VMware为大数据应用铺平道路

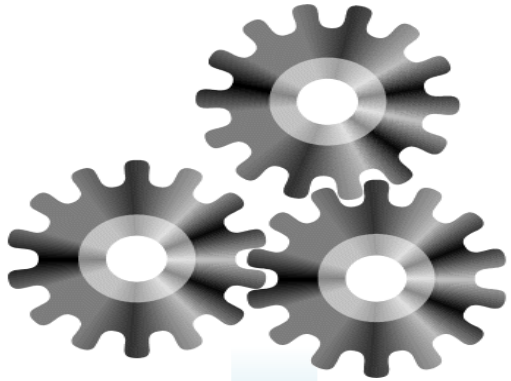
董波
Vmware 高级产品线经理

vmware

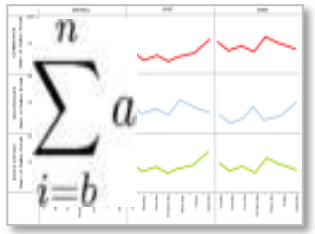
The Emerging Pattern of Big Data Systems: Retail Example

Real-Time Streams

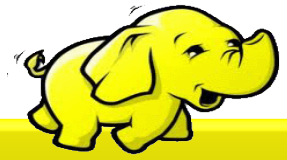
Real-Time Processing



Machine Learning



Data Science



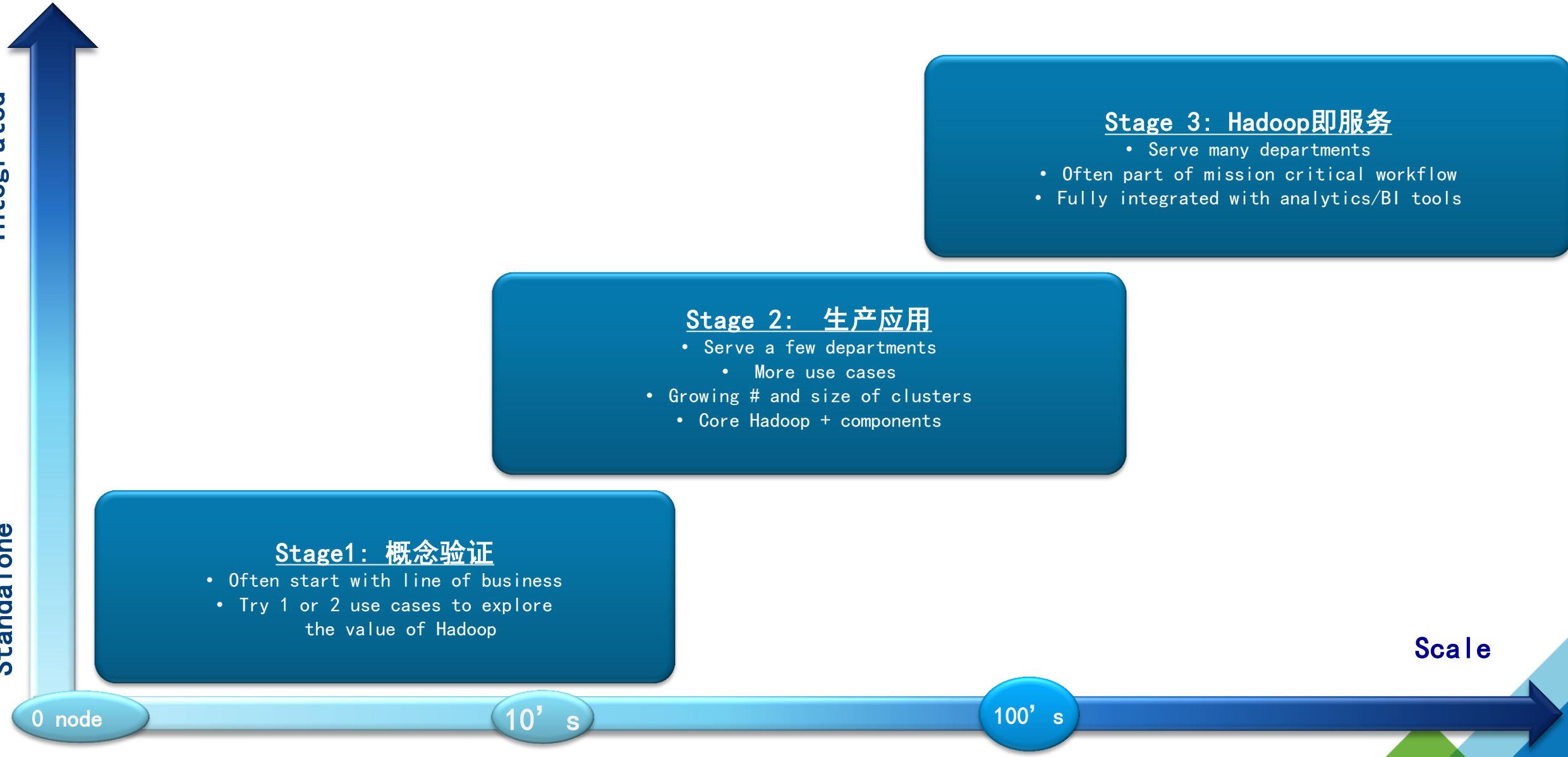
Parallel Data Processing

Exa-scale Data Store

Cloud Infrastructure

vmware

企业大数据应用的三阶段



vmware

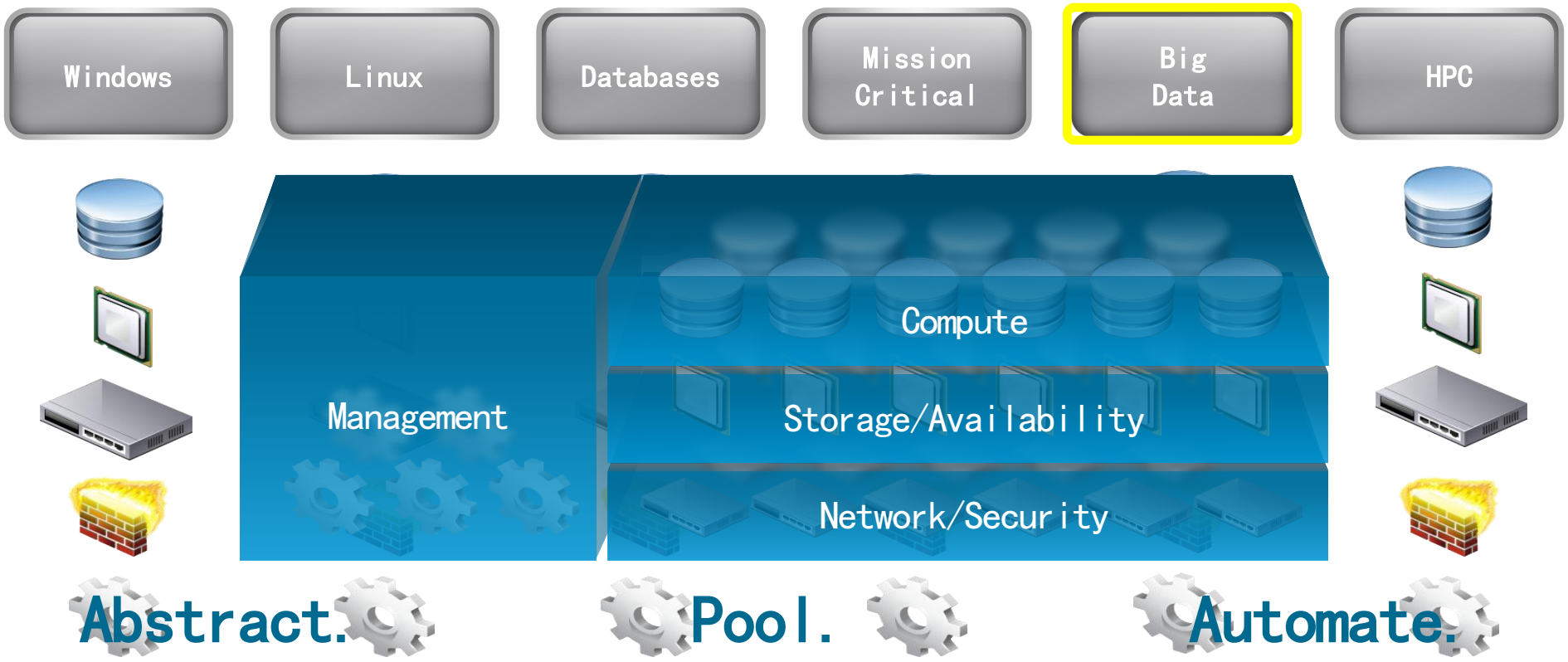
概念验证阶段

快速低成本的验证大数据技术带来的价值

vmware

巧妇难为无米之炊，没有机器怎么办？

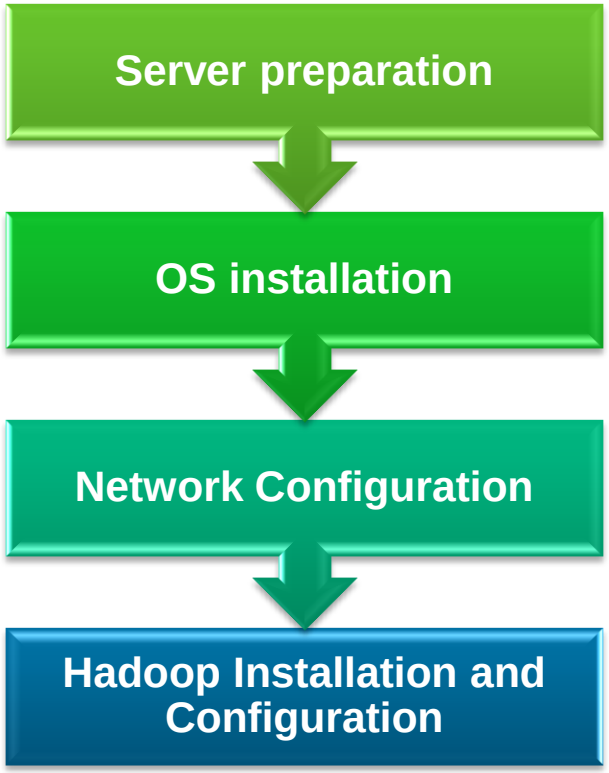
利用软件定义数据中心的共享资源，零起步成本

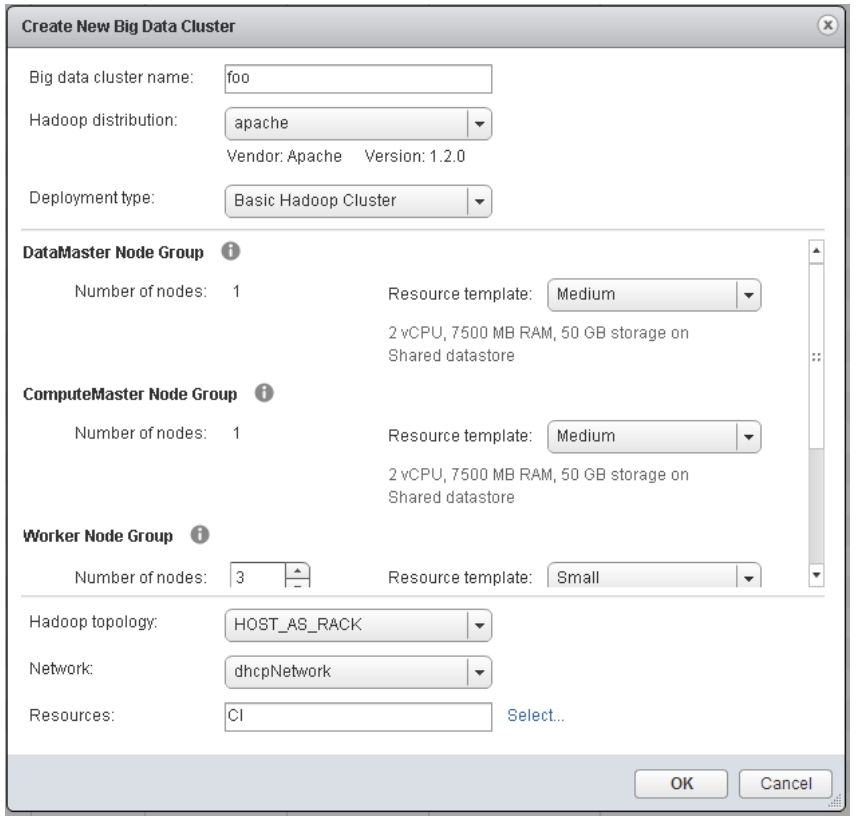


vmware

时间紧任务重，人手不够怎么办？

vSphere Big Data Extension帮你快速简便部署，让你全力关注业务





自动而不失控制

```
{
  "name": "master",
  "roles": [
    "hadoop_namenode",
    "hadoop_jobtracker"
  ],
  "instanceNum": 1,
  "instanceType": "LARGE",
  "cpuNum": 2,
  "memCapacityMB":4096,
  "storage": {
    "type": "SHARED",
    "sizeGB": 20
  },
  "haFlag":"on",
  "rpNames": [
    "rp1"
  ]
}
```

Storage configuration
Choice of shared or local

High Availability option

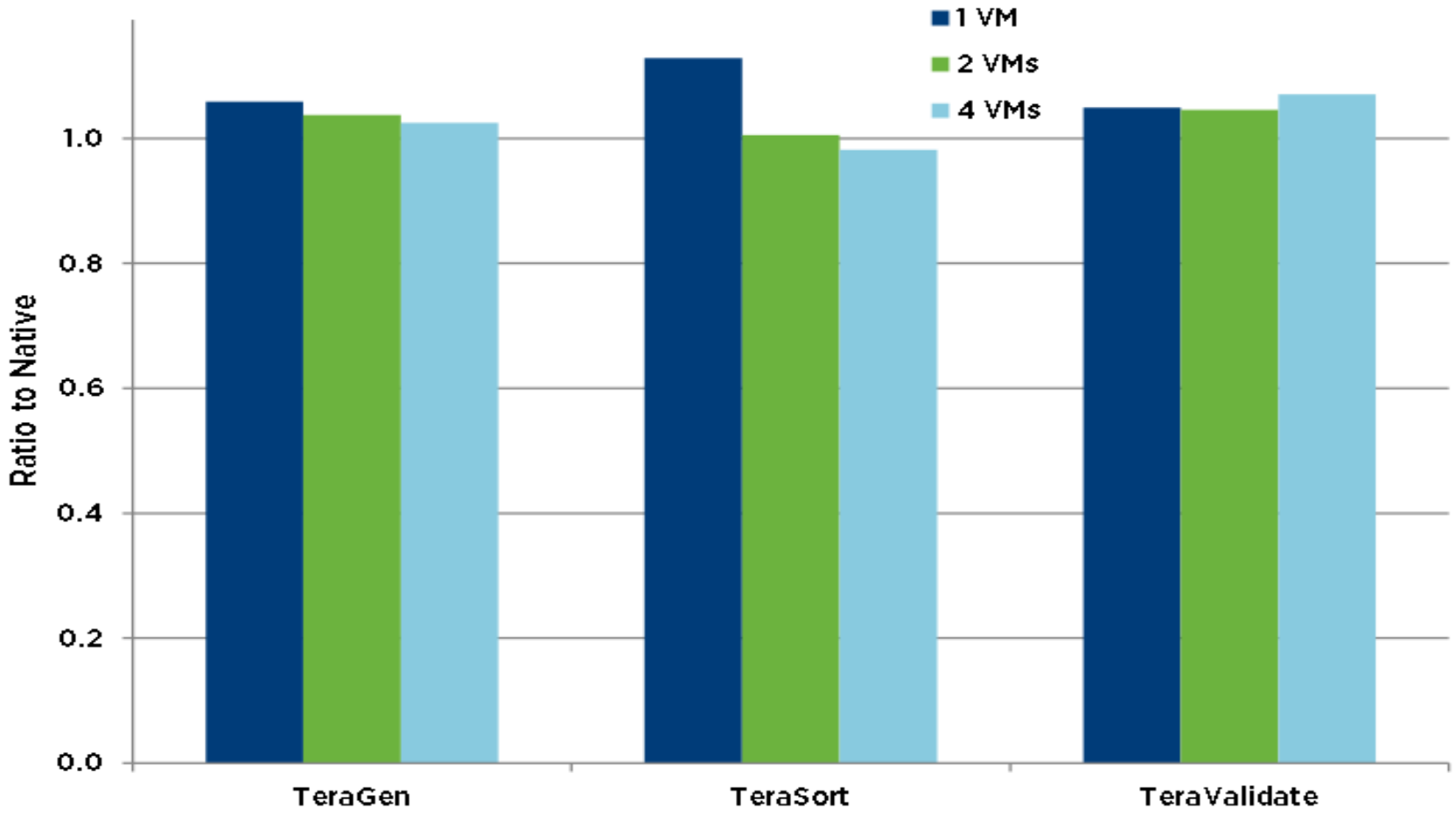
```
{
  "name": "data",
  "roles": [
    "hadoop_datanode"
  ],
  "instanceNum": 3,
  "instanceType": "MEDIUM",
  "cpuNum": 2,
  "memCapacityMB":2048,
  "storage": {
    "type": "LOCAL",
    "sizeGB": 50
  }
  "placementPolicies": {
    "instancePerHost": 1,
    "groupRacks": {
      "type": "ROUNDROBIN",
      "racks": ["rack1", "rack2", "rack3"]
    }
  }
}
```

Number of nodes and resource configuration

VM placement policies

vmware

简便而不失性能



Source: http://www.vmware.com/resources/techresources/10360

vmware

BDE可与第三方管理工具无缝集成

只用BDE

BDE 部署虚机并且
自动安装配置Hadoop软件

两步部署

BDE 部署配置好操作系统，
网络，存储的虚机

使用Hadoop厂商的管理工具
安装Hadoop软件

集成方案

BDE部署虚机然后调用
Hadoop厂商的管理工具的API

Hadoop厂商的管理工具
在后台安装Hadoop软件

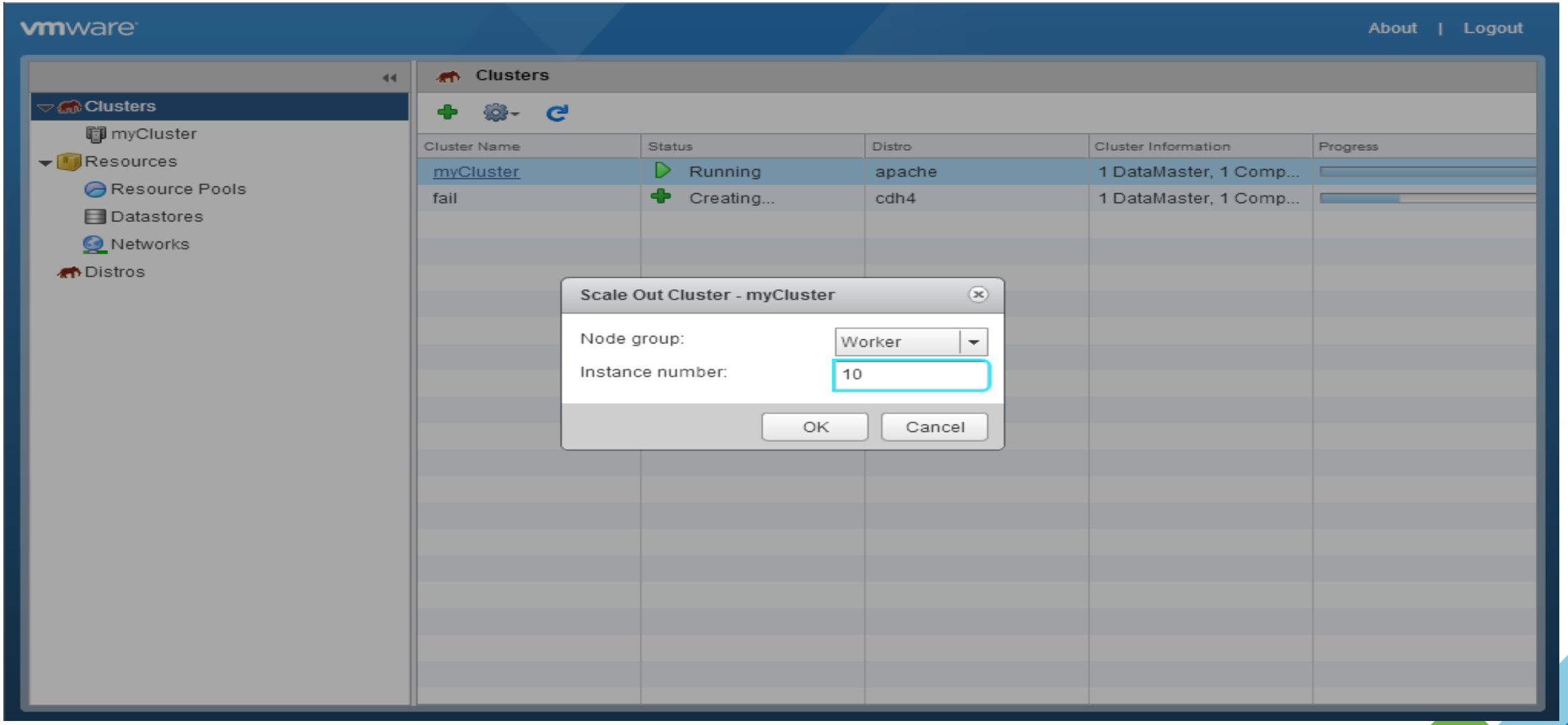
生产应用阶段

满足应用SLA，满足系统扩容需求

vmware

业务负载不明确，怎么规划集群硬件？

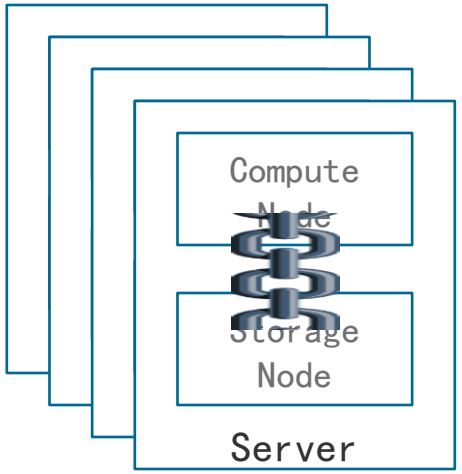
按需弹性伸缩，以变应变



vmware

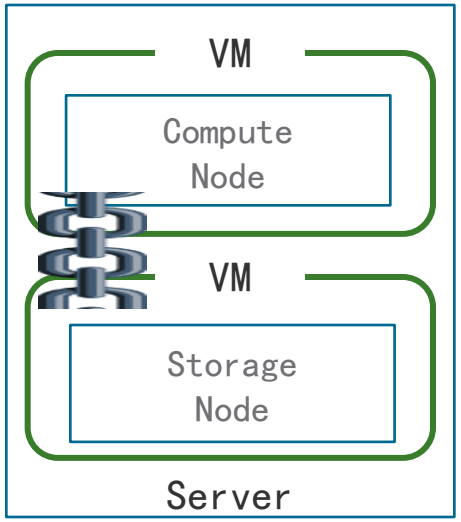
资源消耗不平衡，闲置浪费怎么办？

物理部署Hadoop集群无法充分利用资源

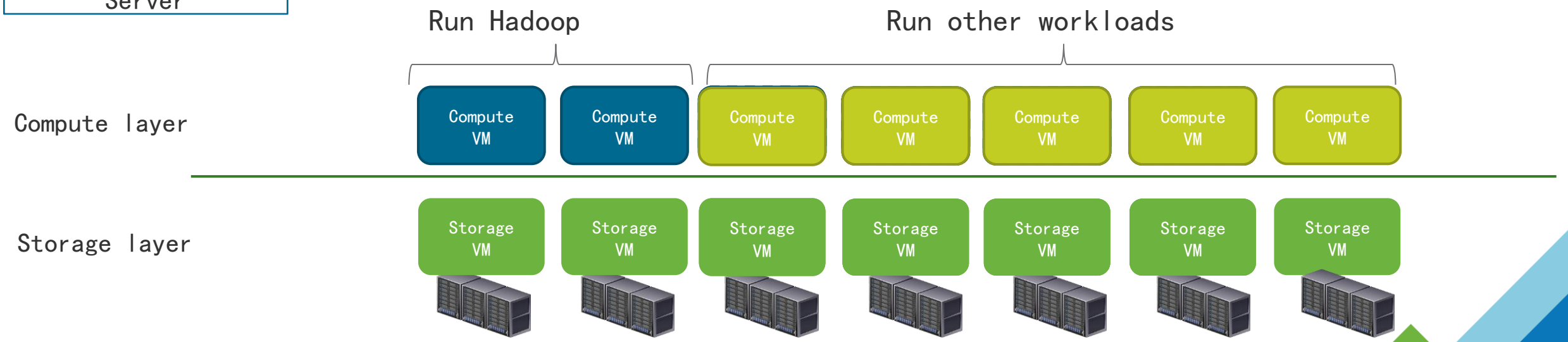


- 特定的计算存储比例导致缺乏灵活性和低利用率
  - 计算和存储容量比决定于硬件规格
  - 但不同的业务有不同的特性（计算密集或存储密集）
  - 僵化的基础设施导致浪费
  - 集群越大越是如此
- 产生的结果是
  - Hadoop集群的整体CPU利用率低
  - 有人为不同的集群采购不同的硬件

灵活的虚拟化Hadoop提高资源利用率



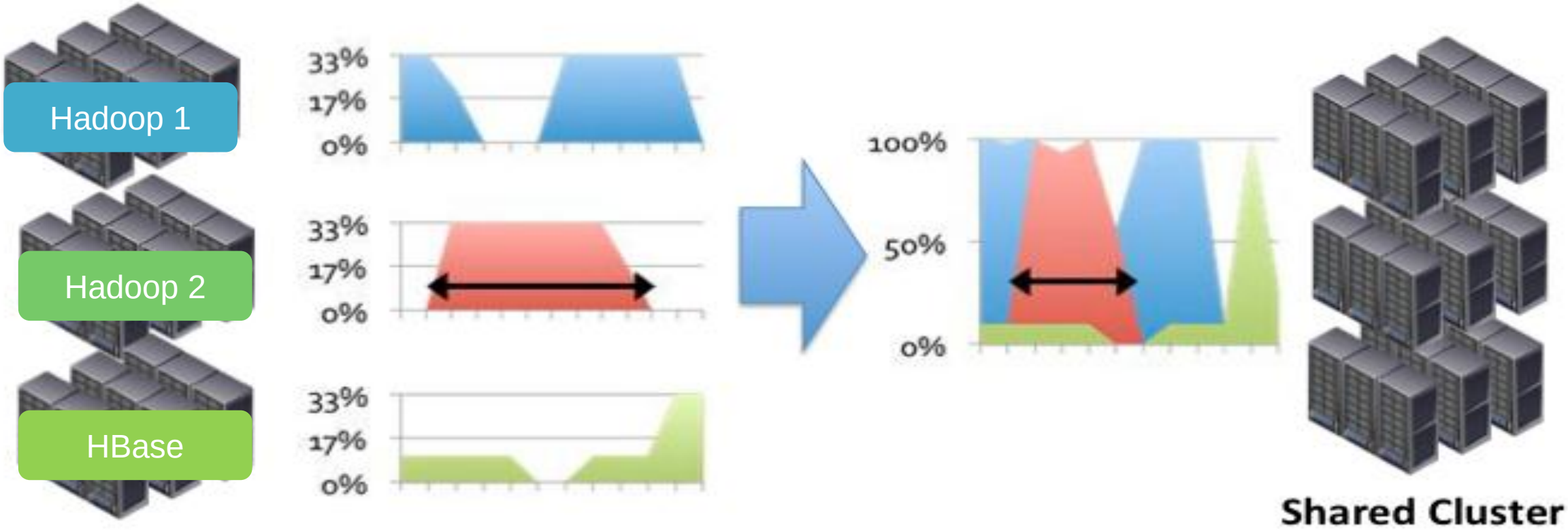
- 把计算和存储分开到不同的虚机里
- 无状态的计算层可以灵活的按需伸缩
- 仍然可以确保Data locality
- 多余的计算资源可以用于其他工作



vmware

业务具有峰谷特性，高峰扩容怎么办？

共享资源，按需快速敏捷扩容



- Consolidated cluster has access to entire pool of physical resources
- Take advantage of multi-tenancy to increase utilization during non-peak hours
- Reduce latency on priority jobs on consolidated cluster

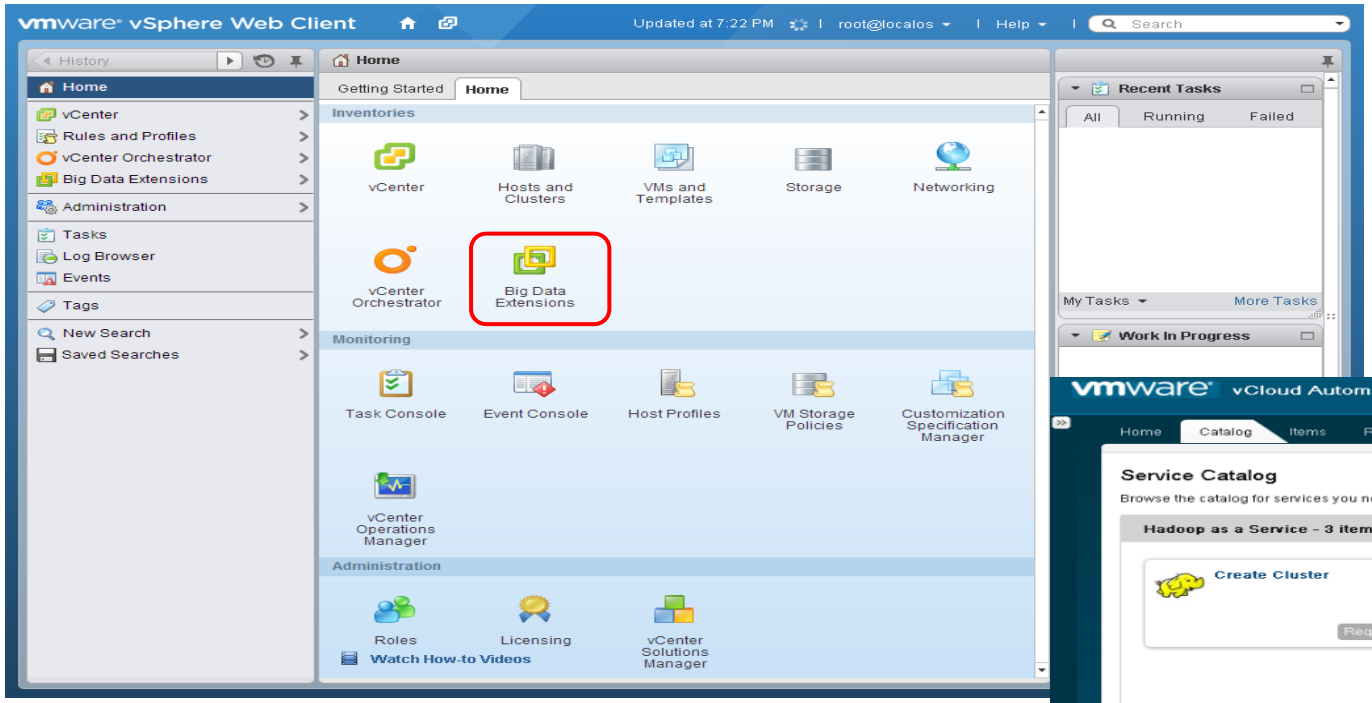
Hadoop即服务阶段

快速敏捷高效的满足各个业务的差异化需求

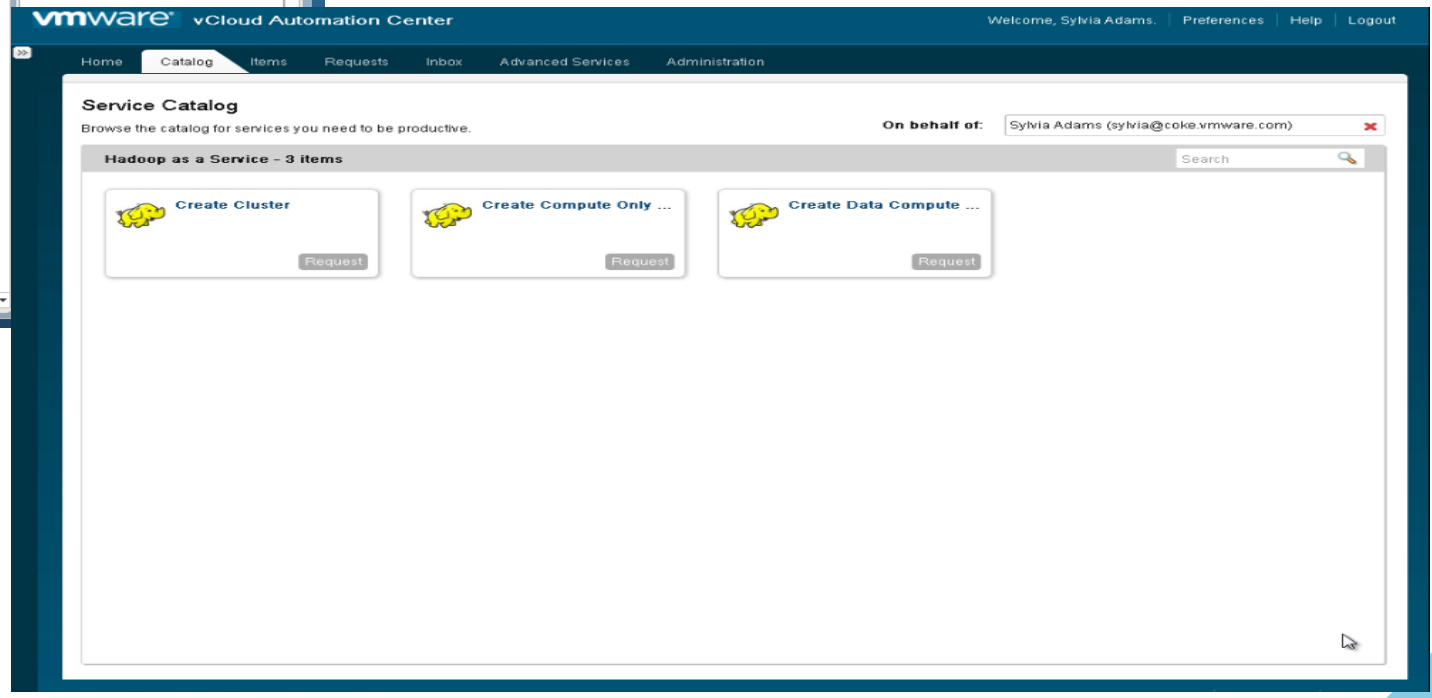
vmware

怎么建设、运维自助服务平台？

自助服务平台实现角色分工



通过vCenter管理基础设施

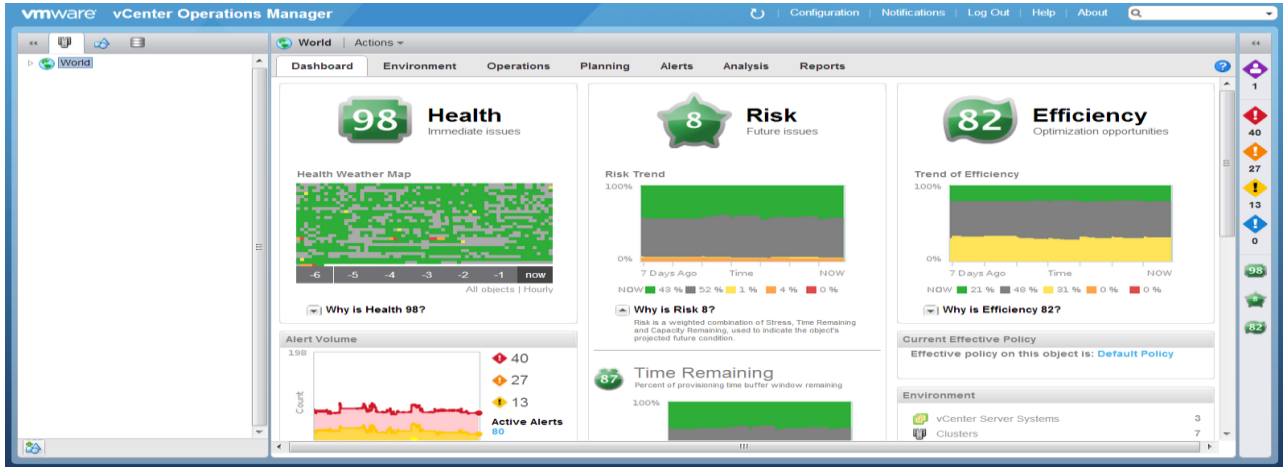


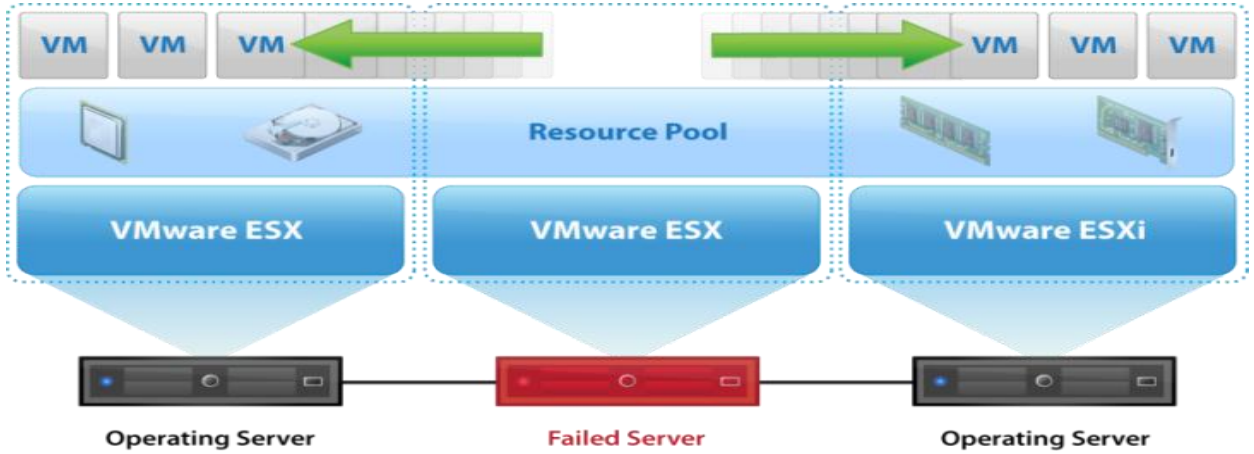
通过vRealize Automation Center
实现Hadoop as a Service

vmware

使用现有工具管理基础设施

- vRealize Operations Manager
  - 实现系统全面监控
  - 智能自动分析管理
  - 基于预测主动运维





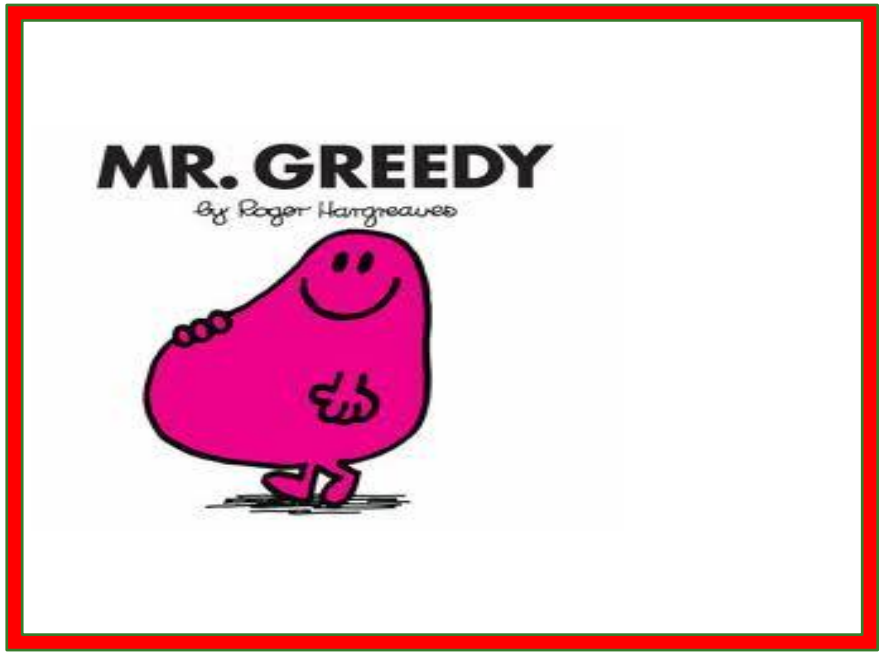
vSphere vMotion, HA, FT

- 消除计划或非计划宕机时间
- 检测失效自动恢复

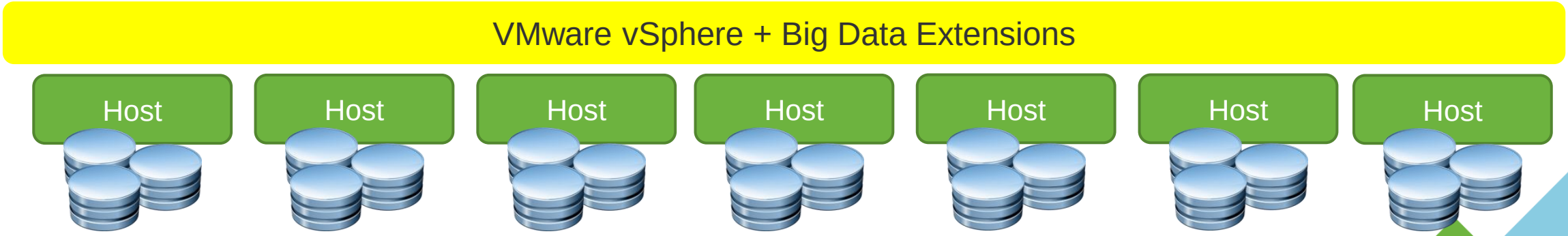
vmware

众口难调，各个业务要求不同怎么办？

共享硬件设施条件下的完全隔离



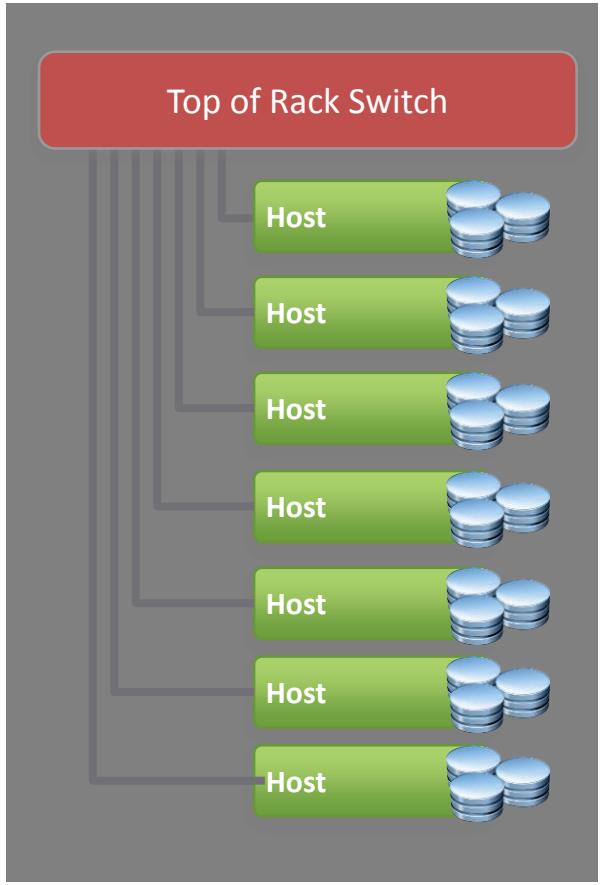
- 资源隔离
  - 约束资源占用
  - 预留资源保证需求
- 版本隔离
  - 允许不同供应商，不同版本共存
- 安全隔离
  - 保证不同用户的私有数据
  - 运行时的完全隔离



vmware

灵活地选择存储

使用本地硬盘



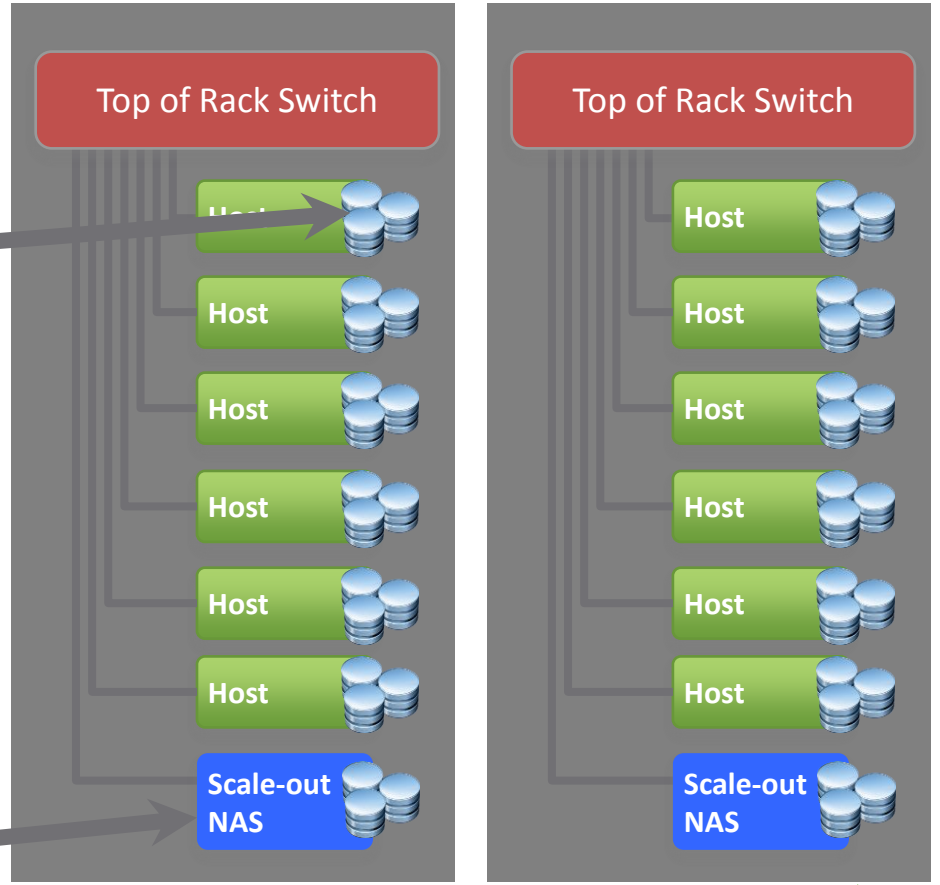
High Performance 10GBE
Switch per Rack

16-24 core server
12-24 SATA 2-4TB Disks
10 GbE adapter
iSCSI/NFS for Shared
Storage for vMotion etc,…

使用Scale-out NAS

本地硬盘
用于临时存储

Isilon
Scale-out NAS
用于持久数据



Temp Data

Shared Data

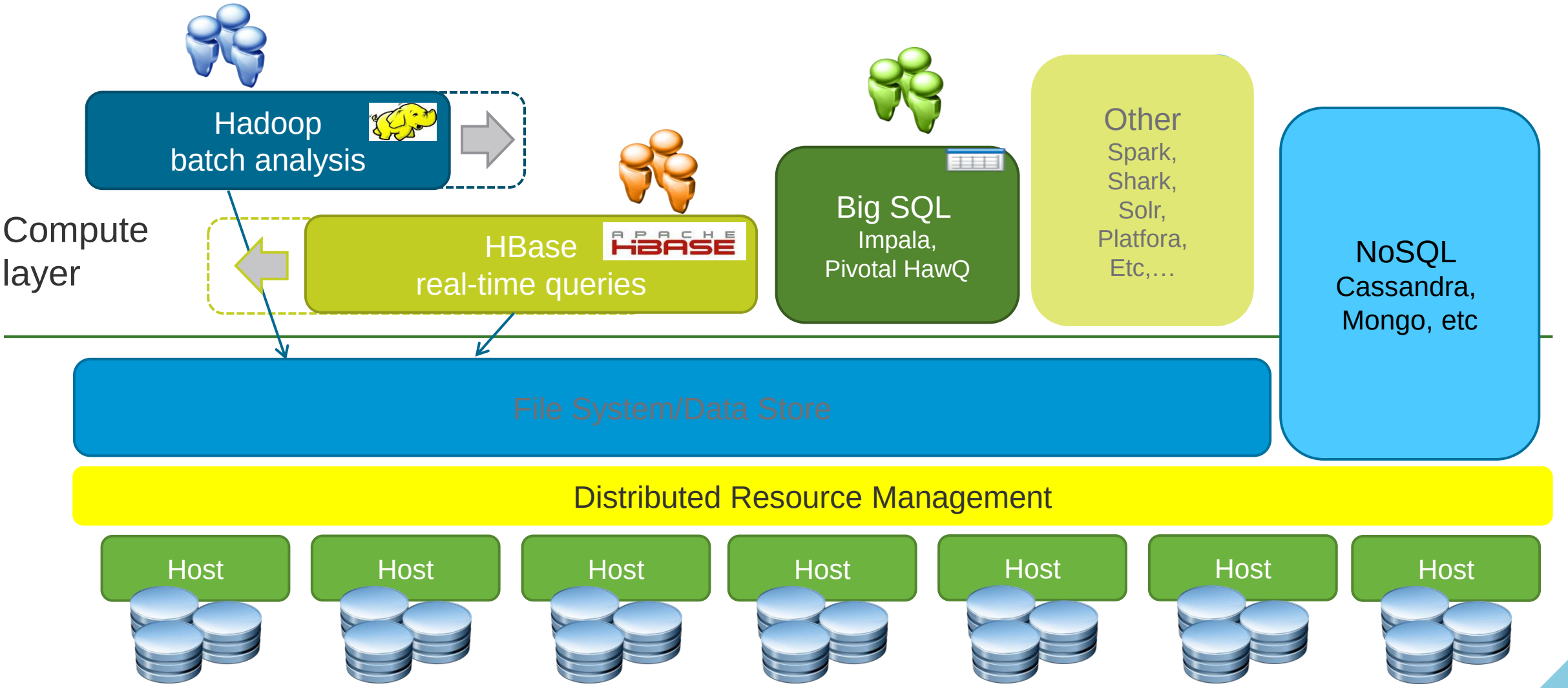
vmware

灵活地选择各种发布版和应用堆栈

- 内置支持各大主流Hadoop发布版的常用模块

- 通过与Hadoop供应商管理工具集成提供全栈支持

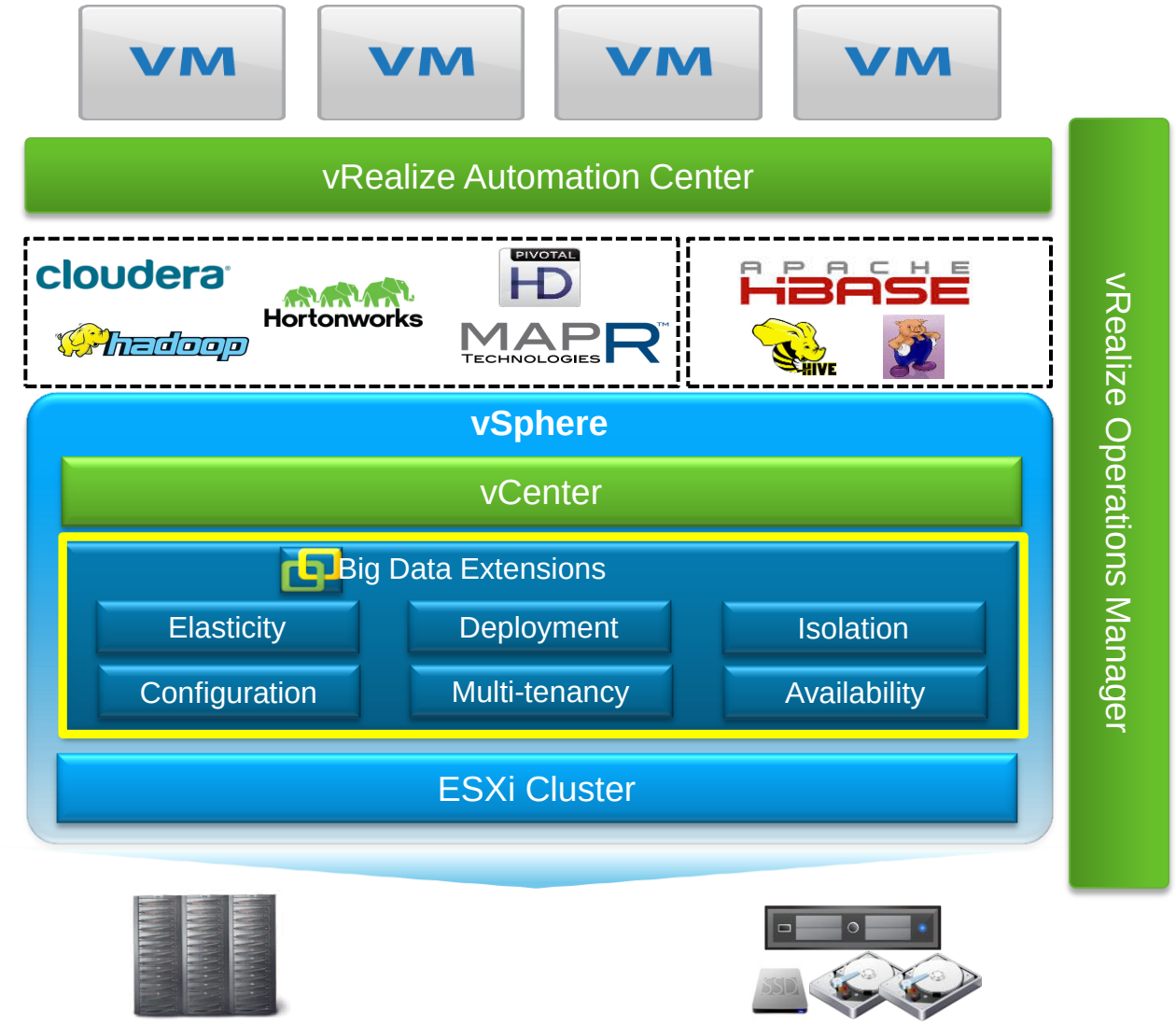
运行各种应用的同一平台



总结

vmware

Vmware为大数据应用提供平台



vmware

Big Data Extensions

操作简便保证性能

- 快速部署
- 自助服务工具
- 内置性能调优最佳实践

资源利用最大化

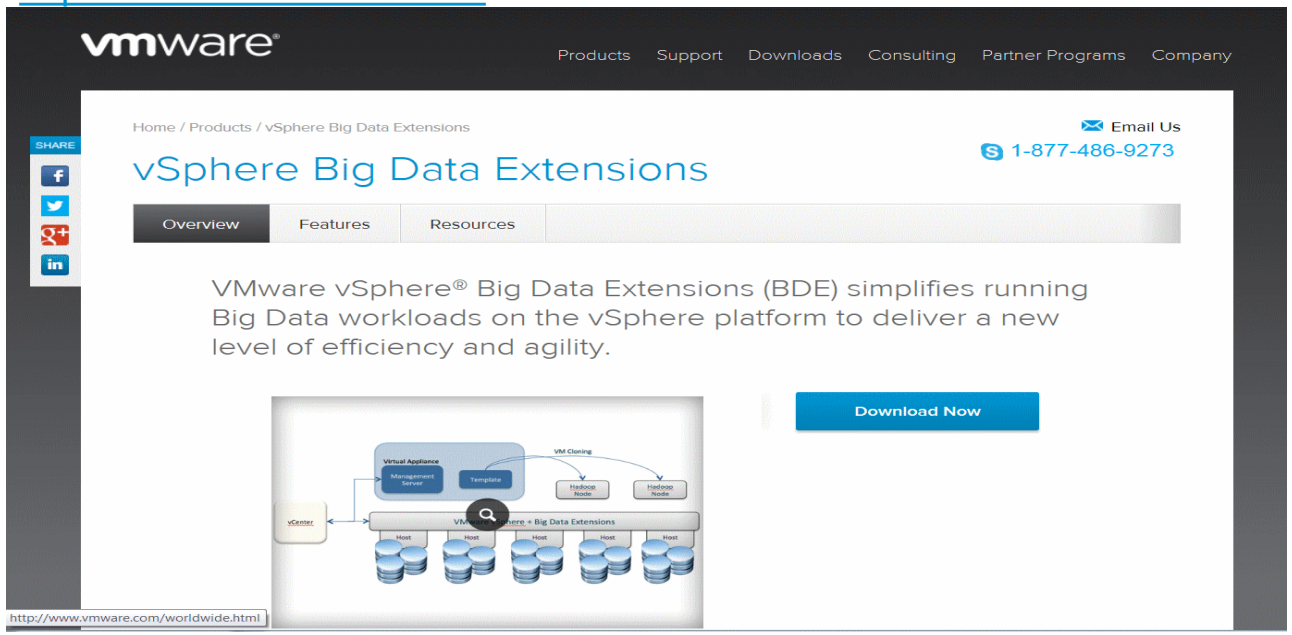
✓ 多租户
✓ 弹性伸缩
✓ 基于虚机的隔离
✓ 提高资源利用率

高度可扩展

✓ 更多部署形式选择
✓ 大规模应用下保证灵活度
✓ 控制成本
✓ 利用vSphere功能

VMware vSphere BDE和相关资源

- VMware vSphere BDE web site
  - http://www.vmware.com/bde



- Virtualized Hadoop Performance with VMware vSphere 5.1
  - http://www.vmware.com/resources/techresources/10220
- Benchmarking Case Study of Virtualized Hadoop Performance on vSphere 5
  - http://vmware.com/files/pdf/VMW-Hadoop-Performance-vSphere5.pdf
- Hadoop Virtualization Extensions (HVE) :
  - http://www.vmware.com/files/pdf/Hadoop-Virtualization-Extensions-on-VMware-vSphere-5.pdf
- Apache Hadoop High Availability Solution on VMware vSphere 5.1
  http://vmware.com/files/pdf/Apache-Hadoop-VMware-HA-solution.pdf

vmware

谢 谢！

vmware®