ANALYTICS BIG DATA SUMMIT

Reference Architecture and Best Practices for Virtualizing Hadoop Workloads

Justin Murray

VMware

Agenda

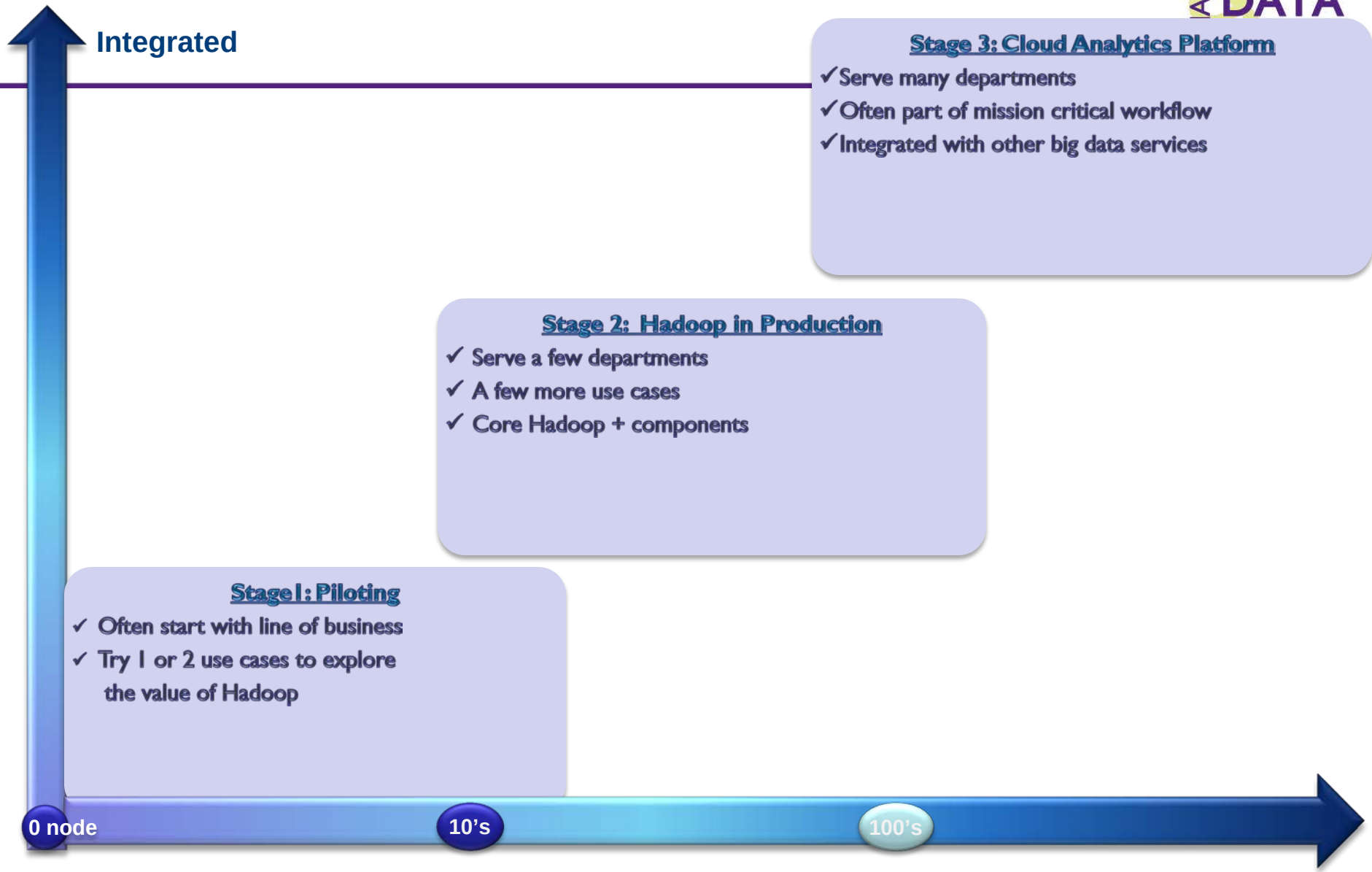
ANALYTICS BIG DATA SUMMIT

- The Hadoop Journey
- Why Virtualize Hadoop?
- Elasticity and Scalability
- Performance Tests
- Storage Reference Architectures
- Isilon Architecture and Benefits
- vSphere Big Data Extensions
- Conclusion and Q&A

The Customer Journey with Hadoop

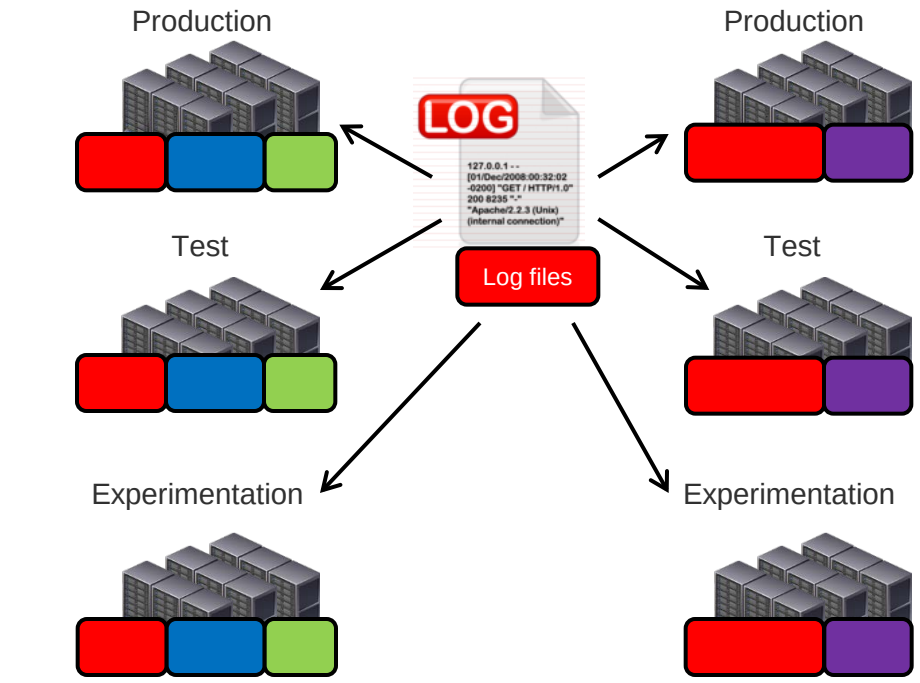
vmware

The Hadoop Journey



Why Virtualize Hadoop?

Customer Example: Enterprise Adoption of Hadoop



Dept A: recommendation engine

Dept B: ad targeting

SLA: Jobs complete in 15 minutes
Bandwidth limited to 30 nodes at peak

Issues:

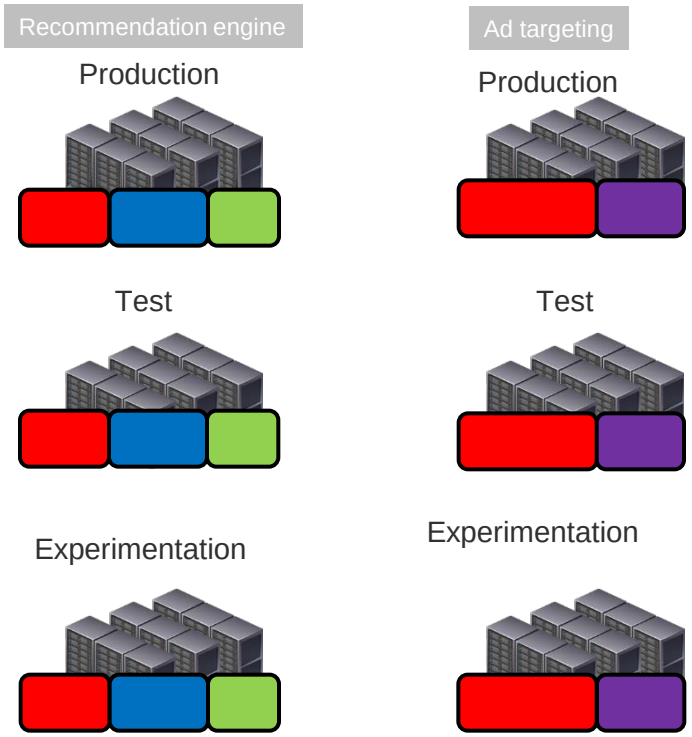
1. Multiple clusters to manage
2. Redundant common data in separate clusters
3. Peak compute and I/O resource is limited to number of nodes in each independent cluster

Transaction data

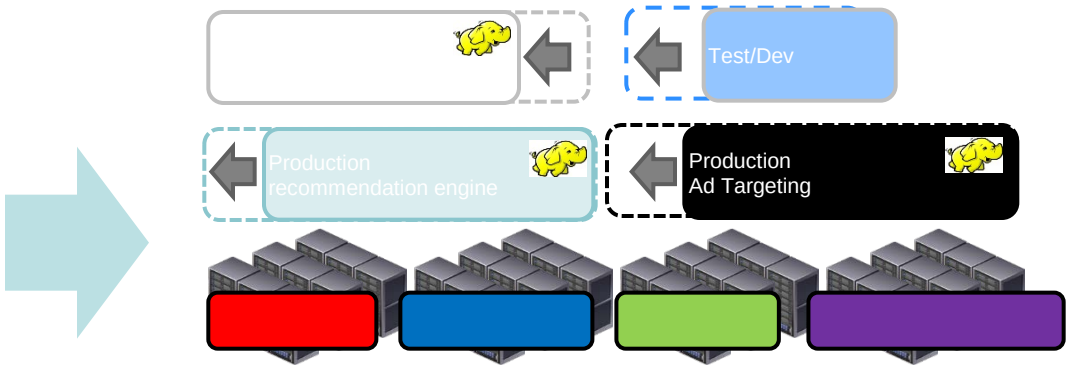
Social data

Historical cust behavior

What if you could…



One physical platform to support multiple virtual big data clusters



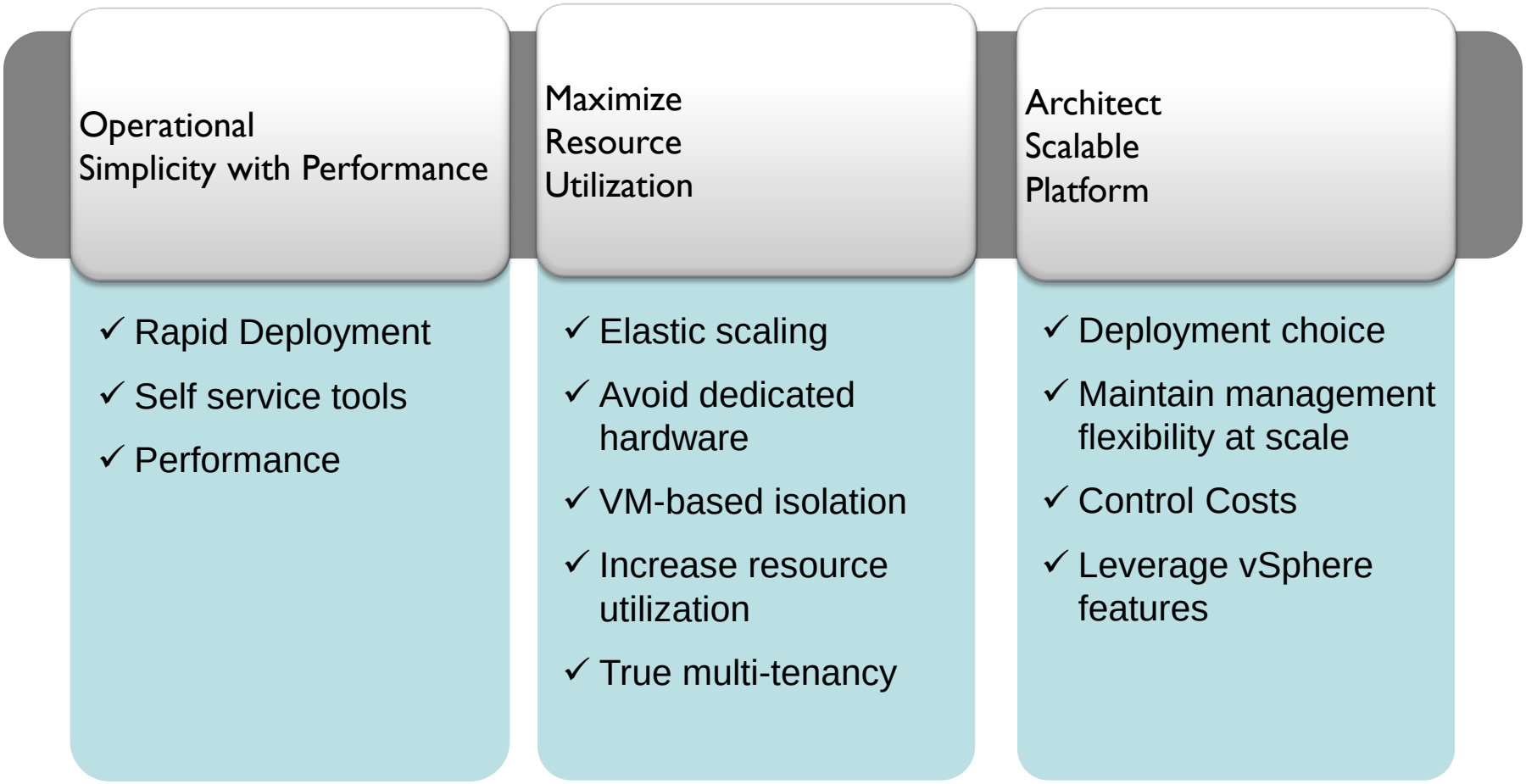
Without Virtualization

- Multiple copies of common data (e.g. historical data, log data etc.) in separate Hadoop clusters

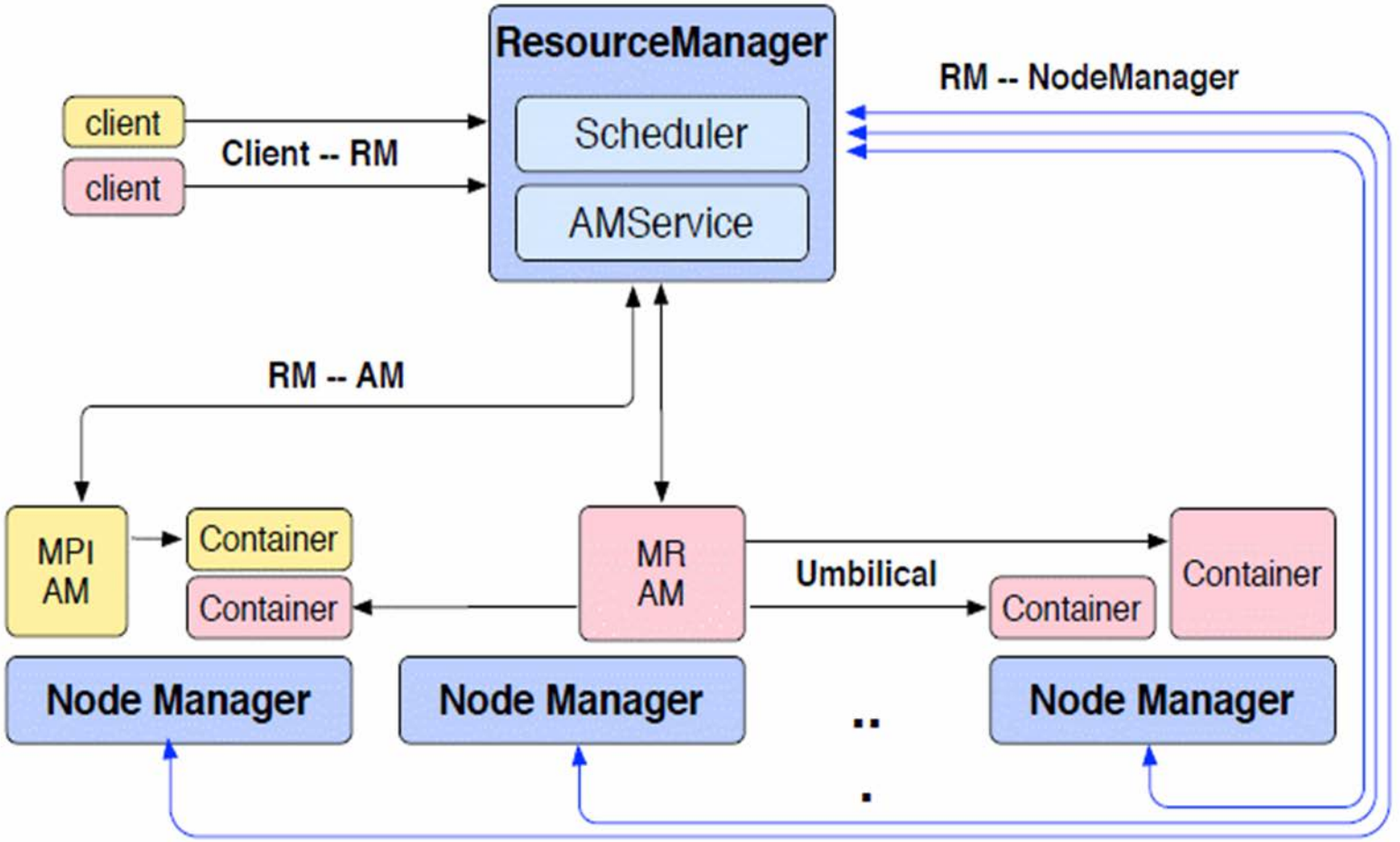
Consolidate and virtualize

- Single copy of common data results in less storage requirements while maintaining good isolation between different MapReduce clusters

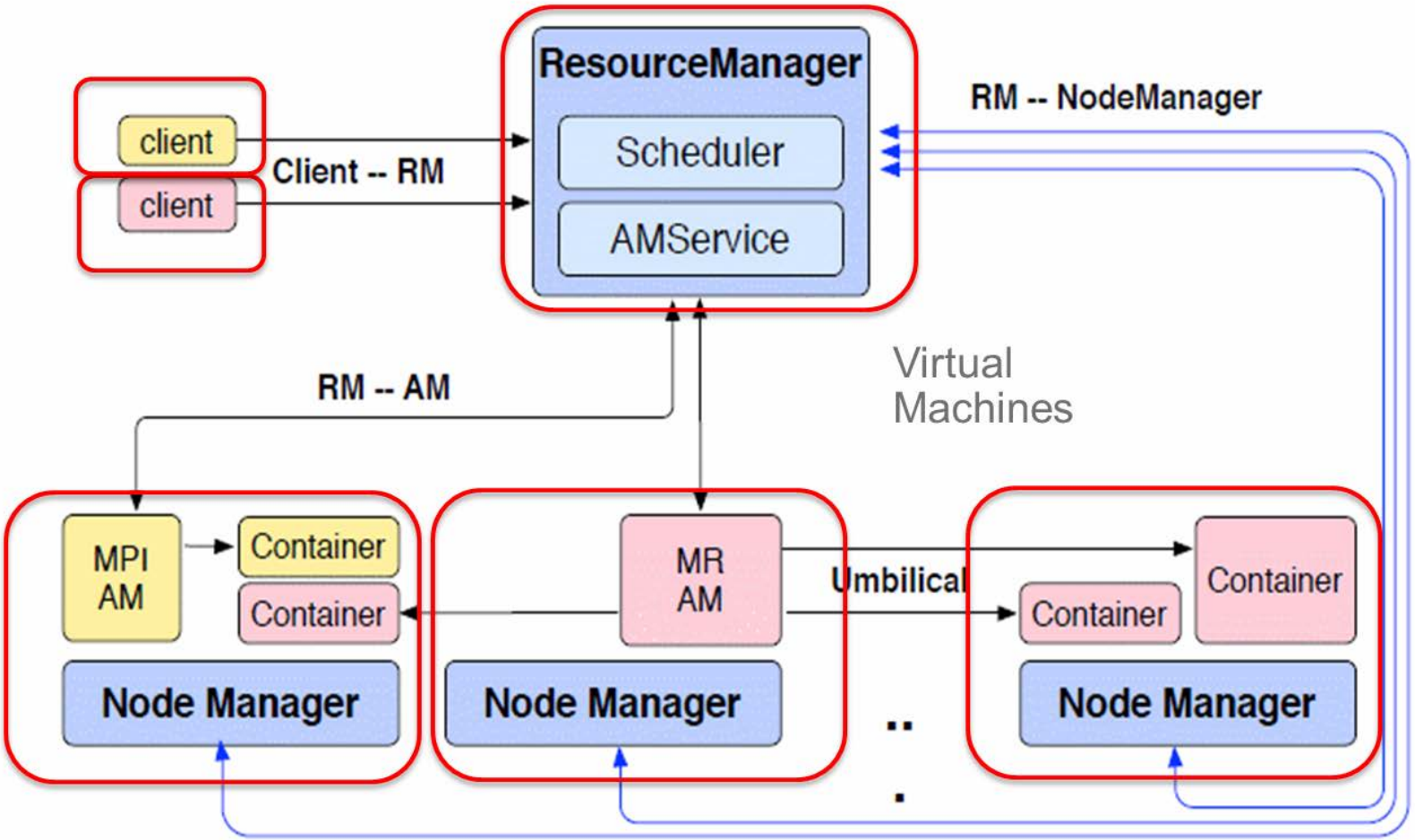
Big Data Extensions Value Propositions



Hadoop 2.0 – Yet Another Resource Negotiator

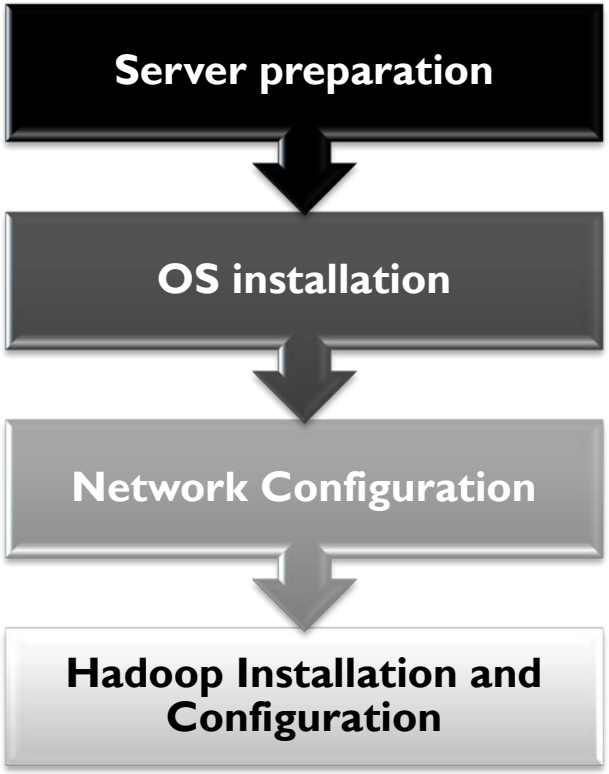


A Virtualized Hadoop 2.0 Cluster

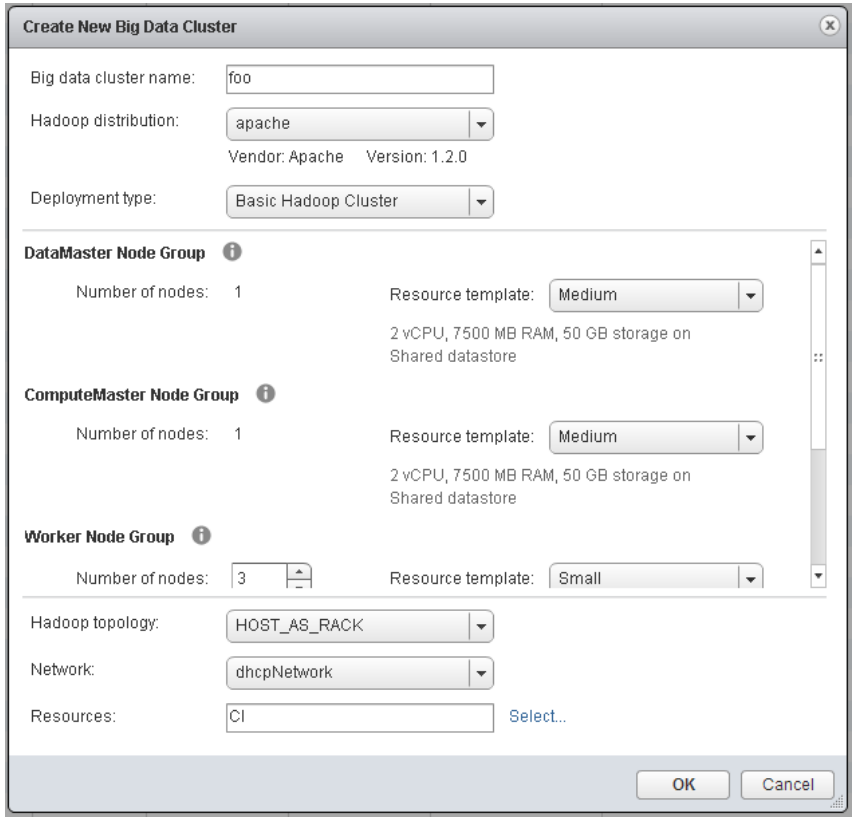


vSphere Big Data Extensions

- Deploy Hadoop Clusters in Minutes



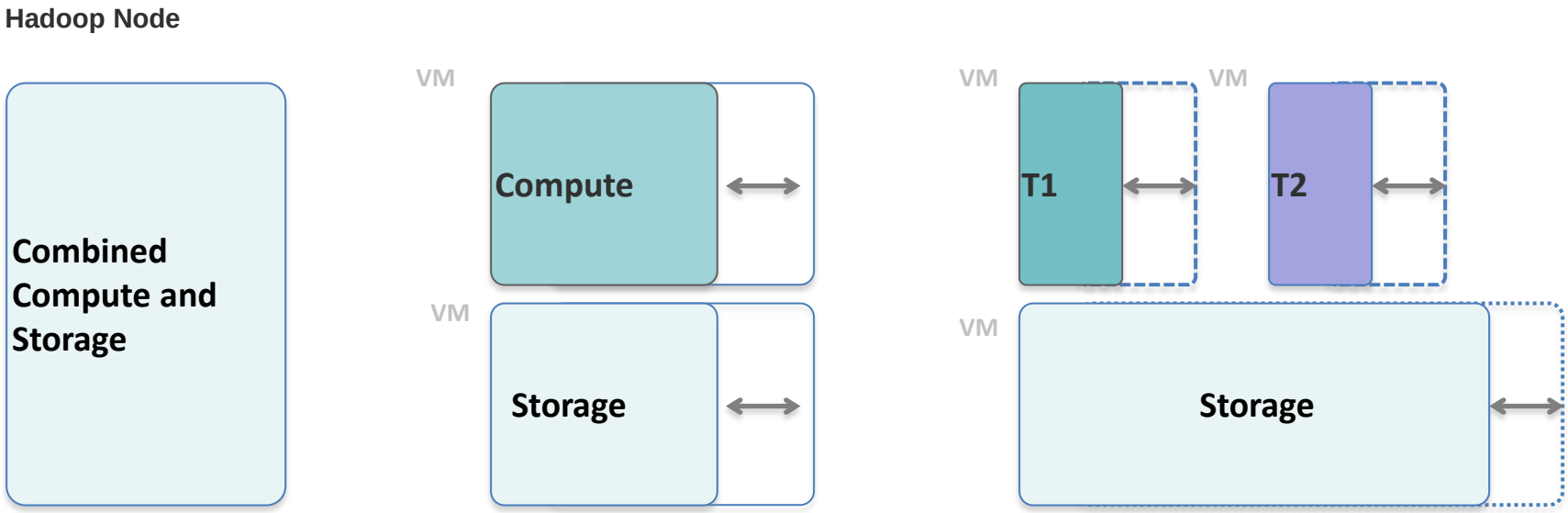
From a manual process



To fully automated, using the GUI

Elastic, Multi-Tenant Hadoop with Virtualization

VM



Unmodified Hadoop node in a VM

- VM lifecycle determined by Datanode
- Limited elasticity

Separate Compute from Storage

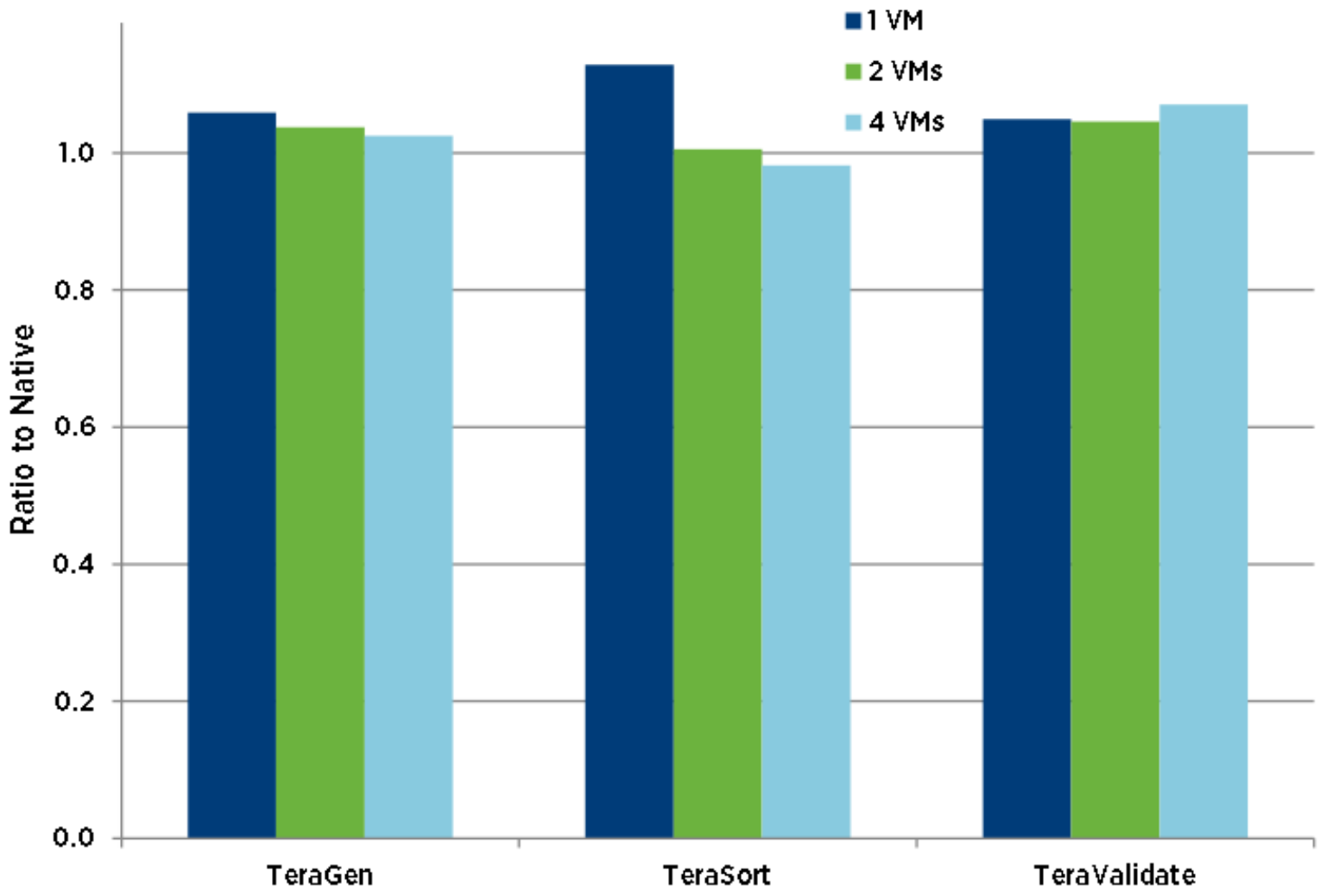
- Separate compute from data
- Stateless compute
- Elastic compute

Separate Virtual Compute Clusters per tenant

- Separate virtual compute
- Compute cluster per tenant
- Stronger VM-grade security and resource isolation

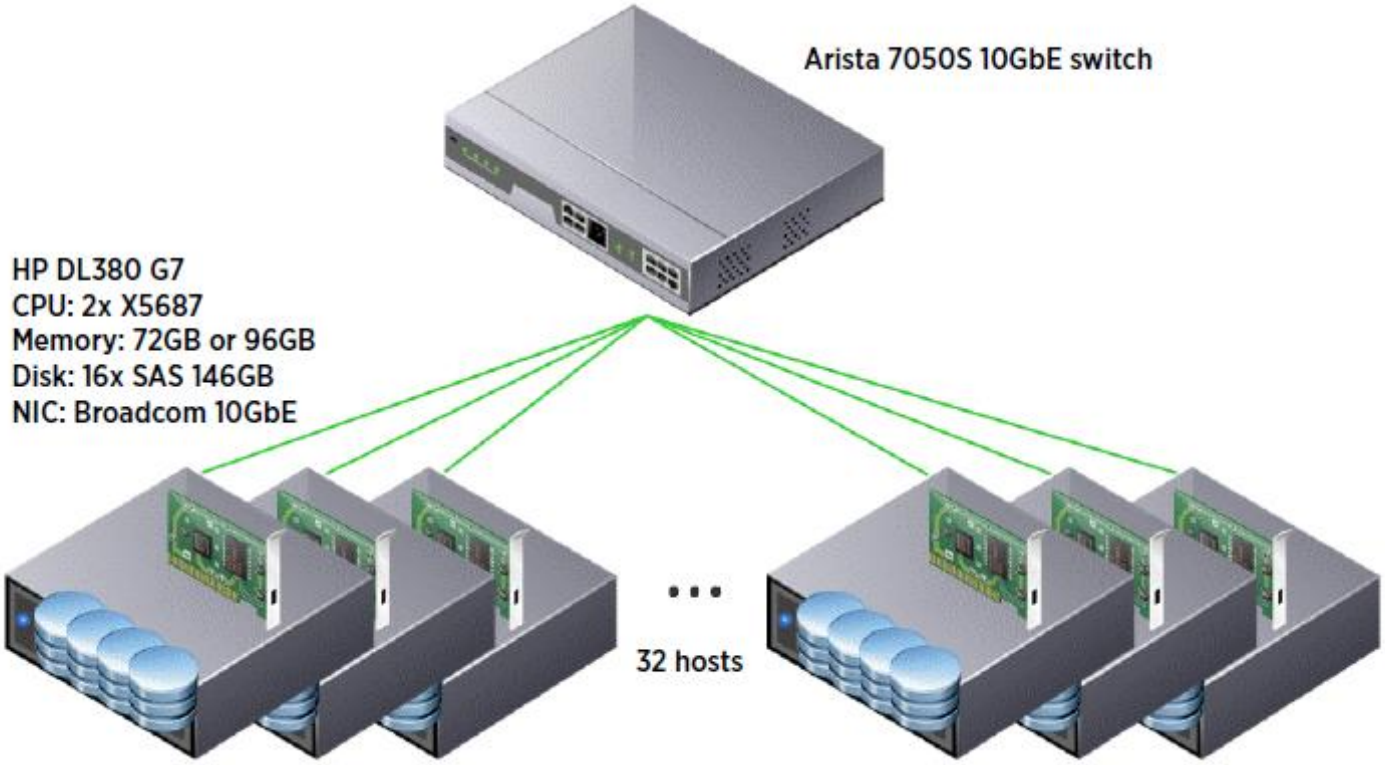
Performance and Reference Architectures

Native vs. Virtual, 32 hosts, 16 disks per host



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

Reference Architecture: 32-Server Performance Test

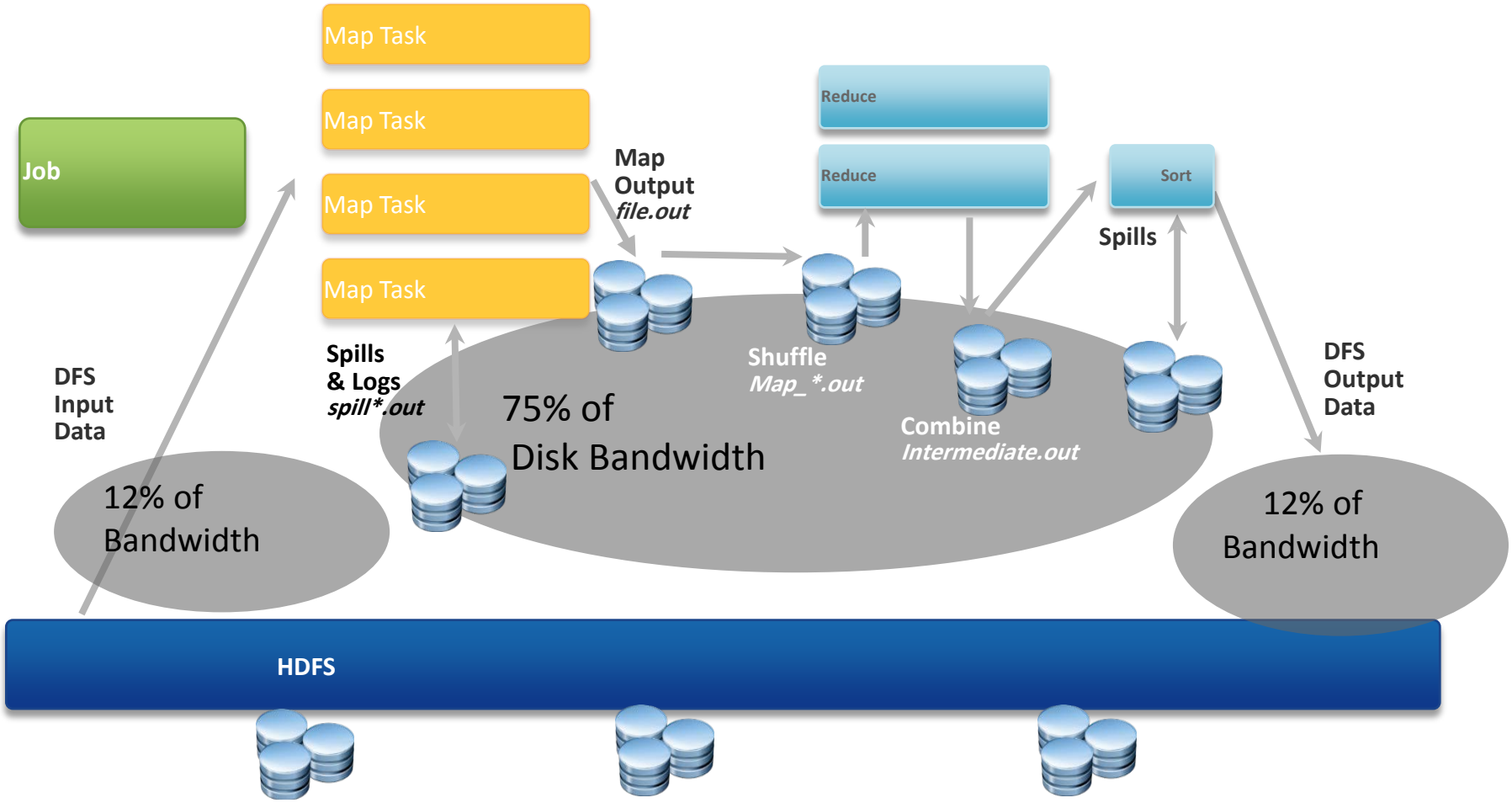


Up to four VMs per server
vCPUs per VM fit within socket size (e.g. 4 VMs x 4 vCPUs, 2 X 8)
Memory per VM - fit within NUMA node size

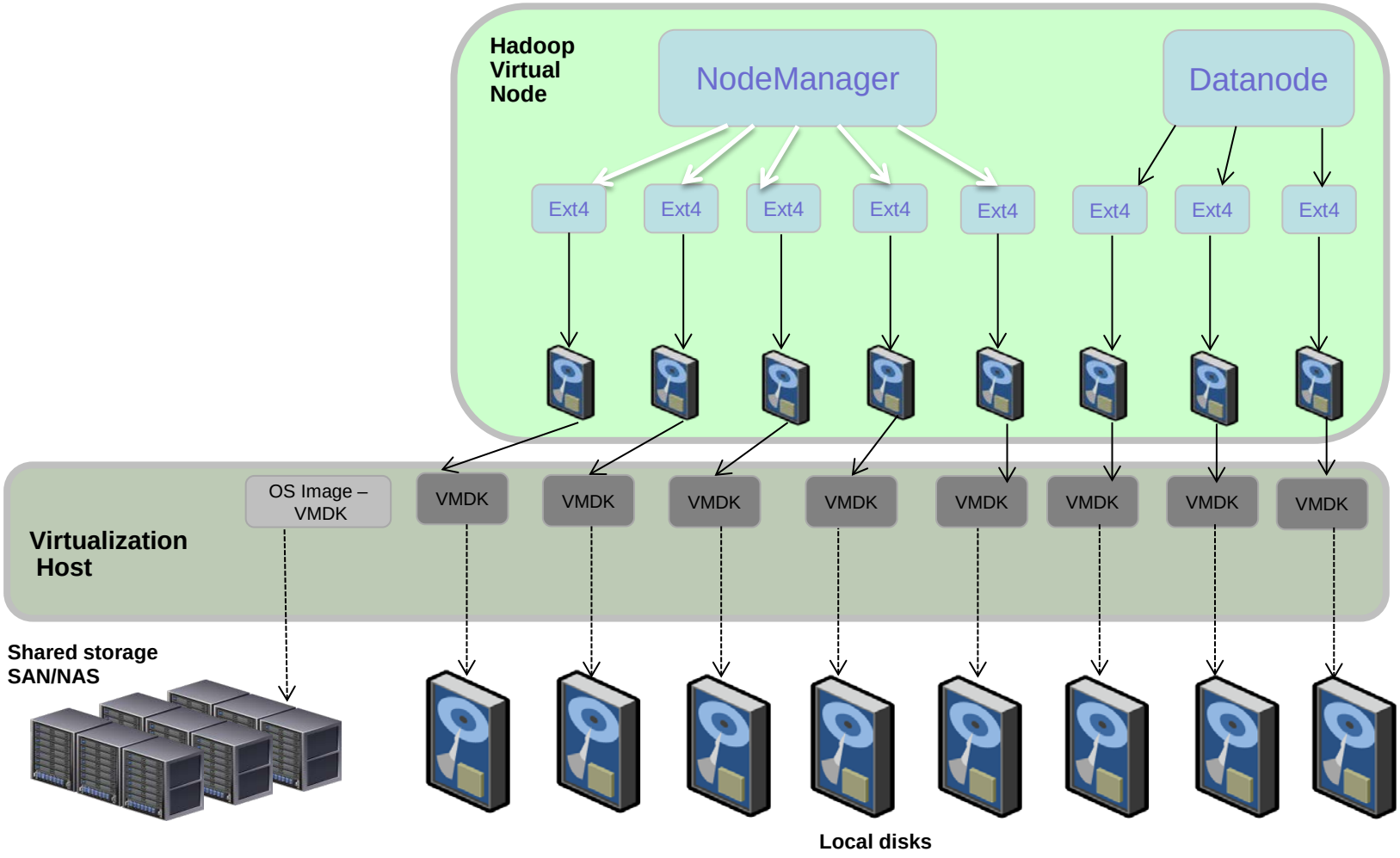
2013 Tests done using Hadoop 1.0

I/O Profile of a Hadoop MapReduce Job

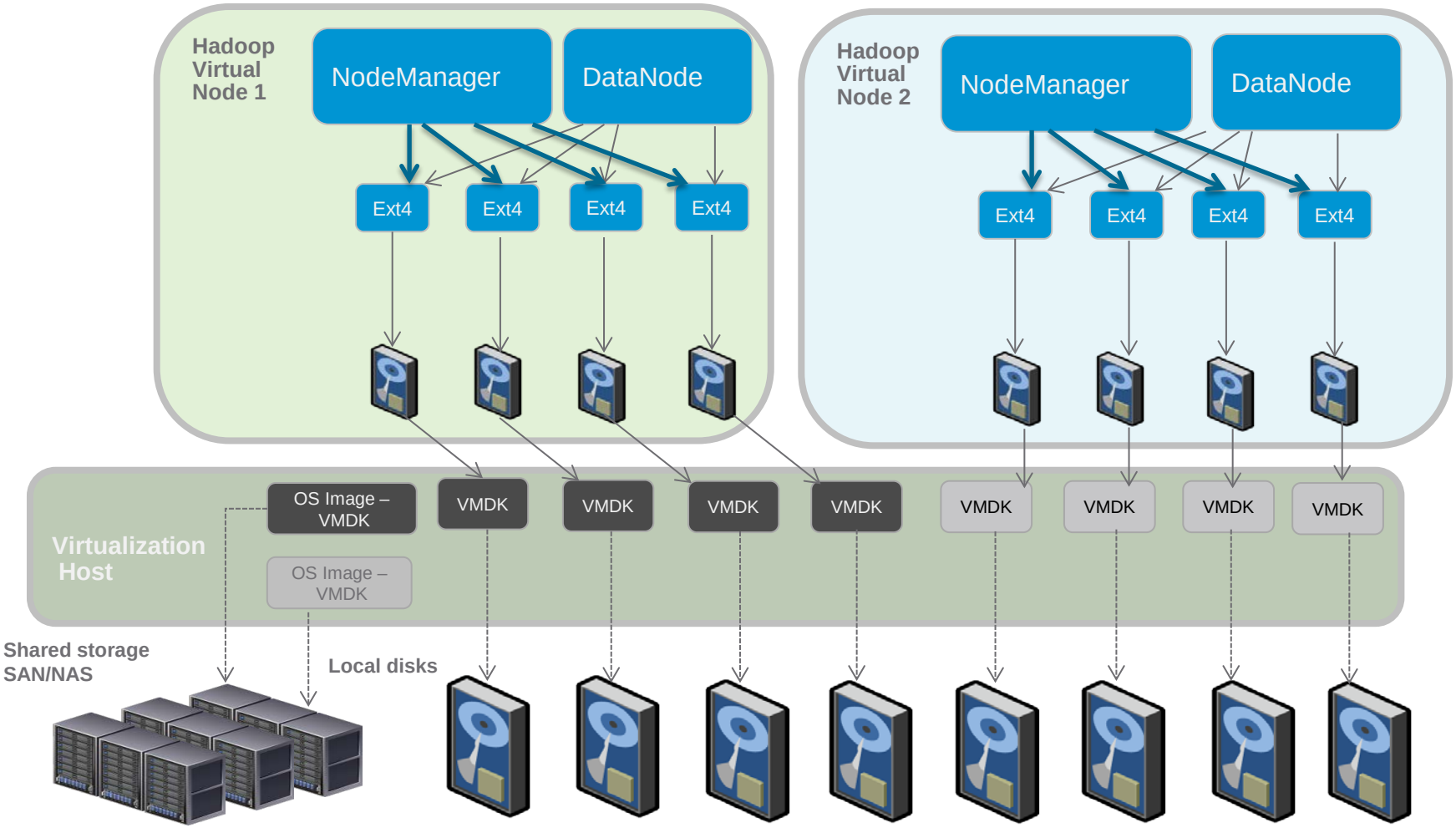
(TeraSort example application)



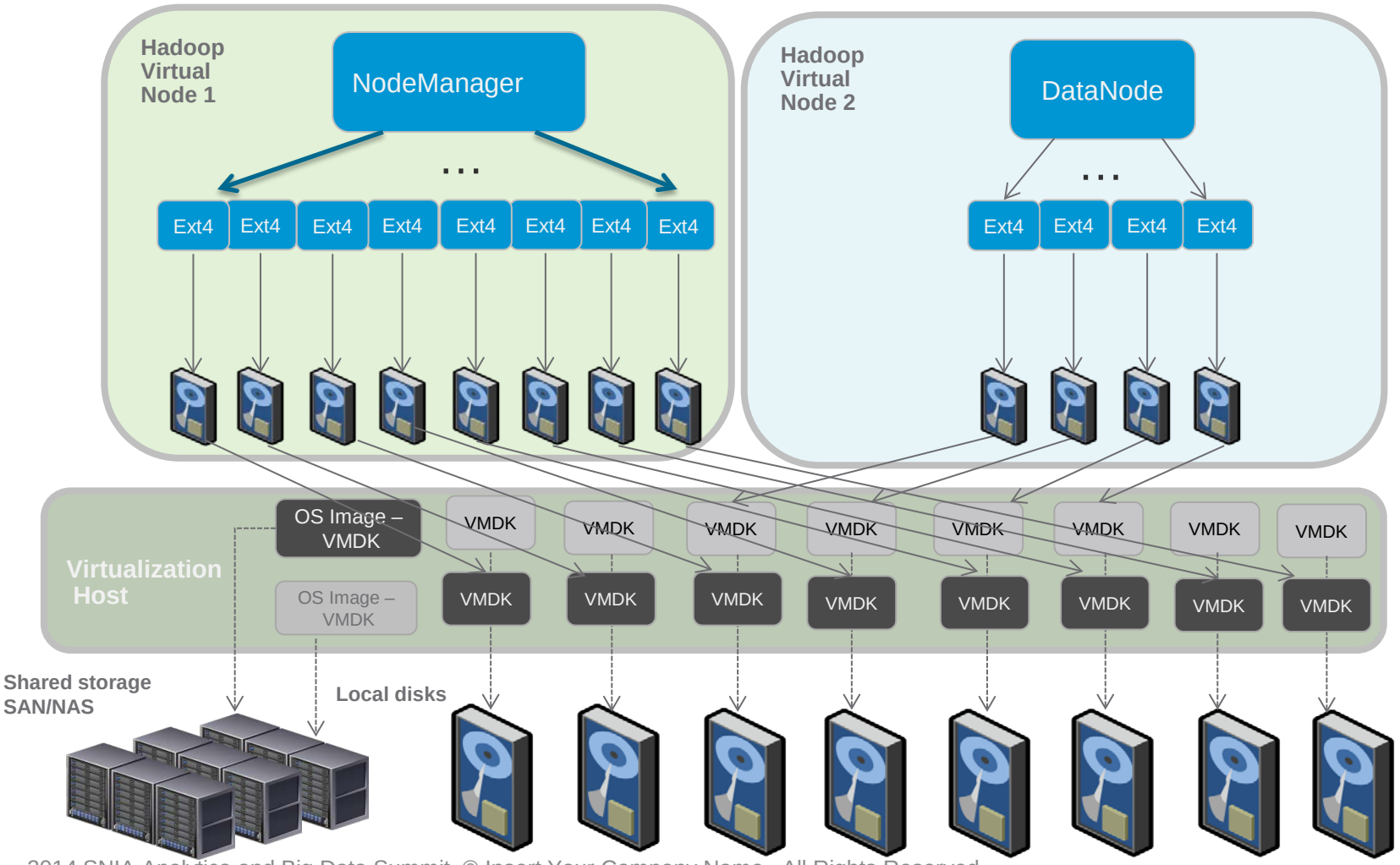
The Combined Model – Standard Deployment



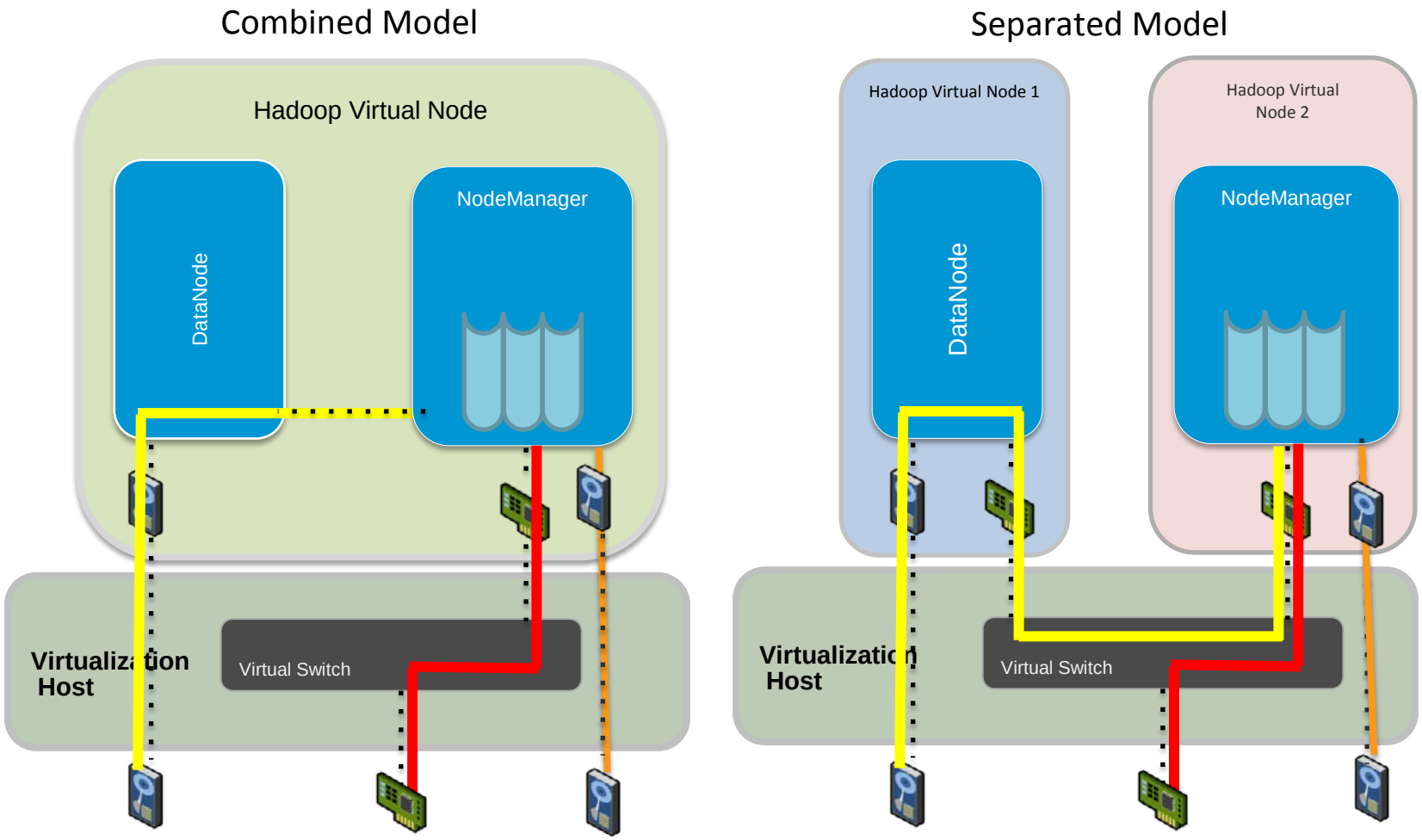
Combined Model – Two Virtual Machines on a Host Server



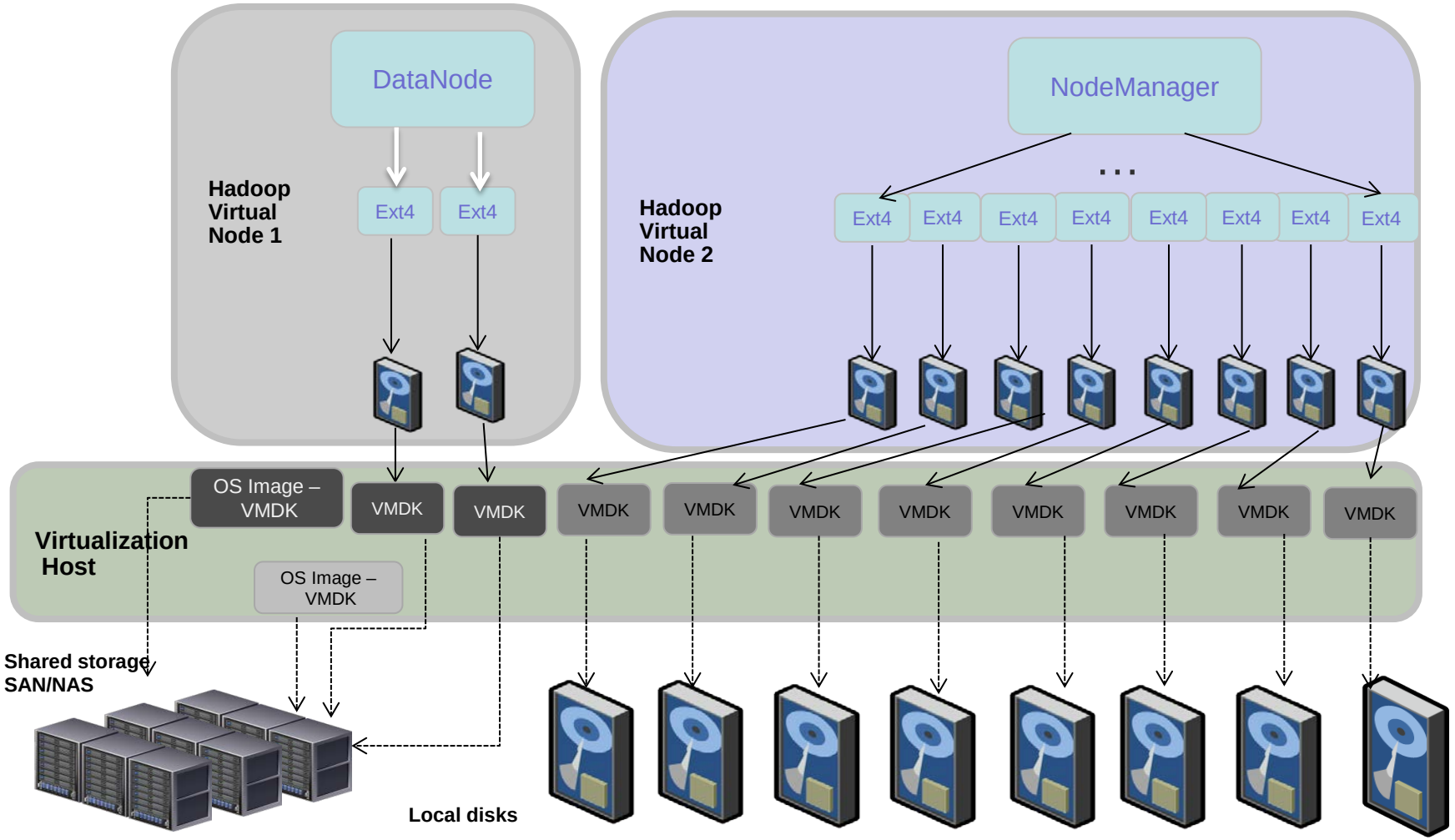
The Data-Compute Separation Deployment Model



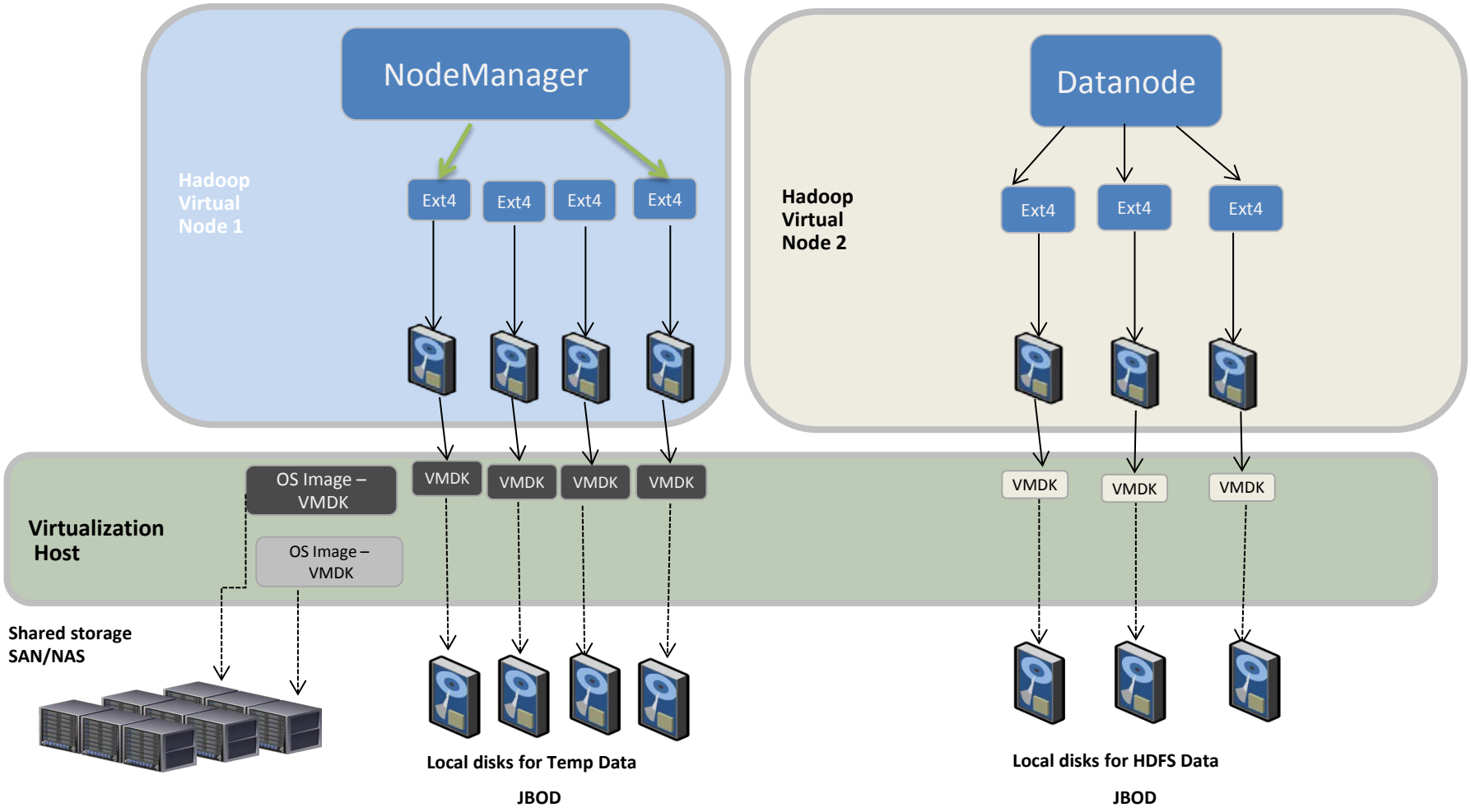
Data Paths: Combined vs Data-Compute Separation



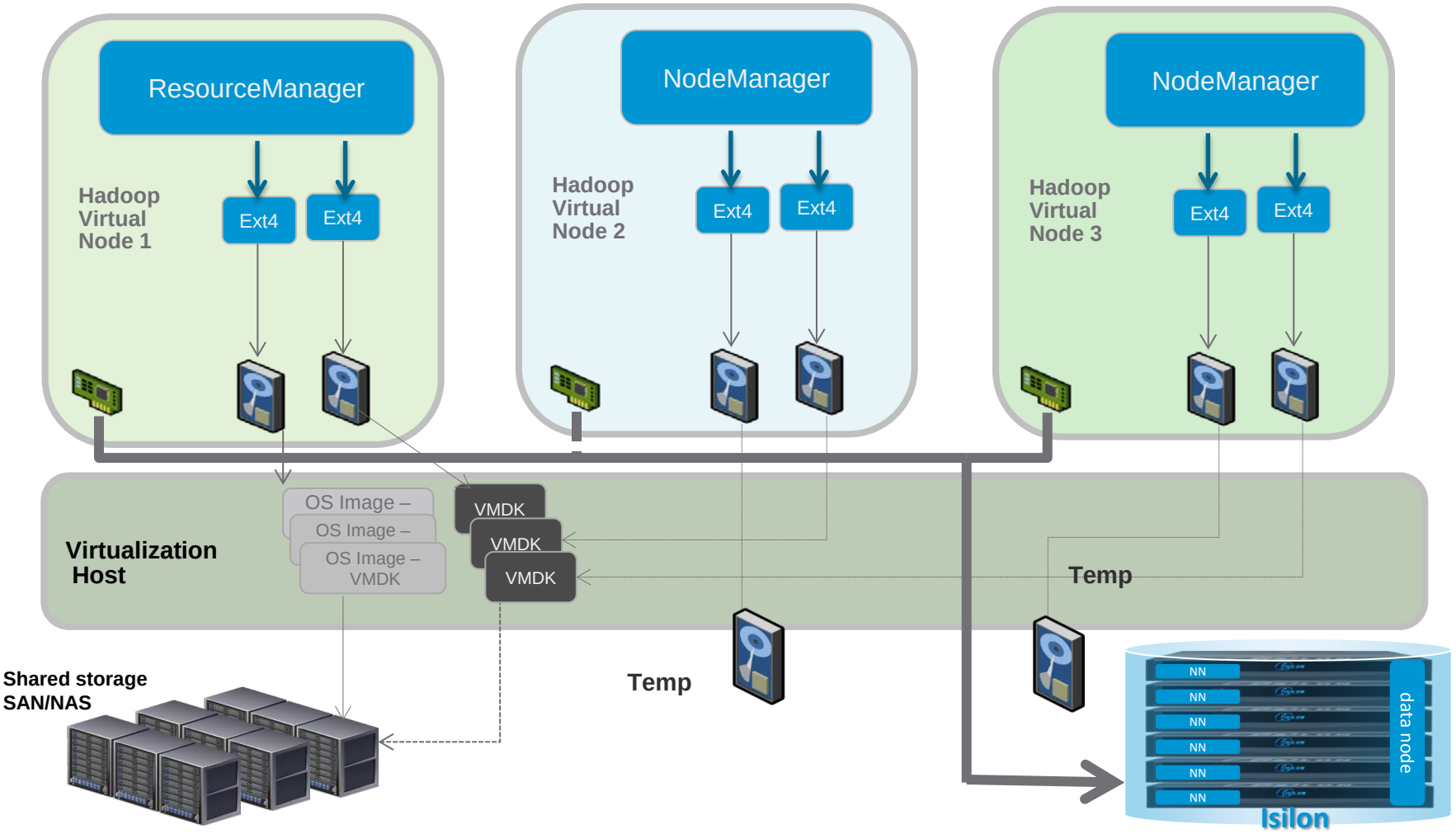
Alternative Storage for Data/Compute Separation



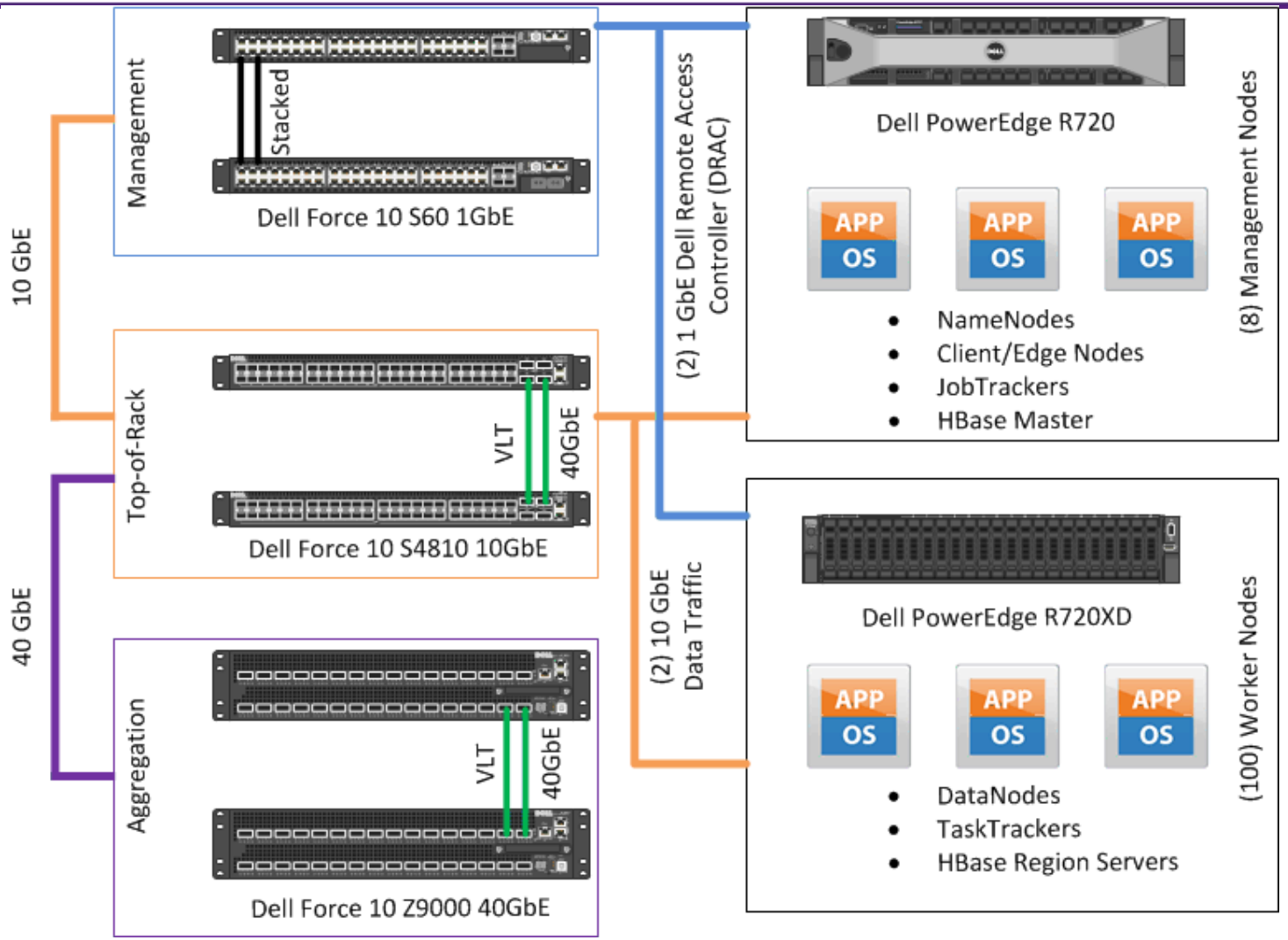
VMDK Isolation for Performance



Data/Compute Separation With Isilon



Larger Architecture with Data Compute Separated



Hybrid storage model - the best of both worlds

Shared Storage

Local Storage

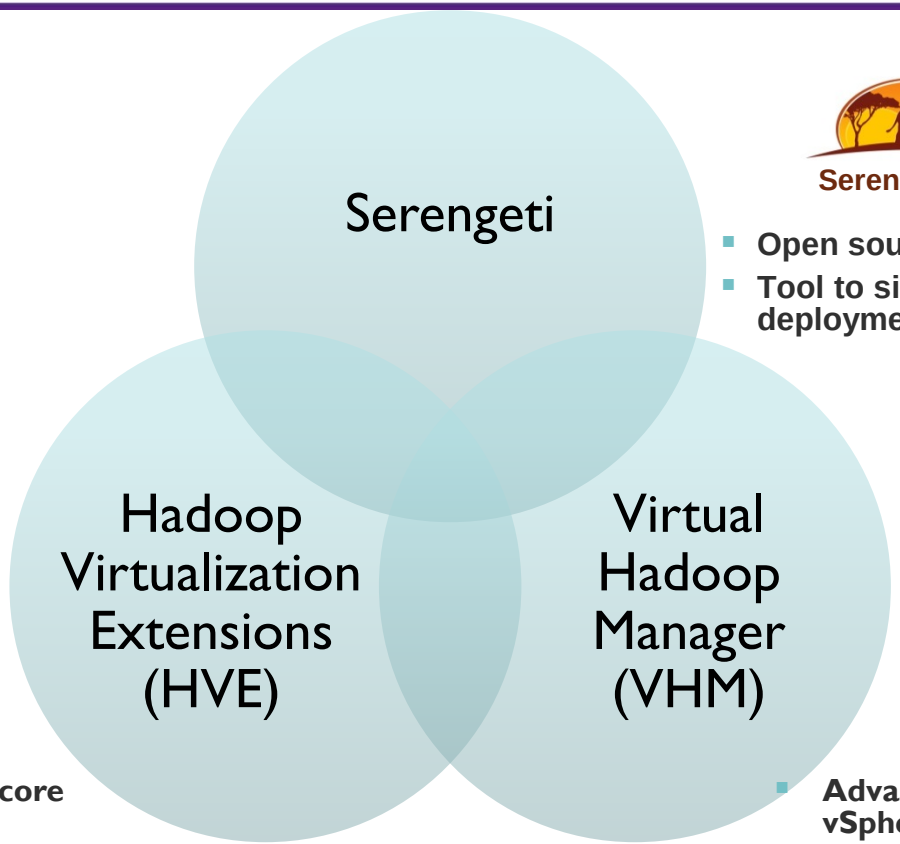
NFS for HDFS

- **Master nodes**
- NameNode, ResourceManager, ZooKeeper etc. on shared storage
  - Leverage vSphere vMotion, HA and FT
- **Worker nodes**
  - NodeManager/DataNode on local storage
  - Lower cost, scalable bandwidth
  - Temp data is written to local storage for best performance
  - NFS storage for HDFS data is a very good alternative to local

vSphere Big Data Extensions and Project Serengeti

vmware®

Big Data Extensions - Highlights



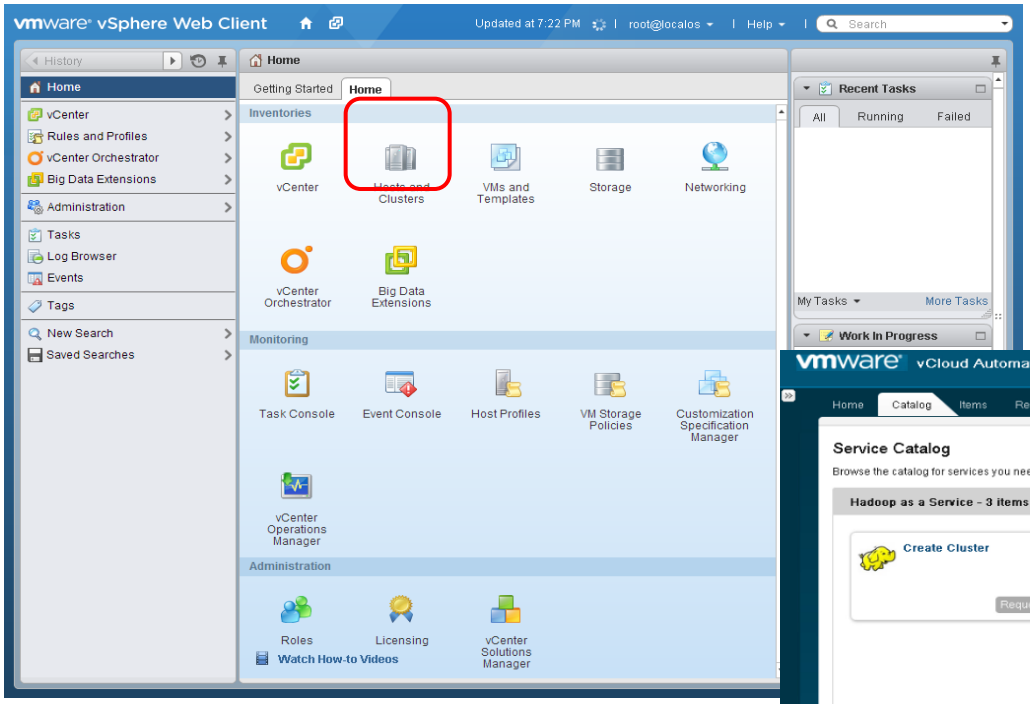
Serengeti

- Open source project
- Tool to simplify virtualized Hadoop deployment & operations

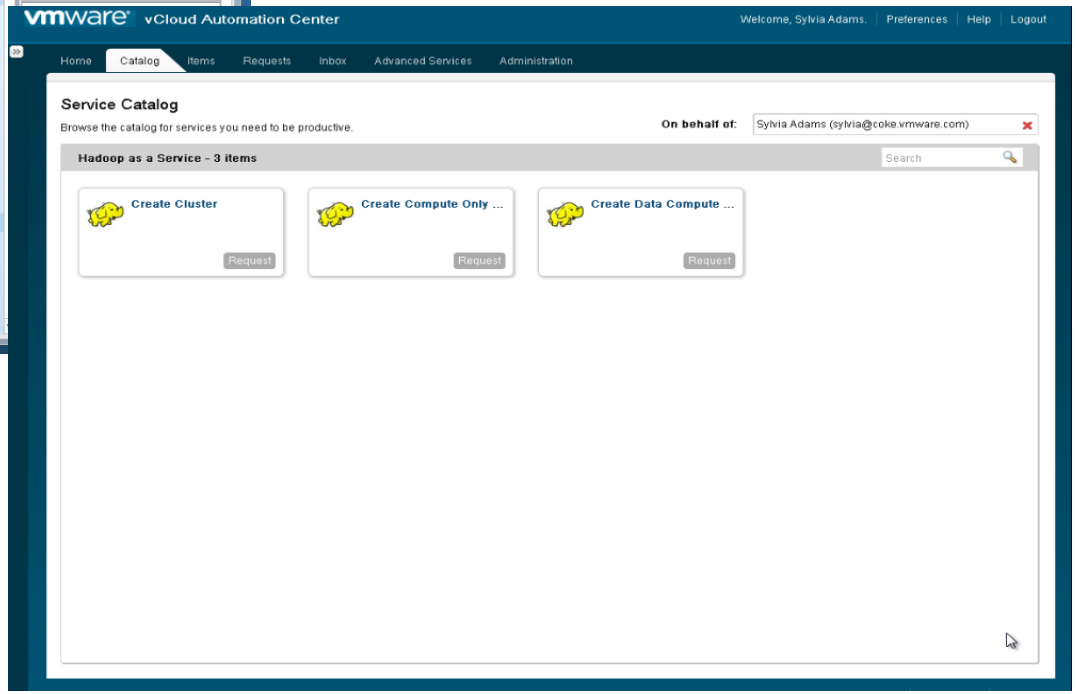
- Virtualization changes for core Hadoop
- Contributed back to Apache Hadoop

- Advanced resource management on vSphere

Introducing vSphere Big Data Extensions (BDE)

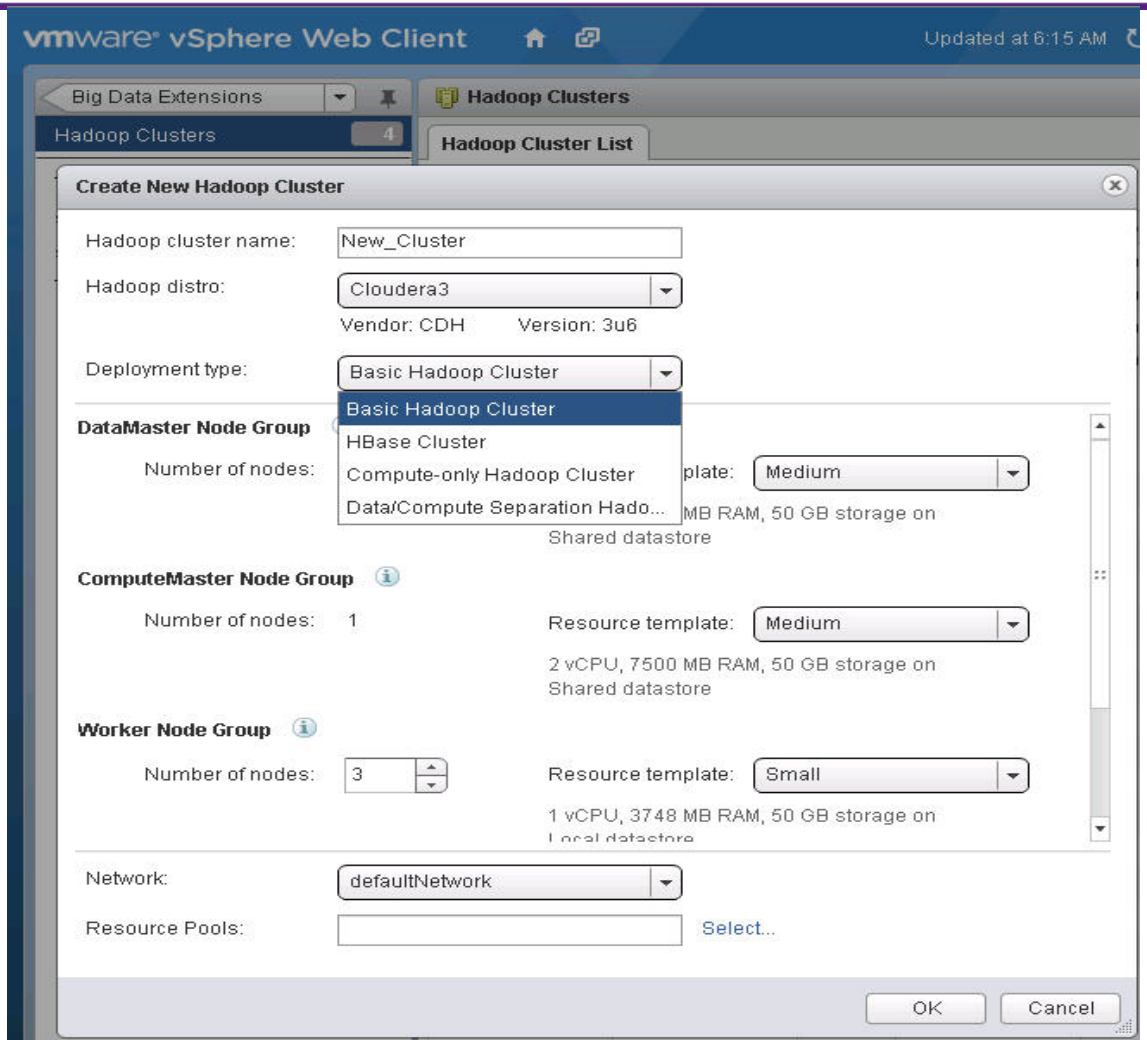


vCenter Plugin

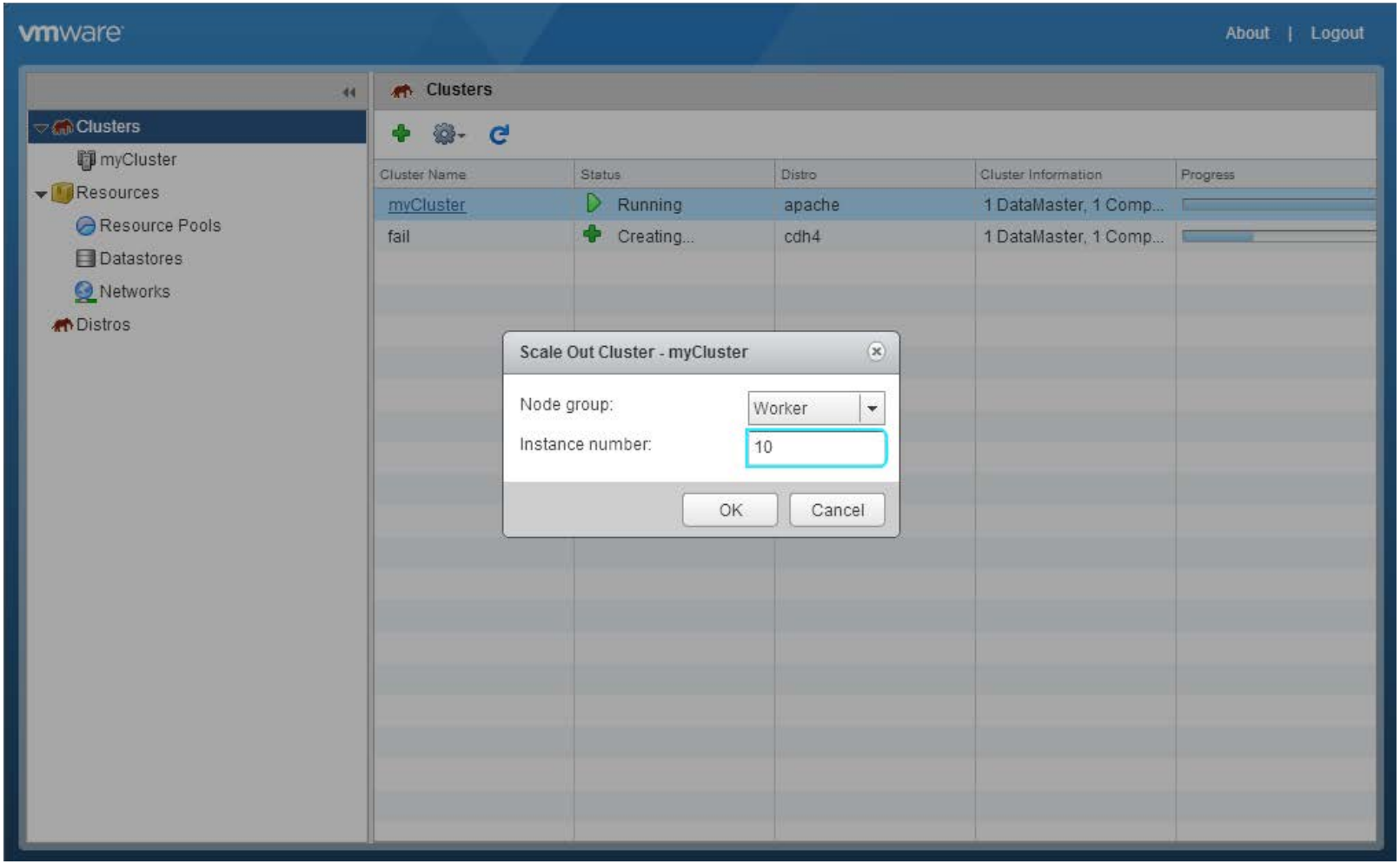


Hadoop as a Service
with
vCloud Automation Center

Brief Tour of Big Data Extensions



One Click to Scale out the Cluster on the Fly



BDE Allows Flexible Configurations

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

Storage configuration
Choice of shared or local

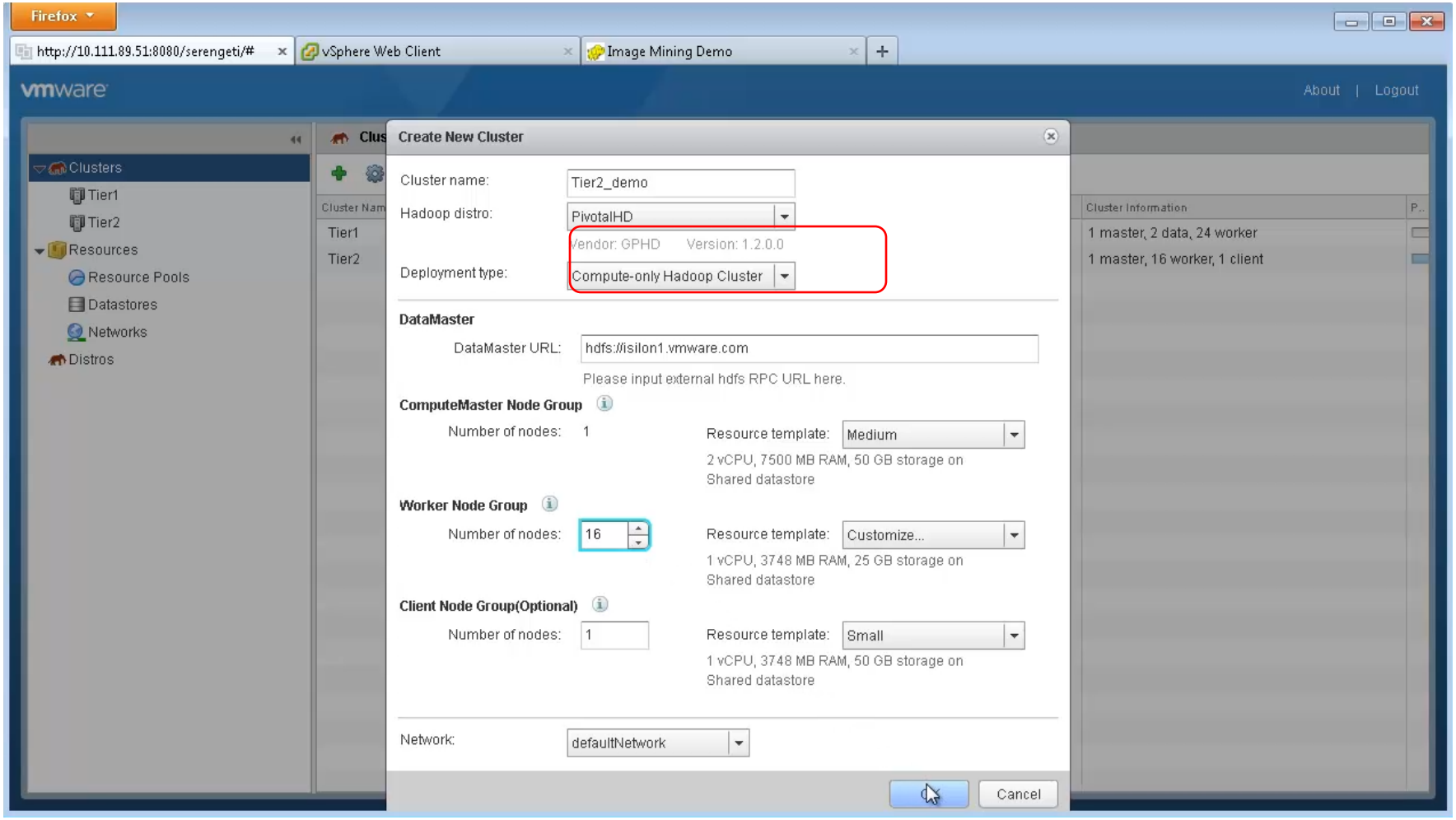
High Availability option

```
{
  "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"]
    }
  }
}
```

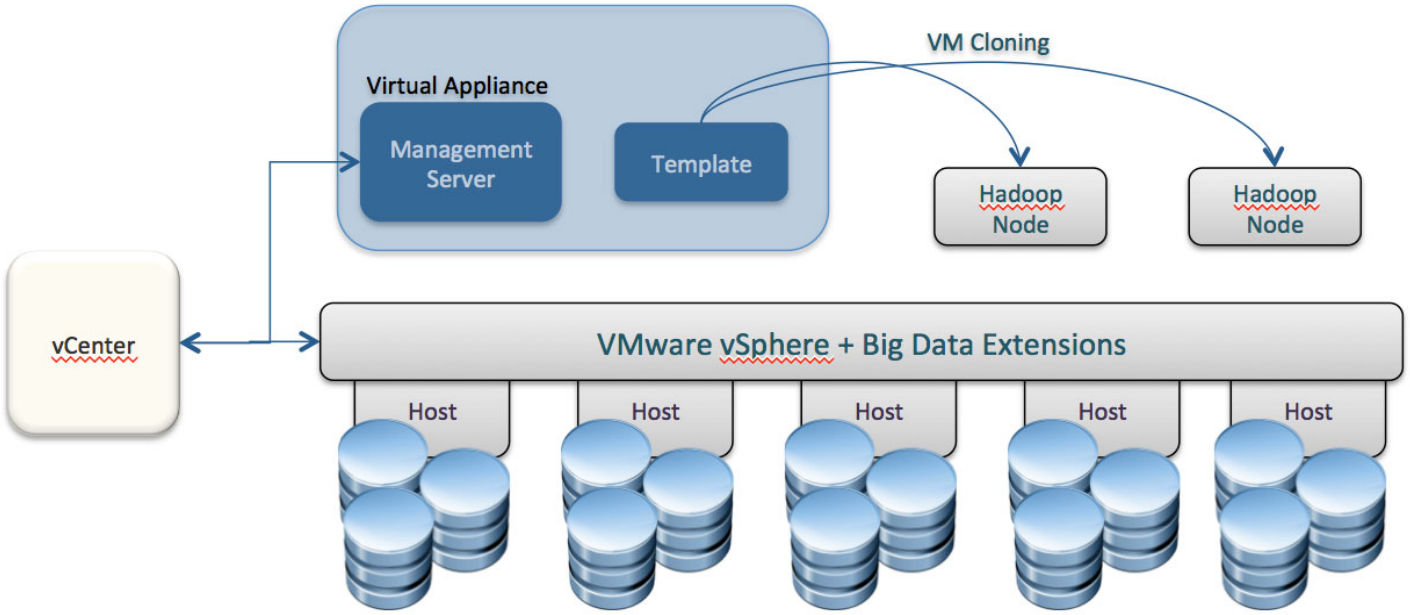
Number of nodes and resource configuration

VM placement policies

External HDFS : Simple to Set Up



How BDE works



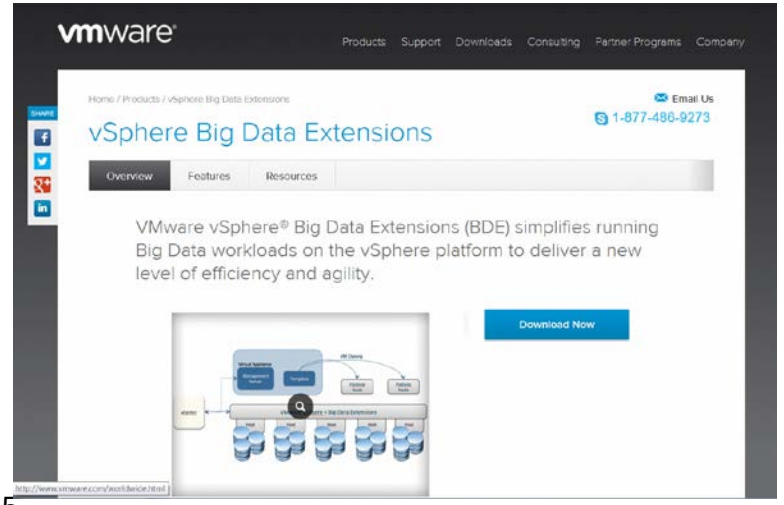
vSphere Configuration

- Provision the virtual machines at the right size
  - Reserve 6% of physical memory on the ESXi Server for vSphere usage
  - Avoid over-commitment
  - Enable NUMA and keep the virtual machine memory and cpu size within the NUMA node

- NUMA scheduler is important for virtualized Hadoop performance
  - Poor configuration can result in performance degradation
  - Data VM preferably should be distributed across NUMA nodes

VMware vSphere BDE and Hadoop Resources

- 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

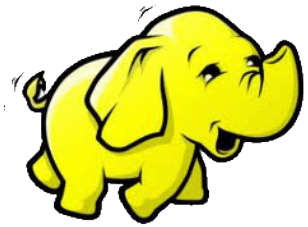
Conclusions

- Hadoop workloads work very well on VMware vSphere
  - Various performance studies have shown that any difference between virtualized performance and native performance is minimal
  - Follow the general best practice guidelines that VMware has published

- vSphere Big Data Extensions enhances your Hadoop experience on the VMware virtualization platform
  - Rapid provisioning tool for deployment of Hadoop components in virtual machines
  - Algorithms for best layout of your Hadoop data and cluster components are built into the BDE HVE components
  - Design patterns such as data-compute separation can be used to provide elasticity of your Hadoop cluster on demand.
  - User self service available with Hadoop using tools such as vCloud Automation Center integrated with BDE

Thank You

jmurray@vmware.com
Justin Murray



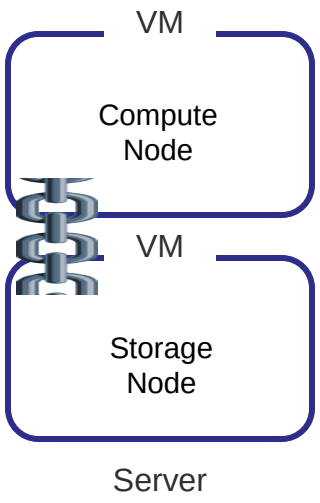
vmware

Backup Slides

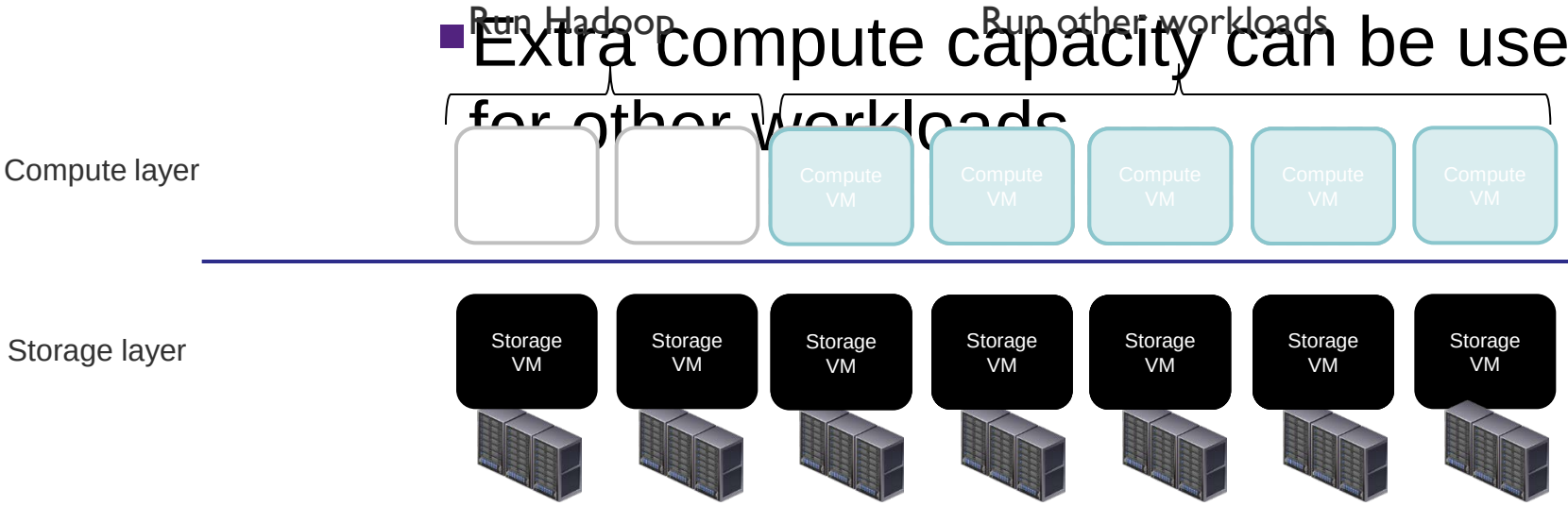
Today's Challenges on Hadoop Infrastructure

- Fixed compute and storage coupling leads to low utilization and inflexibility
  - Compute and storage linked together with fixed ratio based on the hardware specification
  - Not all jobs are created equal (data vs. compute intensive)
  - Inflexible infrastructure leads to waste
    - Too little compute power → slow processing
    - Too much compute power → sitting idle
  - Problem compounds with larger clusters
- So what happens?

Getting more out of your infrastructure



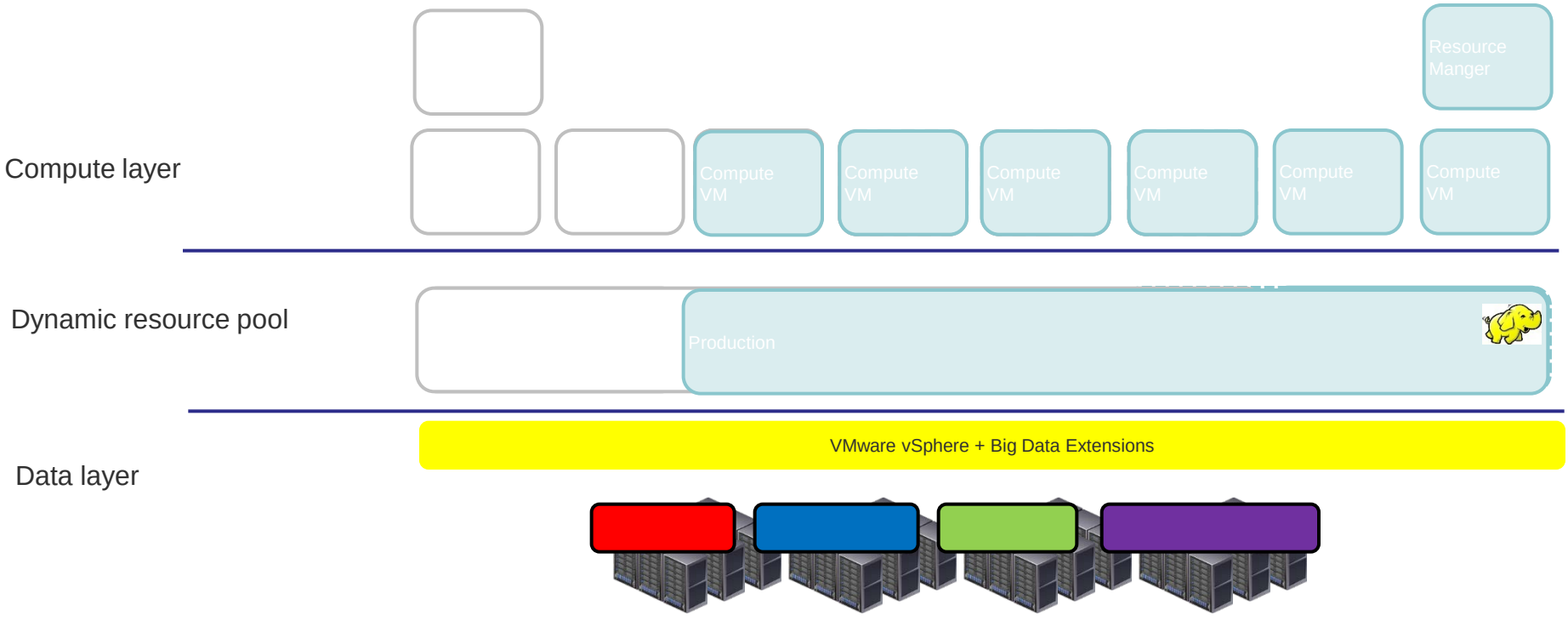
- Decouple the linkage between compute and storage
- Stateless compute can grow and shrink elastically
- Data locality is preserved, place the compute where data resides
- Extra compute capacity can be used for other workloads



Elasticity and Scalability

Elastic Scalability & Multiple Workloads

- Deploy separate compute clusters for different tenants sharing HDFS.
- Commission/decommission compute nodes according to priority and available resources



Hadoop 1.0