

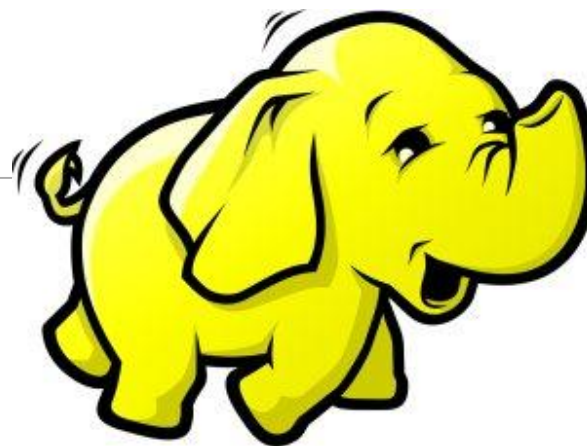
---

# 戴尔大数据解决方案分享



Terry Li

企业解决方案拓展经理



# 戴尔大数据解决方案简介



# 戴尔大数据解决方案的方法论……

简化客户体验

缩短投入生产  
所需的时间

优化解决方案性  
能

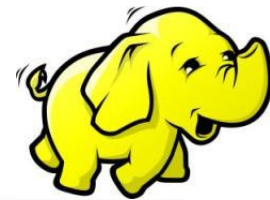
提供最佳的投  
资回报

我们如何实现？

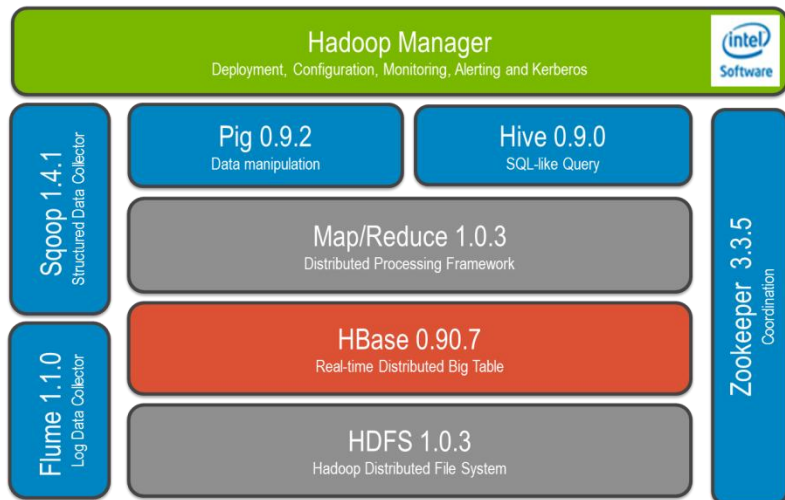
- 与合作伙伴协作
- 结合硬件和服务
- 参考架构和规模确定
- 更加深入的售前咨询
- 整合的售后技术支持



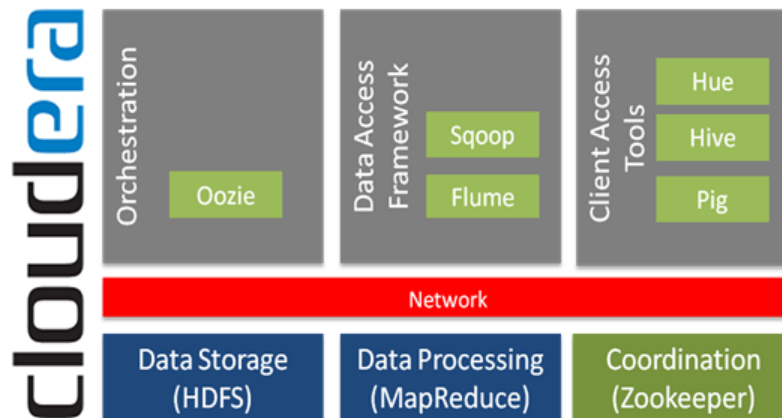
# 戴尔基于Hadoop的大数据解决方案



## Intel Hadoop



## Cloudera Hadoop



或

价值、性能和灵活性



Data Center Solutions



# 戴尔基于Hadoop的大数据解决方案

## Intel Hadoop

- 基于Hadoop底层的大量优化算法，使应用效率更高、计算存储分布更均衡
- 企业级发行版保证长期稳定运行
- Intel Hadoop Manager提供独有的基于浏览器的集群安装和管理界面
- 提供跨数据中心的HBase数据库虚拟大表功能

或

## Cloudera Hadoop

- 在商用和非商用环境中，Apache Hadoop排名第一的发行版本。
- Cloudera的管理工具让管理简单和直接。
- Cloudera雇用了90%以上的CDH的项目初始者和提交者
- 对稳定性和互操作性进行了测试

## 价值、性能和灵活性



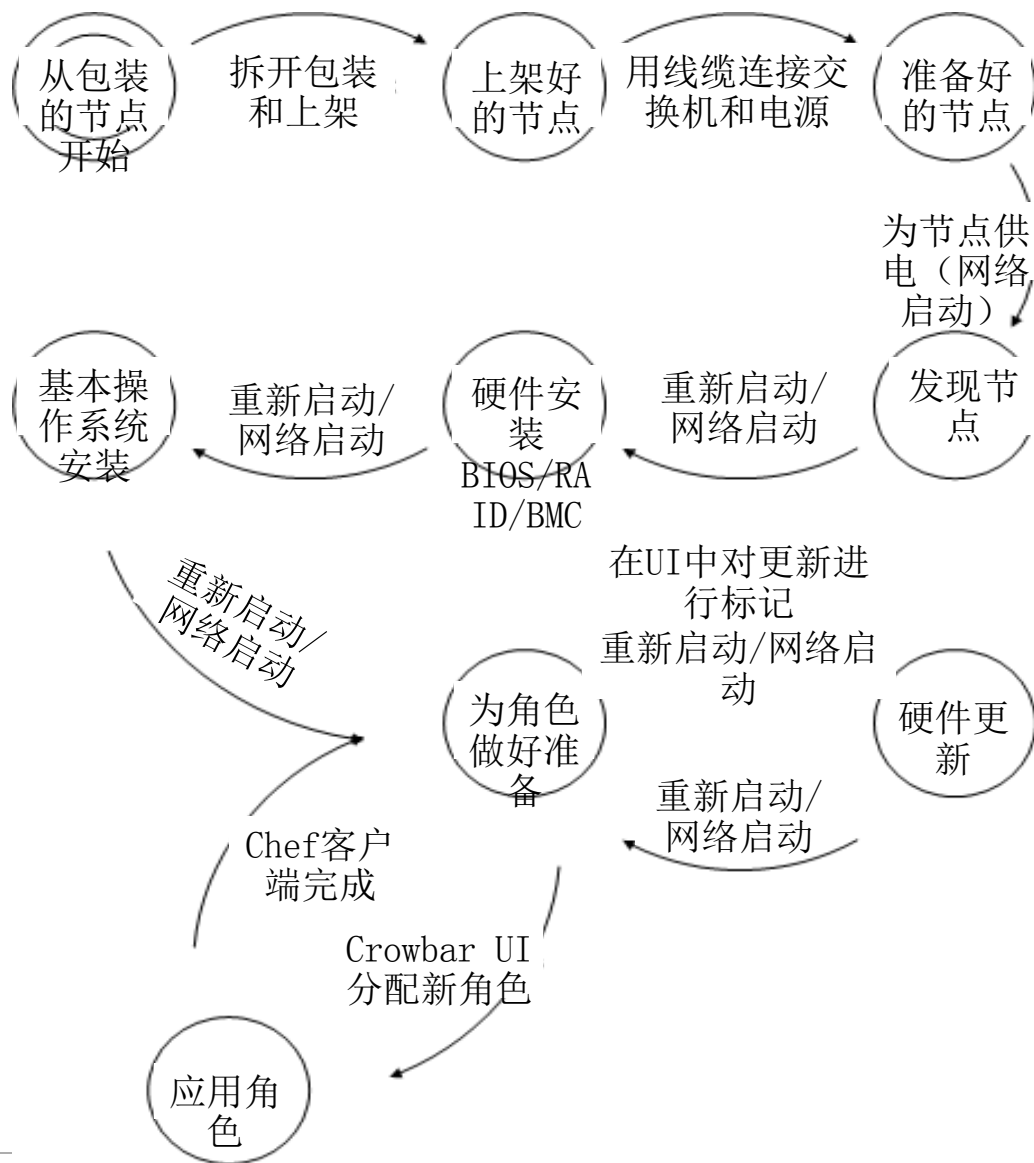
- 验证并合格的参考架构，强大的计算和存储、灵活的网络连接
- Crowbar工具快速安装、部署、监控、运维
- 最大限度的优化 – 能源、性能、密度、总体拥有成本（TCO）



# 戴尔 Crowbar - Hadoop集群部署的自动化

## 从“裸机”开始

- 系统BIOS更新和配置
- RAID/SAS配置
- 操作系统部署
- 部署Hadoop软件
- Hadoop软件配置
- 监控和报警的集成
- Hadoop持续运维



# 典型的Hadoop用户痛点 – 戴尔解决方案的价值

## 缺少集成的配置管理

- 戴尔的Crowbar软件，让您Hadoop环境中的配置变化被快速地部署和保持一致性。

## 默认的Hadoop安装，较差的资源分配

- 戴尔的参考架构提供配置参数，确保在部署中最大限度地利用戴尔硬件。

## 缺少网络架构经验

- 戴尔的参考架构包括一个推荐的、经过测试、经过验证的网络配置。

## 缺少监测、报警和诊断

- 戴尔Crowbar部署Nagios、Ganglia，结合Cloudera Hadoop企业版或Intel Hadoop管理工具，从而提供完整的解决方案，对您的Hadoop集群进行监测、报警、诊断和故障排除。

## 单点故障

- 戴尔推荐的参考架构包括一些配置要求，可以解决常见的与Hadoop Name Node（名称节点）相关的单点故障。

# 戴尔大数据解决方案包括

## Hadoop 发行版

- 英特尔 Hadoop 发行版
- Cloudera Hadoop 发行版

## 戴尔Crowbar工具

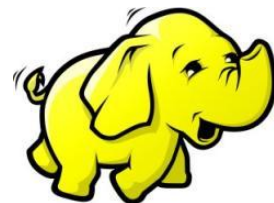
- Chef、Ganglia、Nagios、IPMI

## 戴尔PE-C6220、PE-C8000及PE-R720服务器

## 戴尔PC-6248、Force10 S60千兆以太网交换机

## 解决方案通过以下方式提供:

- 参考架构
- 部署向导
- 戴尔部署服务

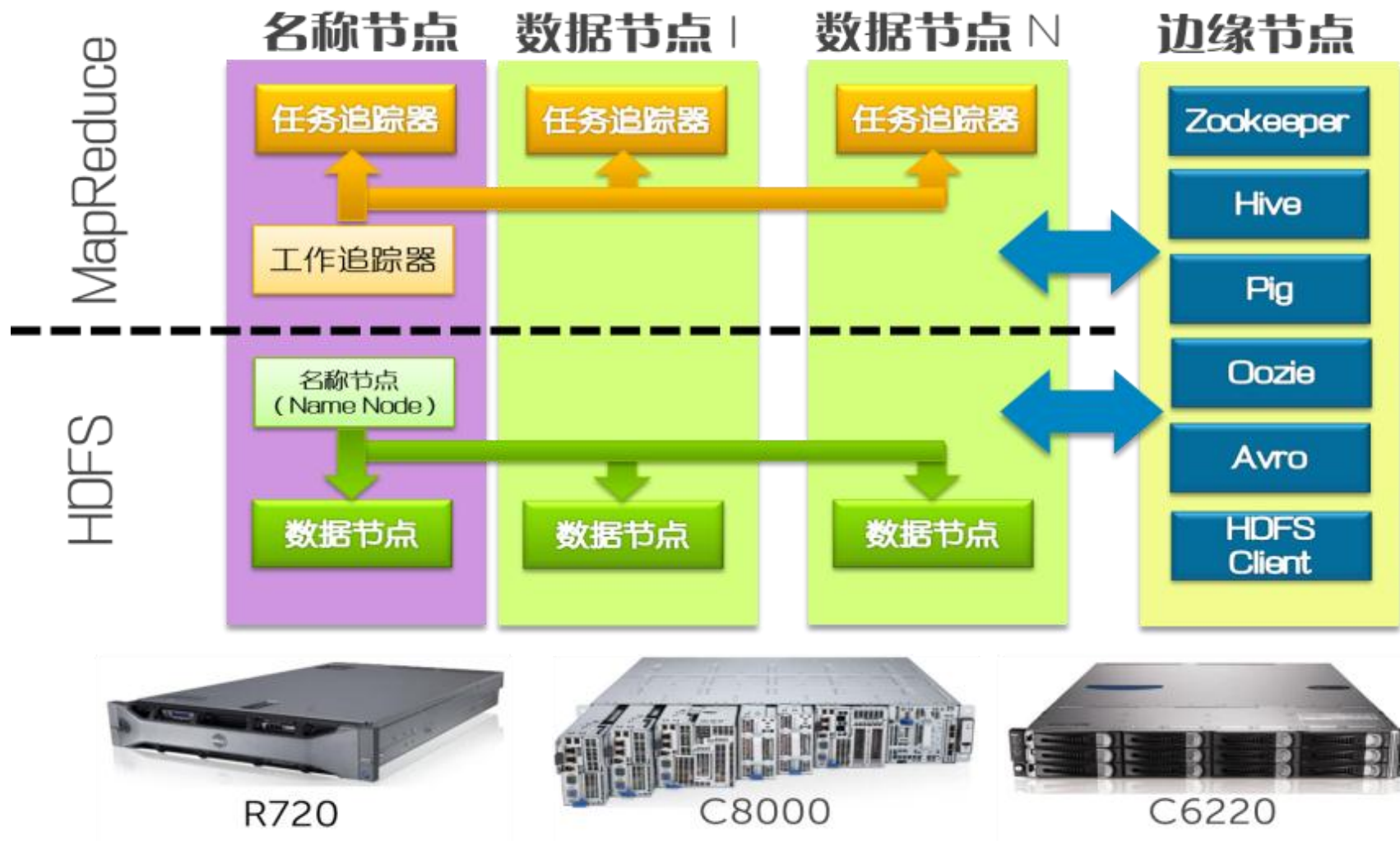




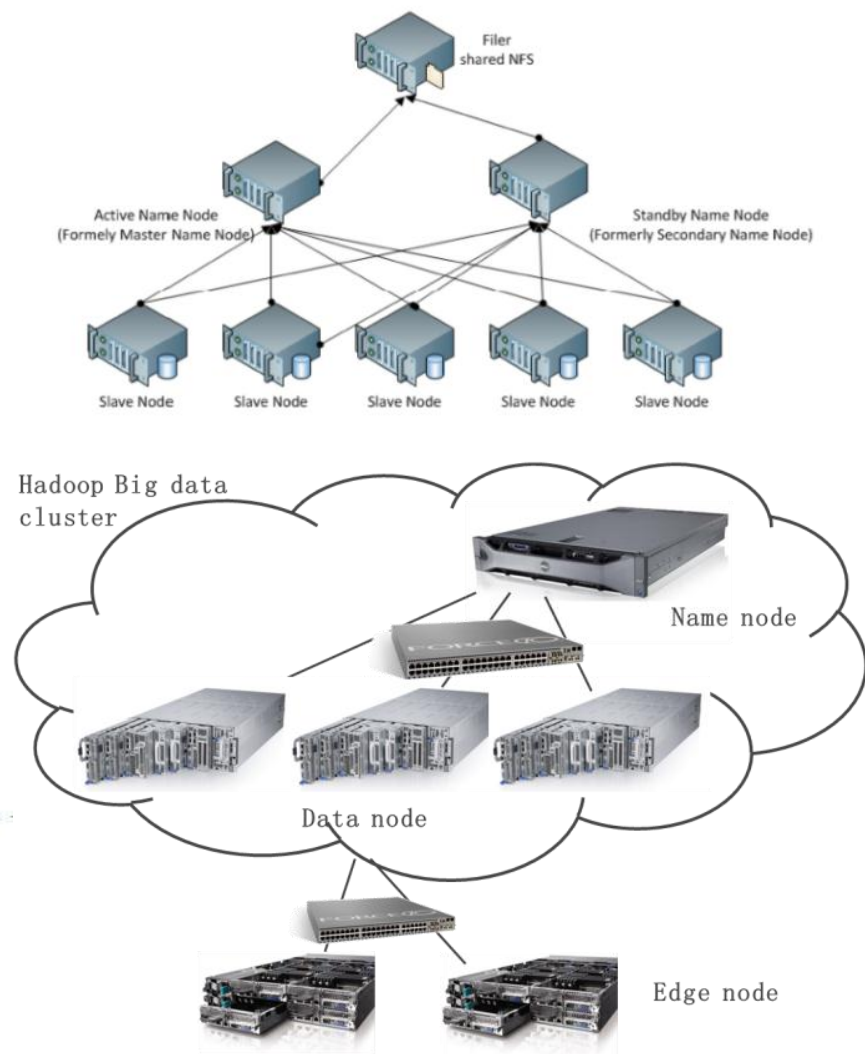
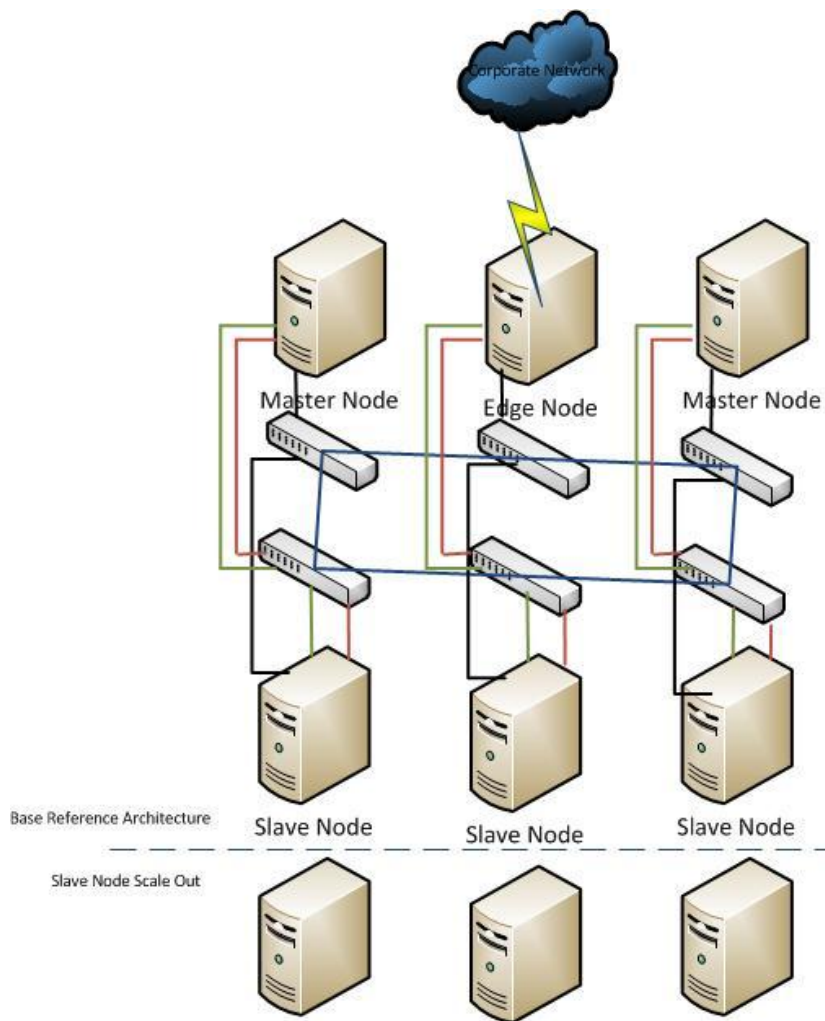
# 戴尔大数据解决方案参考架构



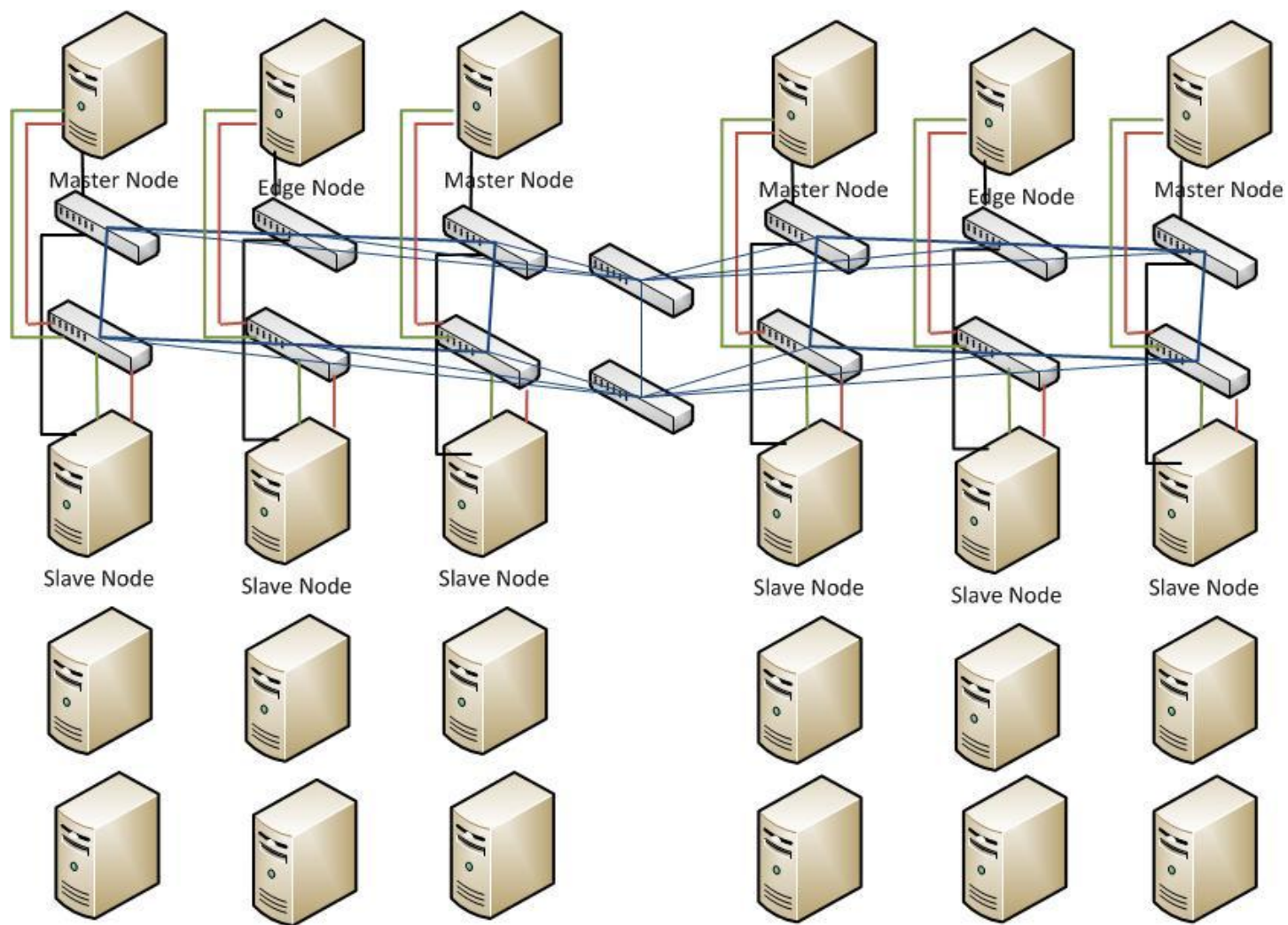
# 戴尔 Hadoop 部署参考架构



# 戴尔 Hadoop 部署参考架构



# 戴尔 Hadoop 部署参考架构 - 扩展



# 为何服务器非常重要

更快的数据分析

更轻松地扩展您的数据

提高灵活性

降低运行支出

对资本支出的巨大影响





# 为Hadoop优化的服务器平台 - PEC8000

通过PowerEdge C平台，从数据存储、报告和分析系统中获得价值，提供规模、速度、丰富性和易用性

- **PowerEdge C8000 - 12代产品**
- **高密度计算配置：**
  - C8000 ( 4U机箱 ) + 8 台服务器节点C8220 + 2 个双冗余电源节点。
  - 2颗 Intel E5-26XX CPU , 16根内存槽位 , 2块 2.5"硬盘。
  - 特点：与一般2U单节点双路机架服务器相比，计算密度为4倍，节点功耗更低。
- **高密度存储配置：**
  - C8000 ( 4U机箱 ) + 1 台服务器节点C8220 + 1 个内置电源节点 + 4 个存储节点C8000XD。
  - 每个双宽槽位存储节点含12块3.5"硬盘或24块 2.5"硬盘，4个存储节点则共有48块3.5"硬盘或 96块2.5"硬盘。
  - 特点：与一般2U12块3.5"硬盘或24块2.5"硬盘存储服务器相比，存储密度为2倍，每TB功耗更低。



C8000高性能计算和存储服务器



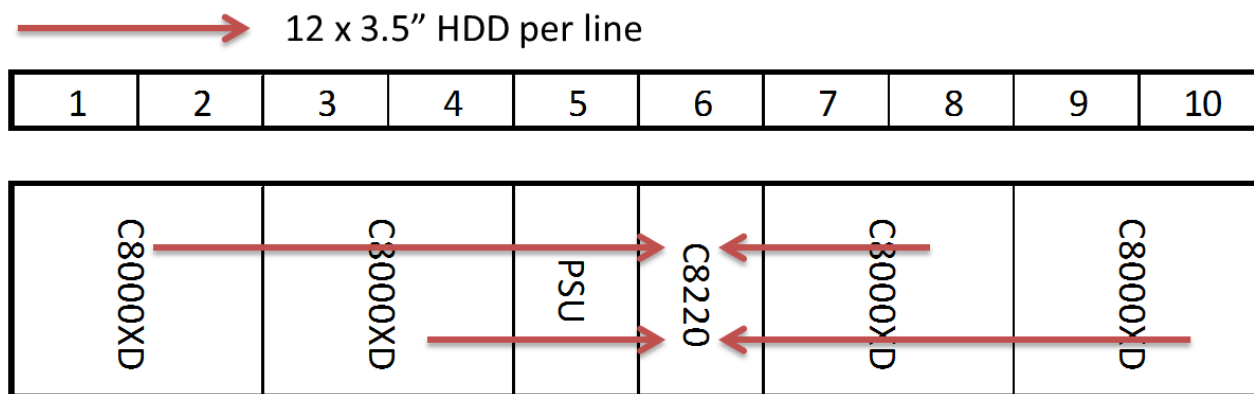
# 数据节点，高性价比存储型配置 一

最优化的每TB成本，每数据节点48块3.5”盘

HDFS存储容量要求非常大，计算要求适中的场景

C8000 4U1Node48HDD:

- 1 x C8000 Chassis with 1 x PSU (1400W \* 2)
- 1 x C8220 : E5-2620 \* 2 / 64G / 500G SATA 2.5” \* 2 / LSI9202
- 4 x C8000XD : 3T 3.5” SATA \* 12 (All 48 HDD connect to C8220 node)



# 运营成本对比 （不含服务器成本）

- 以5 PB Hadoop, 1667 个3TB  
硬盘测算

- 35 \* 4U48HDD 或 139 \*  
2U12HDD

- 3年总节省成本为:

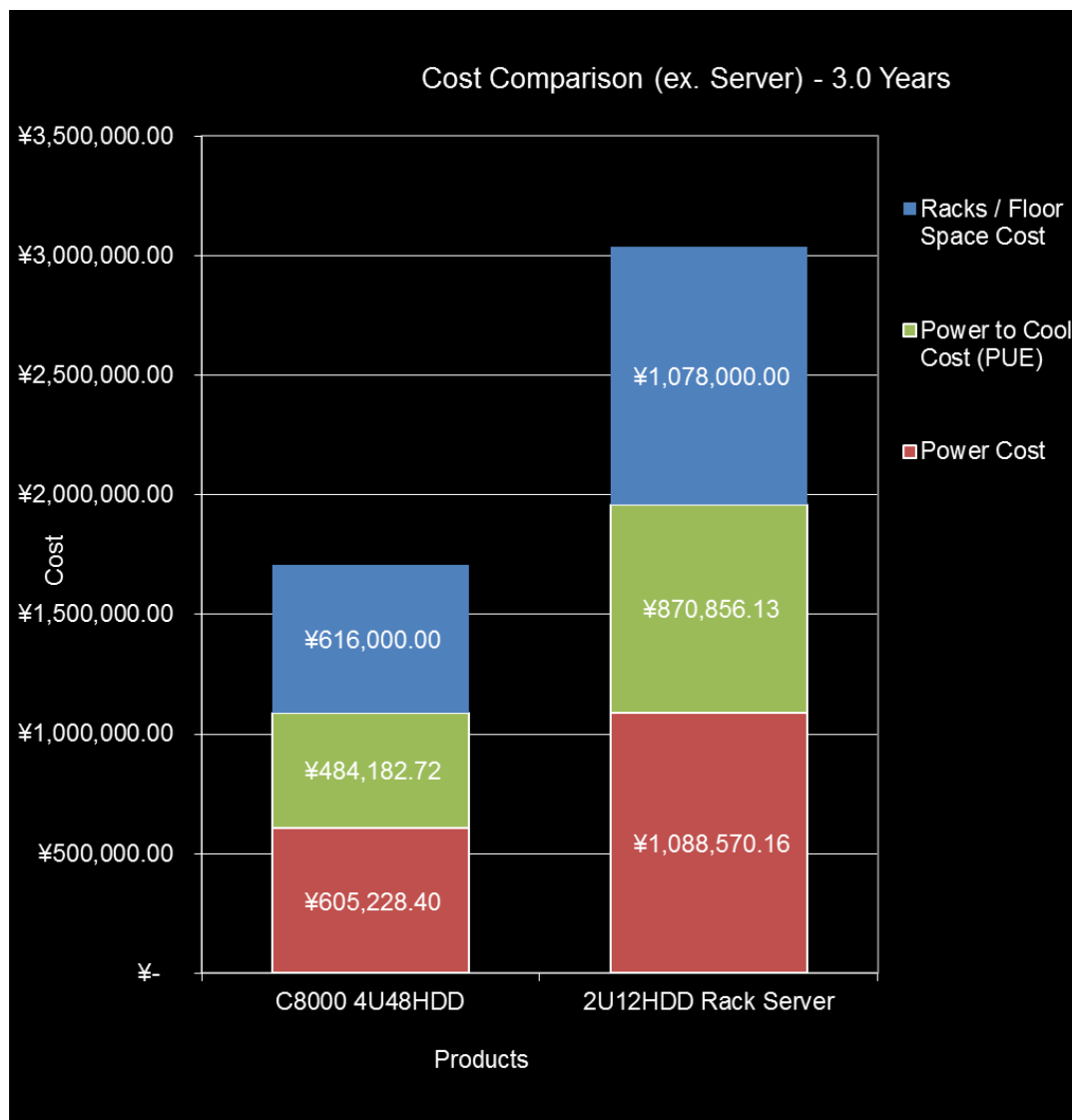
¥1,332,015.00

- 节省的成本可用于另  
购买2U12HDD服务器:

25台

- 节省比例高达:

43.85%





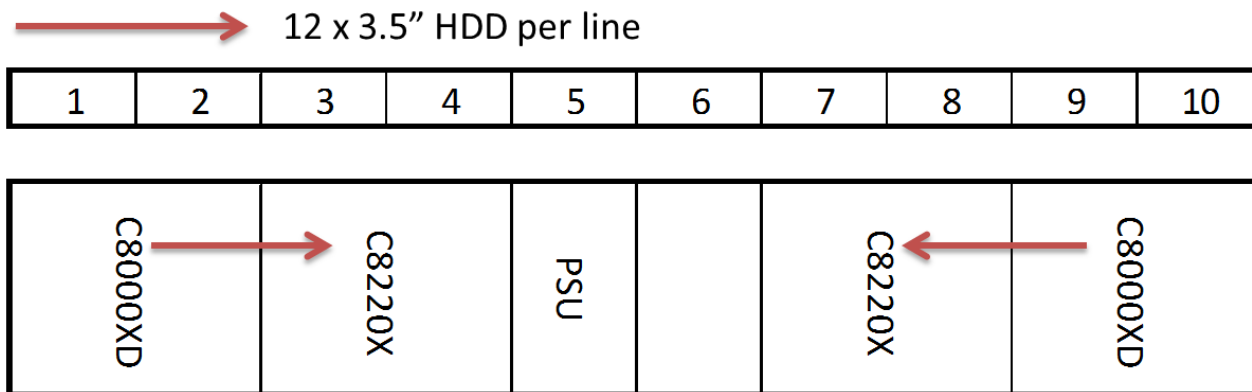
# 数据节点，高性价比均衡配置 二

存储容量与计算密度的均衡搭配，每数据节点16块3.5”盘

HDFS存储容量要求比较大，计算要求适中的场景

C8000 4U2Node32HDD:

- 1 x C8000 Chassis with 1 x PSU (1400W \* 2)
- 2 x C8220X : E5-2620 \* 2 / 64G / 500G SATA 2.5” \* 2 / 3T 3.5” SATA \* 4
- 2 x C8000XD : 3T 3.5” SATA \* 12 (12 HDD connect to each C8220X)

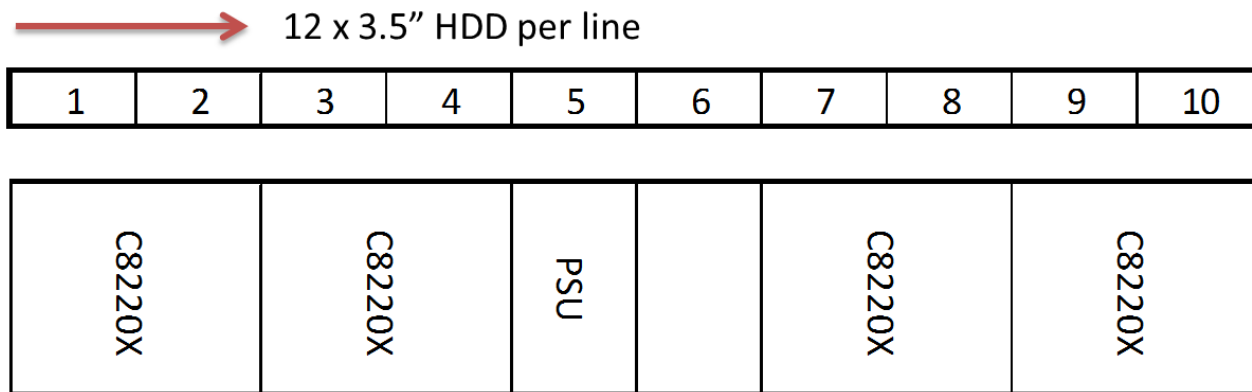


# 数据节点，高性价比计算型配置

较高计算密度和计算能力的配置，每数据节点4块3.5”盘  
HDFS存储容量要求适中，计算复杂的场景

C8000 4U4Node :

- 1 x C8000 Chassis with 2 x PSU (1400W \* 2)
- 4 x C8220X : E5-2650 \* 2 / 128G (16 \* 8G) / 500G SATA 2.5” \* 2 / 3T 3.5" NL SAS \* 4



# 为Hadoop优化的服务器平台 - PEC6220

通过PowerEdge C平台，从数据存储、报告和分析系统中获得价值，提供规模、速度、丰富性和易用性

- **PowerEdge C6220 - 12代产品**
- C6000（2U机箱）+ 4 台服务器节点C6220 + 2 个冗余电源。
- 2颗 Intel E5-26XX CPU，16根内存槽位，12块 3.5” 硬盘 或 24块2.5” 硬盘
- 服务器密度为传统1U服务器的两倍，同时保留热插拔硬盘灵活性
- 所有节点都可进行独立维护，管理员可随时对任何一个节点进行停机维护，不影响其他节点的正常运行
- 区别于传统2U4节点云服务器，更提供前置硬盘灵活分配技术，支持多种类型业务搭配部署



C6220高性能和多节点服务器



# 数据节点、边缘节点，高性价比计算型配置

HDFS存储容量要求较小，计算复杂的场景

C6000 2U4Node (平均分配硬盘):

- 1 x C6000 Chassis with 2 x PSU (1400W \* 2)
- 4 x C6220 : E5-2650 \* 2 / 128G (16 \* 8G) / 1T SATA 2.5" \* 6

C6000 2U4Node (灵活分配硬盘):

- 1 x C6000 Chassis with 2 x PSU (1400W \* 2)
- 2 x C6220 : E5-2650 \* 2 / 128G (16 \* 8G) / 1T SATA 2.5" \* 2，用于边缘节点
- 2 x C6220 : E5-2650 \* 2 / 128G (16 \* 8G) / 1T SATA 2.5" \* 10，用于数据节点

# 名称节点，标准2U机架式配置

- PowerEdge R720 - 12代产品
- 强调性能与扩展性的主流2路/2U机架式服务器
- 2颗 Intel E5-26XX CPU，24根内存槽位，8个3.5” 硬盘（R720），12个3.5” 硬盘（R720XD）

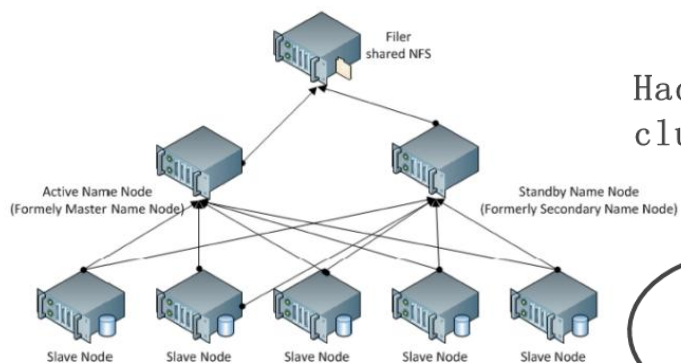


较高计算能力及内存容量的配置，单台服务器可达768GB内存  
HDFS名称节点，高可用、大内存需求的场景

R720 \* 2 （Master Name Node + Standby Name node）：

— E5-2650 \* 2 / 128G (16 \* 8G) / 600G SAS 3.5” \* 6

# Hadoop部署参考



Hadoop Big data cluster

