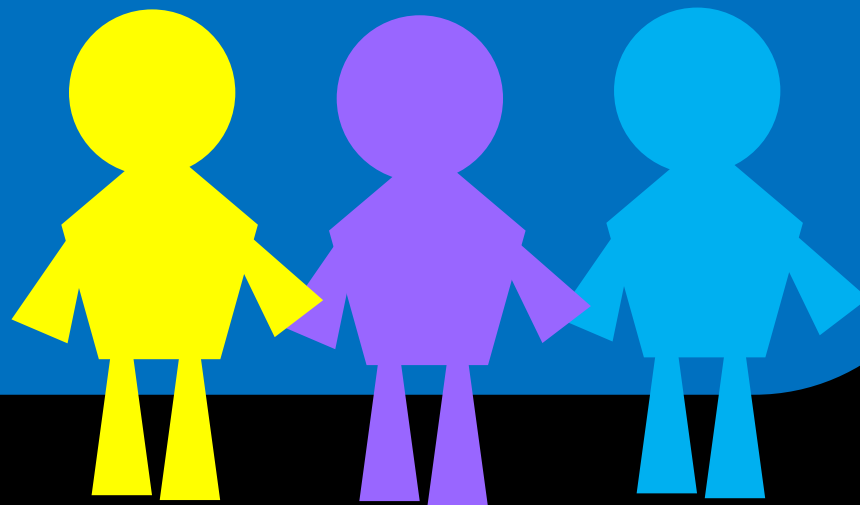
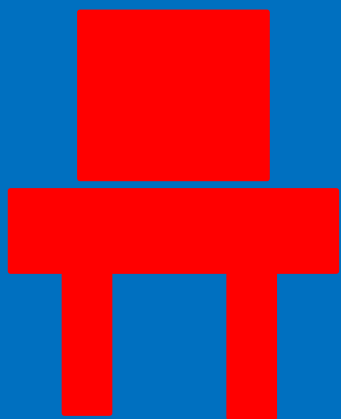
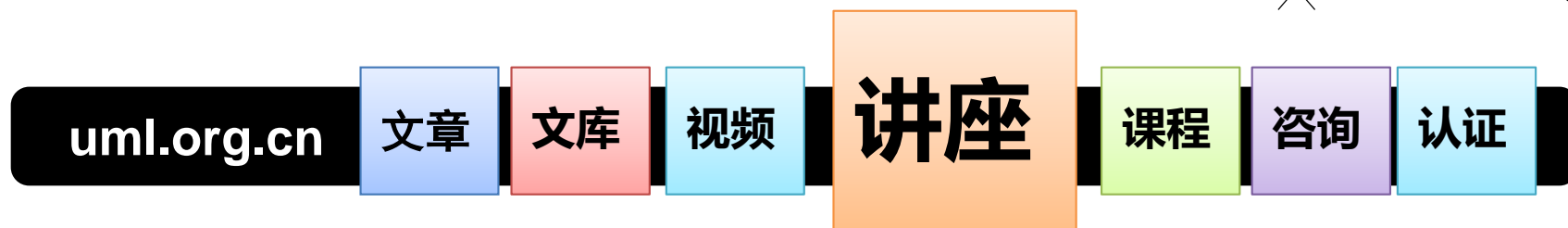




火龙果软件  
uml.org.cn

# 2014年助学计划





## 火龙果讲堂：

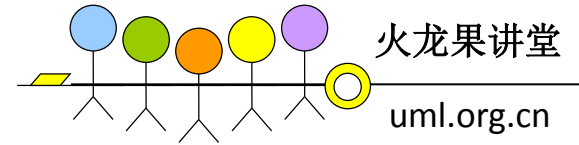
- 一线专家
- 案例回顾
- 经验分享

# 混合云架构及企业应用

杨大江

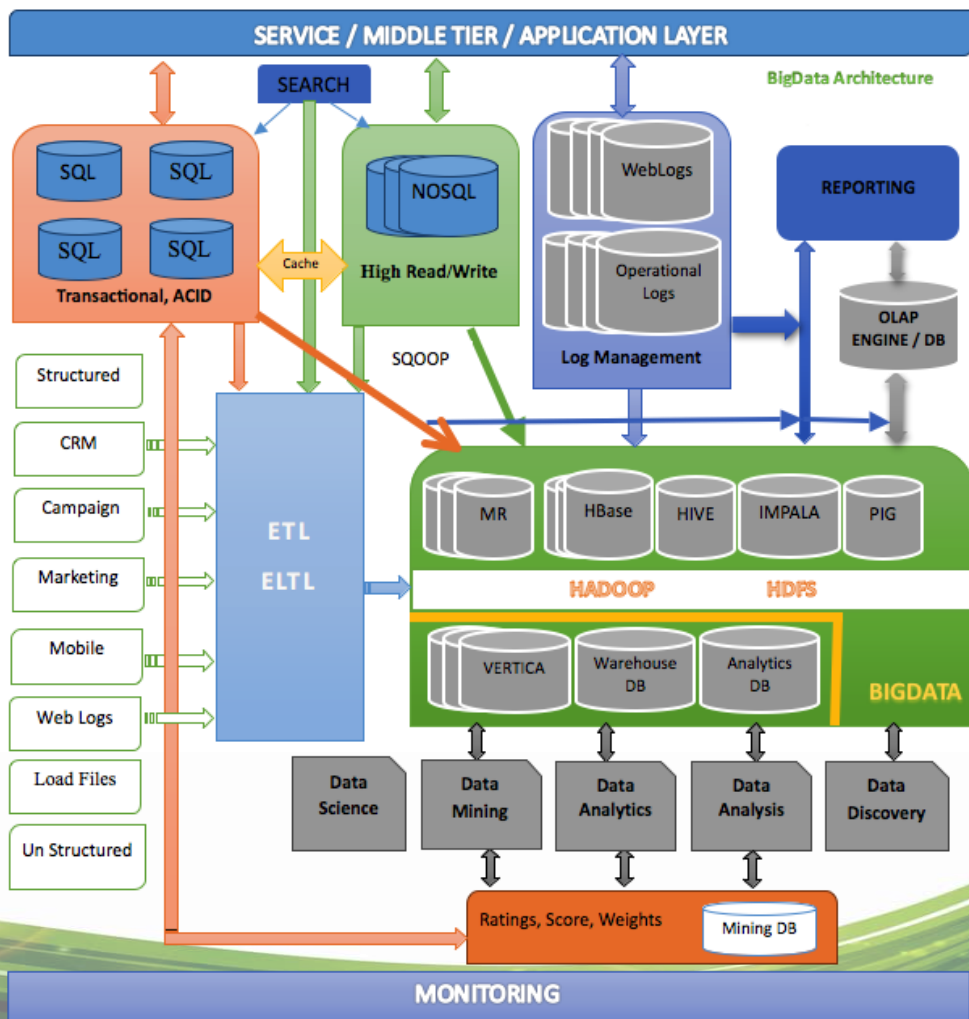
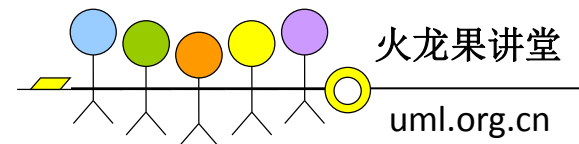
桉树系统公司

# 提纲



- Eucalyptus 公司的Eucalyptus私有云系统
- CLOUDERA公司 Hadoop发行版
- 如何基于IaaS服务开发与部署应用
- 结合两种平台优势补充和提高传统架构
- 实际案例

# 引题：典型的BIGDATA架构

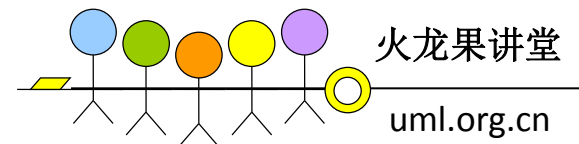


four major logical components:

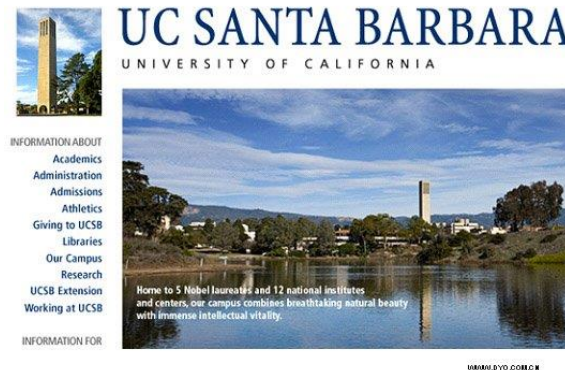
1. Data Source
2. Data Transformation
3. Data Processing or Data Integration
4. Data Consumption

# Eucalyptus 公司的Eucalyptus私有云系统

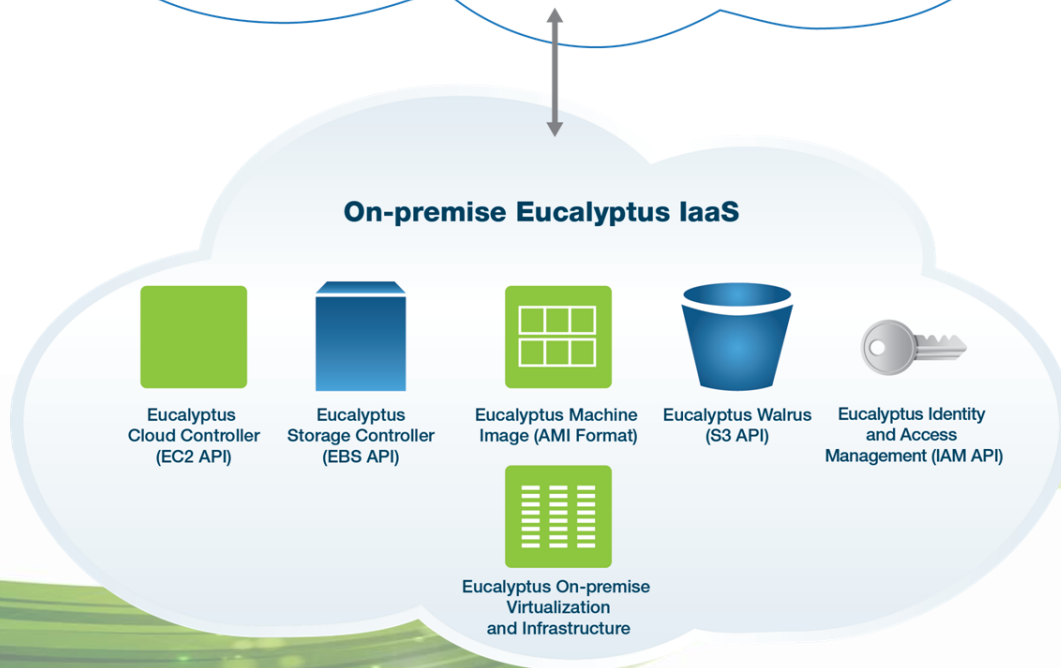
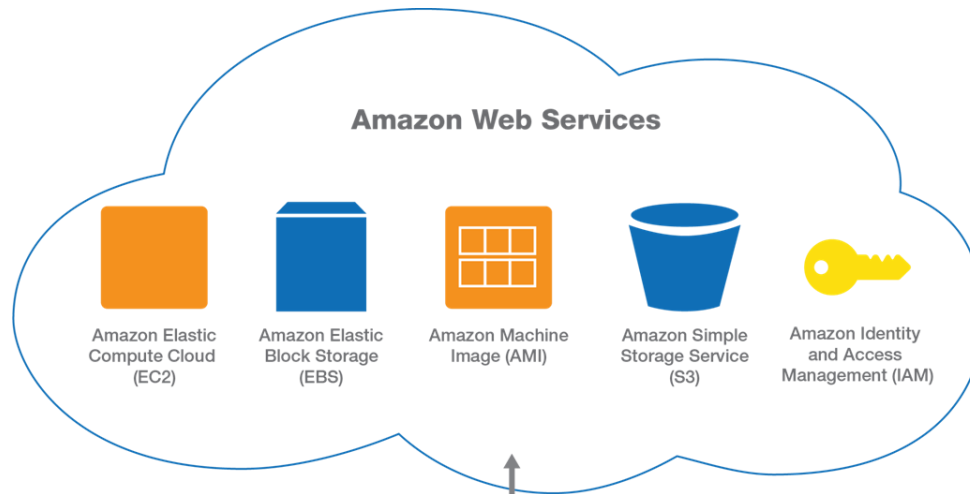
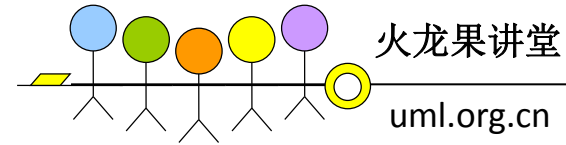
# 桉树公司概况



- 领先的开源私有云平台的创立者
- 成立于**2009年1月**，总部在美国加州圣巴巴拉
- 开源云计算的领导者
- **NEA, Benchmark, BV Capita**投资
- 诞生于加州大学Santa Barbara分校（最新世界大学排名第七位，[www.Leidenranking.com](http://www.Leidenranking.com)），美国政府国家科学基金投资云计算研究项目**EUCALYPTUS**（2008年5月发布）
- 企业版于**2009年**首次发布
- 美国政府财政支出官网和美国偶像投票系统均由**EUCALYPTUS**支持

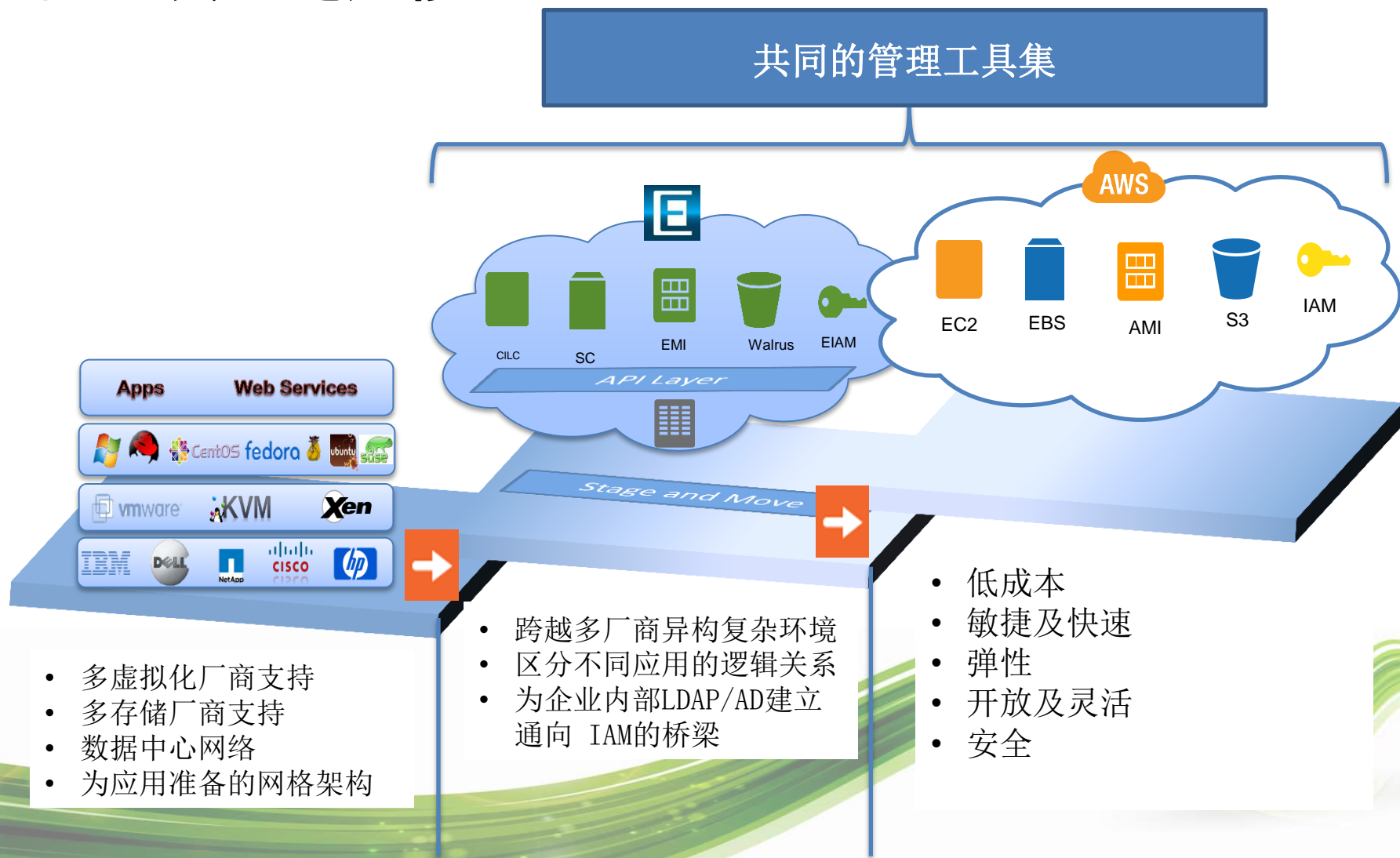


# 亚马逊AWS - 桉树联盟





# 客户应用在企业内部云和亚马逊AWS间无缝迁移



- 多虚拟化厂商支持
- 多存储厂商支持
- 数据中心网络
- 为应用准备的网格架构

- 跨越多厂商异构复杂环境
- 区分不同应用的逻辑关系
- 为企业内部LDAP/AD建立通向 IAM的桥梁

- 低成本
- 敏捷及快速
- 弹性
- 开放及灵活
- 安全

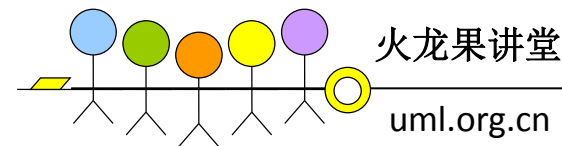
## 桉树适用环境

1. 工作量爆炸性增长: 互联网和移动互联网应用
2. 应用周期短暂: 研发和测试环境
3. 大数据: 分析和大量计算

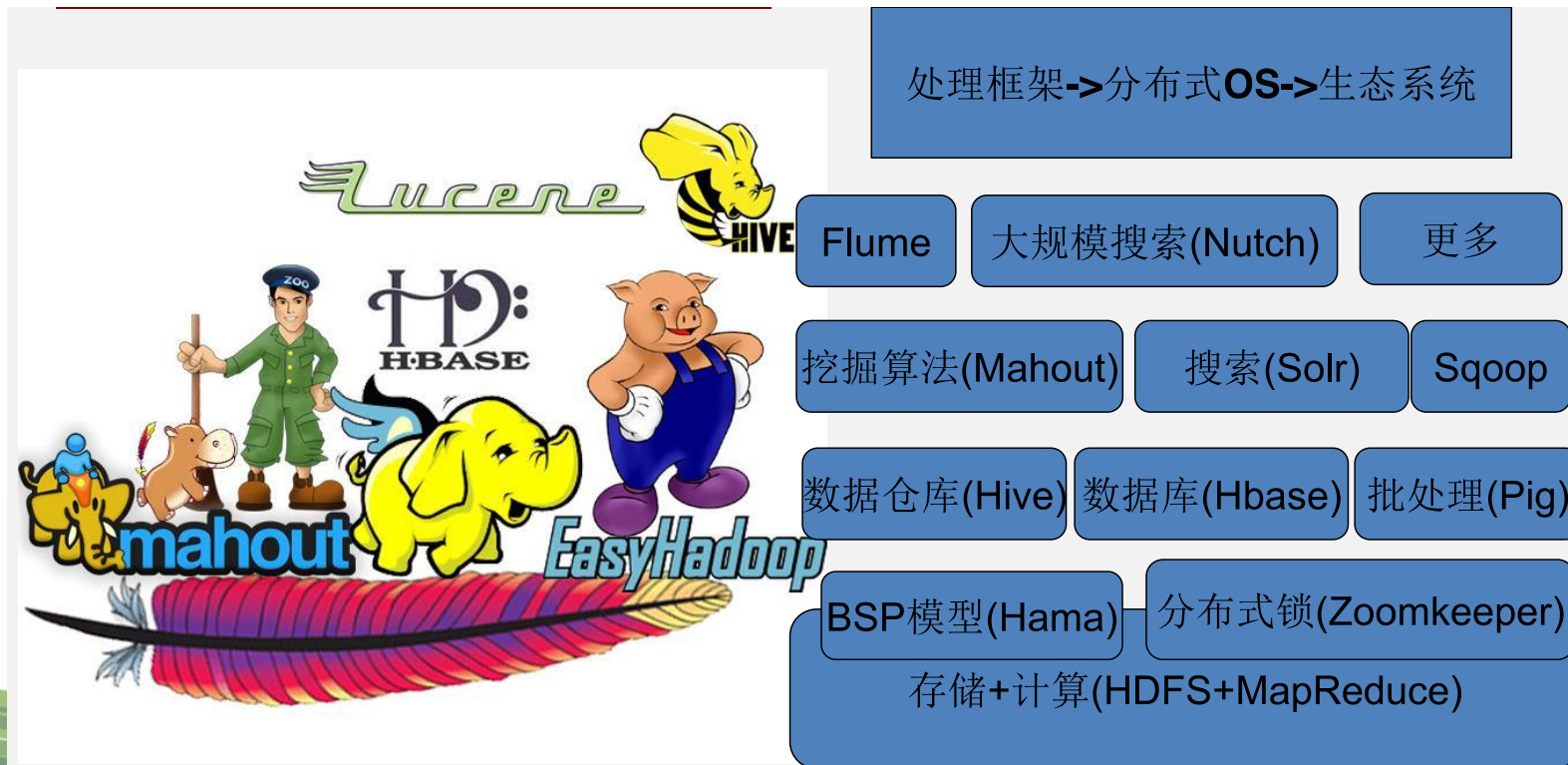
混合云，数字媒体编/解码，大规模复杂数据处理，教育和远程教育，云存储，虚拟桌面，公有云，电信运营商级别云服务，等等

# CLOUDERA公司的Hadoop发行版

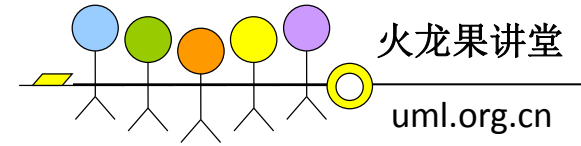
# HADOOP



- 一个分布式系统基础架构，由Apache基金会开发。用户可以在不了解分布式底层细节的情况下，开发分布式程序。充分利用集群的威力高速运算和存储。



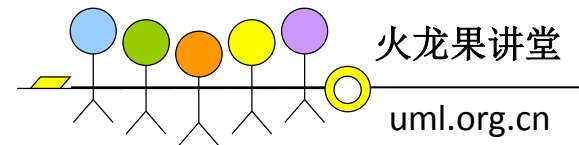
# cloudera



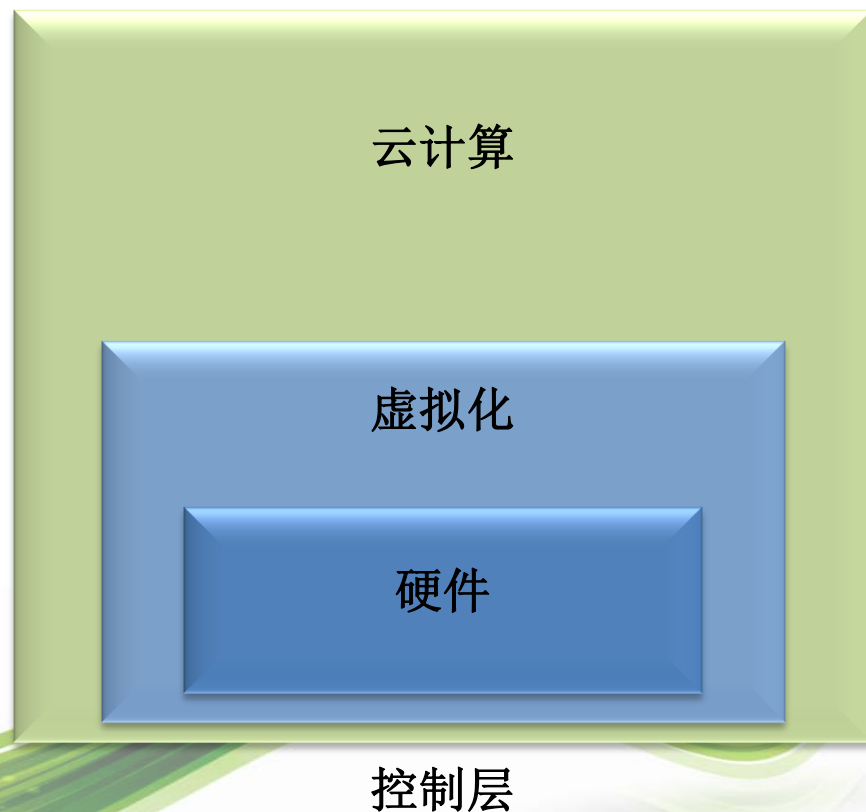
- 由于Hadoop深受客户欢迎，许多公司都推出了各自版本的Hadoop，也有一些公司则围绕Hadoop开发产品。在Hadoop生态系统中，规模最大、知名度最高的公司则是Cloudera公司

# 如何基于私有云IaaS服务开发与部署应用

# 引题：虚拟化与云计算的关系



- 云计算利用虚拟化。
- 云计算不仅仅是虚拟化。
- 虚拟化能够在—台物理主机上运行多个相互隔离的虚拟机。
- 云计算提供每一个认证用户可信的自服务接口，QoS，以及私有、安全的内部VM网络。

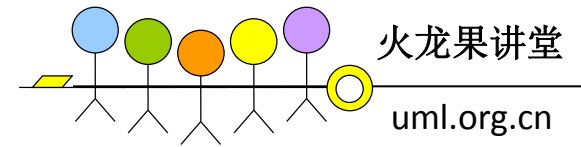


# 虚拟化和云计算之间的差异

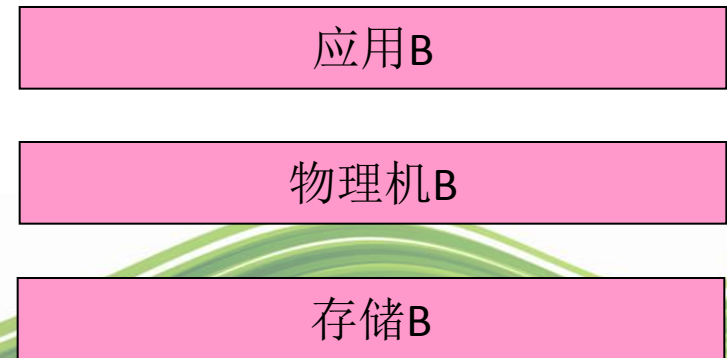
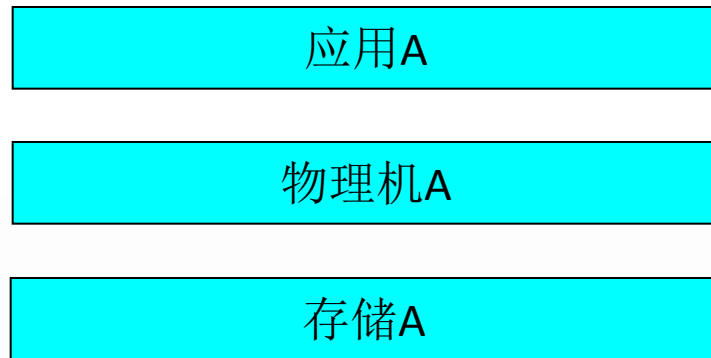
	传统的虚拟化	云计算
<b>VM 位置</b>	在一个特定的主机	在可用资源区中任何主机上
<b>VM 存储</b>	持续的	暂时的(除了使用 volumes)
<b>VM 资源(CPU, 内存)</b>	用户自定义	标准尺寸(小, 中, 大等)
<b>VM资源更改方式</b>	调整现有VM	启动新的VM
<b>提供时间</b>	数分钟到数小时	数分钟
<b>VM故障恢复方式</b>	试图恢复出现故障的VM	丢弃并启动新的虚拟机



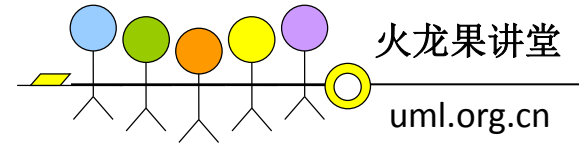
# 传统的应用部署模式



- 每个应用单独配备硬件：计算、存储、网络
- 平均资源利用率比较低
- 业务峰值时段处理能力不够，缺乏扩展能力

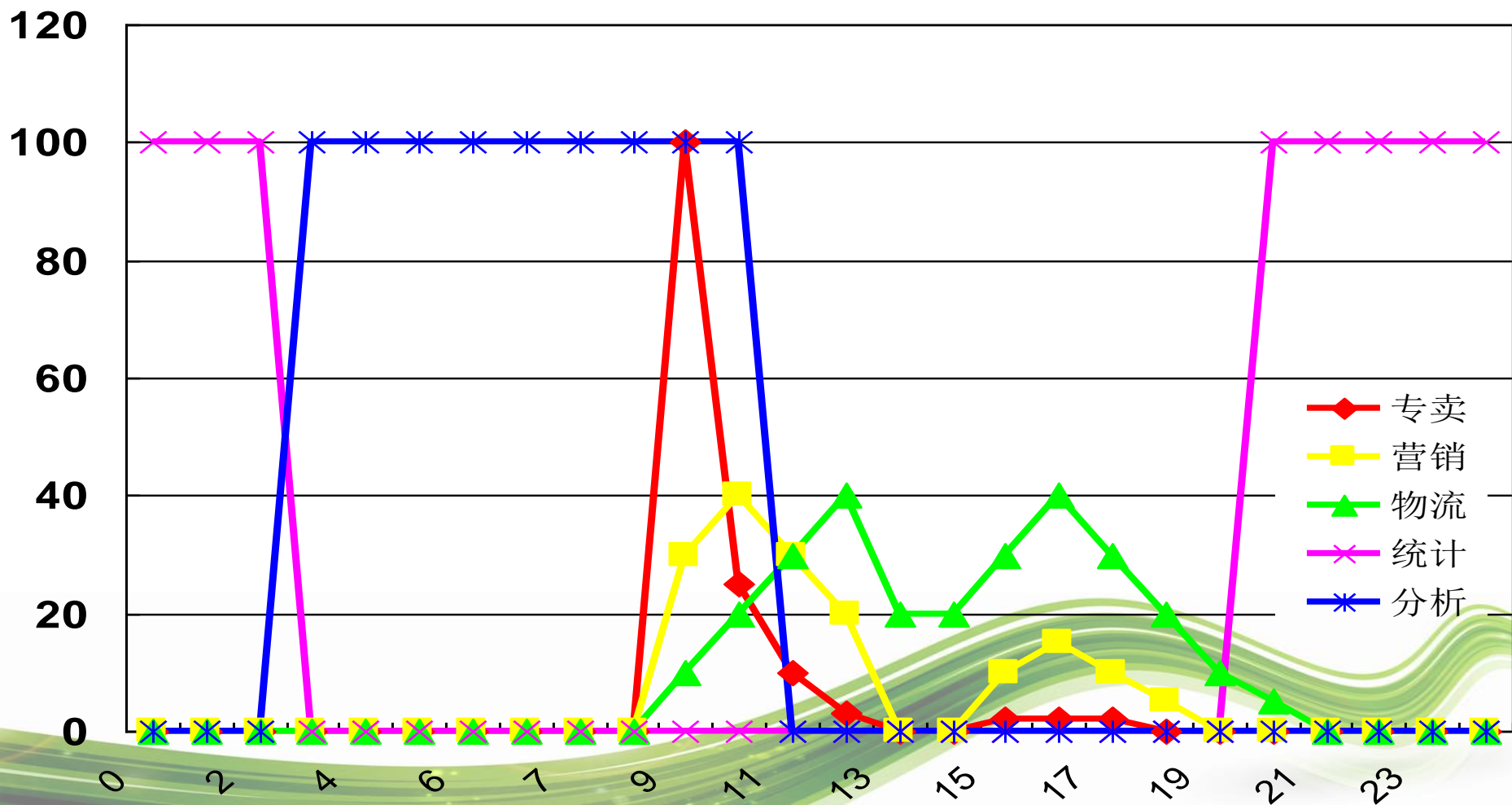


# 传统的应用部署模式

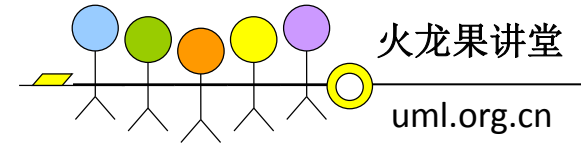


- 行业常见应用：
  - 营销系统、专卖系统、物流系统、客户管理
  - 人事管理、办公自动化
  - 统计分析
- 业务峰值时段处理能力不够，缺乏扩展能力
  - 上午九点（专卖）
  - 促销
  - 红眼时段（统计分析）

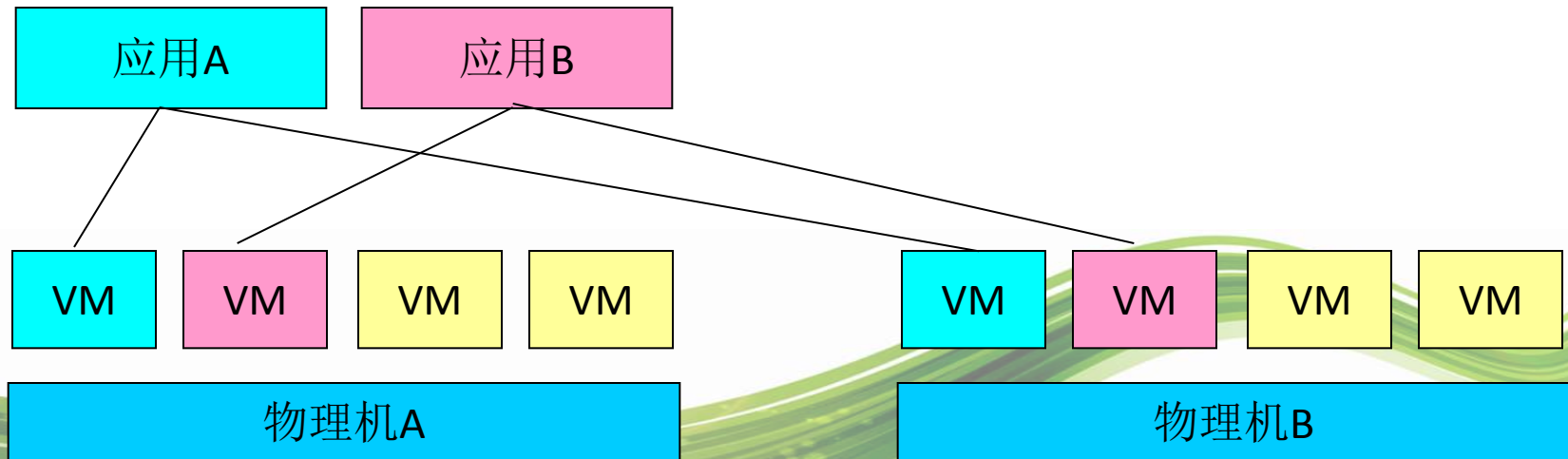
# 传统应用的“负载~时间”曲线



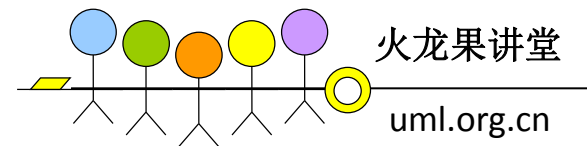
# 云计算应用部署模式



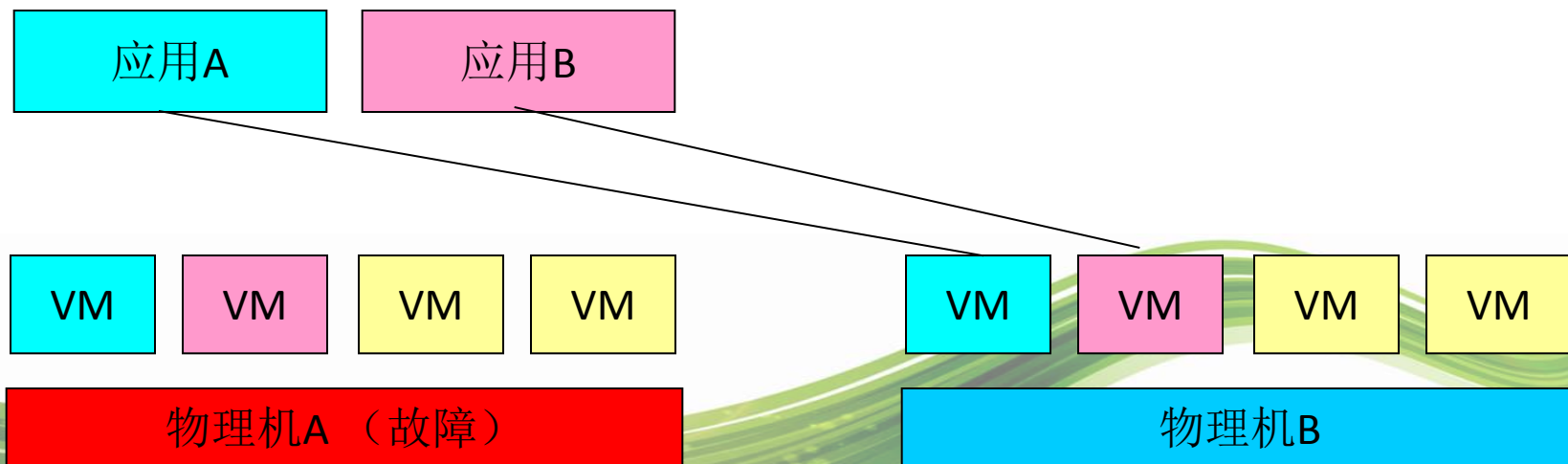
- 单台物理机通过虚拟化技术，提供多台相互独立的虚拟机
- 应用部署到虚拟机上
- 对于关键性应用，通过横向扩展提供高可用性
- 空闲的虚拟机随时可以用于应对爆发性业务



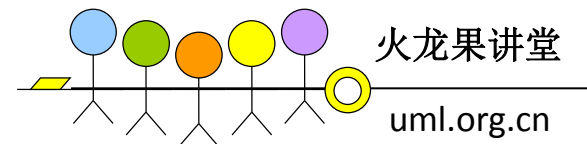
# 云计算应用部署模式



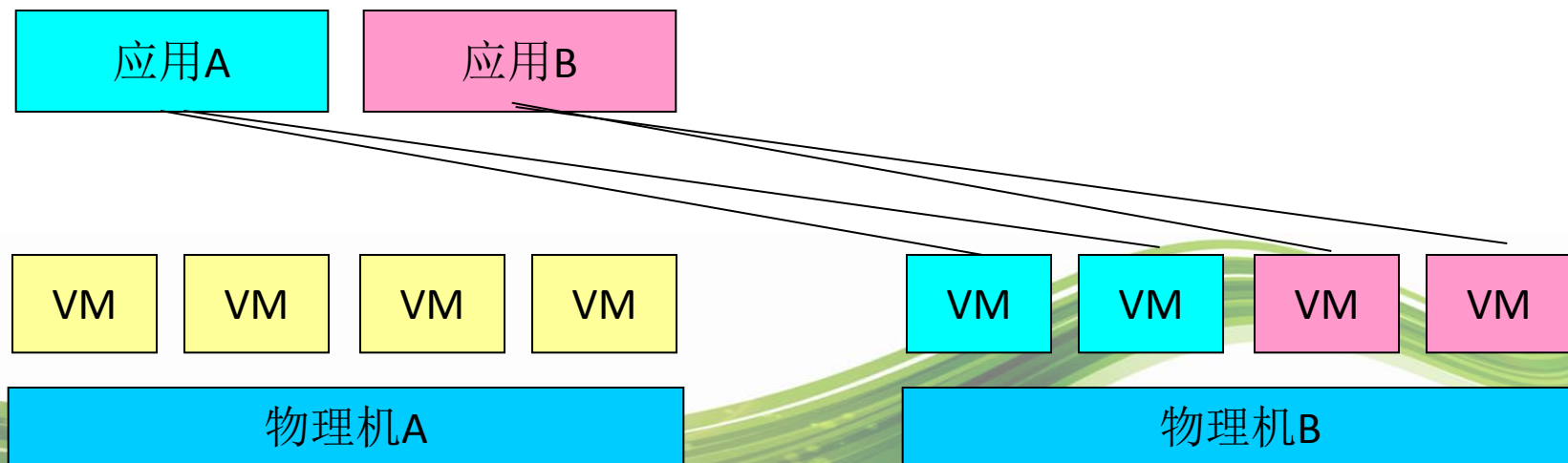
- 部分硬件故障时，其他硬件依然能够保证业务不致中断



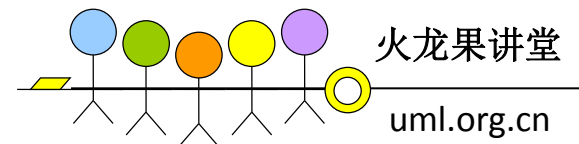
# 云计算应用部署模式



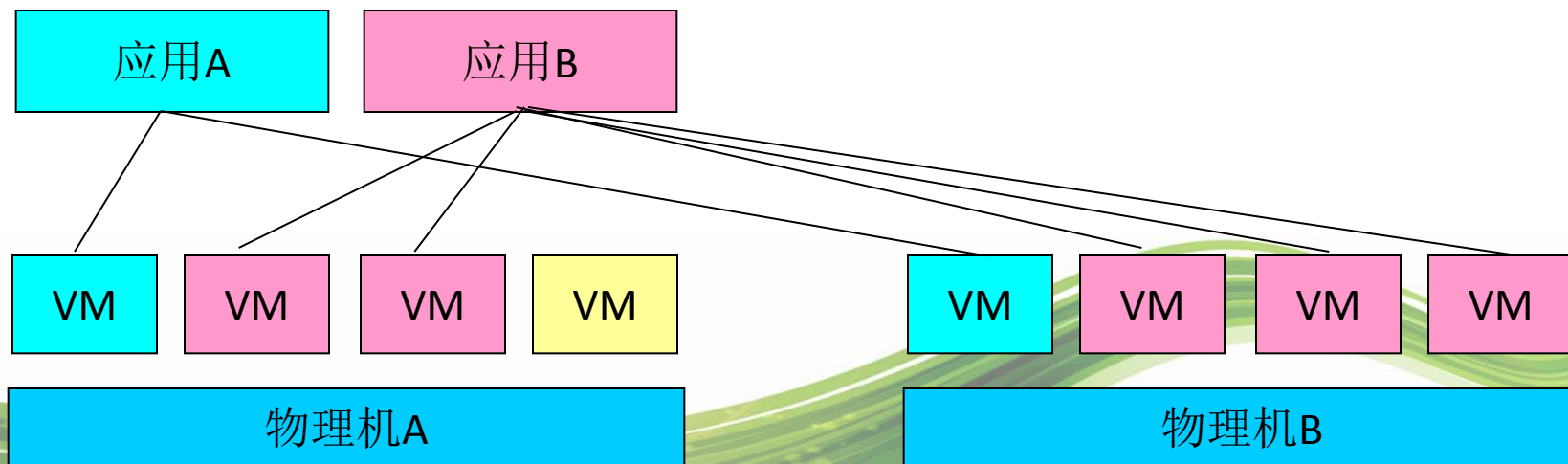
- 对硬件进行维护时，通过动态迁移保障业务正常运行
- 对空闲出来的硬件进行维护后，再把负载回迁



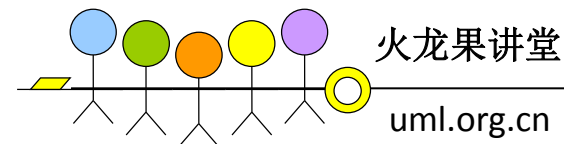
# 云计算应用部署模式



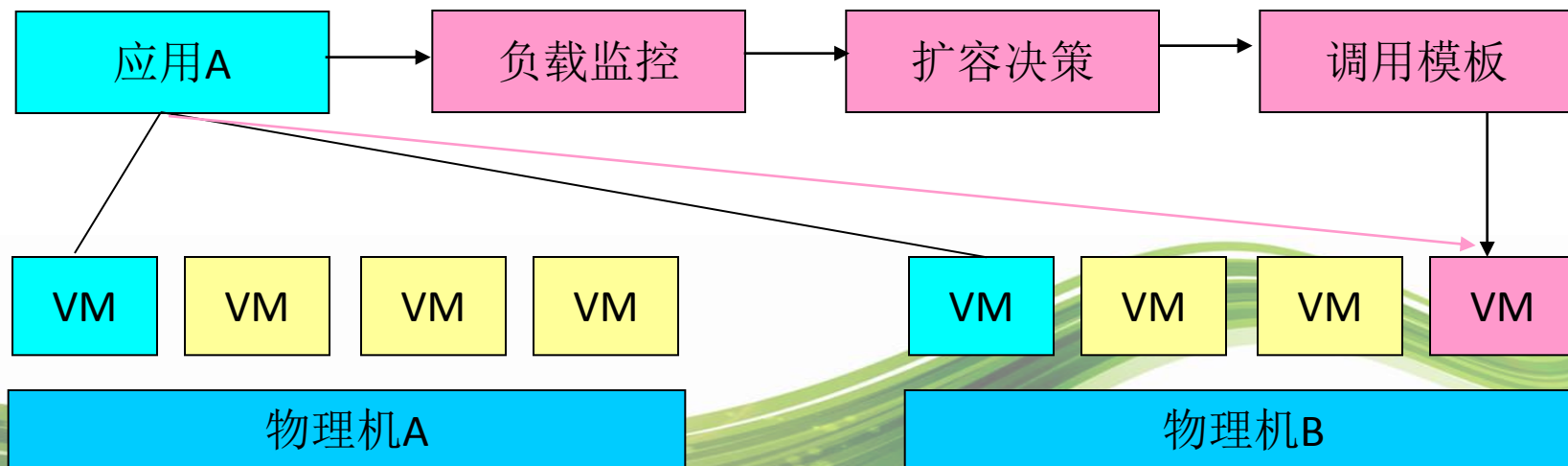
- 应用负载上升时，可以通过横向扩展提高处理能力
- 应用的整体处理能力可以超过原有专有硬件的处理能力
- 整体硬件投资不变，可以通过错峰运行调度业务压力
  - （营销类业务白天处理，统计分析类业务晚上处理）



# “云感应”型应用

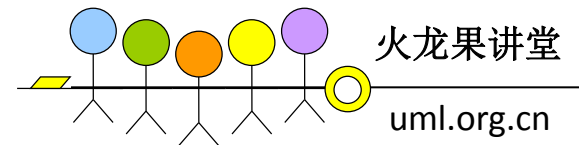


- 应用本身对承载自身各个组件的虚拟机进行负载监控
  - CPU、内存、磁盘IO
- 基于预定策略的动态负载调整
  - 某个参数超过某个阈值，动态增加一台虚拟机
  - 某个参数低于某个阈值，动态减少一台虚拟机



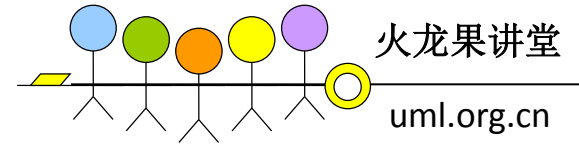


# 充分利用云平台的特性



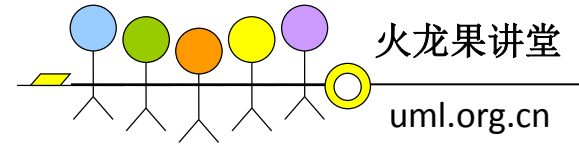
- Eucalyptus提供动态创建、删除虚拟机的API（AWS兼容）
  - 基于预先定义的EMI动态创建、删除虚拟机
- Eucalyptus提供ELB和Auto Scaling功能
  - 对虚拟机负载进行监控（CPU、内存、磁盘IO）
  - 基于预先定义的触发器进行动态负载调整
  - 此功能将在2013年4月发布的3.3版本中提供

# 值得关注的细节



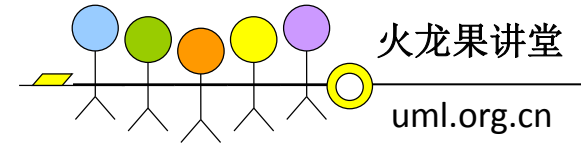
- 所有的虚拟机尽可能统一规格（CPU和内存配置）
- 如何确定虚拟机的规格？
  - 单台虚拟机能够应对日常的业务流量
  - 通过性能测试确定虚拟机的性能瓶颈
- 操作系统盘与数据盘分离
- 使用临时性系统盘，不往系统盘写数据
- 通过EBS卷或者数据库保存持久性数据
- 不要使用巨大型BfEBS卷

# 值得关注的细节



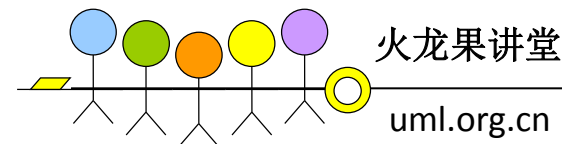
- 尽可能不将负载大的关键业务数据库运行在虚拟机上
- 数据库读写分离（写入Master，读取Slave）
- 数据库性能的瓶颈是磁盘IO
  - 增加磁盘数量
  - 做RAID
- 大部分虚拟化技术（包括VMWare）都无法完美支持数据库
  - 主要的问题来自虚拟化导致的磁盘IO性能损失
  - 使用基于逻辑卷的磁盘映像能够提升磁盘IO性能
  - 使用基于文件的磁盘映像大幅度降低磁盘IO性能
- 在数据库压力不大的情况下，可以用虚拟机承载数据库业务
  - 即使如此，也要将数据保存在EBS卷上

# 值得关注的细节

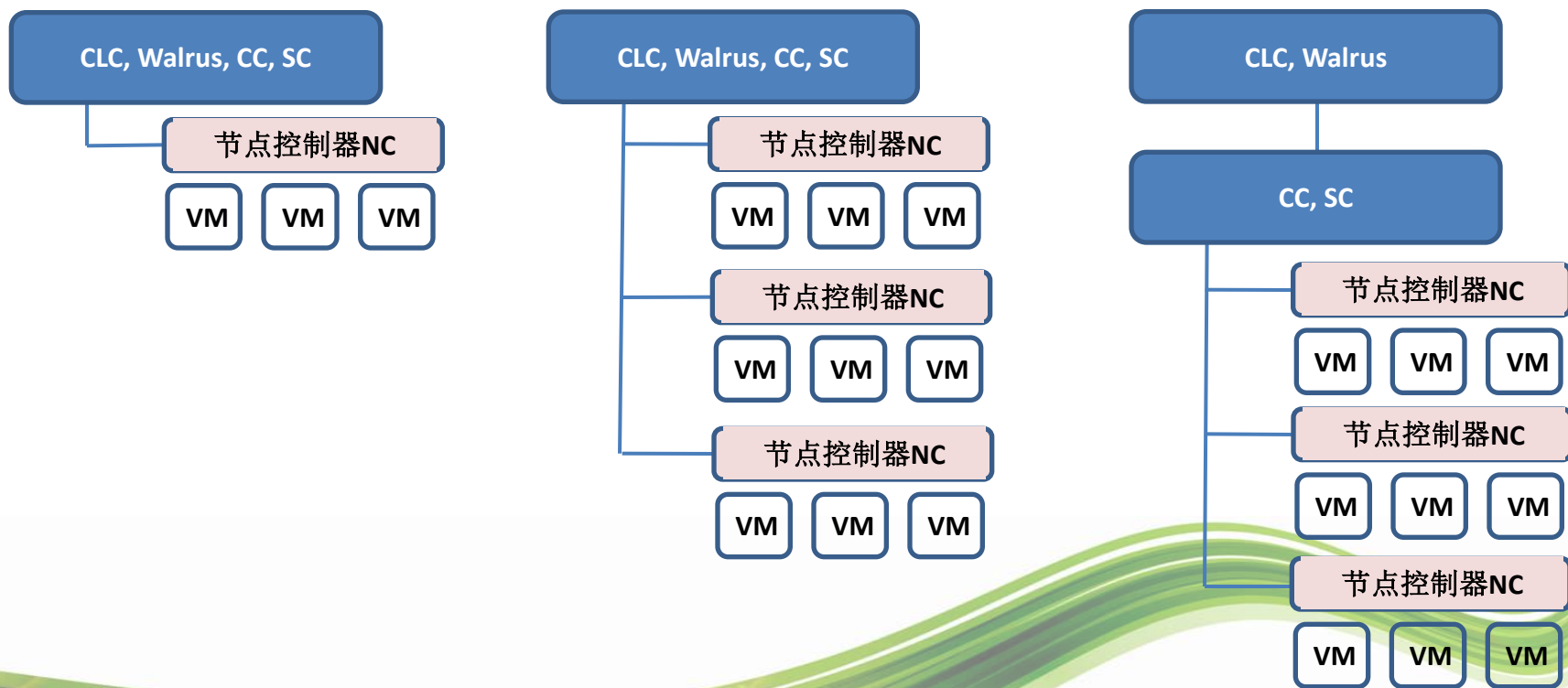


- 在开发测试阶段就开始使用云平台
  - 基于云平台搭建开发测试环境
  - 测试新的软件版本时，创建新的虚拟机，销毁旧的虚拟机
  - 尽可能使用API而不是图形用户界面
- 充分测试

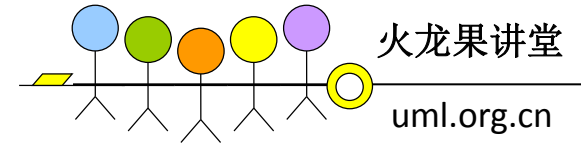
# 搭建Eucalyptus云平台



- 规模从小到大，从简单到复杂，应用逐步迁移



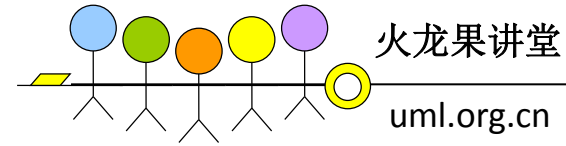
# 搭建Eucalyptus云平台



- 现在就去下载Eucalyptus云平台的最新版本
  - <http://www.eucalyptus.com/>
- 准备两台服务器，刻盘，安装，测试
- 加入邮件列表，获得社区支持
  - <http://lists.eucalyptus.com/>

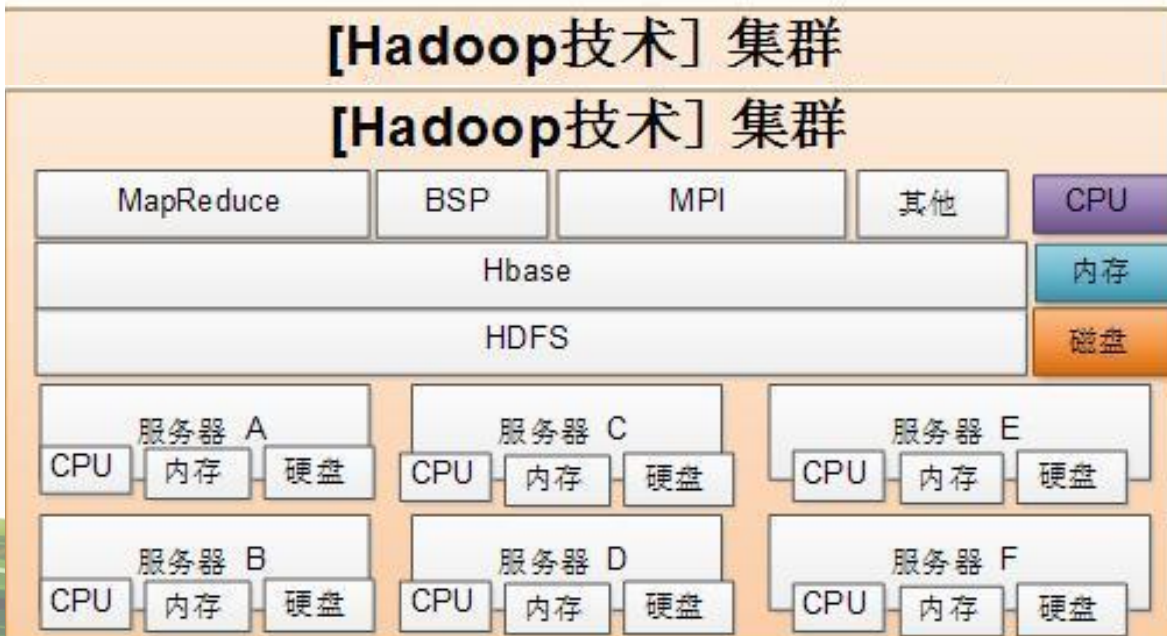
# 结合两种平台优势补充和提高传统架构

# Euca & Cloudera



# 分

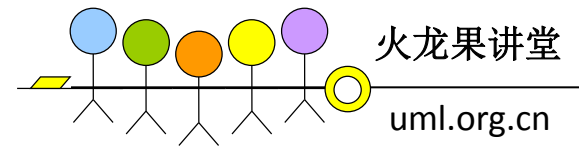
## EUCA



## Cloudera

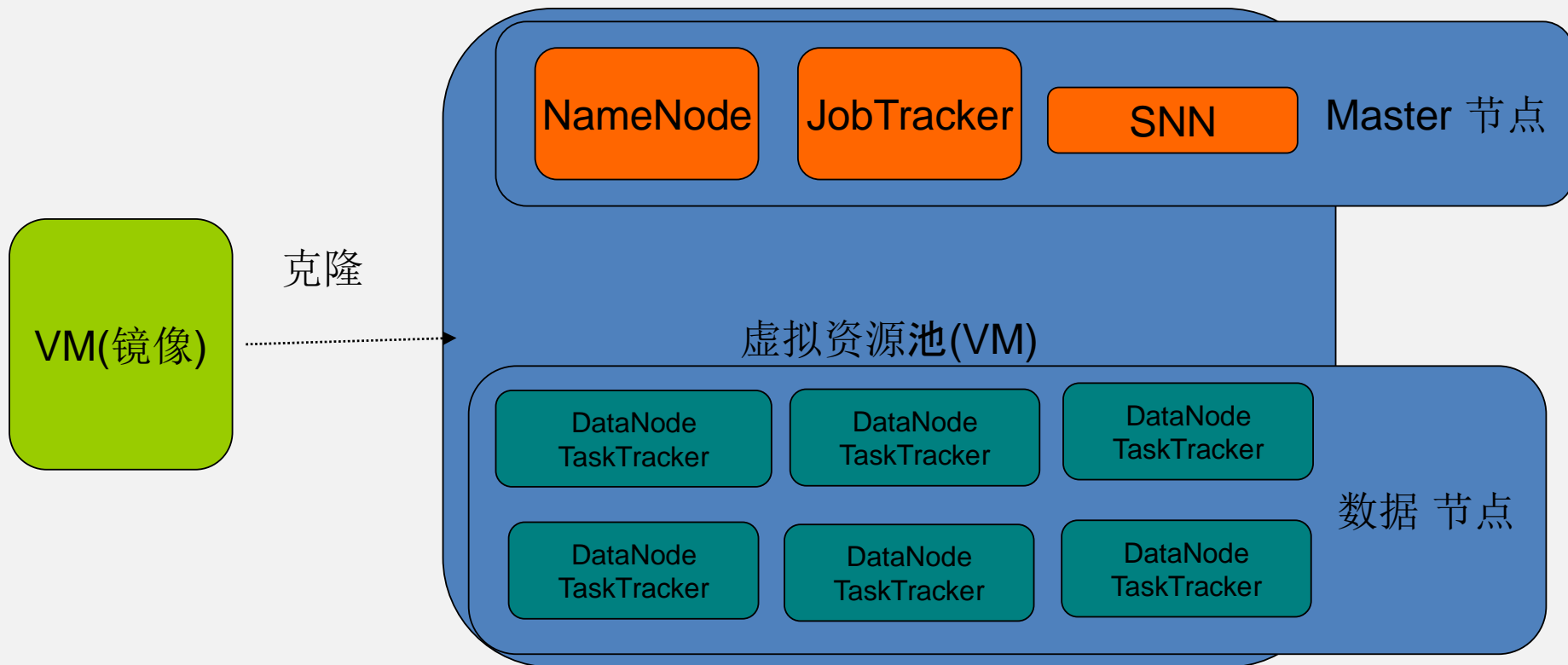


# 结合两种子平台优势 构建大数据系统

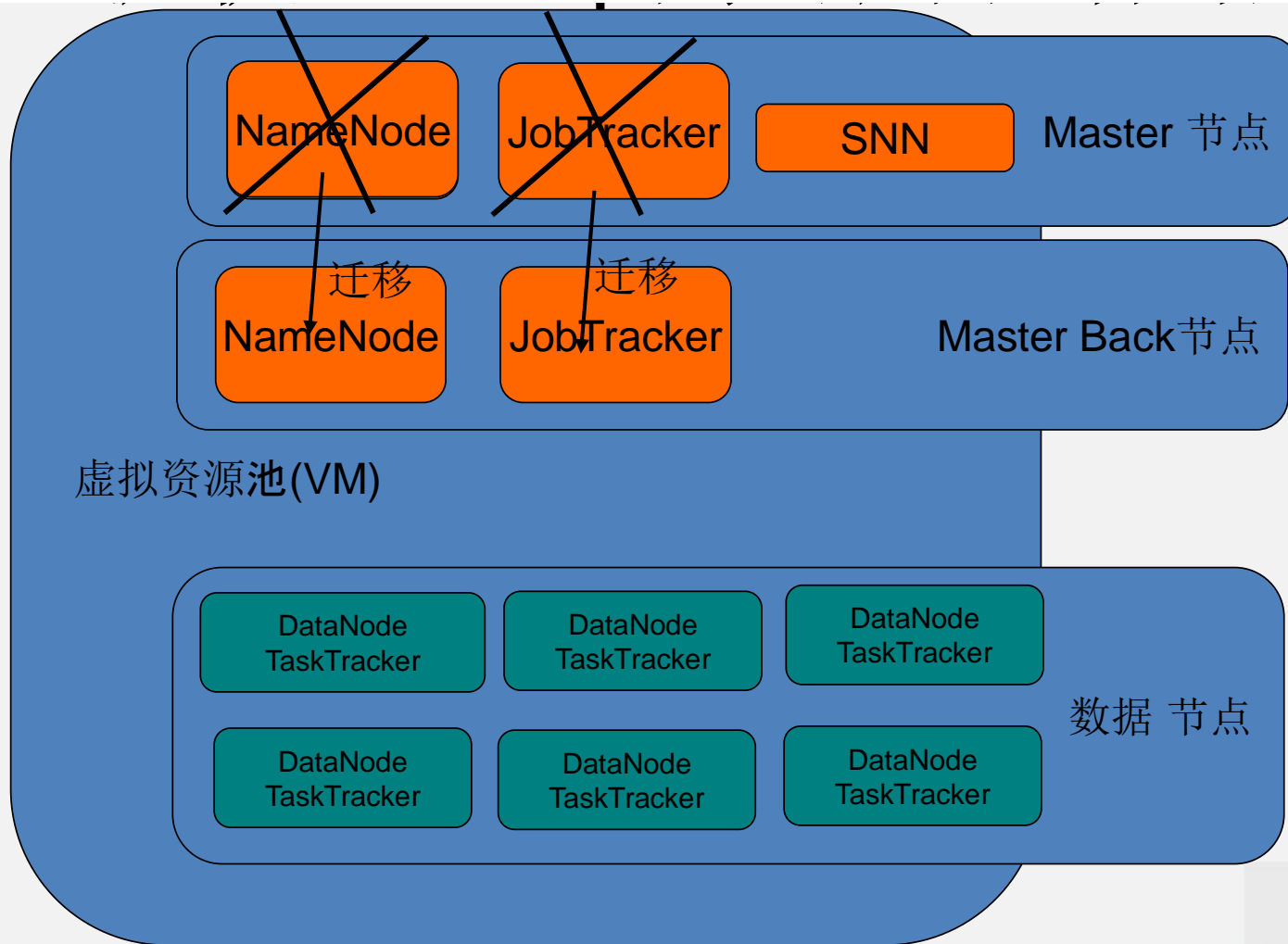


- 提升Hadoop部署速度
- 提供Hadoop高可用和容错能力
- 提升Hadoop环境资源利用率
- 安全隔离让Hadoop落地更安全
- 实现Hadoop云端多租户
- 集群易于维护和迁移
- 使用异构集群实现高密度存储和计算

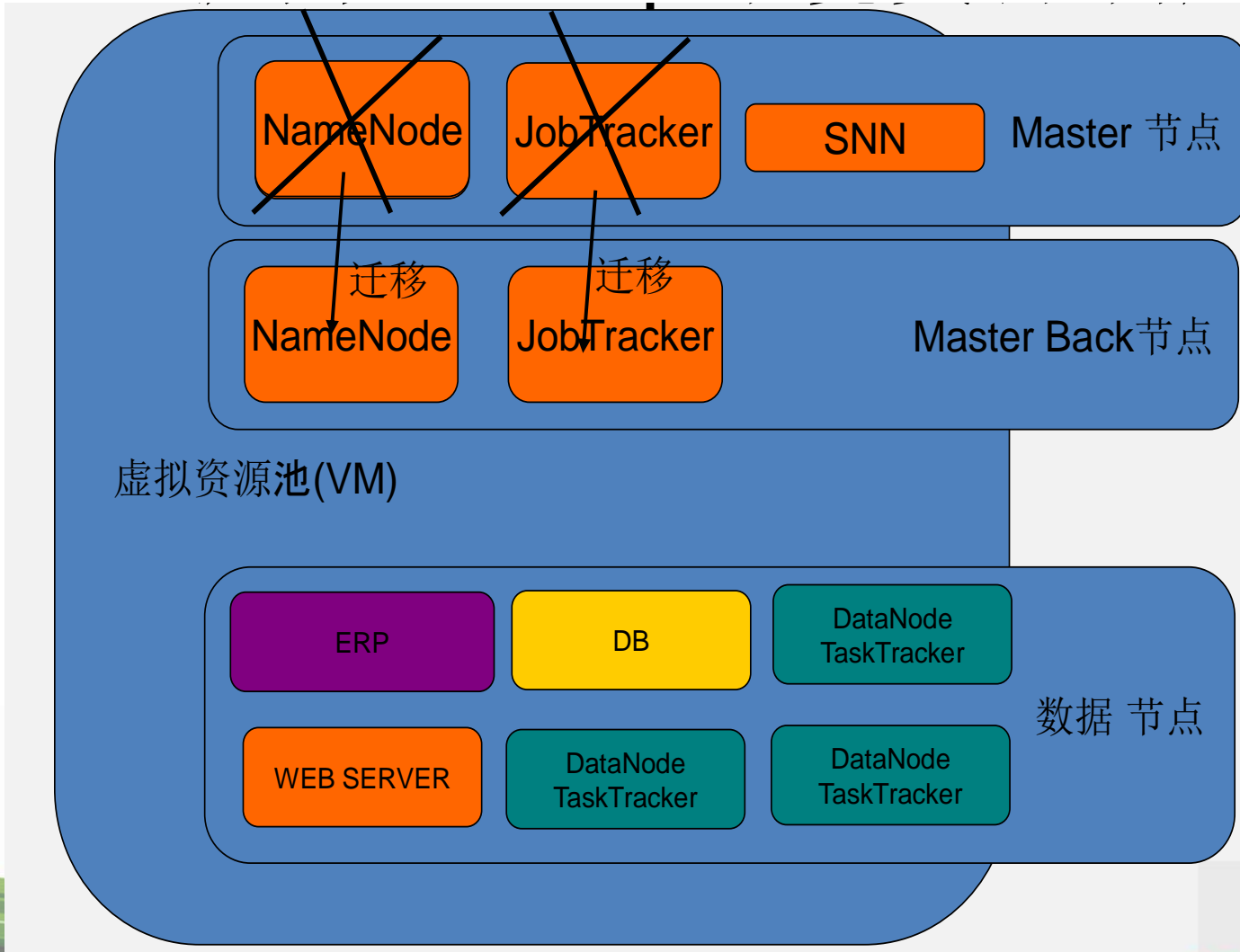
# 提升Hadoop部署速度



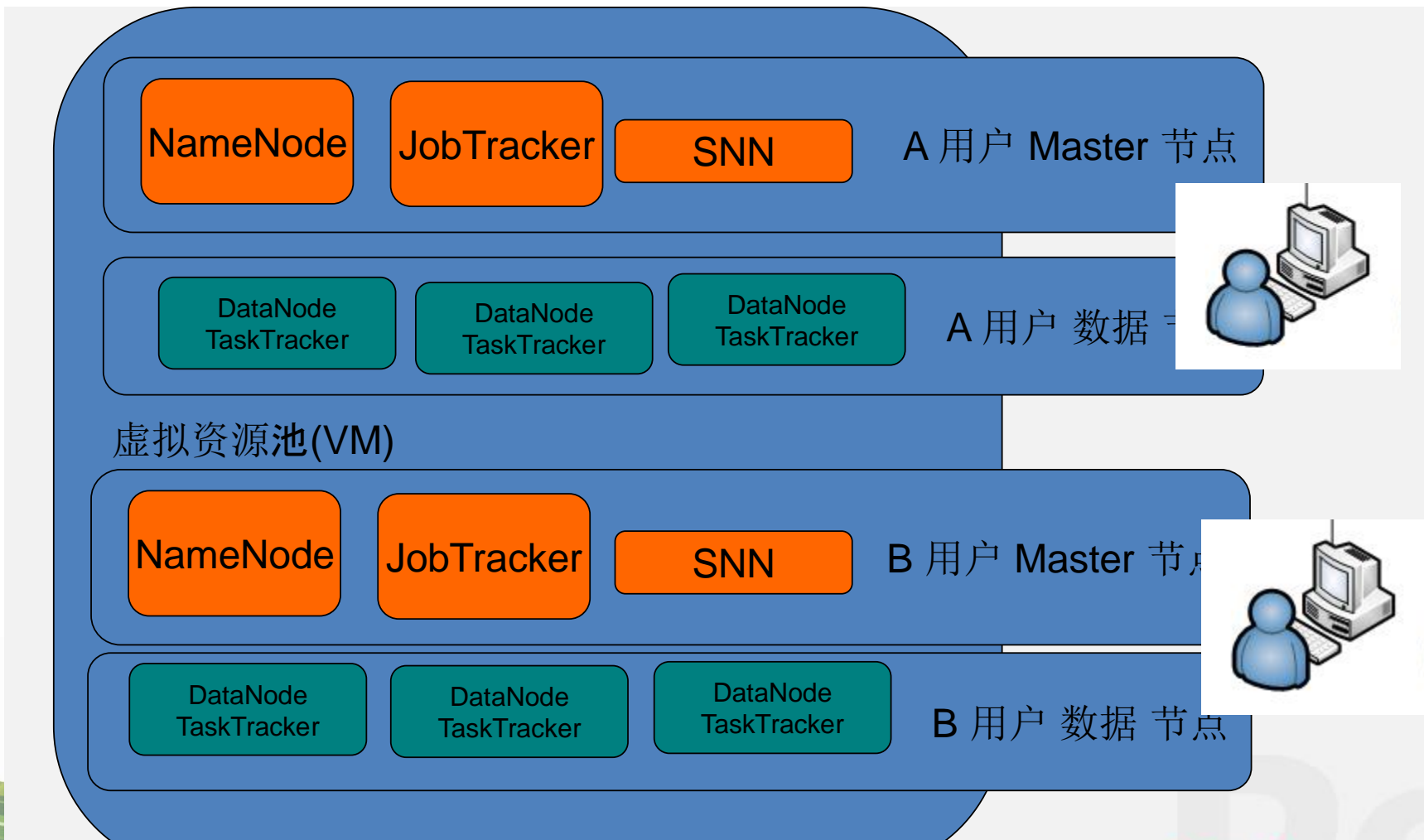
# 提供Hadoop高可用和容错能力



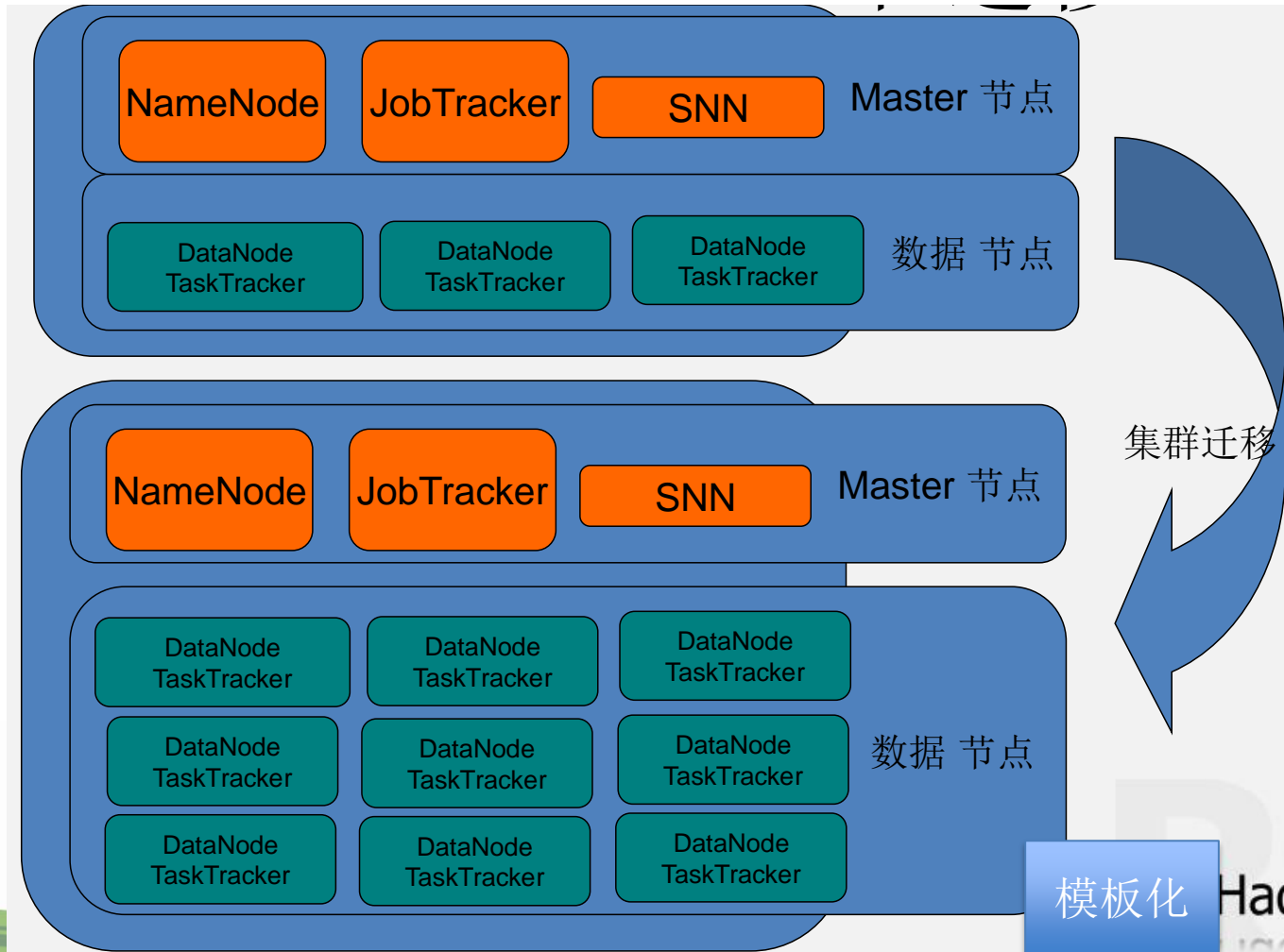
# 提升Hadoop环境资源利用率



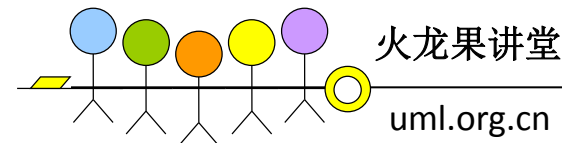
# 实现安全隔离以及Hadoop云端多租户



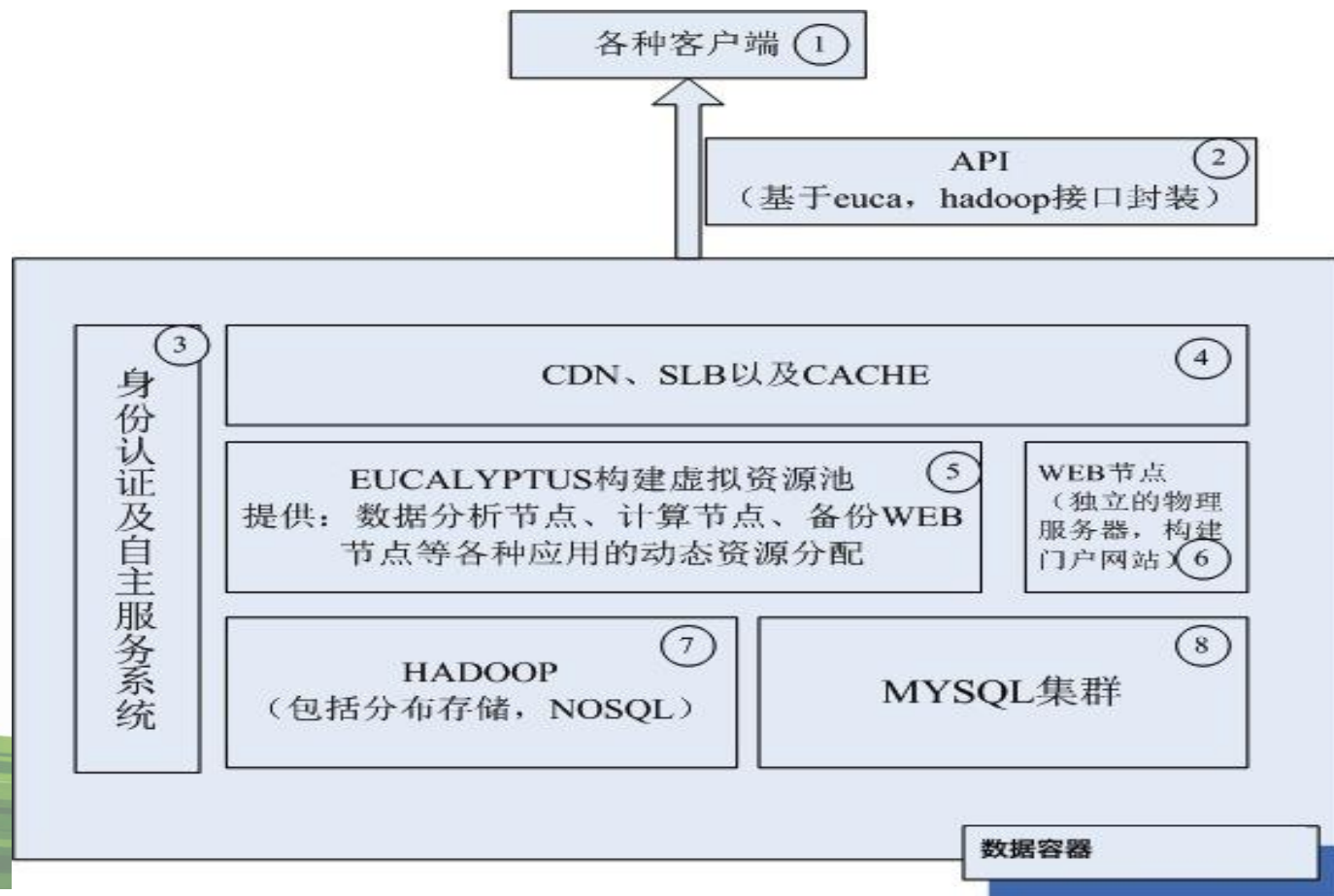
# 集群易于维护和迁移



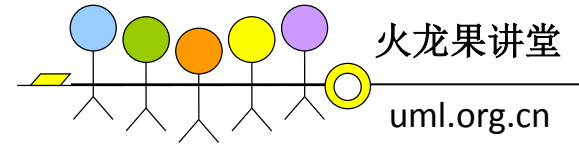
# 案例分享：郑州某金融公司



- 建立这个决策系统，是为了能够实时分析用户的交易行为，预测其结果，规避可能带给公司的风险。



# 如何落地混合云与大数据项目



- 1.从解决现有架构瓶颈开始,使用云计算及Hadoop等架构解决原有问题为落脚点。(例如报表计算慢?服务器资源利用不合理?)
- 2.从非核心,离线非交易的业务开始。
- 3.从小规模集群(3—4个)开始。
- 4.从依赖外部团队,到自主培养人才。



# 交流

- Q & A





讲座	2014年3月29日 产品设计方法与流程
----	----------------------

讲座	2014年4月19日 web安全防护与设计
----	-----------------------

课程	云平台架构与应用 <a href="http://uml.org.cn/course">uml.org.cn/course</a>
----	---

QQ群	284013037 (云计算)
-----	-----------------

更多...	<a href="http://uml.org.cn">uml.org.cn</a> 文章, 文档, 视频
-------	---