BC10

VMware HA Guidelines and Best Practices

Luis Robles
Product Manager

Nitin Suri
Systems Engineer

Sridhar Rajagopal
Staff Engineer

VMware

This session may contain product features that are currently under development.

This session/overview of the new technology represents no commitment from VMware to deliver these features in any generally available product.

Features are subject to change and must not be included in contracts, purchase orders, or sales agreements of any kind.

Technical feasibility and market demand will affect final delivery.

Pricing and packaging for any new technologies or features discussed or presented have not been determined.

Agenda

- **Introduction**
  - Overview of VMware HA
  - Customer Examples
- **Best Practices**
  - Setup and Networking
  - Resource Management
  - Troubleshooting
- **Technology Directions**

Clustering 101

- **Clustering has been around since the 1980's**
  - In many forms, incarnations, and platforms
- **Defined by Wikipedia as:**
  - "A group of tightly coupled computers that work together closely so that in many respects they can be viewed as though they are a single computer… usually deployed to improve performance and/or availability…"
- **2 fundamental principles: Redundancy and Mobility**
  - When something fails, the combination of both of these principles allows processing to be quickly restored on an alternate system.
  - **In VMware Infrastructure 3: DRS & HA clusters satisfy these needs**

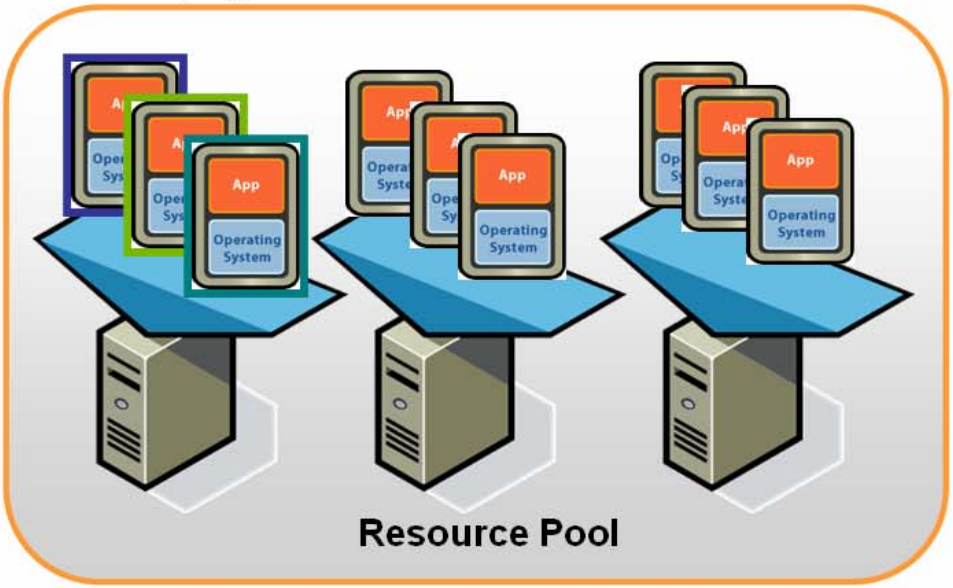
Virtualization Disrupts the Status Quo

Virtualization is disruptive to traditional OS-based clustering, and can provide better ways of achieving redundancy and mobility:

	Redundancy	Mobility
Legacy Clustering	○ Redundant pairs of identical physical hardware ○ Multiple OS & applications installed & maintained	○ Complex mechanisms for restarting workloads, incur downtime, sometimes recover
Virtualization	○ Redundant pools of similar physical hardware ○ Single OS & application installed & maintained	○ Simple mechanisms move applications freely ○ Can avoid downtime and recover quickly

High Availability with VMware HA

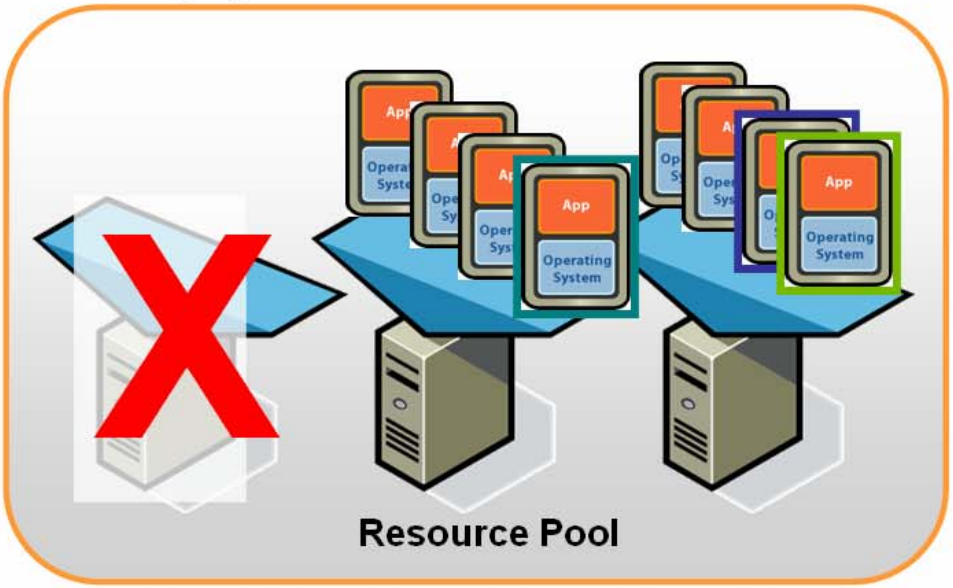
- **What is VMware HA?**
  - Automatic restart of virtual machines in case of unexpected physical server failures



- **Customer Benefits:**
  - High availability with minimal capital & operational costs
    - Reduced need for passive stand-by hardware & dedicated administrators
  - Broadly applicable
    - All x86 workloads virtualized are automatically protected without agents or scripting
  - Manageability & Flexibility
    - Very simple to deploy and maintain
    - Application manageability unchanged
    - N-to-N failovers, hosts and VMs easily moved in and out of clusters

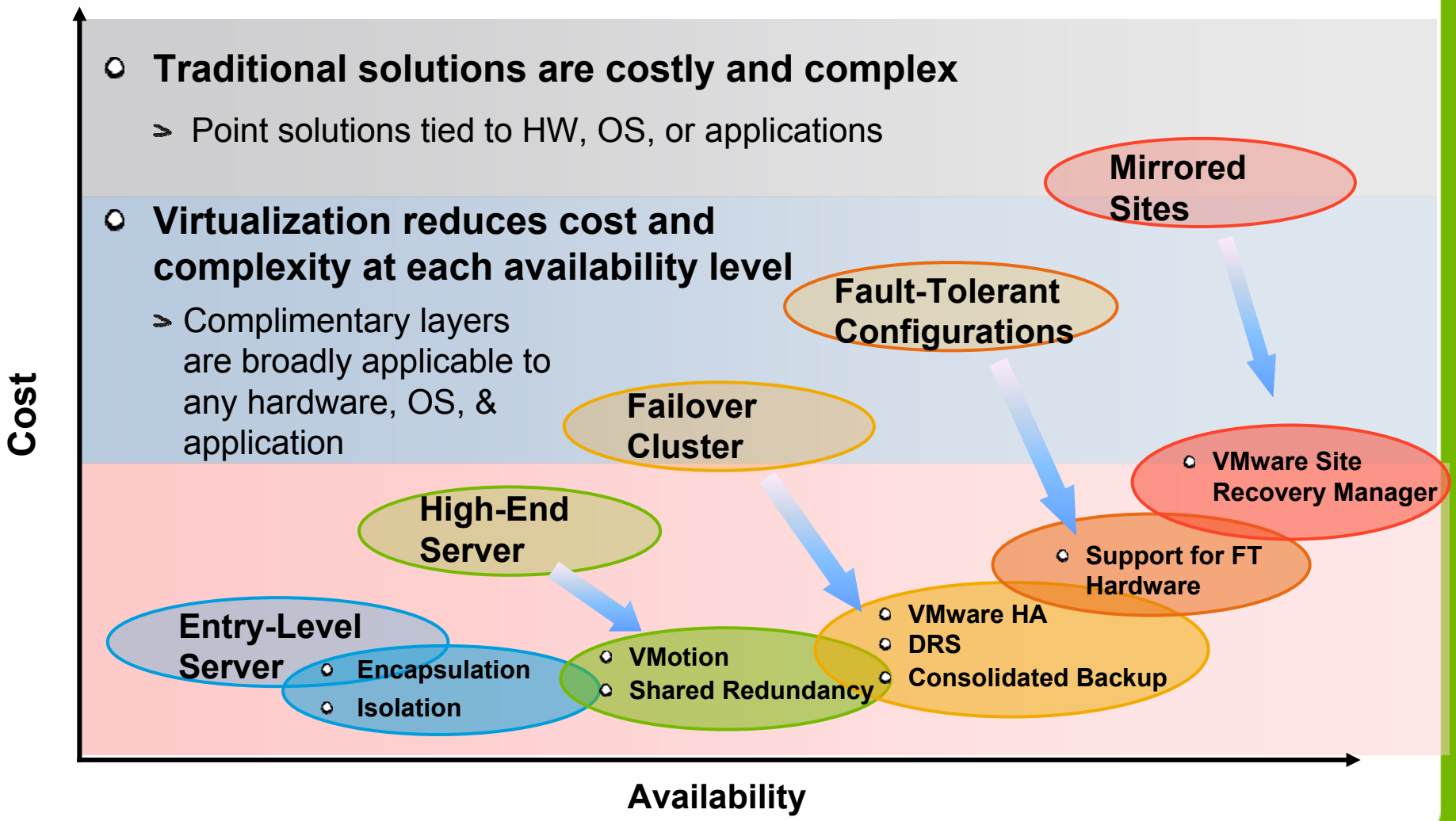
High Availability with VMware HA

- **What is VMware HA?**
  - Automatic restart of virtual machines in case of unexpected physical server failures



- **Customer Benefits:**
  - High availability with minimal capital & operational costs
    - Reduced need for passive stand-by hardware & dedicated administrators
  - Broadly applicable
    - All x86 workloads virtualized are automatically protected without agents or scripting
  - Manageability & Flexibility
    - Very simple to deploy and maintain
    - Application manageability unchanged
    - N-to-N failovers, hosts and VMs easily moved in and out of clusters

Business Continuity in a Virtualized World



VMware Availability Products And Features

	Avoid planned outages	Quick recovery from unplanned outages
Component	NIC Teaming, Multipathing	
Server	VMotion, DRS + Maintenance Mode	VMware HA
Storage	Storage VMotion	Encapsulation, VCB
Data	N/A	Encapsulation, VCB
Site	Encapsulation, boot from shared storage, instant reprovisioning, HW independence, resource pools, snapshots, VLANs	

VMware Availability Products And Features

	Avoid planned outages	Quick recovery from unplanned outages
Component	NIC Teaming, Multipathing	
Server	VMotion, DRS + Maintenance Mode	VMware HA
Storage	Storage VMotion	Encapsulation, VCB
Data	N/A	Encapsulation, VCB
Site	VMware Site Recovery Manager	

VMware HA Customer: Wyse Technology

"VMware Infrastructure 3 reduced the frequency of outages for core infrastructure services by 100 percent, saving 416 man hours per year in unplanned maintenance. VMware Distributed Resource Scheduler (DRS) and VMware High Availability (HA) are fully automated and performing flawlessly.

VMware HA is really a simpler and cost-effective alternative to complex, traditional clustering technologies."

Faan DeSwardt, Director of Enterprise Architecture,
Wyse Technology

Agenda

- **Introduction**
  - Overview of VMware HA
  - Customer Examples
- **Best Practices**
  - HA Under the Covers
  - Setup and Networking
  - Resource Management
  - Troubleshooting
- **Technology Directions**

nitin

suri

minus "g"

stating

the

obvious

Tomātō TomA[b]to


TARGET
Firestone
Firestone
TARGET
TARGET

Pres-101

3 X 3

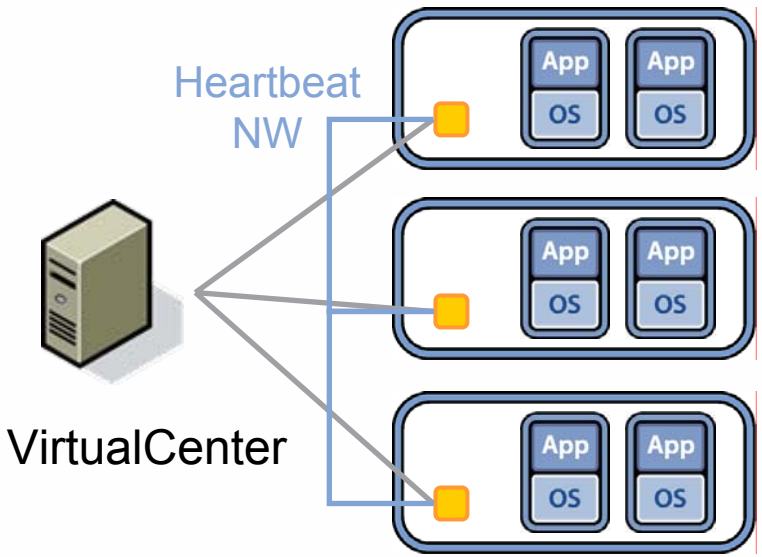
The quick young fox jumped over the lazy dog. The quick young fox jumped over the lazy. The quick young fox jumped over the lazy dog. The quick young fox jumped over the lazy dog. The quick young fox jumped over the lazy dog. The quick young fox jumped over the lazy dog. The quick young fox jumped over the lazy dog. The quick young fox jumped over the lazy. The quick young fox jumped over the

Agenda

- **Introduction**
  - Overview of VMware HA
  - Customer Examples
- **Best Practices**
  - HA Under the Covers
  - Setup and Networking
  - Resource Management
  - Troubleshooting
- **Technology Directions**

VMware HA Under the Covers

HA agents are installed & configured on ESX Hosts via VirtualCenter



- **Post configuration, HA agents maintain a heartbeat and communication network**
  - Ability to perform failovers is independent from VirtualCenter availability
- **VMFS / shared storage allows hosts to access & power-on virtual machines**
  - Distributed locking prevents simultaneous access to protect data integrity
- **During a failover, quick restart is the primary goal**
  - DRS algorithms balance workloads after HA has recovered virtual machines

Redundancy, Redundancy, Redundancy

- **Server Hardware**
  - Opt for proven models, redundant power supplies & cooling fans
- **Shared storage**
  - Redundant HBAs, failover paths in storage fabric pre-tested
- **Network**
  - Redundant NICs, switch failure scenarios pre-tested
- **Management**
  - VMware HA agents operate independently of VirtualCenter, yet centralized views & inventory are critical during failure conditions
  - Consider additional protection mechanisms for VirtualCenter
    - See session “BC23 Highly Available and Recoverable VirtualCenter”

Best Practices - Setup & Networking

1. **Proper DNS & Network settings are needed for initial configuration**
   - After configuration DNS resolutions are cached to /etc/FT_HOSTS (minimizing the dependency on DNS server availability during an actual failover)
   - DNS on each host is preferred (manual editing of /etc/hosts is error prone)

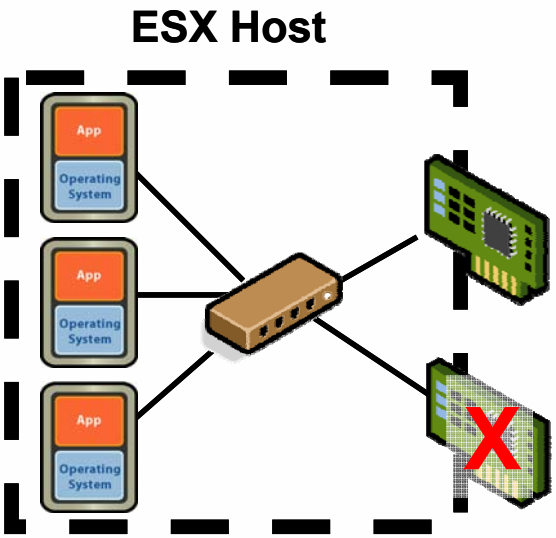
2. **Redundancy to ESX Service Console networking is essential (several options)**
   - Choose the option that minimizes single points of failure
   - Gateways/isolation addresses should respond via ICMP (ping)
   - Enable PortFast (or equivalent) on network switches to avoid spanning tree related isolations

3. **Network maintenance activities should take into account dependencies on the ESX Service Console network(s)**
   - VMware HA can be temporarily disabled through the Cluster->Edit Settings dialog

4. **Valid VM network label names required for proper failover**
   - Virtual machines use them to re-establish network connectivity upon restart

Network Configuration

Network redundancy between the ESX service consoles is essential for reliable detection of host failures & isolation conditions

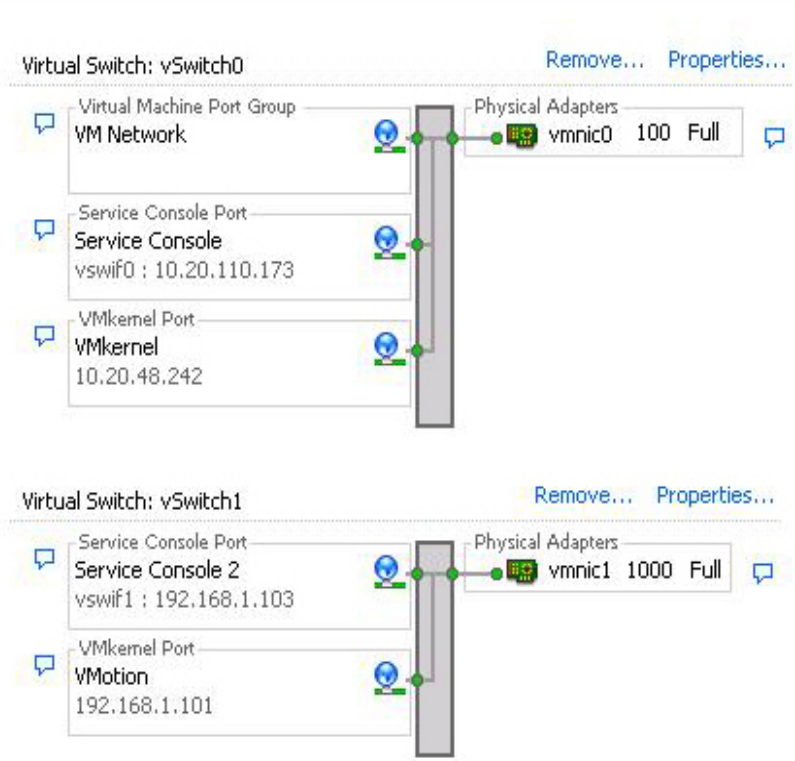


- **A single service console network with underlying redundancy is usually sufficient:**
  - Use a team of 2 NICs connected to different physical switches to avoid a single point of failure
  - Configure vNics in vSwitch for Active/Standby configuration (rolling failover = “yes”, default load balancing = route based on originating port ID)
- **Consider extending timeout values & adding multiple isolation addresses (*see appendix)**
  - Timeouts of 30-60 seconds will slightly extend recovery times, but will also allow for intermittent network outages

Network Configuration (Continued)

Beyond NIC teaming, a secondary service console network can be configured to provide redundant heartbeating & isolation detection

- **HA will detect and use a secondary service console network**
  - Adding a secondary service console portgroup to an existing VMotion vSwitch avoids having to dedicate an additional subnet & NIC for this purpose
  - Also need to specify an additional isolation address for the cluster to account for the added redundancy (*see appendix)
- **Continue using the primary service console network & IP address for management purposes**
  - Be careful with network maintenance that affects the primary service console network and the secondary / VMotion network



Best Practices – Resource Management

1. **Larger groups of homogenous servers will allow higher levels of utilization across an HA/DRS enabled cluster (on average)**
   - More nodes per cluster (current maximum is 16) can tolerate multiple host failures while still guaranteeing failover capacities
   - Admission control heuristics are conservatively weighted (so that large servers with many VMs can failover to small servers)

2. **To define the sizing estimates used for admission control, set reasonable reservations as the minimum resources needed**
   - Admission control will exceed failover capacities when reservations are not set; otherwise HA will use largest reservation specified as the “slot” size.
   - At a minimum, set reservations for a few virtual machines considered “average”

3. **Admission control may be too conservative when host and VM sizes vary widely**
   - Perform your own capacity planning by choosing “Allow virtual machines to be powered on even if they violate availability constraints”. HA will still try to restart as many virtual machines as it can.

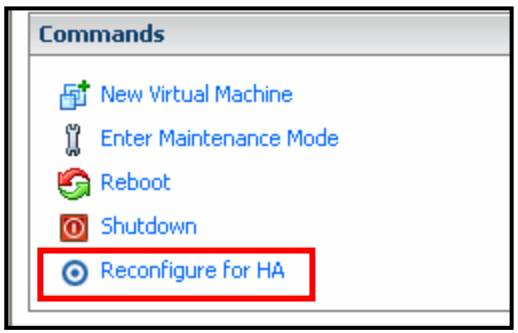
Troubleshooting VMware HA

1. **Verify the following**
   - IP connectivity, DNS resolution, shared storage and networks are visible throughout the cluster, service consoles have valid & reachable gateways

2. **Pay attention to cluster warnings, task details, and events**
   - Clusters having a YELLOW or RED icon are indicative of some issue or problem. The problem or issue at hand is specified using config issues – the big yellow boxes you sometimes see in the summary page.
   - Task details will often explain why an HA task has failed. Many pre-checks such as hostname resolutions are performed as of VC 2.0.2.
   - VirtualCenter events provide a history of VMware HA related actions

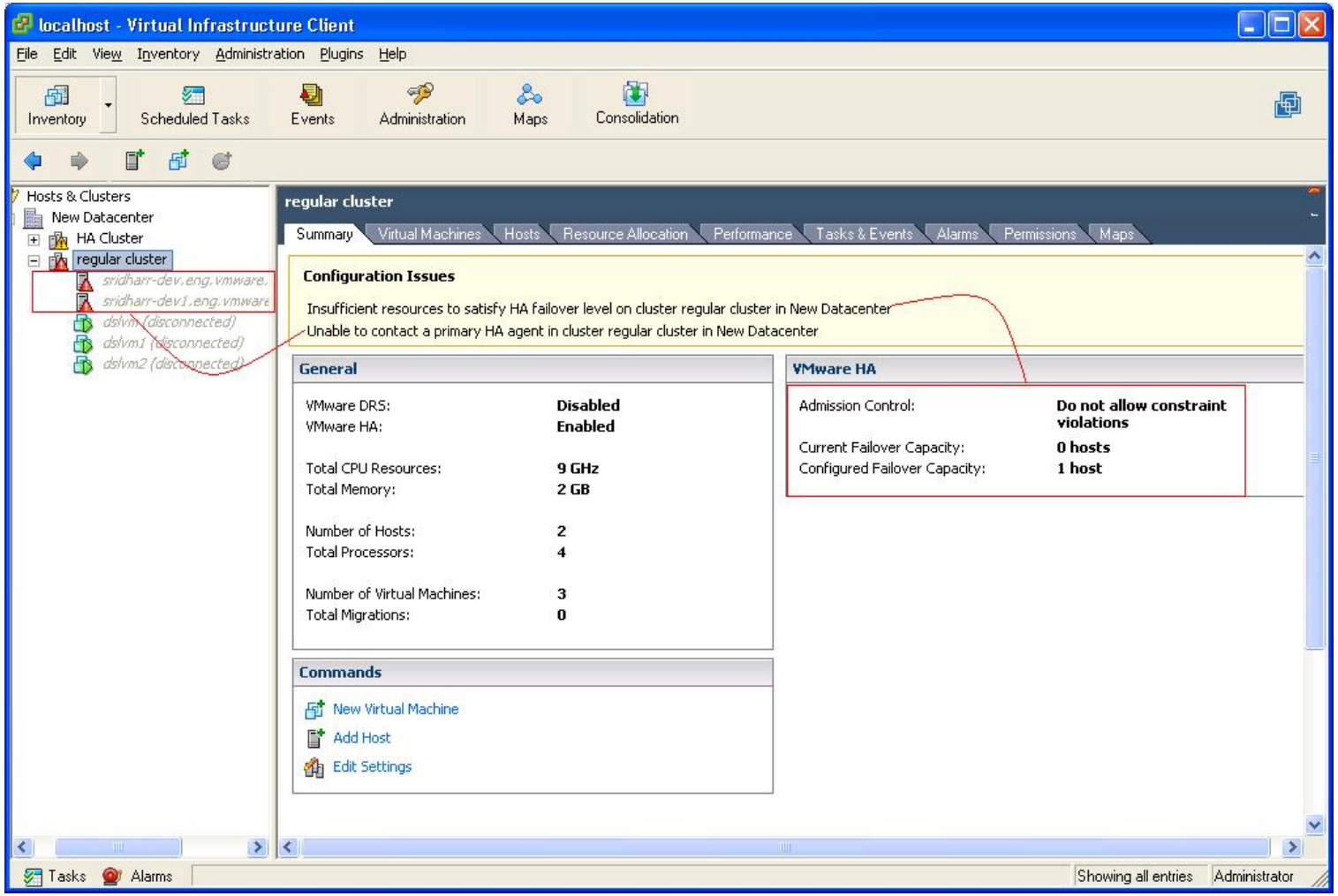
Troubleshooting (continued)

3. **Inspect ESX Server logs and HA agent configuration files:**
   - Check for ESX Service Console networking errors first, HA agents next
   - HA agent logs: /opt/LGTOaam512/log/* & /opt/LGTOaam512/vmsupport/*
   - /opt/LGTOaam512/config/vmware-sites contains a list of cluster nodes

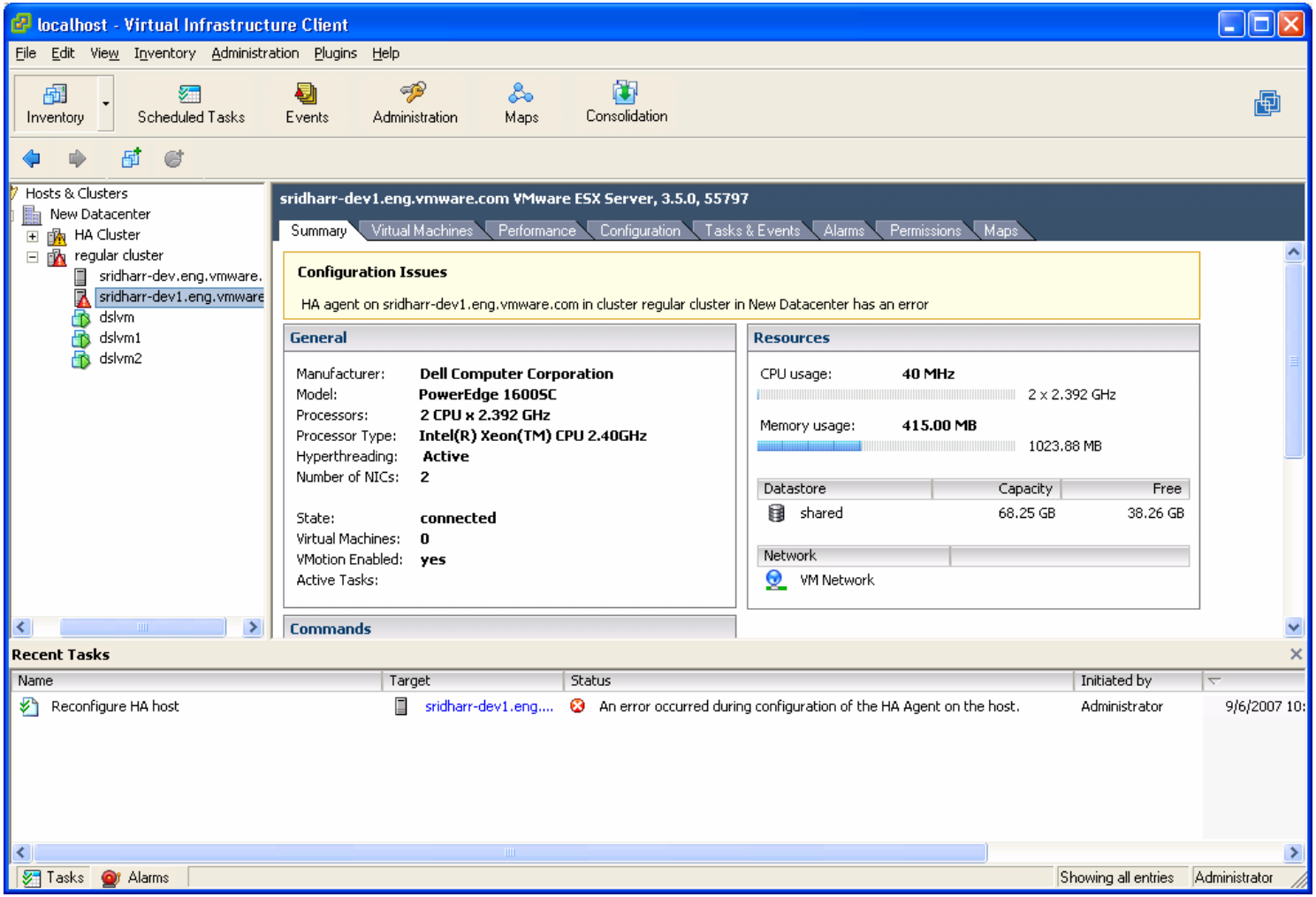


4. **Re-initialize HA cluster configuration (as a last resort)**
   - Per host: Select ESX Host, Summary Tab, Reconfigure for HA
   - Per cluster: Select Cluster, Edit Settings, Uncheck HA enabled, wait for reconfiguration task to complete, and then check to re-enable.
   - Not a fix-all.

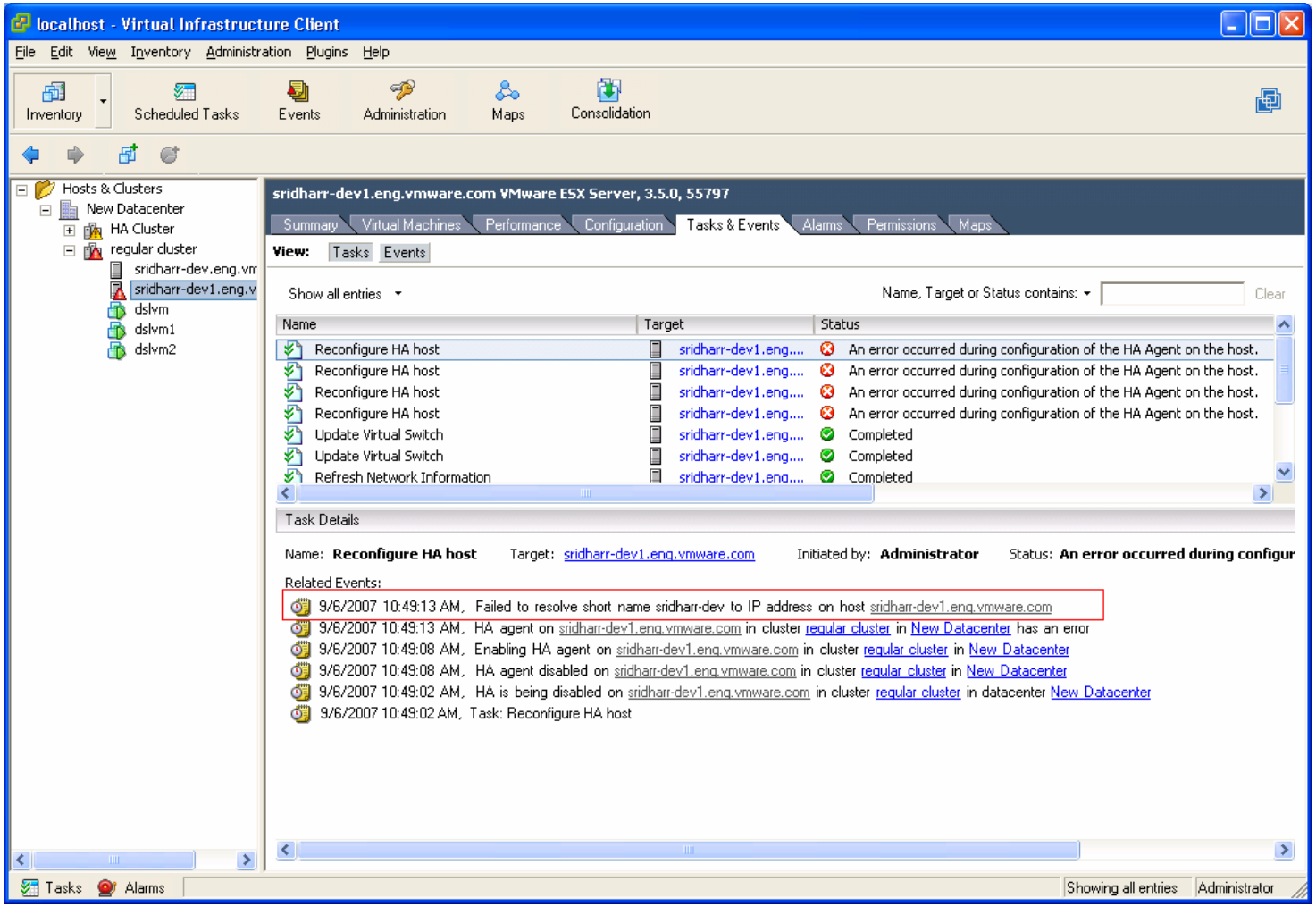
Troubleshooting Example (Cluster Configuration)



Troubleshooting Example (Host Configuration)



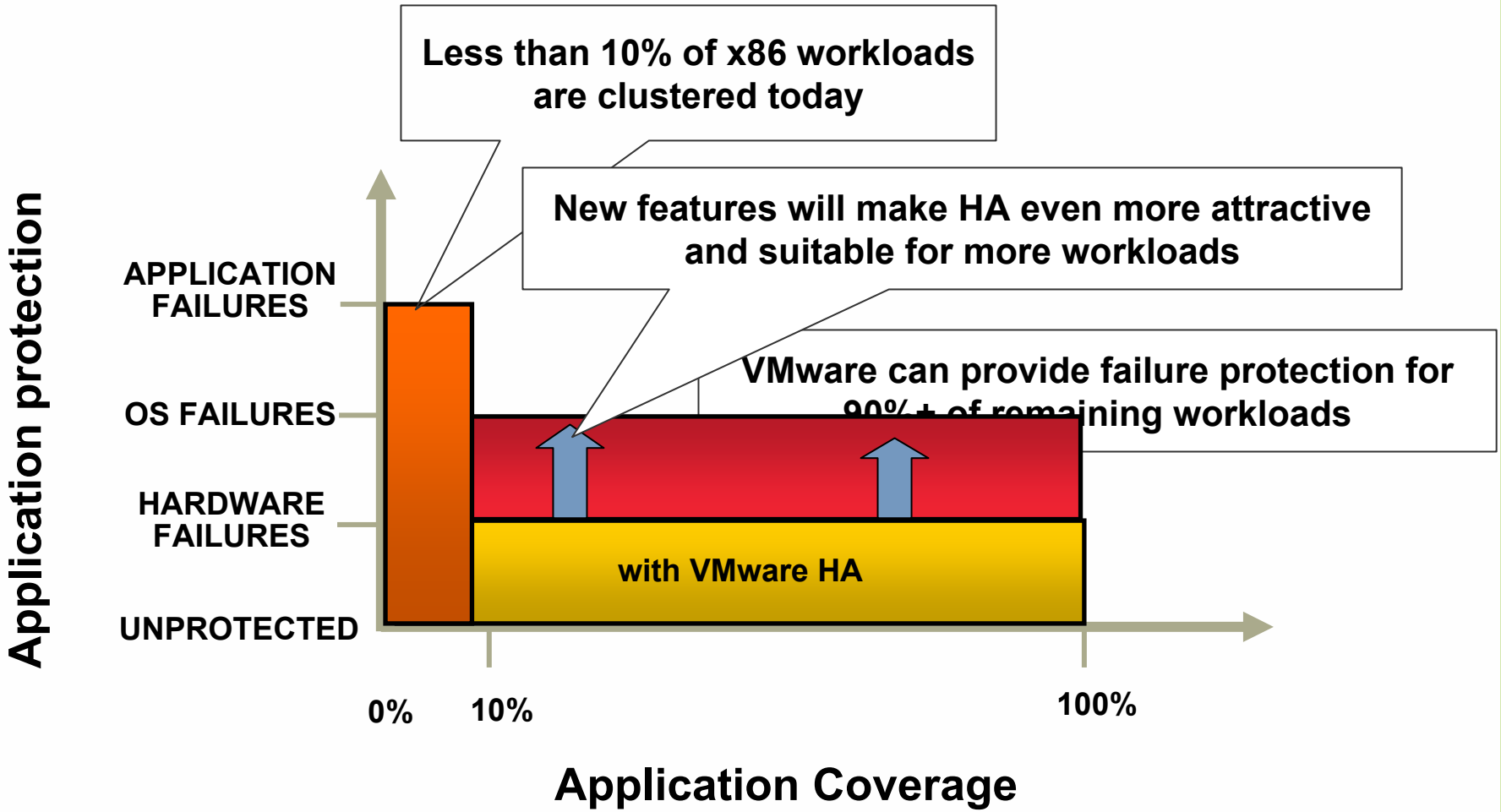
Troubleshooting Example (Task Details)



Agenda

- **Introduction**
  - Overview of VMware HA
  - Customer Examples
- **Best Practices**
  - Setup and Networking
  - Resource Management
  - Troubleshooting
- **Technology Directions**

Evolution of VMware HA



Short-term Technology Directions

- **Increased scalability**
  - Support for up to 32 nodes per cluster (currently 16)
- **Improved usability & Configuration Validation**
  - Host name resolution checks
  - Lack of redundant service console networking
  - Unreachable isolation addresses
- **Support for VMware ESX Server 3i**
  - HA functionality equivalent to ESX Server 3.x

Medium-term Technology Directions

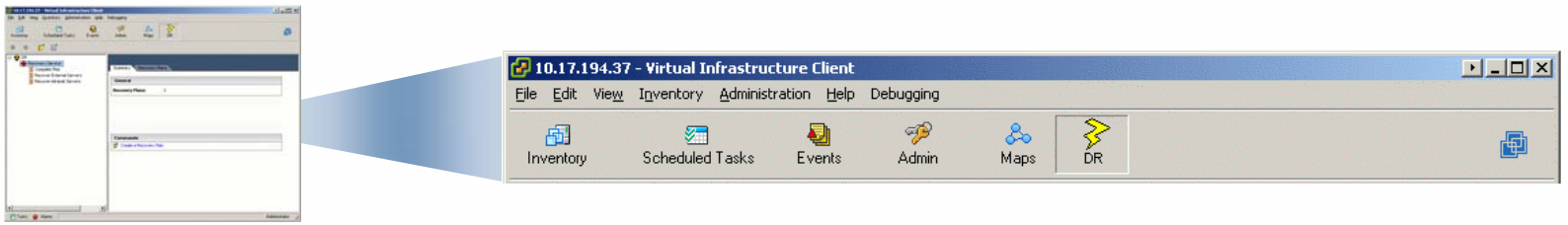
- **Individual Virtual Machine Failure Monitoring**
  - Utilize VMware Tools heartbeats to gauge VM health
    - Resets VMs when there is a problem
  - Controlled via HA advanced options
  - User specified policy, with defaults for each
    - Disabled by default
    - Loss of heartbeats for entire time window declared a failure
    - Max. number of failures after which automated response is stopped
    - Time window for failures – failures outside window get aged out (e.g. max of 3 failures in the last day).
    - Time window for guest-OS start-up & stabilization of heartbeats

Medium-term Technology Directions

- **HA Recovery in spite of total cluster / environmental failures**
  - Current version: at least one host assumed to survive, requiring manual intervention in case of total network & power outages
  - Future version: do not require a surviving host, recover from total cluster failures caused by temporary outages, and automatically restore resources

VMware Site Recovery Manager

- **New VMware DR Workflow Automation product**
- **Simplifies disaster recovery across multiple datacenters, lowers costs, and improves business continuity**
  - Integrates VMware Infrastructure with storage replication technologies
  - Provides disaster recovery protection for all workloads virtualized
  - Ensures disaster scenarios are executed according to plan
- **Plugs into VirtualCenter and orchestrates configurations, failover, failback, and testing across sites**
- **Check out "BC31 New Trends in Disaster Recovery" for details!**



Questions?

BC10

VMware HA Guidelines and Best Practices

Sridhar Rajagopal

Luis Robles

Nitin Suri

VMware

Appendix

- **Please refer to the following KB article for details on applying advanced HA customizations:**
  - Setting preferred failover hosts for individual virtual machines
  - Adjusting the default timeout used for failure & isolation detection (New in VC 2.0.2)
  - Changing the default isolation response address
  - Setting more than one isolation response address (New in VC 2.0.2)

http://kb.vmware.com/kb/1002080

VMWORLD 2007

EMBRACING YOUR VIRTUAL WORLD

BREAKOUT SESSION