource.readServices(DataView.FULL)) {
        System.out.println("\t" + service.getName());
        RolesResourceV2 rolesResource = servicesResource.getRolesResource(service.getName());
        for (ApiRole role : rolesResource.readRoles()) {
            System.out.println("\t\t" + role.getName());
        }
    }
}
}

```

请参考 CM 客户端 API 开源项目（[链接](#)，[链接](#)）。

2.4.3. SNMP 集成

SNMP 是一个标准的消息转发协议。通常在大型的企业级系统中都会进行部署。CM 提供了方便的 SNMP 集成能力，只需要进行简单的部署，就能够将 SNMP 进行集成，并且将集群中的告警信息进行转发。



3. 高级功能

以下功能只包含在商业版本中，在免费版本中不提供。有关商业版本许可证

信息，请参见[链接](#)。

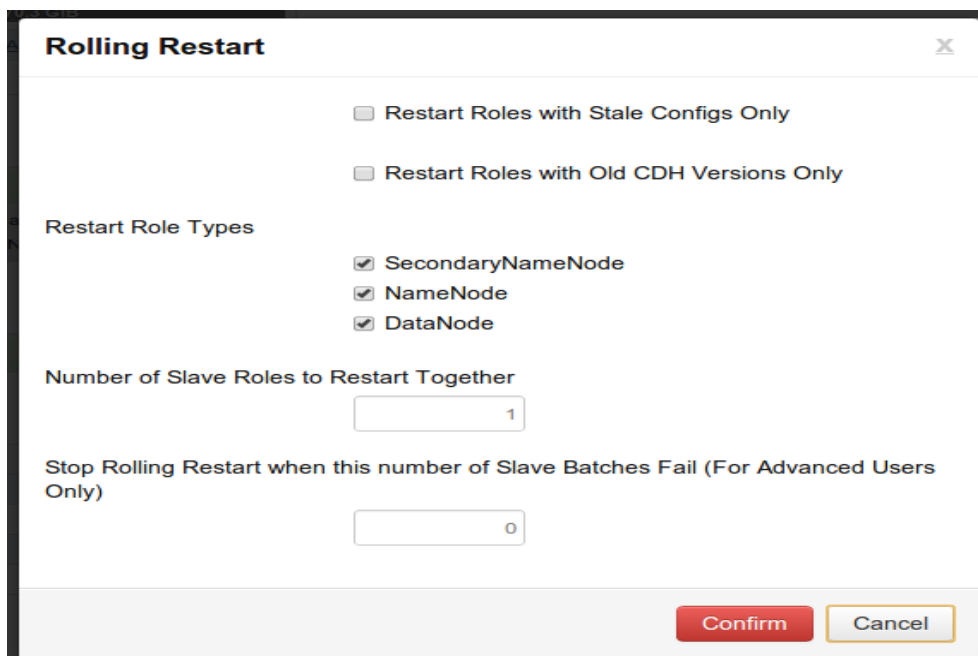
3.1. 软件滚动升级

Hadoop 是一个正在快速发展的新技术，版本正在迅速的衍进，对于部署了 Hadoop 的系统，无论出于稳定性还是功能性的原因，版本升级是不可避免的。但是，通常业务系统都是在线的生产系统，需要 24 小时不间断的提供对外服务。

没有滚动升级之前，Hadoop 平台，要进行组件升级时，需要停止集群运行，然后进行升级，再启动集群，这样带来的问题是，在进行升级时，需要停集群，停服务，停应用。这样会严重影响持续的服务能力。整个过程如果由系统管理员手工操作，一个规模为 100 台服务器的集群通常需要 1 周左右，300 台服务器的集群通常需要 3 周时间。由于 Hadoop 产品开源，需要不断升级来保证新近发现重大 bug 及时修复，因此通常每 3 个月需要一次升级或打补丁。

因此，CM 提供了滚动升级的功能，支持在平台进行升级时，能够继续对外提供服务，以及应用。其原理是在升级时，首先停掉一个节点，然后进行升级，然后再启动，然后循环的进行下一个节点的升级。最后完成整个集群的升级。通过逐个节点升级的方式，保证集群的对外服务功能，并且最终完成整个集群的升级。由于分布式系统的可扩展性，在升级过程中，集群依然能够持续运行，持续对外提供服务。

在 Cloudera 管理界面中提供了滚动升级的向导，方便管理员进行操作：



The image shows a 'Rolling Restart' dialog box from the Cloudera management interface. It contains several configuration options for performing a rolling restart of Hadoop roles. At the top, there are two unchecked checkboxes: 'Restart Roles with Stale Configs Only' and 'Restart Roles with Old CDH Versions Only'. Below these is a section titled 'Restart Role Types' which contains three checked checkboxes: 'SecondaryNameNode', 'NameNode', and 'DataNode'. Further down is a field labeled 'Number of Slave Roles to Restart Together' with a value of '1'. At the bottom, there is a field labeled 'Stop Rolling Restart when this number of Slave Batches Fail (For Advanced Users Only)' with a value of '0'. The dialog box has a title bar with a close button and two buttons at the bottom right: 'Confirm' and 'Cancel'.

Rolling Restart

☐ Restart Roles with Stale Configs Only

☐ Restart Roles with Old CDH Versions Only

Restart Role Types

☒ SecondaryNameNode

☒ NameNode

☒ DataNode

Number of Slave Roles to Restart Together

1

Stop Rolling Restart when this number of Slave Batches Fail (For Advanced Users Only)

0

Confirm **Cancel**

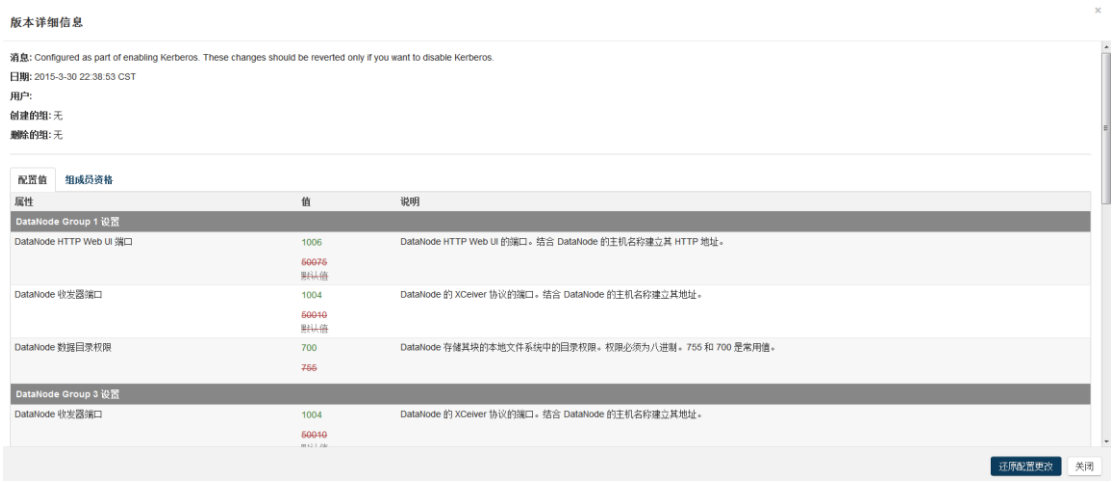
3.2. 参数版本控制

3.2.1. 配置历史查看



3.2.2. 配置回滚

CM 提供了配置的历史记录与回滚，用户可以查看参数的修改历史，可以方便的回滚到某一个具体的阶段。



3.3. 备份及容灾系统 BDR

Cloudera 为 Hadoop 平台提供了一个集成的易用的灾备的解决方案（Cloudera BDR）。BDR 为灾备方案提供了丰富的功能。对于企业级数据中心来说，通常存储的是非常关键的企业数据，BDR 能够保证数据始终是可用的，哪怕面对整个数据中心断电的情况。

Cloudera Manager 为 BDR 提供了完整的用户界面，包括如下的功能：

1. 用户可以自由选择与业务相关的关键数据进行灾备
2. 创建数据备份的调度策略。由于数据备份通常需要较大的网络带宽，如果没有良好的调度，将会影响业务系统的运行，需要对于备份的时间，备份的带宽等加以限制。
3. 通过一个集中的控制台跟踪数据备份任务的进度，以及备份过程中出错的文件。
4. 针对出错的任务，或者异常中断的任务，提供告警机制，以及时让用户感知。

这些功能都能够无缝的和 HDFS, Hive 等存储的备份进行结合，并且在备份的过程中，并不需要对存储于 Hive 表中的数据进行人为的转换。数据的备份同时也包括元数据的备份，因此，备份出来的数据仍然能够体现表格的元数据信息。

开源的 Hadoop 默认提供了集群间的数据拷贝方案“distcp”，Disctp 是一个专门的 Mapreduce 任务，用于并行的进行文件复制。Cloudera 的数据备份方案建立在 Distcp 之上，进行了性能上的优化，提供更高效，更快速的文件复制。

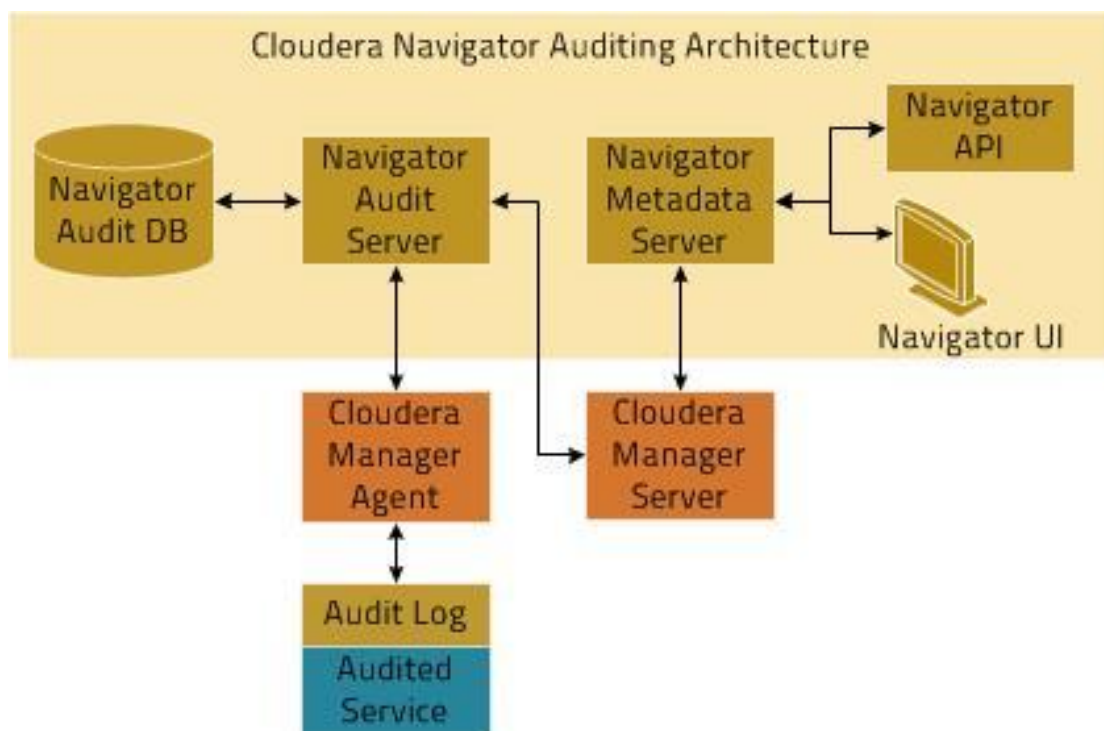
在新版本中，BDR 还提供了一个新功能，能够在运行备份之前，估算整个过程需要的资源和开销，帮助用户评估灾备的风险。

目前通常的灾备解决方案是，一方面利用数据快照，对存储于 Hadoop 的数据进行增量快照，同时利用 BDR 所提供的页面，对数据备份进行定制化，通过将数据快照备份到另一个 Hadoop 集群。当目前运行的 Hadoop 集群出现了严重的问题，数据丢失或者无法恢复时，可以通过存储在备用集群的数据，进行灾难恢复。



3.4. 数据审计

Cloudera Navigator 的审计功能支持对于数据的审计和访问，其架构如下：



一旦配置了 Cloudera Navigator 审计功能，收集和过滤审核事件的插件将会被打开并插入到 HDFS, HBase 和 Hive(也就是 HiveServer2 和 Beeswax 服务器) 服务。该插件负责将审计事件写入到本地文件系统的审计日志中。而 Cloudera Impala 和 Sentry 则自己收集和过滤审核事件，并直接将其写入审计日志文件。支持的审计事件比如集群资源管理操作节点操作、用户操作、资源服务的启动和停止、应用程序操作等。

运行在每台服务器上的 CM 代理 (Cloudera Manager Agent) 会监视审计日志文件，并将这些事件发送到 Navigator 审计服务器 (Navigator Audit Server)。如果任何消息在发送时失败，CM 代理将会自动重试。由于没有使用临时内存缓冲区，一旦审计事件被写入审计日志文件，就可以保证被交付 (只要文件系统可用)。CM 代理记录并跟踪审计日志中当前成功发送的的审计事件，因此及时出现任何崩溃或重新启动，只需找到上次成功发送审计事件的位置，然后就可以重新开始发送。审核日志一旦被成功发送到 Navigator 审计服务器，CM 代理会负责清除。如果一个插件无法将审计事件写入审计日志文件，它可以丢弃事件或停止正在运行的进程 (取决于配置的策略)。

Navigator 审计服务器执行以下功能：

- 跟踪和合并事件
- 存储事件给审计数据库

3.5. 安全集成

3.5.1. Kerberos 集成

Hadoop 支持多种安全机制，例如 Kerberos。由于使用这些安全机制的配置工作以及管理工作都非常繁琐且复杂，CM 在界面上提供了启用 Kerberos、基于角色的访问控制等功能的途径，用户只需要通过简单的页面点击就能够完成复杂的配置工作。

步骤一：

启用 Kerberos 用于 Cluster 1

欢迎

此向导将带您完成配置 Cloudera Manager 和 CDH 以使用 Kerberos 进行身份验证的步骤。群集以及 Cloudera Management Service 中的所有服务将作为向导的一部分重启。在继续使用向导前，请阅读有关启用 Kerberos 的[文档](#)。

使用向导前，请确保已执行以下步骤：

设置正在运行的 KDC。Cloudera Manager 支持 MIT KDC 和 Active Directory。

☒ 是的，我已设置正在运行的 KDC。

KDC 应配置为拥有非零票证生存期和可更新的生存期。如果票证不可更新，则 CDH 不能正常工作。

☒ 是的，我已检查 KDC 允许可更新的票证。

如果想使用 Active Directory，OpenLdap 客户端库应安装在 Cloudera Manager Server 主机中。另外，Kerberos 客户端库应安装在所有主机中。

☒ 是的，我已安装客户端库。

Cloudera Manager 需要有权限在 KDC 中创建其他帐户的帐户。

☒ 是的，我已为 Cloudera Manager 创建适当的帐户。

步骤二：

启用 Kerberos 用于 Cluster 1

KDC 信息

指定有关 KDC 的信息。Cloudera Manager 使用下面的属性生成在群集中运行的 CDH 守护程序的主体。

KDC 类型

MIT KDC

Active Directory

用于在 CDH 群集中进行身份验证的 KDC 类型。

KDC Server 主机

sb-node1.example.com

C

kdc

KDC Server 所在的主机。

Kerberos 安全领域

EXAMPLE.COM

C

default_realm

Kerberos 安全所用领域。注：更改该设置将清除 Cloudera Manager 中现有的所有凭据和 Keytab。

Kerberos 加密类型

rc4-hmac

+ -

KDC 支持的加密类型。备注：如果想使用 AES 加密，确保按[此处](#)的说明部署 JCE 无限强度政策文件。

Kerberos Principal 最大可更新生命期

7

天

C

Cloudera Manager 生成的 Kerberos 主体最大可更新生存期。只有在使用 MIT KDC 时才使用该属性。如果 KDC 应该提供最大可更新生存期，则将该属性设为零。备注：建议不要采用票证不可更新的主体，因为会妨碍 Hadoop 服务运行。

步骤三：

KRB5 配置

指定生成群集的 krb5.conf 所需的属性。可以使用安全词字段指定高级 KDC 设置的配置，例如，使用跨领域身份验证。

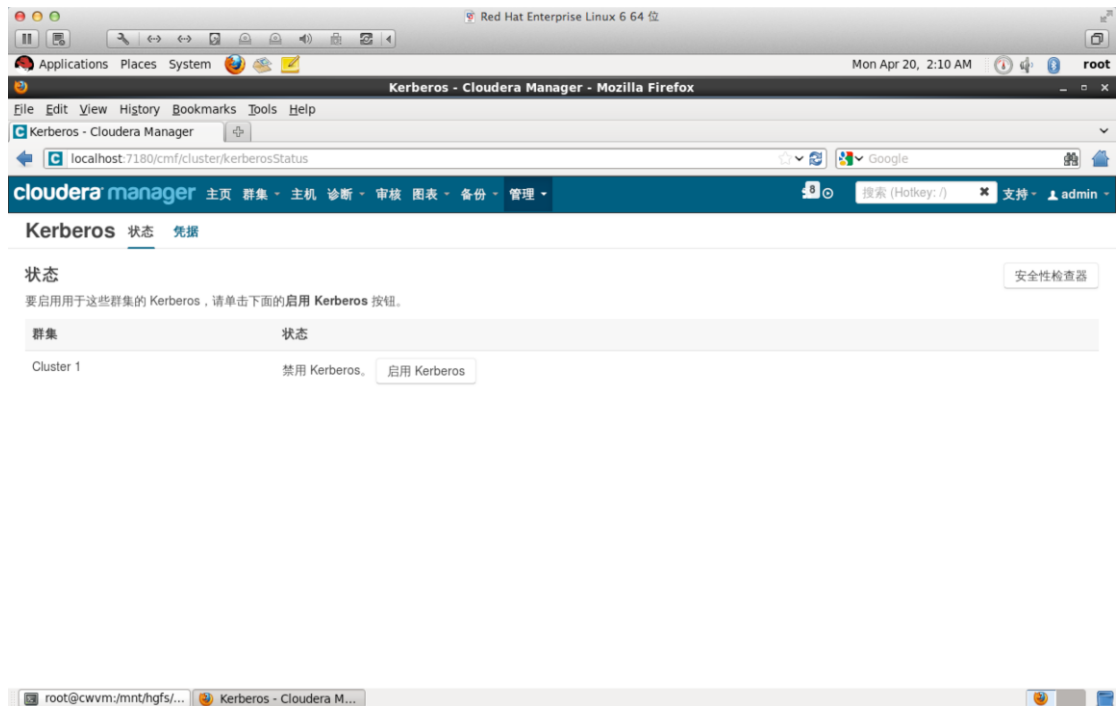
通过 Cloudera Manager 管理 krb5.conf

Cloudera Manager 是否应该在安全群集中配置和部署 krb5.conf。如果没有选中该属性，则必须确保在安全群集的主机中以及 Cloudera Manager Server 的主机中部署 krb5.conf。

步骤四：执行并完成

进度 				
命令	上下文	状态	开始日期	结束于
 启用 Kerberos	Cluster 2	正在进行	2015-3-31 13:24:18 UTC	
命令进度				
已完成 4 个步骤（共 7 个）。 <div><div></div></div>				
<div><div><div>✔</div><div>停止群集</div><div>All services successfully stopped.</div><div>详细信息</div></div></div>				
<div><div><div>✔</div><div>停止 Cloudera Management Service</div><div>Command completed with 6/6 successful subcommands</div><div>详细信息</div></div></div>				
<div><div><div>✔</div><div>将所有服务配置为使用 Kerberos</div><div>已成功完成 16 个步骤。</div></div></div>				
<div><div><div>✔</div><div>等待生成凭据</div><div>已成功完成命令 (829)</div></div></div>				
<div><div><div><div><div><div></div></div></div><div>部署客户端配置</div><div>详细信息</div></div></div></div>				
<div>启动 Cloudera Management Service</div>				
<div>启动群集</div>				

步骤五：服务监控及配置



3.5.2. 外部安全认证集成

Cloudera Manager 支持通过内部数据库或这外部服务进行用户认证。具体支持的外部服务如下，详细信息参见官方文档（[链接](#)）。

3.5.2.1. Active Directory

3.5.2.2. OpenLDAP 兼容服务器

Settings - Cloudera Manager

10.0.57.2:7190/cmf/settings

cloudera manager


Home Clusters Hosts Diagnostics Audits Charts Back Administration

Search (Hotkey: /) Support admin

Settings

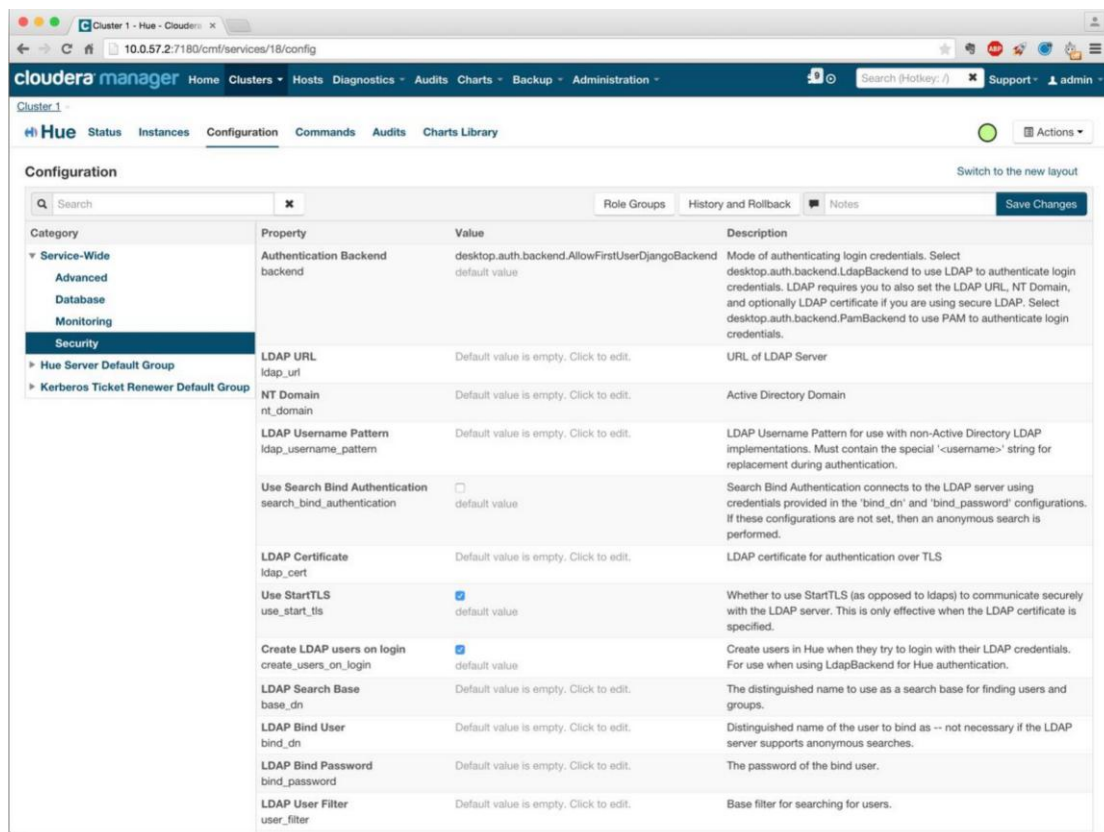
Switch to the new layout

Search

Settings marked  will not take effect until after the Cloudera Manager Server has restarted.

Save Changes

Category	Property	Value	Description
Performance	Authentication Backend Order	Database Only	The order in which authentication backends are used for authenticating a user.
Advanced	External Authentication Type	Active Directory	The type of external authentication to use.
Monitoring	LDAP URL	Default value is empty. Click to edit.	URL of the LDAP server to authenticate against
Security	LDAP Bind User Distinguished Name	Default value is empty. Click to edit.	Distinguished name of the user to bind as for searches.
Ports and Addresses	LDAP Bind Password	Default value is empty. Click to edit.	Password for the bind user
Kerberos	Active Directory NT Domain	Default value is empty. Click to edit.	Active Directory NT Domain to authenticate against
Other	LDAP User Groups	Default value is empty. Click to edit.	A list of LDAP group names. If a user is not a member of one of the configured groups, that user is prevented from logging into Cloudera Manager. If this is left empty, all LDAP users will be able to log in.
Support	LDAP User Search Filter	Default value is empty. Click to edit.	A search filter for finding users. Typically, this will be (uid=[0]), where [0] will be replaced by the username that was used at the login screen.
External Authentication	LDAP User Search Base	Default value is empty. Click to edit.	A base distinguished name for searching for users. This may be used as a fallback mechanism if the DN pattern does not match any user.
Parcels	LDAP Group Search Filter	Default value is empty. Click to edit.	A search filter for finding groups. Typically, this will be (member=[0]), where [0] will be replaced by the DN of a successfully authenticated user.
Network	LDAP Group Search Base	Default value is empty. Click to edit.	A base distinguished name for searching for groups.
Custom Service Descriptors	LDAP Full Administrator Groups	Default value is empty. Click to edit.	A list of LDAP group names. If a user is a member of one of the configured groups, they will be granted Full admin access upon logging in.
	LDAP User Administrator Groups	Default value is empty. Click to edit.	A list of LDAP group names. If a user is a member of one of the configured groups, that user is granted User admin access on login.
	LDAP Cluster Administrator Groups	Default value is empty. Click to edit.	A list of LDAP group names. If a user is a member of one of the configured groups, that user is granted Cluster admin access on login.
	LDAP BDR Administrator Groups	Default value is empty. Click to edit.	A list of LDAP group names. If a user is a member of one of the configured groups, that user is granted BDR admin access on login.
	LDAP Configurator Groups	Default value is empty. Click to edit.	A list of LDAP group names. If a user is a member of one of the



3.5.2.3. 外部程序

3.5.2.4. SAML

4. 监控指标

Cloudera Manager 提供多达 102 类监控指标(见下表),覆盖所有的服务及功能,包括集群硬件使用情况(网络、CPU、内存以及硬盘等)、服务状态等,同时指标按集群级别、主机级别、用户级别以及表/目录级别等等分级统计,总指标项上万,比如:集群指标超过 3000 个、HBase 系统级指标就超过 1000 个、HDFS 系统级指标超过 300 个等等。具体指标列表可查阅如下在线网页资料:

<http://www.cloudera.com/content/cloudera/en/documentation/core/latest/to>

pics/cm_metrics.html

Accumulo	History Server	Navigator Metadata Server
Accumulo 1.6	Hive	Network Interface
Activity	Hive Metastore Server	NodeManager
Activity Monitor	HiveServer2	Oozie
Agent	Host	Oozie Server
Alert Publisher	Host Monitor	Rack
Attempt	HttpFS	RegionServer
Beeswax Server	Hue	Reports Manager
Cloudera Management Service	Hue Server	ResourceManager
Cloudera Manager Server	Impala	SecondaryNameNode
Cluster	Impala Catalog Server	Sentry
DataNode	Impala Daemon	Sentry Server
Directory	Impala Daemon Pool	Server
Disk	Impala Llama ApplicationMaster	Service Monitor
Event Server	Impala Pool	Solr
Failover Controller	Impala Query	Solr Collection
Filesystem	Impala StateStore	Solr Replica
Flume	Isilon	Solr Server
Flume Channel	JobHistory Server	Solr Shard
Flume Channel Tier	JobTracker	Spark
Flume Sink	JournalNode	Spark (Standalone)
Flume Sink Tier	Kerberos Ticket Renewer	Sqoop 1 Client
Flume Source	Key Management Server	Sqoop 2
Flume Source Tier	Key Management Server Proxy	Sqoop 2 Server
Garbage Collector	Key Trustee	Tablet Server
HBase	Key-Value Store Indexer	TaskTracker
HBase Namespace	Lily HBase Indexer	Time Series Table
HBase REST Server	Logger	Tracer
HBase Region	MapReduce	User
HBase Table	Master	WebHCat Server
HBase Thrift Server	Monitor	Worker
HDFS	NFS Gateway	YARN (MR2 Included)
HDFS Cache Directive	NameNode	YARN Pool
HDFS Cache Pool	Navigator Audit Server	ZooKeeper

Cloudera Manager 监控指标分类索引表

4.1. 部分样例状态指标

4.1.1. HDFS 指标

KPI 名称	KPI 描述	单位	级别
CPU 占用率	CPU 平均占用率	%	系统级/节点级
内存占用率	内存平均占用率	%	系统级/节点级
系统空间	总空间	MB	系统级/节点级
已用空间	已用空间	MB	系统级/节点级
可用空间	剩余空间	MB	系统级/节点级
空间使用率	已用空间与系统空间的比值	%	系统级/节点级
读流量	统计周期内读流量统计	MB	系统级/节点级
写流量	统计周期内写流量统计	MB	系统级/节点级
读 IOPS	每秒进行读 (I/O) 操作的次数	个/s	系统级/节点级
写 IOPS	每秒进行写 (I/O) 操作的次数	个/s	系统级/节点级

4.1.2. MapReduce 指标

KPI 名称	KPI 描述	单位	级别
提交作业数	提交作业数	个	系统级
完成作业数	完成作业数	个	系统级
失败作业数	失败作业数	个	系统级
正在运行的作业数	正在运行的作业数	个	系统级
Map 总任务数	Map 总任务数	个	系统级
Reduce 总任务数	Reduce 总任务数	个	系统级
Map 任务完成数	Map 任务完成数	个	系统级
Reduce 任务完成数	Reduce 任务完成数	个	系统级
正在执行的 Map 任务数	正在执行的 Map 任务数	个	系统级

正在执行的 Reduce 任务数	正在执行的 Reduce 任务数	个	系统级
平均 Map 任务执行时间	平均 Map 任务执行时间	秒	系统级
平均 Reduce 任务执行时间	平均 Reduce 任务执行时间	秒	系统级
最小 Map 任务执行时间	最小 Map 任务执行时间	秒	系统级
最小 Reduce 任务执行时间	最小 Reduce 任务执行时间	秒	系统级
最大 Map 任务执行时间	最大 Map 任务执行时间	秒	系统级
最大 Reduce 任务执行时间	最大 Reduce 任务执行时间	秒	系统级
Map 任务执行失败数	Map 任务执行失败数	个	系统级
Reduce 任务执行失败数	Reduce 任务执行失败数	个	系统级
Map 任务执行成功数	Map 任务执行成功数	个	系统级
Reduce 任务执行成功数	Reduce 任务执行成功数	个	系统级

4.1.3. HBase 指标

KPI 名称	KPI 描述	单位	级别
系统空间	系统总空间	MB	系统级/
已用空间	NoSQL 系统已用数据量（系统级）； RegionServer 的 Region 个数（节点级）	MB	系统级/节点级
可用空间	NoSQL 系统可用数据量	MB	系统级/
空间使用比率	已使用的空间与总空间的比值	%	系统级/
CPU 平均使用率	NoSQL 系统平均 CPU 使用率	%	系统级/节点级
CPU IOWait 使用率	NoSQL 系统平均 CPU IOWait 使用率	%	系统级/节点级
压缩合并队列长度	数据分区服务器中的压缩合并（Compaction）队列大小，当存储文件数量达到该数值时将启动压缩合并。	个	系统级
请求时延 10 毫秒次数	请求响应时间小于或等于 10 毫秒的请求次数	次	系统级/
请求时延 2000 毫秒次数	请求响应时间小于或等于 2000 毫秒的请求次数	次	系统级/

请求时延 2000 毫秒以上次数	请求响应时间大于 2000 毫秒的请求次数	次	系统级/
读 IO 次数	读 IO 次数	次	节点级
写 IO 次数	写 IO 次数	次	节点级
IO 次数	节点 IO 次数（读写 IO 次数总和）	次	节点级

4.2. 部分样例报警指标

4.2.1. Hive

4.2.1.1. 系统级

告警名称	告警描述	定位信息
空间使用比率超阈值告警	存储资源空间占用比超过门限	系统标识
	存储系统全部节点 CPU 平均使用率超过门限	系统标识
	存储系统全部节点内存平均使用率超过门限	系统标识
系统状态异常	存储系统不可用、只读告警	系统标识

4.2.1.2. 节点级

告警名称	告警描述	定位信息
节点状态异常	节点故障	节点 IP/主机名

空间使用比率超阈值告警	存储资源空间占用比超过门限	节点 IP/主机名
CPU 使用率超阈值告警	存储系统单节点 CPU 平均使用率超过门限	节点 IP/主机名
内存使用率超阈值告警	存储系统单节点内存平均使用率超过门限	节点 IP/主机名
进程异常	系统重要进程异常	节点 IP/主机名、进程标识
元数据故障	元数据写入异常或丢失	节点 IP/主机名、表标识、索引标识、元数据条目

4.2.2. HDFS

4.2.2.1. 系统级

告警名称	告警描述	定位信息
空间使用比率超阈值告警	存储资源空间占用比超过门限	系统标识
CPU 使用率超阈值告警	存储系统全部节点 CPU 平均使用率超过门限	系统标识
内存使用率超阈值告警	存储系统全部节点内存平均使用率超过门限	系统标识

系统状态异常	存储系统不可用、只读告警	系统标识
--------	--------------	------

4.2.2.2. 节点级

告警名称	告警描述	定位信息
节点状态异常	节点故障	节点 IP/主机名
空间使用比率超阈值告警	存储资源空间占用比超过门限	节点 IP/主机名
CPU 使用率超阈值告警	存储系统单节点 CPU 平均使用率超过门限	节点 IP/主机名
内存使用率超阈值告警	存储系统单节点内存平均使用率超过门限	节点 IP/主机名

4.2.3. MapReduce

4.2.3.1. 系统级

告警名称	告警描述	定位信息
CPU 使用率超	存储系统全部节点 CPU 平	系统标识

阈值告警	均使用率超过门限	
内存使用率 超阈值告警	存储系统全部节点内存平 均使用率超过门限	系统标识
系统状态异常	系统不可用	系统标识

4.2.3.2. 节点级

告警名称	告警描述	定位信息
节点状态异常	节点故障	节点 IP/主机名
空间使用比率超 阈值告警	存储资源空间占用 比超过门限	节点 IP/主机名
CPU 使用率超阈 值告警	存储系统单节点 CPU 平均使用率超 过门限	节点 IP/主机名
内存使用率超阈 值告警	存储系统单节点内 存平均使用率超过 门限	节点 IP/主机名
任务失败	Task 执行失败	节点 IP/主机名、用 户标识、作业标识、 任务标识

4.2.4. HBase

4.2.4.1. 系统级

告警名称	告警描述	定位信息
空间使用比率 超阈值告警	存储资源空间占用比超 过门限	系统标识
CPU 使用率超 阈值告警	存储系统全部节点 CPU 平 均使用率超过门限	系统标识
内存使用率超 阈值告警	存储系统全部节点内存 平均使用率超过门限	系统标识
系统状态异常	存储系统不可用、只读告 警	系统标识

4.2.4.2. 节点级

告警名称	告警描述	定位信息
节点状态异常	节点故障	节点 IP/主机名
	存储资源空间占用 比超过门限	节点 IP/主机名
CPU 使用率超 阈值告警	存储系统单节点 CPU 平均使用率超	节点 IP/主机名

	过门限	
内存使用率超 阈值告警	存储系统单节点内 存平均使用率超过 门限	节点 IP/主机名

Cloudera Manager 产品特性

Cloudera Manager 作为 Hadoop 大数据平台的管理工具，旨在帮助用户更容易的使用 Hadoop。其功能主要可以分为四大模块：1 管理功能 2 监控功能 3 诊断功能 4 集成功能。

1. 管理功能：

由于 Hadoop 是一个分布式的系统，由多个服务器组成一个集群，相对于传统的单节点运行的应用系统，由于节点数的增多。Hadoop 的管理更加复杂。CM 提供的管理功能如下：

1.1. 方便的集群安装

Cloudera Manager (以下简称 CM) 提供了强大的 Hadoop 集群部署能力，能够批量的自动化部署节点。安装一个 Hadoop 集群大体只需要 3 个步骤。

添 加 需 要 安 装 的 节 点 ：

cloudera manager

Welcome adminHelpLogout

Specify hosts for your CDH3 cluster installation.

There are currently 12 hosts being managed.

<input checked="" type="checkbox"/>	Expanded Query	Hostname (FQDN)	IP Address	Currently Managed	Result
<input checked="" type="checkbox"/>	name1.cloudera.local	name1.cloudera.local	192.168.1.10	Yes	✓ Host ready: 0 ms response time.
<input checked="" type="checkbox"/>	name2.cloudera.local	name2.cloudera.local	192.168.1.11	Yes	✓ Host ready: 1 ms response time.
<input checked="" type="checkbox"/>	server1.cloudera.local	server1.cloudera.local	192.168.1.101	Yes	✓ Host ready: 1 ms response time.
<input checked="" type="checkbox"/>	server10.cloudera.local	server10.cloudera.local	192.168.1.110	Yes	✓ Host ready: 0 ms response time.
<input checked="" type="checkbox"/>	server2.cloudera.local	server2.cloudera.local	192.168.1.102	Yes	✓ Host ready: 0 ms response time.
<input checked="" type="checkbox"/>	server3.cloudera.local	server3.cloudera.local	192.168.1.103	Yes	✓ Host ready: 0 ms response time.
<input checked="" type="checkbox"/>	server4.cloudera.local	server4.cloudera.local	192.168.1.104	Yes	✓ Host ready: 0 ms response time.
<input checked="" type="checkbox"/>	server5.cloudera.local	server5.cloudera.local	192.168.1.105	Yes	✓ Host ready: 0 ms response time.
<input checked="" type="checkbox"/>	server6.cloudera.local	server6.cloudera.local	192.168.1.106	Yes	✓ Host ready: 1 ms response time.
<input checked="" type="checkbox"/>	server7.cloudera.local	server7.cloudera.local	192.168.1.107	Yes	✓ Host ready: 0 ms response time.
<input checked="" type="checkbox"/>	server8.cloudera.local	server8.cloudera.local	192.168.1.108	Yes	✓ Host ready: 1 ms response time.
<input checked="" type="checkbox"/>	server9.cloudera.local	server9.cloudera.local	192.168.1.109	Yes	✓ Host ready: 0 ms response time.

Add new hosts to cluster

You can install CDH3 on new hosts and add them to your cluster at this point if you desire. Search for hostnames and/or IP addresses using the form below. Use [patterns](#) to enter alphanumeric ranges.

server1[1-9].cloudera.local, server20.cloudera.local

Find Hosts12 hosts scanned, 12 running SSH. Please select desired hosts and press Continue.

QuitBackContinue

安装需要的组件

cloudera manager

Welcome adminHelpLogout

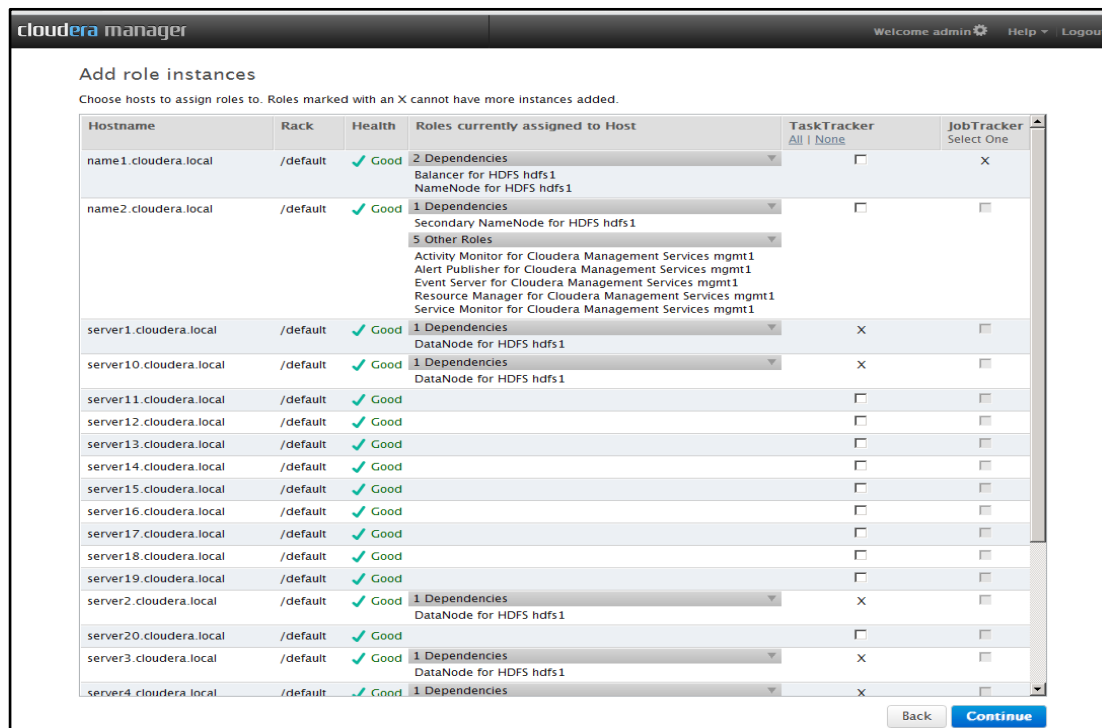
Installation in progress.

8 of 10 host(s) completed successfully.

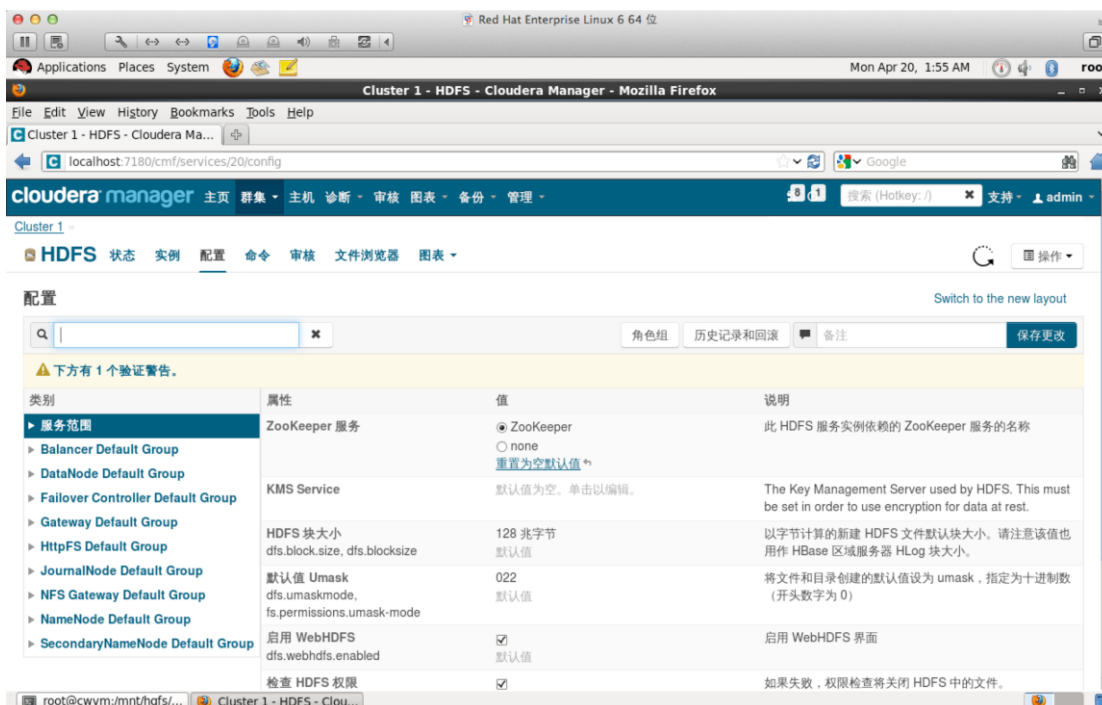
Hostname	IP Address	Progress	Status	
server11.cloudera.local	192.168.1.111	<div></div>	✓ Install completed successfully.	Details
server12.cloudera.local	192.168.1.112	<div></div>	✓ Install completed successfully.	Details
server13.cloudera.local	192.168.1.113	<div></div>	⚙ Installing JDK package...	Details
server14.cloudera.local	192.168.1.114	<div></div>	✓ Install completed successfully.	Details
server15.cloudera.local	192.168.1.115	<div></div>	✓ Install completed successfully.	Details
server16.cloudera.local	192.168.1.116	<div></div>	✓ Install completed successfully.	Details
server17.cloudera.local	192.168.1.117	<div></div>	⚙ Installing JDK package...	Details
server18.cloudera.local	192.168.1.118	<div></div>	✓ Install completed successfully.	Details
server19.cloudera.local	192.168.1.119	<div></div>	✓ Install completed successfully.	Details
server20.cloudera.local	192.168.1.120	<div></div>	✓ Install completed successfully.	Details

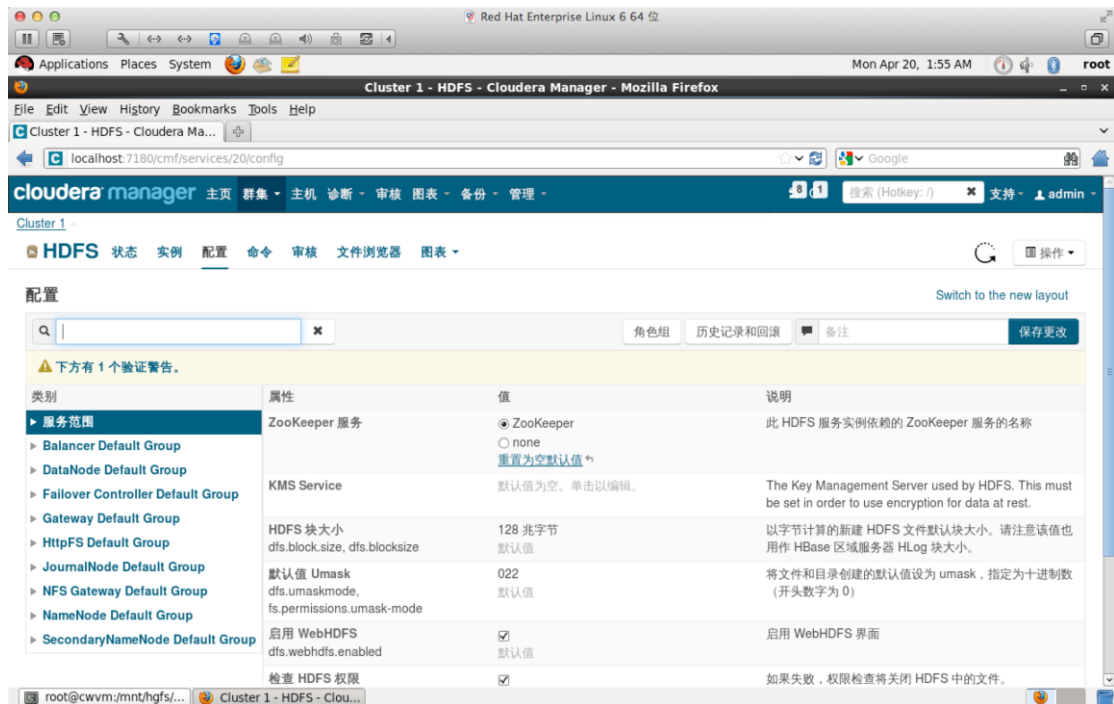
Abort Installation

分配角色



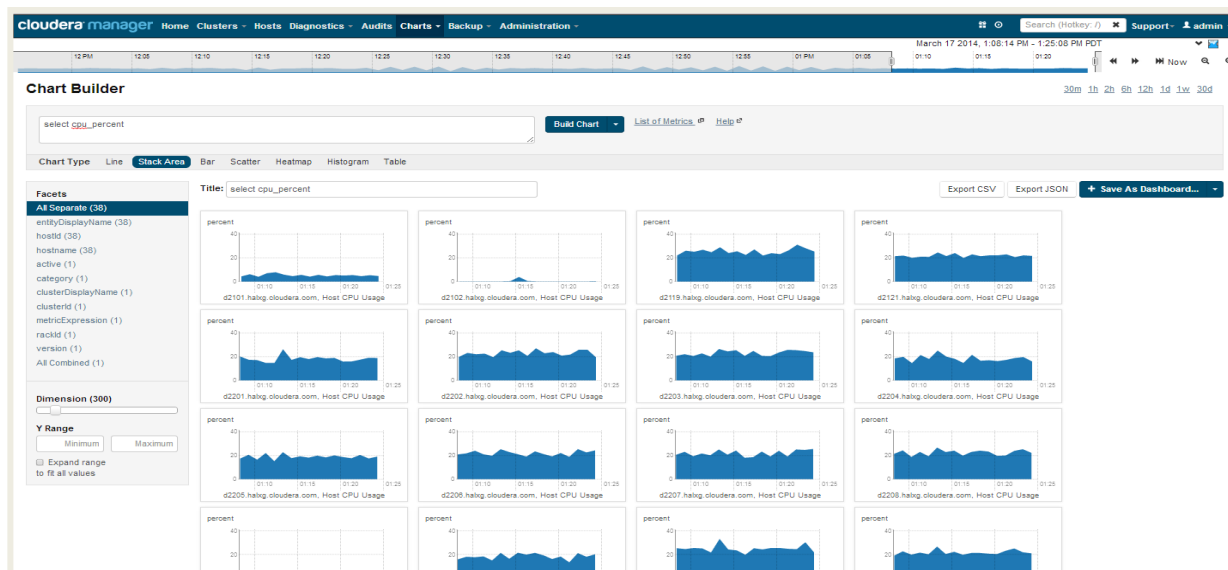
1.2. 配置的修改与回滚





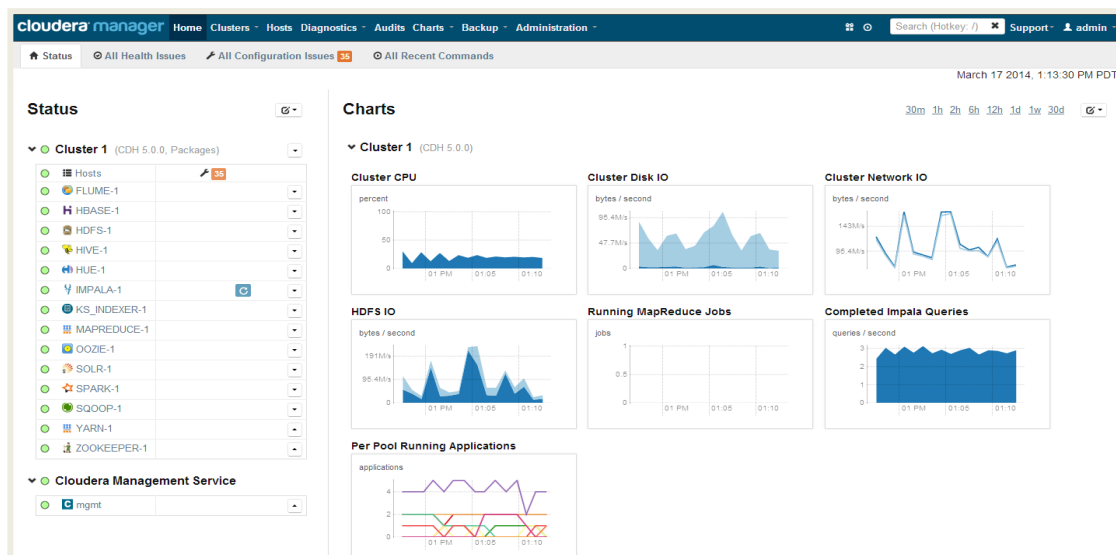
Hadoop 包含许多的组件，不同的组件都包含各种各样的配置，并且分布于不同的主机之上。CM 针对这种情况提供了界面化的参数配置功能，并且能够自动的部署到每个节点。除此以外，CM 还提供了配置的历史记录与回滚，用户可以查看参数的修改历史，可以方便的回滚到某一个具体的阶段。而且，CM 还提供了参数配置组的概念，配置组主要是应对在实际情况中，可能存在同一个集群，节点之间存在不同的硬件配置，或者相同配置承担不同作用的情况，这种情况下不同的节点可能需要进行不同的参数配置。

1.3. 自定义图表



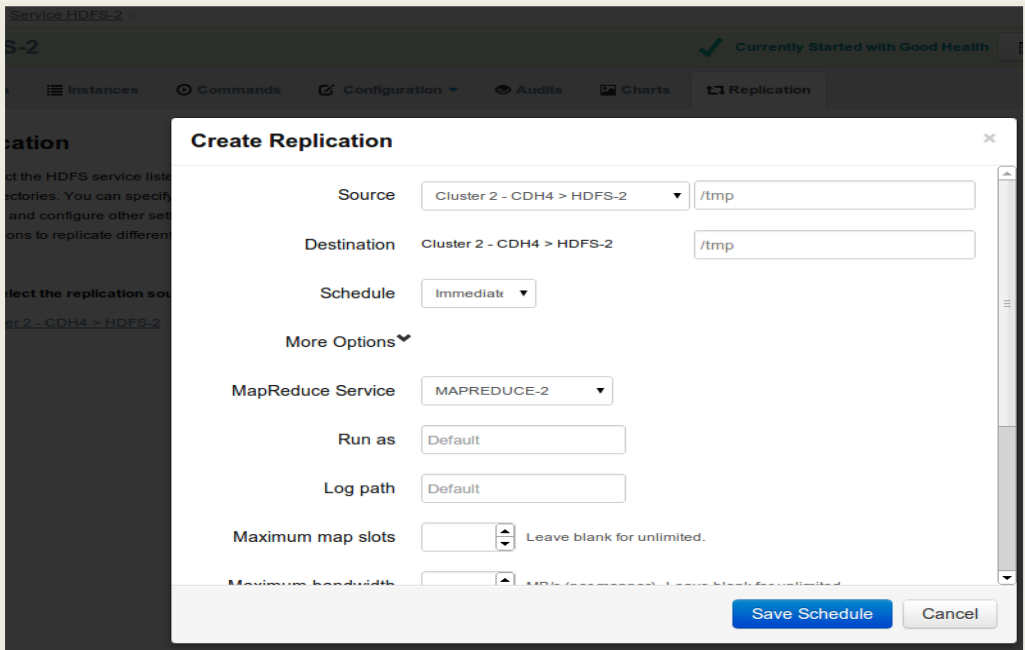
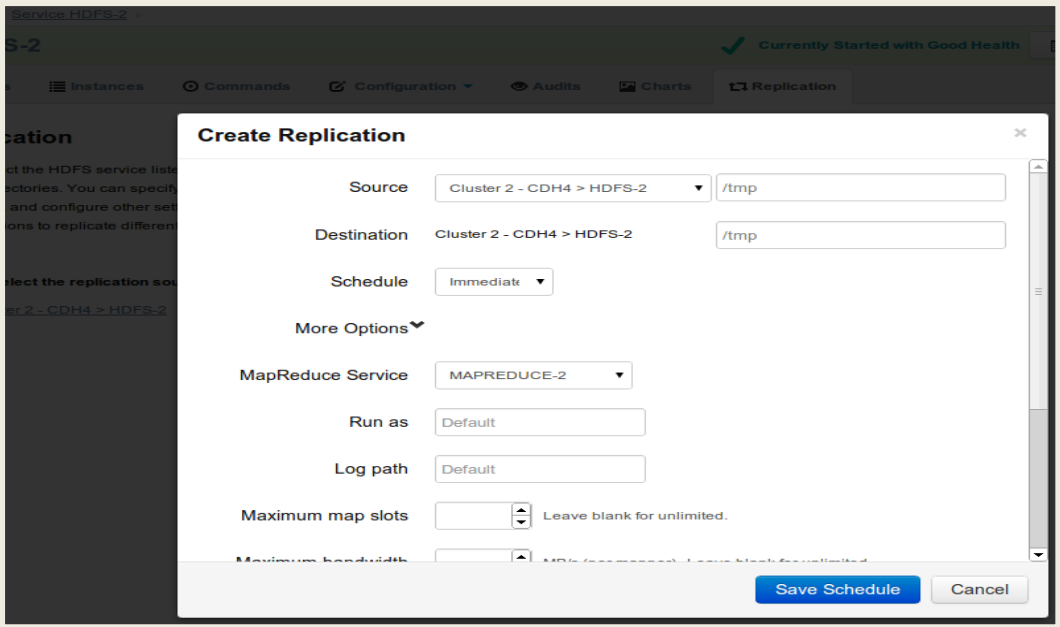
除了 CM 默认提供的性能图表以外，还允许用户自定义图表，使得用户能够查看自己希望的性能图表。CM 提供了一套类似的 SQL 的语言，用户只需要使用这个语言，简单的编写类似 SQL 的语句，就能够生成自己想要的图表。

1.4. 页面自定义



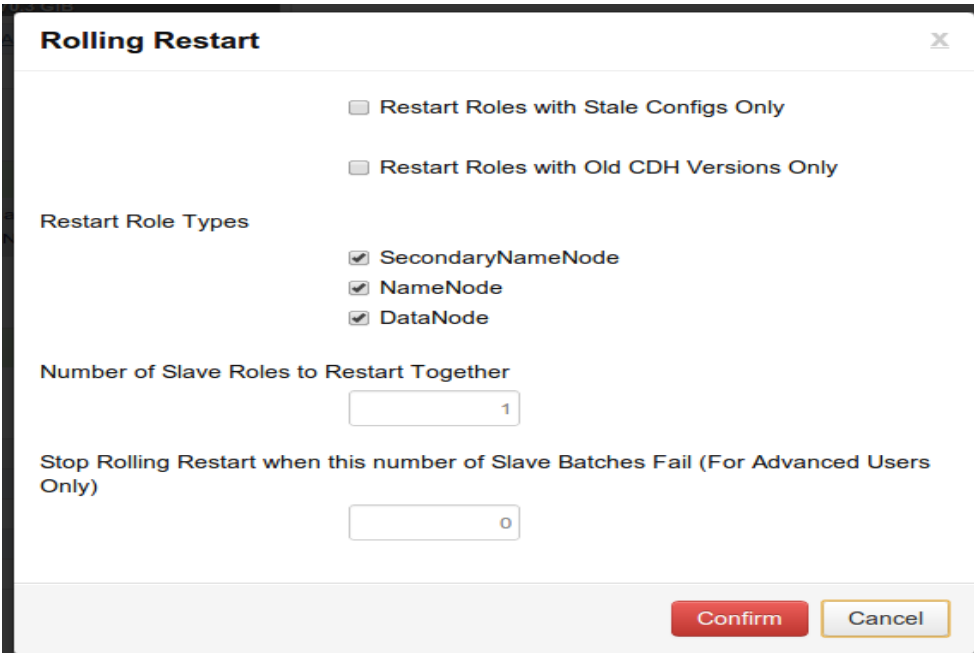
CM 的主页能够允许用户自定义主界面的风格，可以通过简单的拖拽就能够调整主界面上组件的位置，也能够自定义添加需要的组件。

1.5. 数据备份与恢复



数据的备份与恢复是企业级系统非常关键的功能,能够保障异常故障情况下,数据的完整性。Hadoop 能够实现集群间的数据备份与灾难恢复工作,但是本身的配置以及实现非常的复杂与繁琐,不易于使用。CM 提供了界面化的数据备份与灾难恢复。用户只需要在界面上进行简单的配置,就能够达到预期的效果。并且可以设置,数据备份的周期,带宽等参数,以保证在不影响业务的情况下,周期性的进行自动备份。

1.6. 滚动升级

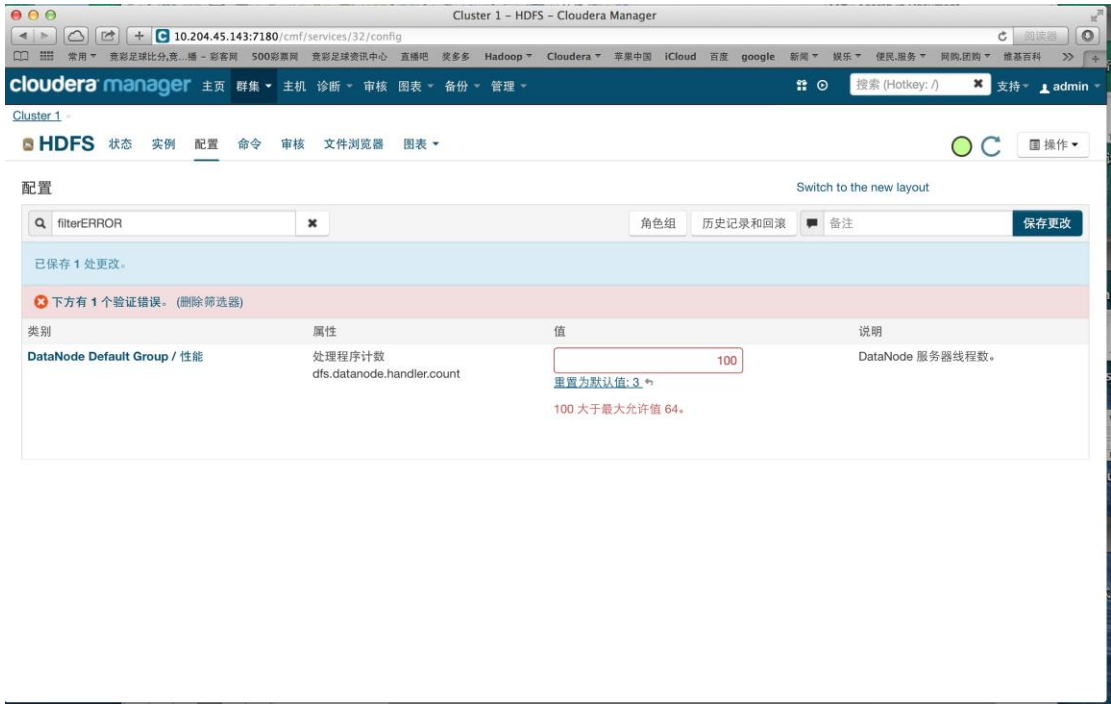


The image shows a 'Rolling Restart' configuration window. It contains several options and input fields. At the top, there are two unchecked checkboxes: 'Restart Roles with Stale Configs Only' and 'Restart Roles with Old CDH Versions Only'. Below these is a section titled 'Restart Role Types' with three checked checkboxes: 'SecondaryNameNode', 'NameNode', and 'DataNode'. Underneath is a field for 'Number of Slave Roles to Restart Together' with the value '1'. At the bottom, there is a field for 'Stop Rolling Restart when this number of Slave Batches Fail (For Advanced Users Only)' with the value '0'. At the very bottom right, there are 'Confirm' and 'Cancel' buttons.

Option	Value
Restart Roles with Stale Configs Only	<input type="checkbox"/>
Restart Roles with Old CDH Versions Only	<input type="checkbox"/>
Restart Role Types	
SecondaryNameNode	<input checked="" type="checkbox"/>
NameNode	<input checked="" type="checkbox"/>
DataNode	<input checked="" type="checkbox"/>
Number of Slave Roles to Restart Together	1
Stop Rolling Restart when this number of Slave Batches Fail (For Advanced Users Only)	0

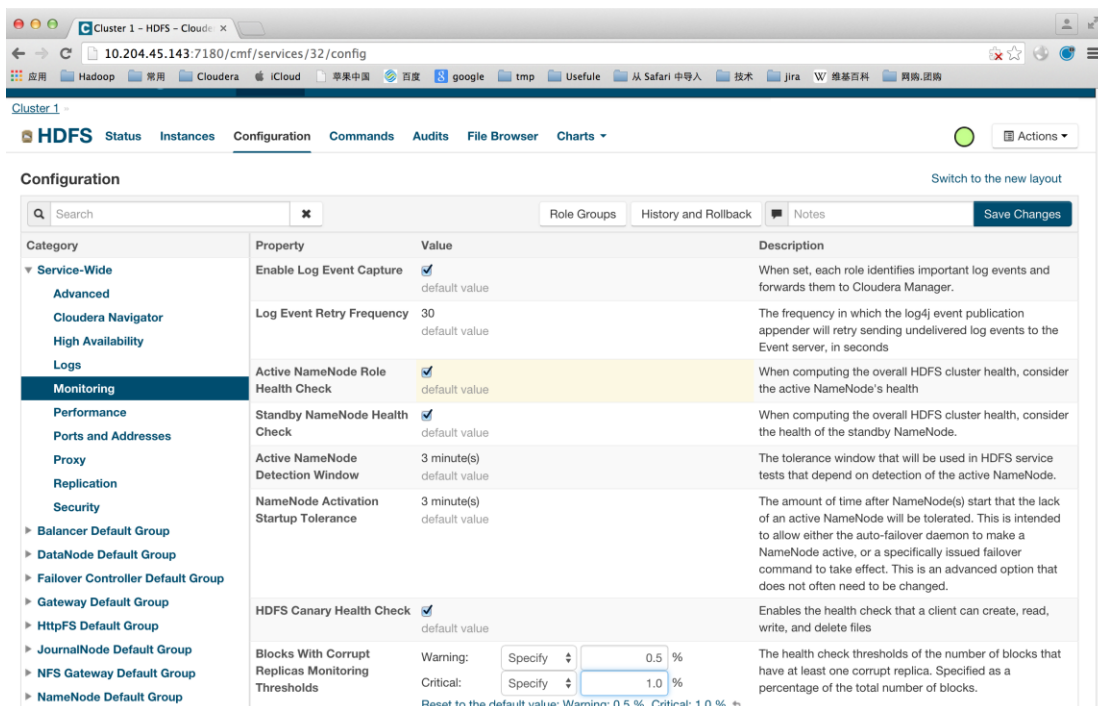
Hadoop 是一个正在快速发展的新技术,版本正在迅速的行进,对于部署了 Hadoop 的系统,版本升级是不可避免的。但是,通常业务系统都是在线的生产系统,需呀 24 小时不间断的提供对外服务。因此,CM 提供了滚动升级的功能,通过逐个节点升级的方式,保证集群的对外服务功能,并且最终完成整个集群的升级。

1.7. 参数验证



当用户在配置过程中，部分参数配置值有问题时，Cloudera Manager 会产生错误提示，以帮助用户更合理的修改参数。

1.8. 告警阈值可配置：

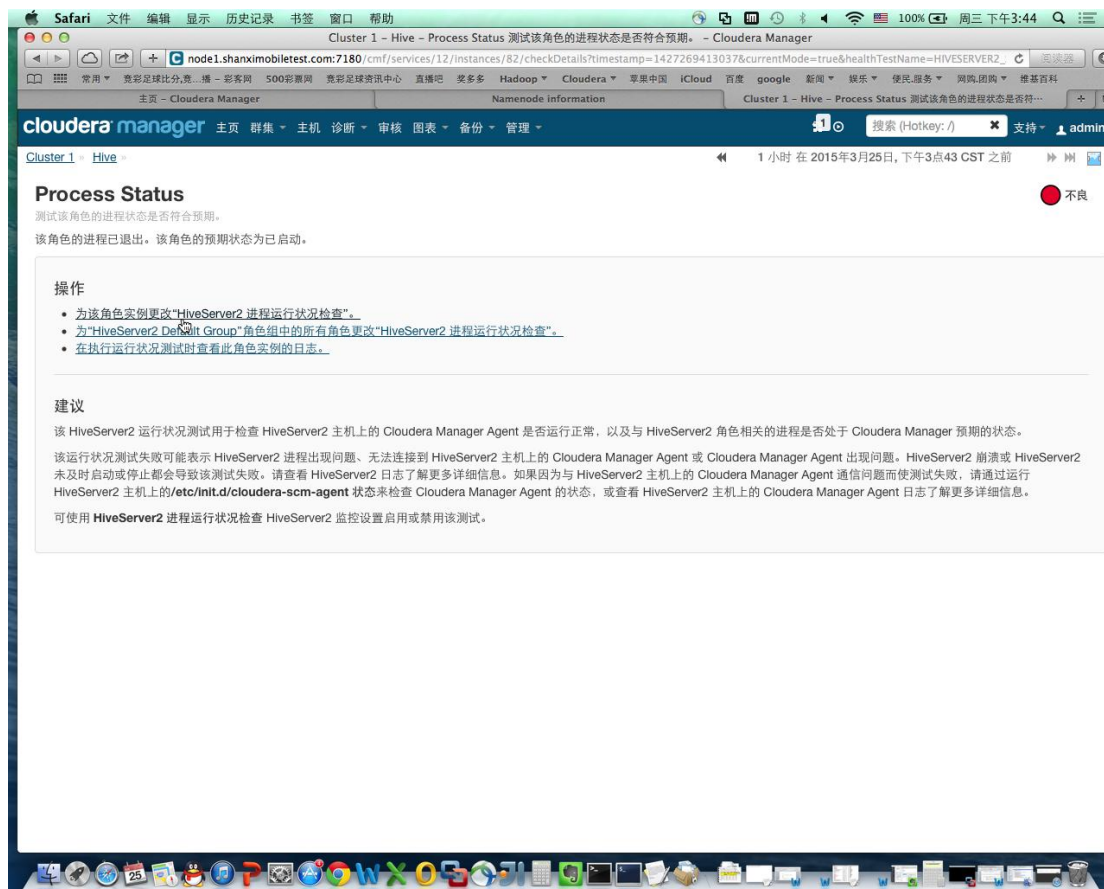


The screenshot shows the Cloudera Manager interface for configuring HDFS. The 'Configuration' tab is selected, and the 'Monitoring' category is chosen from the left sidebar. The 'HDFS Canary Health Check' is enabled, and the 'Blocks With Corrupt Replicas Monitoring Thresholds' are configured.

Category	Property	Value	Description
Service-Wide	Enable Log Event Capture	<input checked="" type="checkbox"/> default value	When set, each role identifies important log events and forwards them to Cloudera Manager.
	Log Event Retry Frequency	30 default value	The frequency in which the log4j event publication appender will retry sending undelivered log events to the Event server, in seconds
	Active NameNode Role Health Check	<input checked="" type="checkbox"/> default value	When computing the overall HDFS cluster health, consider the active NameNode's health
	Standby NameNode Health Check	<input checked="" type="checkbox"/> default value	When computing the overall HDFS cluster health, consider the health of the standby NameNode.
	Active NameNode Detection Window	3 minute(s) default value	The tolerance window that will be used in HDFS service tests that depend on detection of the active NameNode.
	NameNode Activation Startup Tolerance	3 minute(s) default value	The amount of time after NameNode(s) start that the lack of an active NameNode will be tolerated. This is intended to allow either the auto-failover daemon to make a NameNode active, or a specifically issued failover command to take effect. This is an advanced option that does not often need to be changed.
	HDFS Canary Health Check	<input checked="" type="checkbox"/> default value	Enables the health check that a client can create, read, write, and delete files
	Blocks With Corrupt Replicas Monitoring Thresholds	Warning: Specify 0.5 % Critical: Specify 1.0 %	The health check thresholds of the number of blocks that have at least one corrupt replica. Specified as a percentage of the total number of blocks.
	Reset to the default value: Warning: 0.5 %, Critical: 1.0 %		

Cloudera Manager 对于每个组件，都提供了监控的参数配置，例如可以设置达到多少阈值的时候，产生 warning 告警，达到多少阈值的时候产生 Critical 告警。

1.9. 操作提示

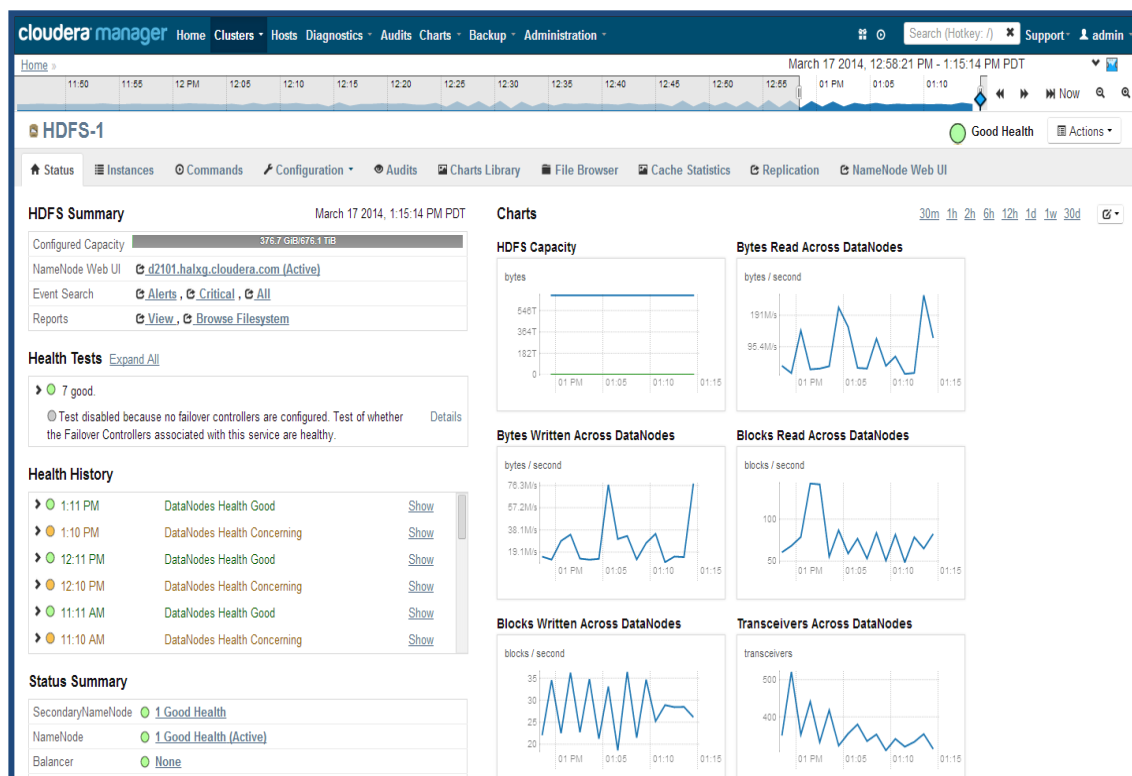


当集群中出现异常状态时, Cloudera Manager 会进行操作提示, 例如下图中, hive server 异常退出, CM 提示“该角色的进程已退出, 预期状态为已启动。”并且, CM 会进行操作提示, 以及建议, 包括针对这种情况, 可以采取的操作, 如何定位问题的原因等等。

2. 监控功能

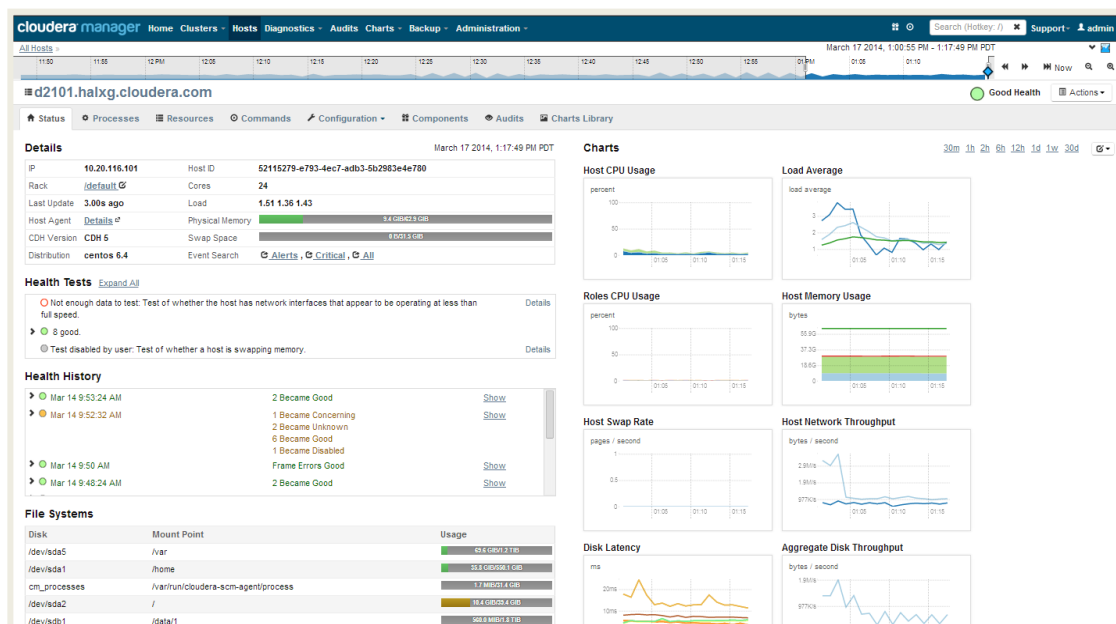
监控功能主要包括对集群各项资源, 以及服务的运行状态的监控, 一旦出现异常进行告警。

2.1. 实时监控集群以及服务的状态



如上图所示，CM 会实时的对集群的各项服务进行监控，一旦服务出现了异常，将会产生告警。同时，对集群的各项性能数据进行监控，例如 CPU，网络等，一方面能够监控集群的性能状态，一方面能够帮助性能调优。

2.2. 节点级别监控

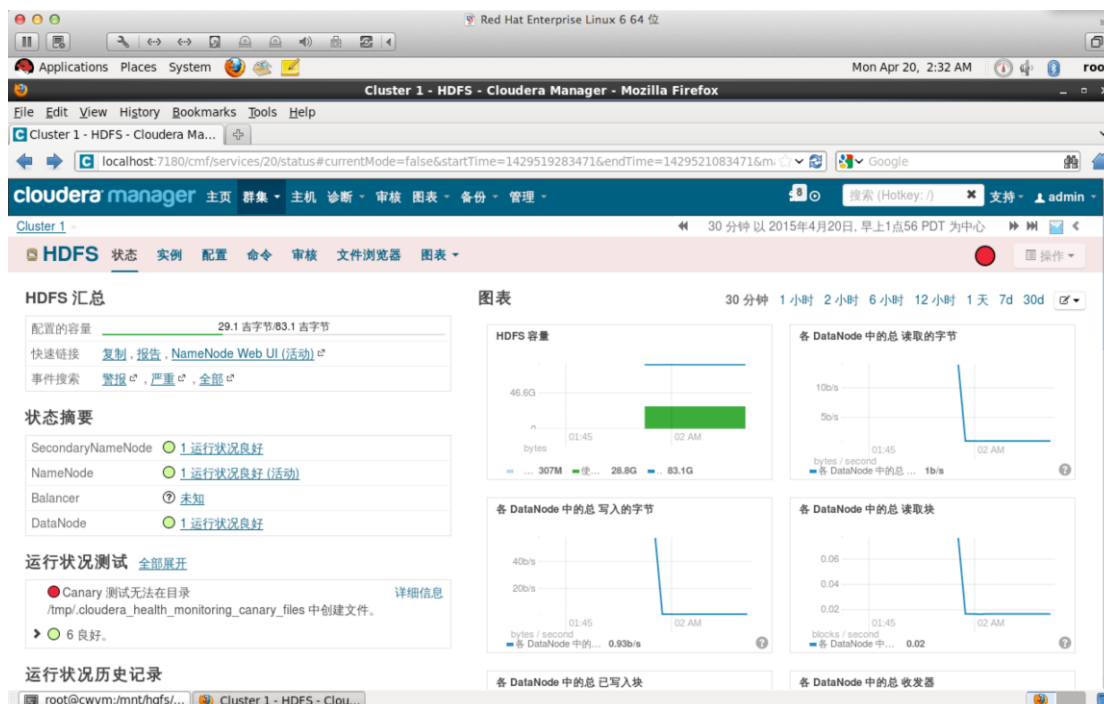


CM 能够支持节点级别的详细信息查看，包括该节点的硬件信息：网络，硬盘等。软件信息：运行的各项服务的状态，并对这些状态进行实时监控

3. 诊断功能

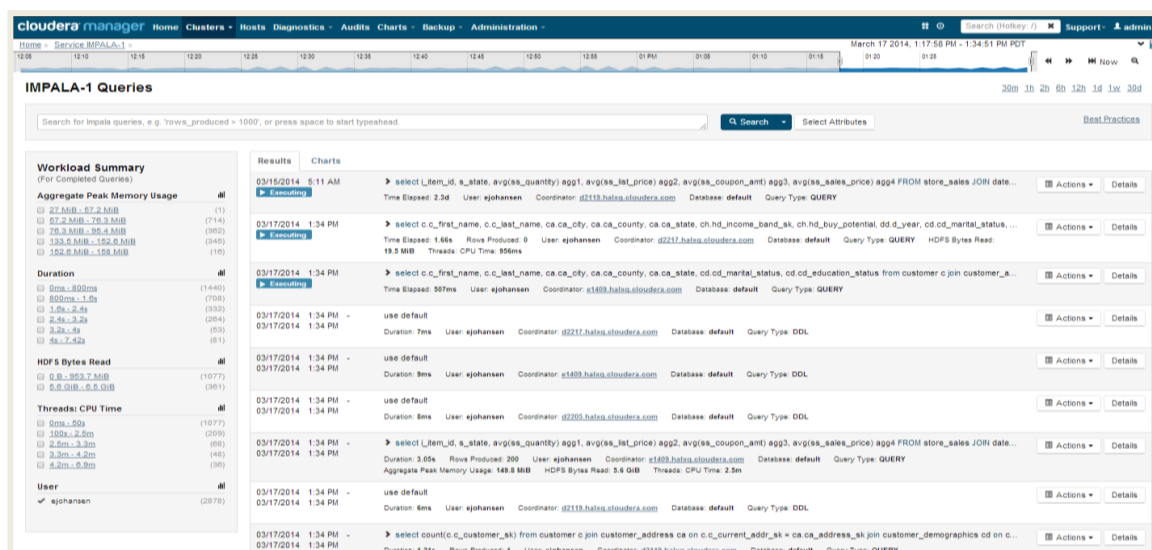
诊断功能主要是指能够帮助用户根据收集的数据查看潜在的或者存在的风险与问题。

3.1. 周期性的服务诊断



CM 会对集群中运行的服务进行周期性的运行状况测试。以检测这些服务的状态是否正常。如果有异常情况会进行告警，有利于更早的让客户感知集群服务存在的问题。

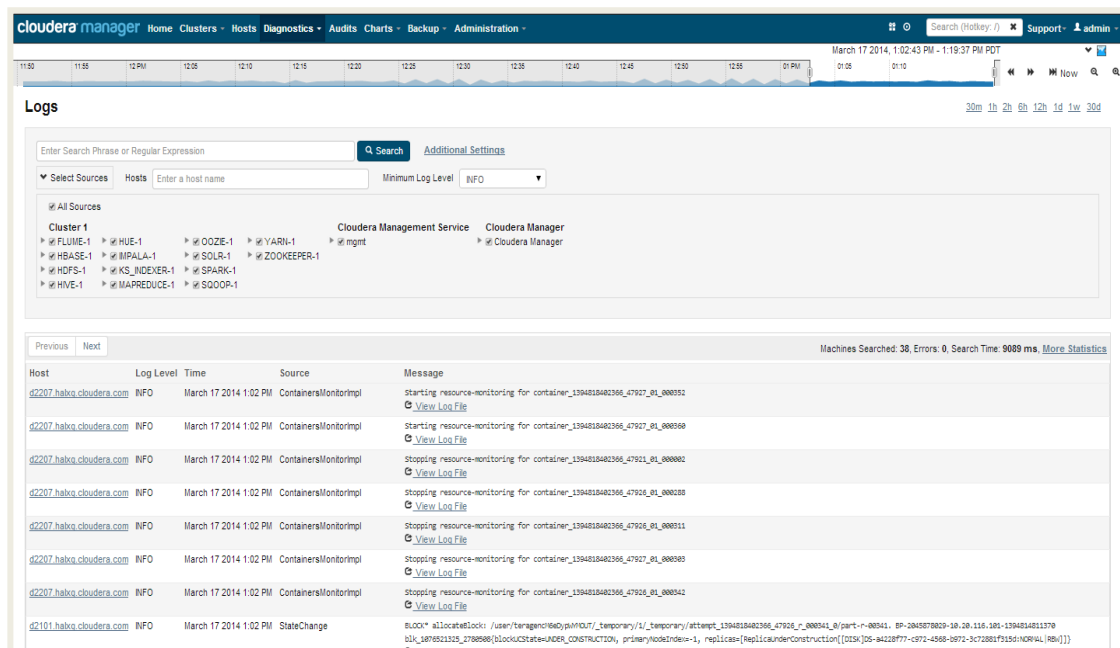
3.2. 集群诊断



CM 在 Hadoop 集群运行期间，会记录集群中发生的事件，例如集群服务

的启停，用户的操作，性能数据等等。CM 会将这些数据都记录到一个文件中。Cloudera 专门提供了一个集群，能够对这些数据进行分析，查找集群中的潜在风险，以及问题。用户可以选择定期的发送诊断数据给 Cloudera，Cloudera 会对这些数据进行分析并及时反馈集群中存在的问题。

3.3. 日志的收集及检索

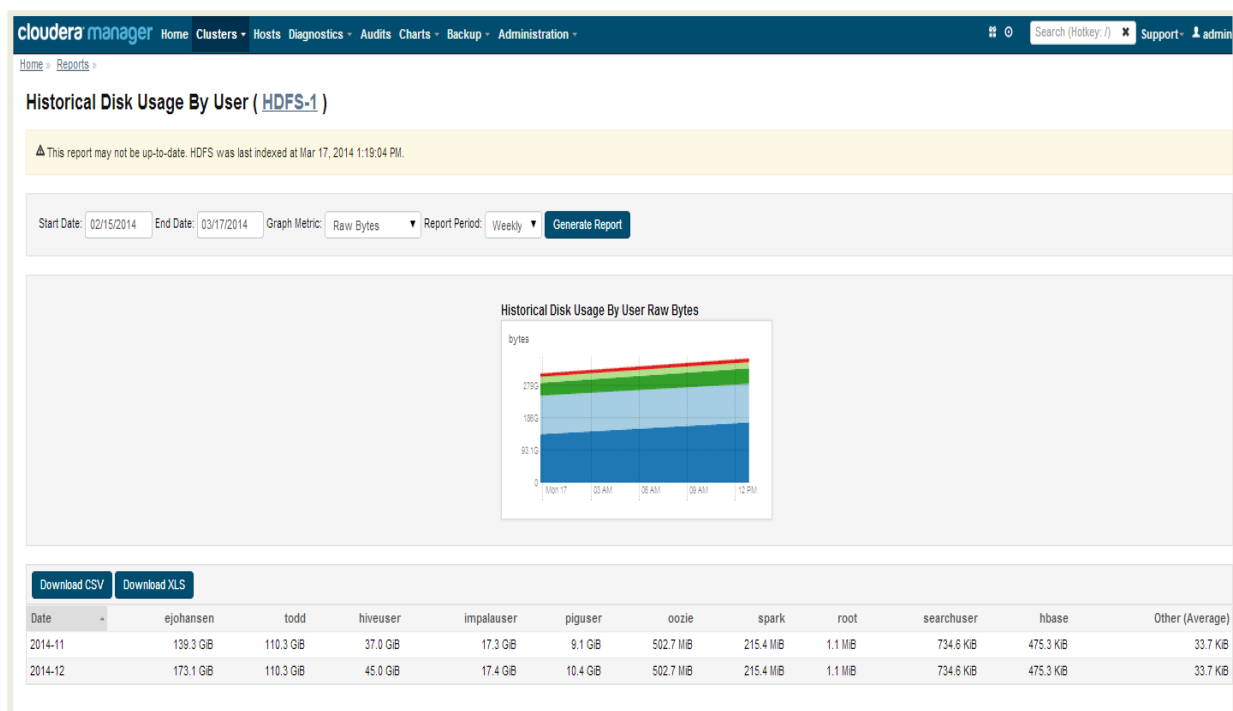


The screenshot displays the Cloudera Manager web interface, specifically the 'Logs' section. The top navigation bar includes links for Home, Clusters, Hosts, Diagnostics, Audits, Charts, Backup, and Administration. A search bar is visible in the top right corner. Below the navigation bar, there's a timeline view showing log activity from 11:50 to 12:50. The main area features a search interface with a text input field, a 'Search' button, and a 'Minimum Log Level' dropdown set to 'INFO'. Under 'Select Sources', there's a tree view showing the hierarchy of log sources, including 'Cluster 1' and various services like 'Flume-1', 'Hue-1', 'HBase-1', etc. Below the search interface, a table lists log entries with columns for Host, Log Level, Time, Source, and Message. The table shows several entries from 'd2207.haiku.cloudera.com' at 'March 17 2014 1:02 PM' with 'INFO' level, all from the 'ContainersMonitorImpl' source. The messages indicate 'Starting resource-monitoring' and 'Stopping resource-monitoring' for various containers. The last entry is a 'StateChange' at the same time, with a detailed message about allocating locks.

Host	Log Level	Time	Source	Message
d2207.haiku.cloudera.com	INFO	March 17 2014 1:02 PM	ContainersMonitorImpl	Starting resource-monitoring for container_13948318482366_47927_01_000552 View Log File
d2207.haiku.cloudera.com	INFO	March 17 2014 1:02 PM	ContainersMonitorImpl	Starting resource-monitoring for container_13948318482366_47927_01_000568 View Log File
d2207.haiku.cloudera.com	INFO	March 17 2014 1:02 PM	ContainersMonitorImpl	Stopping resource-monitoring for container_13948318482366_47927_01_000582 View Log File
d2207.haiku.cloudera.com	INFO	March 17 2014 1:02 PM	ContainersMonitorImpl	Stopping resource-monitoring for container_13948318482366_47926_01_000588 View Log File
d2207.haiku.cloudera.com	INFO	March 17 2014 1:02 PM	ContainersMonitorImpl	Stopping resource-monitoring for container_13948318482366_47926_01_000511 View Log File
d2207.haiku.cloudera.com	INFO	March 17 2014 1:02 PM	ContainersMonitorImpl	Stopping resource-monitoring for container_13948318482366_47927_01_000583 View Log File
d2207.haiku.cloudera.com	INFO	March 17 2014 1:02 PM	ContainersMonitorImpl	Stopping resource-monitoring for container_13948318482366_47926_01_000542 View Log File
d2101.haiku.cloudera.com	INFO	March 17 2014 1:02 PM	StateChange	BLUOX* allocateLock: /user/teragenc1660/pinOUT/_temporary/_attempt_13948318482366_47926_r_000541_0/part-r-005411_00-20116-101-13948318481370 [U_4878521515_2788988](blockState=NODE_CONSTRUCTING, primaryNodeInDe=1, replicas=[replicandConstruction[0054105-4628777-c972-4598-8972-3c72881f3156:NDWAL/0b/]])

Hadoop 集群的每个组件在运行过程中都会存储日志，并且散落在每个节点的不同目录，对于一个大规模的集群，日志的查看是非常麻烦的，需要一个个节点去进行查看。CM 针对这种情况，提供了日志的收集功能，能够通过统一的界面，查看集群中每台机器，各项服务的日志，并且能够根据日志级别等不同的条件进行检索。

3.4. 系统性能使用报告



Hadoop 集群的性能调试尤为重要，因此能够方便精确的监控集群性能尤为重要。CM 能够生产系统性能使用报告，包括集群的 CPU 使用率，单节点的 CPU 使用率，单个进程的 CPU 使用率等等各项性能数据。