CDH版本安装部署

1、Apache Hadoop 不足之处

　　• 版本管理混乱  
　　• 部署过程繁琐、升级过程复杂  
　　• 兼容性差  
　　• 安全性低

2、Hadoop 发行版

• Apache Hadoop  
　　• Cloudera’s Distribution Including Apache Hadoop（CDH）  
　　• Hortonworks Data Platform (HDP)  
　　• MapR  
　　• EMR  
　　• …

1.Apache社区版

优点：  
完全开源免费 社区活跃 文档、资料详实  
  
缺点：  
版本管理比较混乱，各种版本层出不穷，很难选择，选择生态组件时需要大量考虑兼容性问题、版本匹配问题、组件冲突问题、编译问题等。  
集群的部署安装配置复杂，需要编写大量配置文件，分发到每台节点，容易出错，效率低。  
集群运维复杂，需要安装第三方软件辅助。

2.第三方发行版（CDH/HDP/MapR）

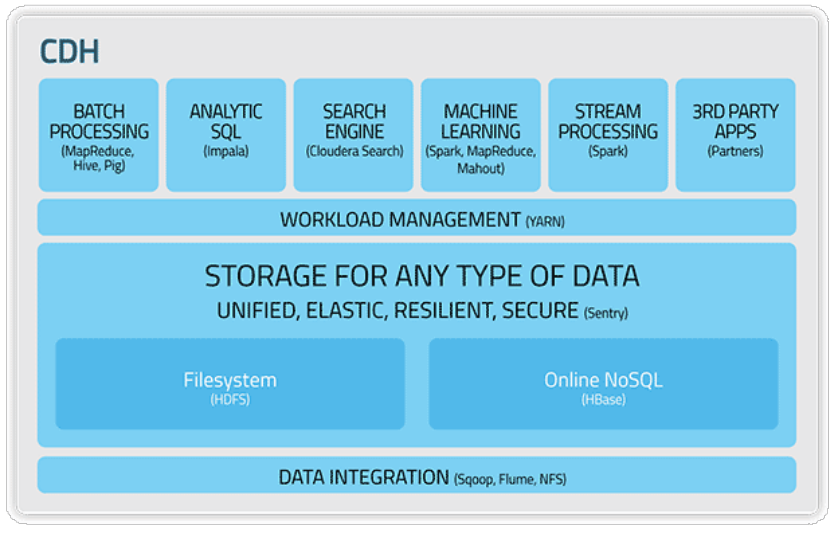
优点：  
基于Apache协议，100%开源。  
版本管理清晰  
提供了部署、安装、配置工具，大大提高了集群部署的效率  
运维简单。提供了管理、监控、诊断、配置修改的工具，管理配置方便，定位问题快速、准确，使运维工作简单，有效。  
  
缺点：  
涉及到厂商锁定的问题。

3、CDH能解决哪些问题

　　• 1000台服务器的集群，最少要花费多长时间来搭建好Hadoop集群，包括Hive、Hbase、Flume、Kafka、Spark等等  
　　• 只给你一天时间，完成以上工作？  
　　• 对于以上集群进行hadoop版本升级，你会选择什么升级方案，最少要花费多长时间？  
　　• 新版本的Hadoop，与Hive、Hbase、Flume、Kafka、Spark等等兼容？

4、CDH简介

　　• Cloudera's Distribution, including Apache Hadoop  
　　• 是Hadoop众多分支中的一种，由Cloudera维护，基于稳定版本的Apache Hadoop构建  
　　• 提供了Hadoop的核心  
　　　　– 可扩展存储  
　　　　– 分布式计算  
　　• 基于Web的用户界面



5、CDH的优点

　　• 版本划分清晰  
　　• 版本更新速度快  
　　• 支持Kerberos安全认证  
　　• 文档清晰  
　　• 支持多种安装方式（Cloudera Manager方式）

6、CDH安装方式

　　• Cloudera Manager  
　　• Yum  
　　• Rpm  
　　• Tarball

7、CDH下载地址

　　• CDH5.4  
　　　　<http://archive.cloudera.com/cdh5/cdh/5/>  
　　•Cloudera Manager5.4.3：  
　　　　 http://www.cloudera.com/downloads/manager/5-4-3.html

建议：首先在 <http://archive.cloudera.com/cdh5/cdh/5> 版本库下找到自己想要得版本(一般cdh版本要一致)，后加上 .tar.gz 下载即可

1. cdh比原生的Apache发行版本包含了更多的补丁，用于增强稳定性，改善功能，有时候还增加功能特性  
   2、cdh版本是由cloudera公司开源的，可以使用cm平台进行管理，比原生的Apache版本安装、维护更加省力  
   3、但是对技术人员的要求更高，必须对原生apache版本的各个组件理解清晰  
   4、在cm管理平台中，cdh的parcel包不包含某些组件，需要自己下载对应的parcel包，比如说kafka  
   5、对hdfs部署过程中，对磁盘进行lvm卷轴或者是磁盘目录统一，对于多台机器，否则之后维护成本高

Cloudera Manager 5（以下简称CM）默认采用在线安装的方式，给不能联互联网或者网络不畅的用户带来了不便，由于网络问题，屡装屡败的挫折感，让初次学习CDH用户真心体验到万事开头难。本文讲解了在CentOS 6环境下搭建本地Yum软件源，局域网内的用户在完全不连互联网的情况下，完成CM的安装。让安装CM的用户体验到一路next，然后finish的顺畅感，速度是用户的第一体验。