HA Design Guidelines for HP Superdomes

and Other Partitioned Servers

Phil Noyes / Peter Chung

Pre-sales Solution Architects

ESG Competitive Sales & Presales

HP WORLD 2003
Solutions and Technology Conference & Expo

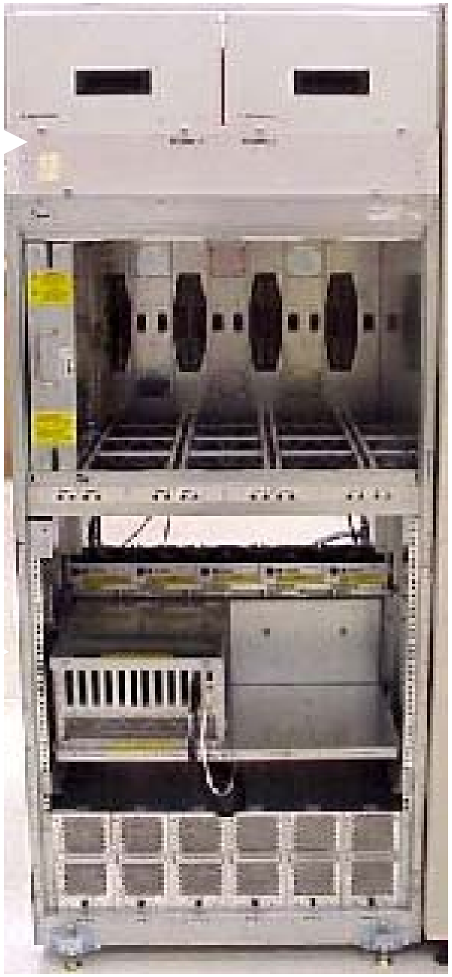
Agenda

- General server high availability design principles
- Superdome-specific high availability design principles
- Clustered Superdome design principles
- Serviceguard cluster arbitration
- vPar design considerations
- Design principles applicable to other partitioned servers

- Focus is **architecture design**, not implementation

Superdome Cabinet View

Front View

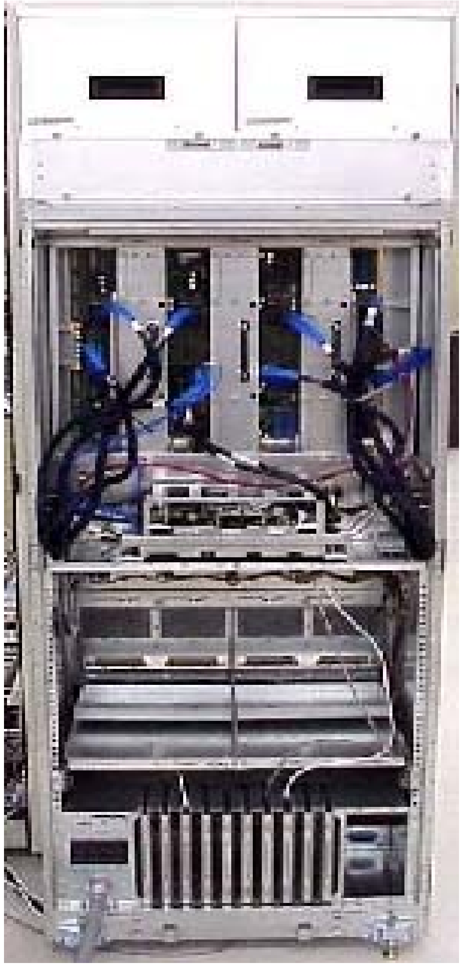


Cell Board (8)

I/O Chassis (4)

Power Supplies

Rear View



I/O Chassis

PDCA

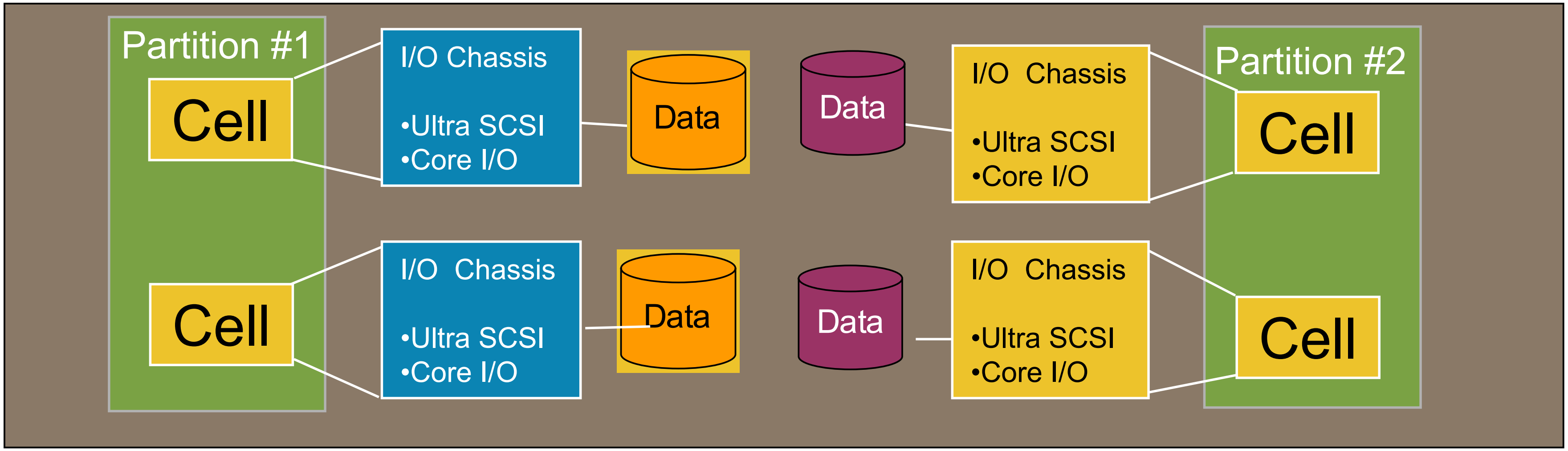
SPOF: Single Points of Failure

Component	Protection Mechanism(s)
Disk	–Multiple host adapters + link failover software –Data protection (software-based or hardware-based)
CPU	–Multiple processors –Dynamic Processor Resilience (online deallocation)
Network	–Multiple host adapters, switches –Link failover software
Power	–Redundant power supplies –Multiple power circuits
Software / App.	–Dependent on specific software product –Serviceguard may restart application on same system or migrate it to adoptive system
System	–Multiple systems, configured in a cluster

Design Principles – Standalone Superdome

- Cell board:
  - Each Superdome partition configured with at least two cell boards
  - Each cell board configured with at least two active CPUs
  - Each cell board configured with at least 4GB RAM (PA arch.)
- Power:
  - Redundant PDCA (Power Distribution Control Assembly)
- I/O Chassis:
  - Redundant I/O chassis per Superdome partition
  - Each Superdome partition configured with two core I/O cards

Example: Superdome Partition I/O



Key Design Principles:

- Redundant I/O chassis per partition
- Redundant Core I/O cards per partition
- Redundant I/O paths to storage
- Data is protected (software mirrored or hardware RAID)
- Diagram can be extended to network configuration

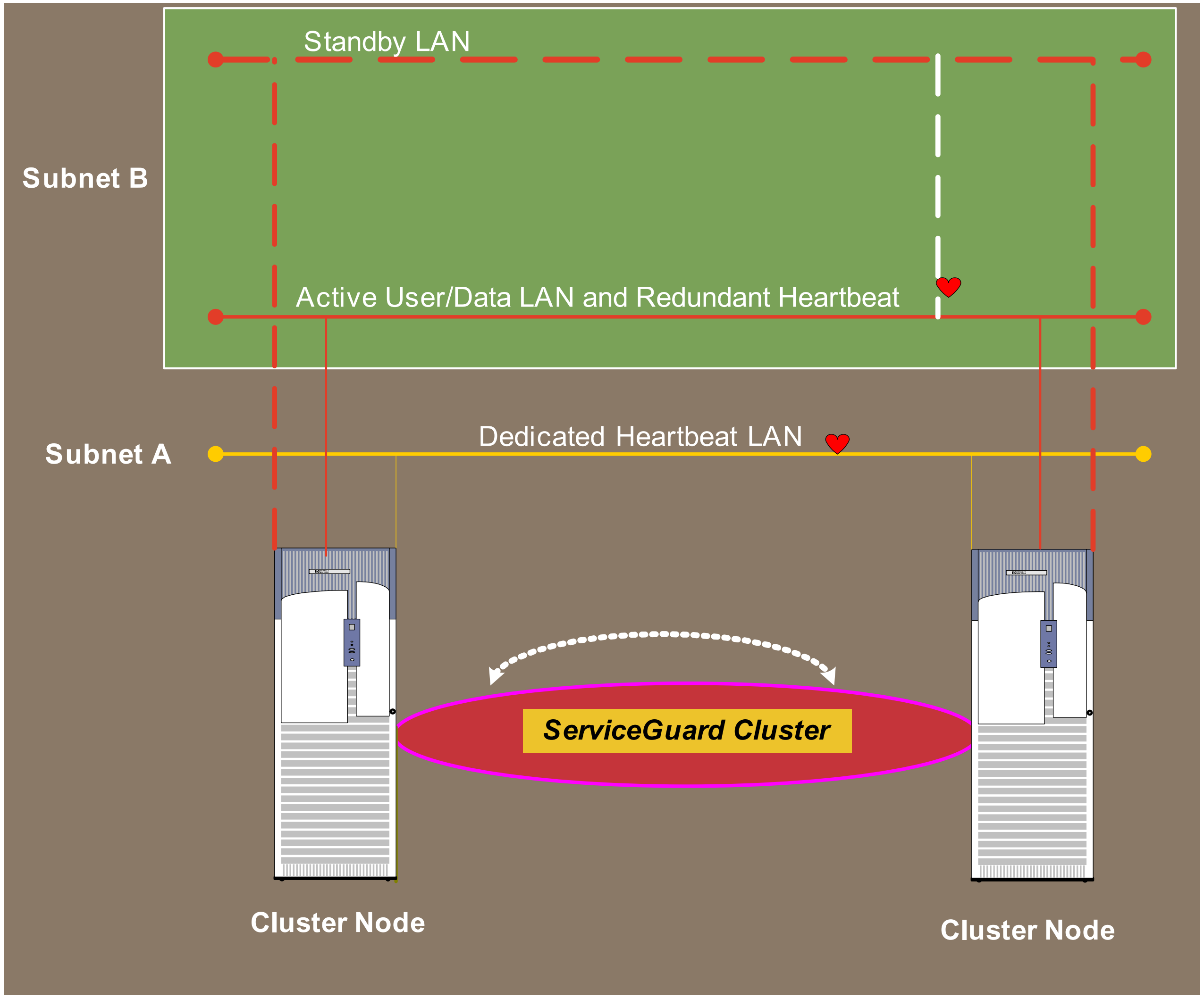
What is a Valid Cluster?

- A independent copy of HP-UX can be configured on any of the following:
  - A hard partition (nPar)
  - A soft partition (vPar)
  - A non-partitioned HP9000 server
- Valid Serviceguard clusters can include the following types of “nodes”:
  - nPars within the same server
  - nPars from different servers
  - Non-partitioned HP9000 servers
  - vPars

Design Principles – Cluster LAN

- Serviceguard cluster “nodes” should include the following network design principles:
  - Cluster heartbeat must be configured with redundant LANs
  - Dedicate one LAN interface to cluster heartbeat
  - User/data subnets should be configured with a standby
  - Primary/standby LAN’s must be the same network stack
- Cluster LAN variables defined in the Serviceguard configuration file (/etc/cmcluster/clusterconf.ascii):
  - *HEARTBEAT_IP* specifies cluster heartbeat traffic
  - *STATIONARY_IP* specifies data traffic only
  - <*blank*> definition specifies a standby LAN

Example: Cluster LAN Design



Key Design Principles

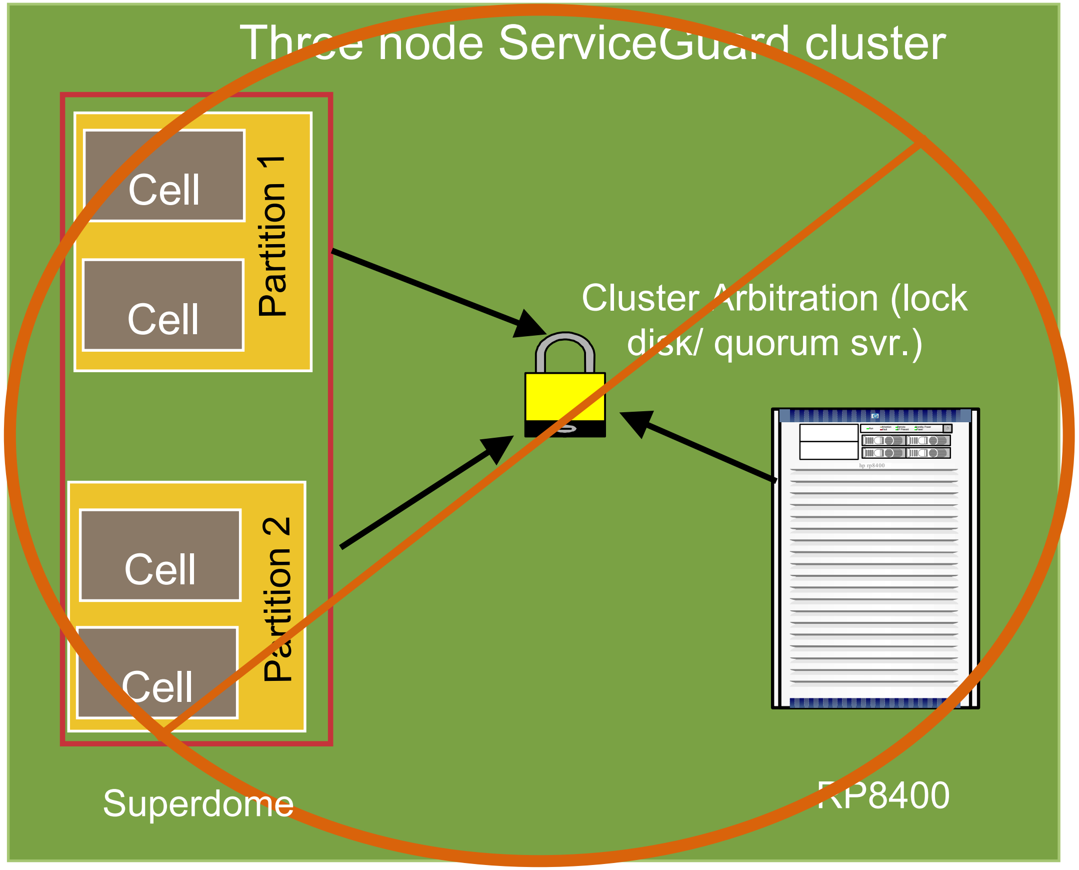
- Dedicated cluster HB
- Redundant HB
- Standby LAN

Serviceguard Cluster Design Principles

- Clusters should never lose > 50% of the nodes due to a single failure
- An arbitrator is required if a cluster can lose exactly 50% of the nodes from a single failure
- Cluster arbitrators must be powered independently of the cluster nodes
- A Superdome (or any partitioned server) cluster configuration should extend beyond a single cabinet

Example: >50% of cluster nodes lost during failure

A cluster should never lose > 50% of the cluster nodes due to a single failure


Three node ServiceGuard cluster
Cell
Cell
Partition 1
Cell
Cell
Partition 2
Cluster Arbitration (lock disk/ quorum svr.)
Superdome
RP8400

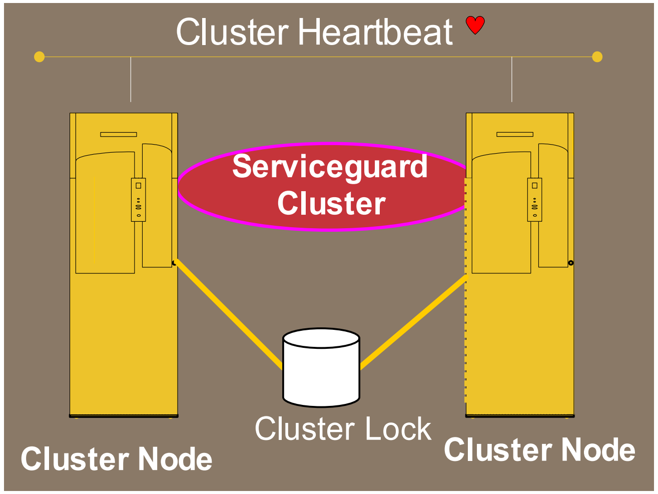
Cluster Arbitration – Cluster Lock Disk

- Cluster arbitration is required in certain situations to ensure cluster reformation.

•Three methods of arbitration: cluster lock disk, quorum server, and arbitrator node

Cluster lock disk:

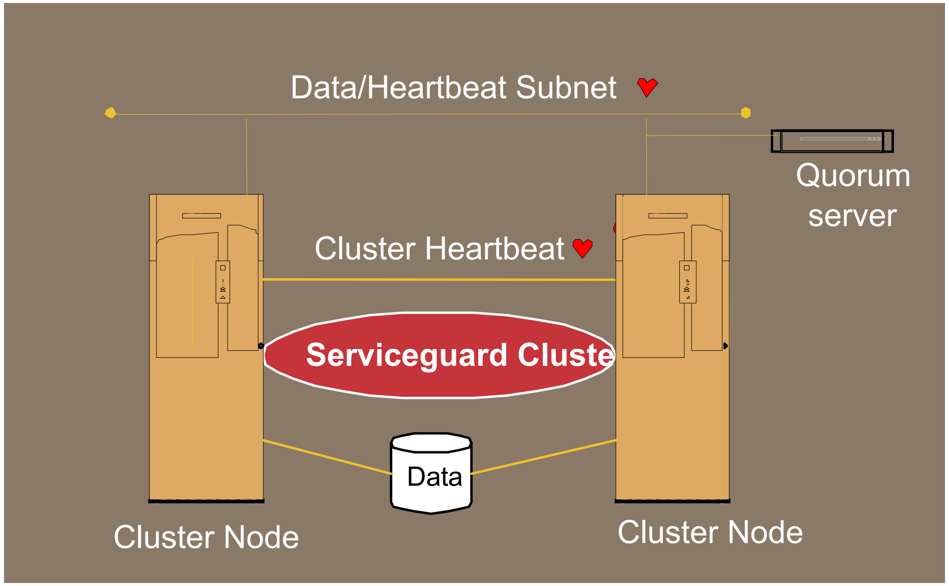
- A volume group which is accessible by all cluster nodes
- Can arbitrate a single cluster, up to four nodes in size.
- Required for a two-node cluster, optional for larger clusters
- *FIRST_CLUSTER_LOCK_VG* defined in /etc/cmcluster/clusterconf.ascii



Cluster Arbitration – Quorum Server

Quorum server:

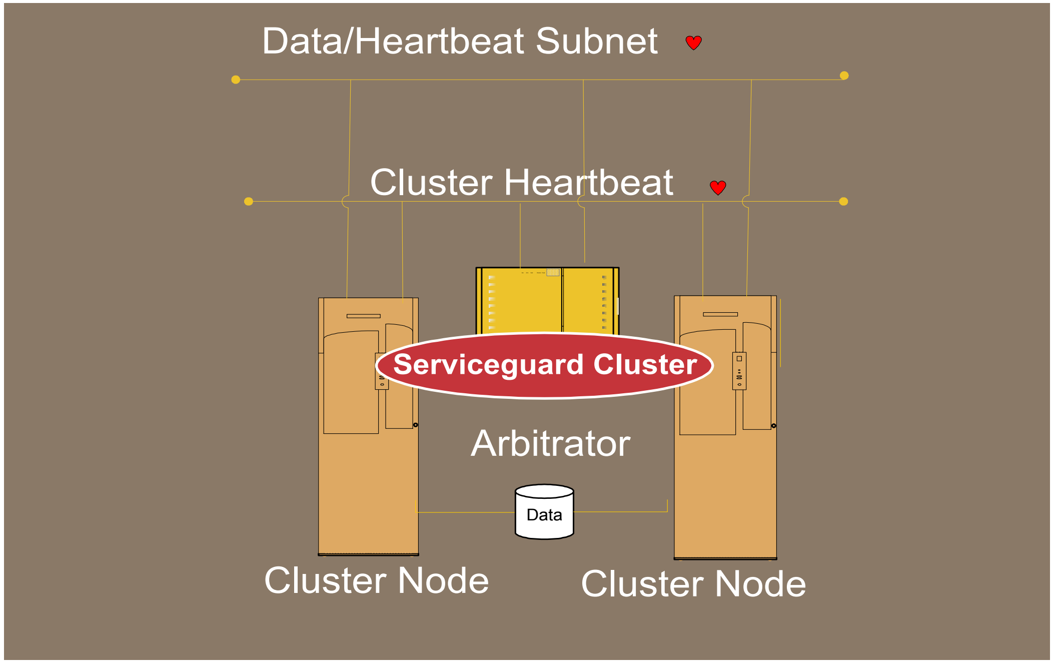
- Can arbitrate up to fifty separate clusters, or up to one hundred nodes
- Is not a Serviceguard cluster node



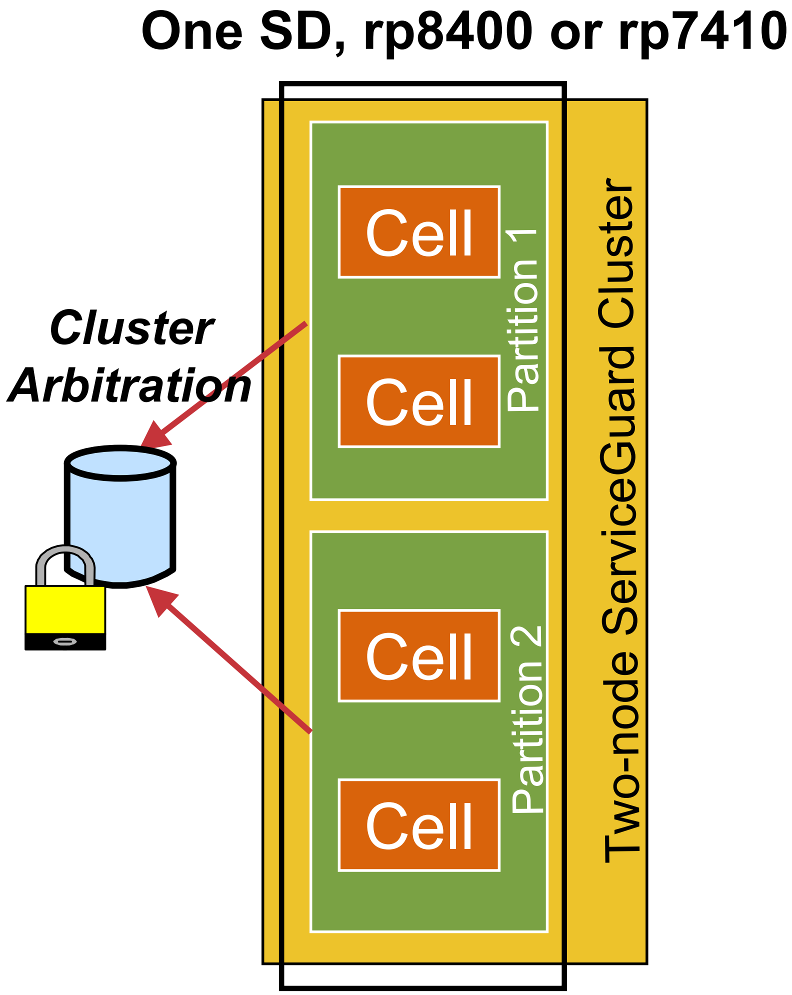
Cluster Arbitration – Arbitrator Node

Arbitrator node:

- Serviceguard cluster node. Doesn't need to be connected to shared storage.
- Serviceguard LAN guidelines still apply
- Required for clusters > four nodes, and three-site disaster tolerant architectures



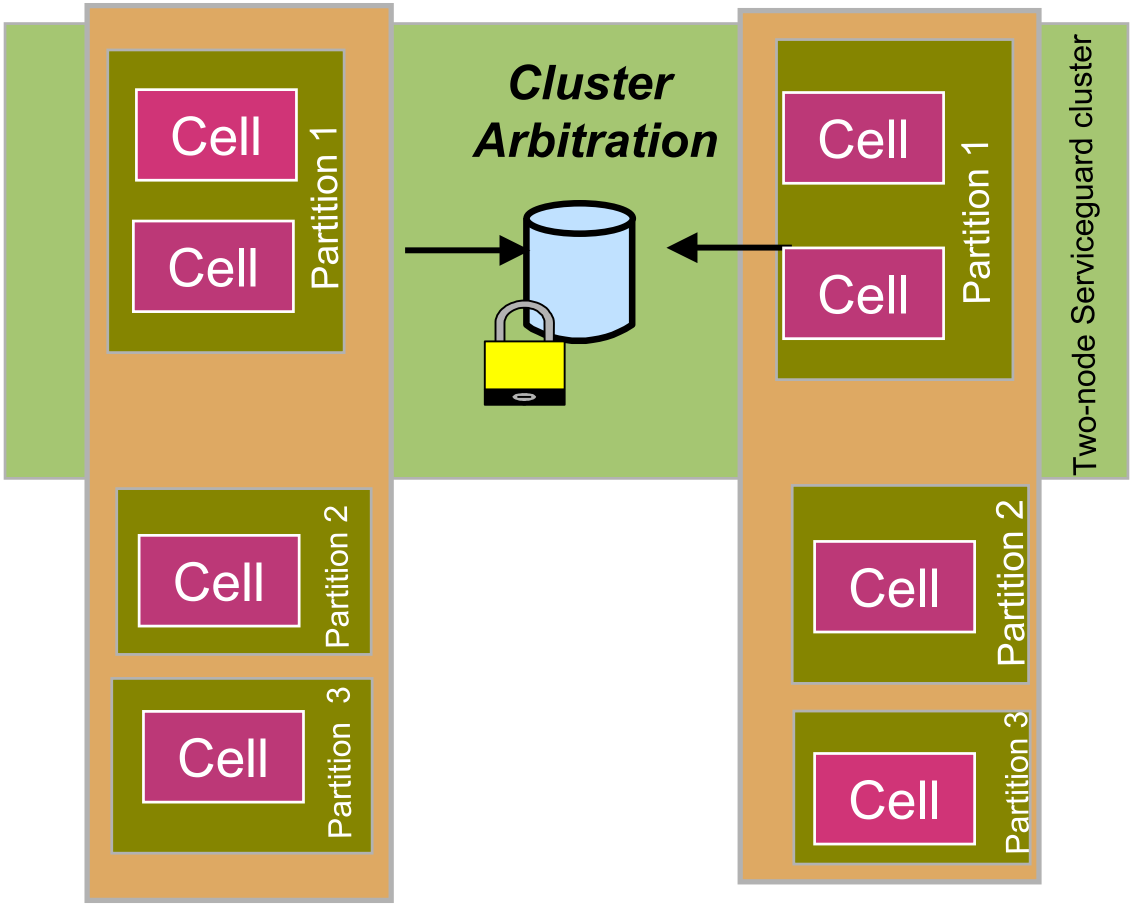
Example: "Cluster in a Box"



- Entire cluster is susceptible to standalone node SPOF's
- Preferred design is to spread cluster nodes among independent cabinets:
  - 2 x SD 32-way cabinets are preferred to 1x 64-way cabinet
- Power inputs should be connected to independent power circuits
  - Arbitrator should be powered independently of the cluster nodes
  - Root mirror should be on separate circuit from root volume

Design Principle: Cluster should extend beyond a single cabinet

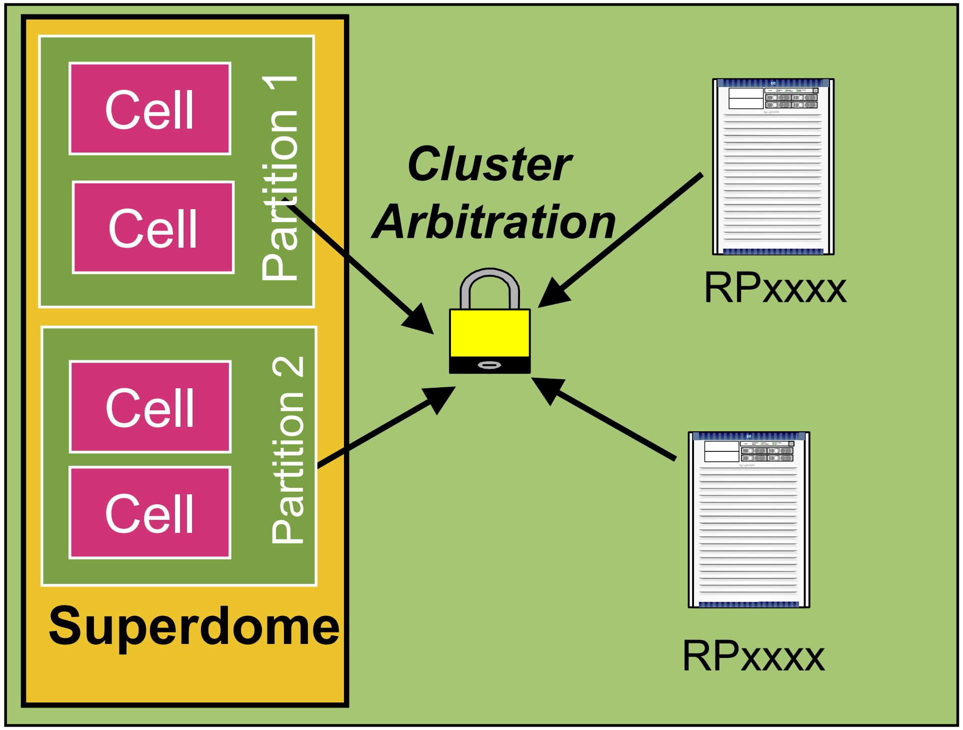
Example: Multi-Cabinet Cluster



Design Principle: Cluster should extend beyond a single cabinet

Example: Mixed Node Cluster

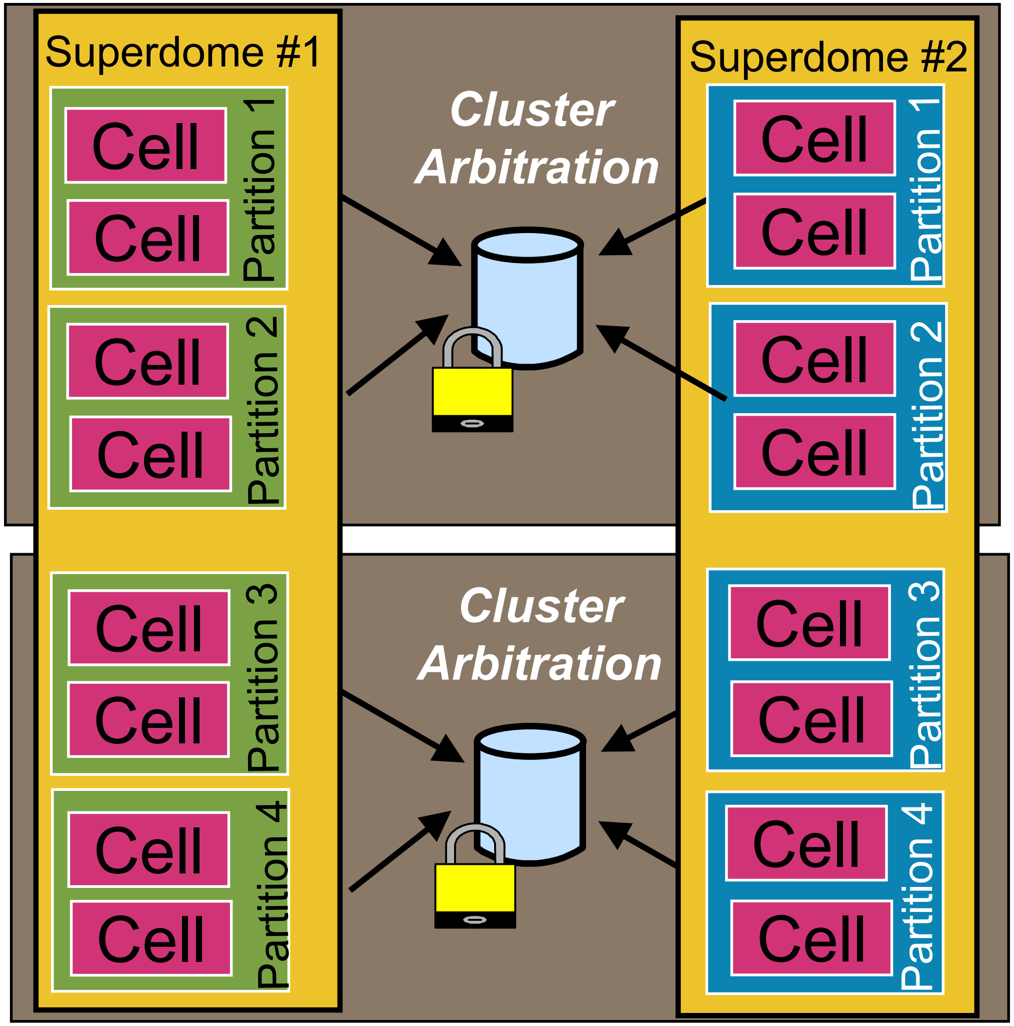
Four-node Serviceguard cluster



- Single cabinet should not contain > 50% of cluster
- Arbitrator needed if single cabinet = 50% of cluster

Example: Multiple Clusters

2 x 4-node clusters within (2) Superdome cabinets



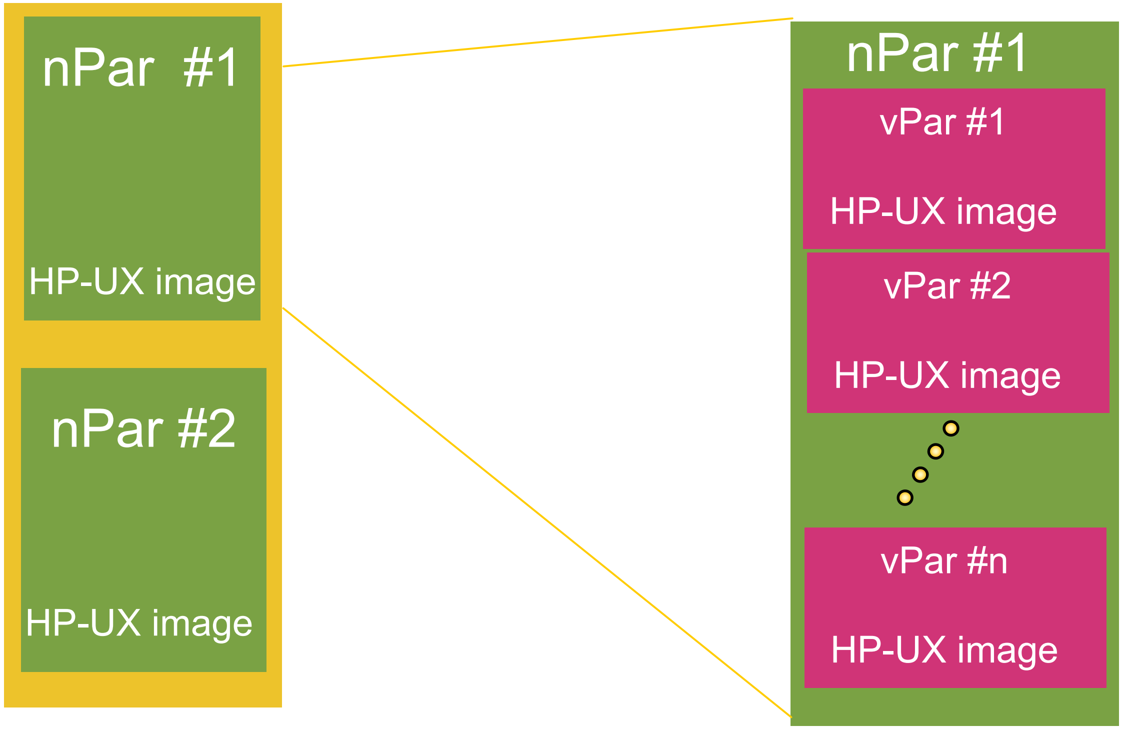
- Cluster lock cannot be shared by multiple clusters
- Single quorum server can support multiple clusters

Cluster #1

Cluster #2

nPars vs. vPars

nPar: hard partition within a cabinet **vPar:** Soft partition within a nPar



Partitionable Server

Single nPar

More electrical separation

More dynamic flexibility

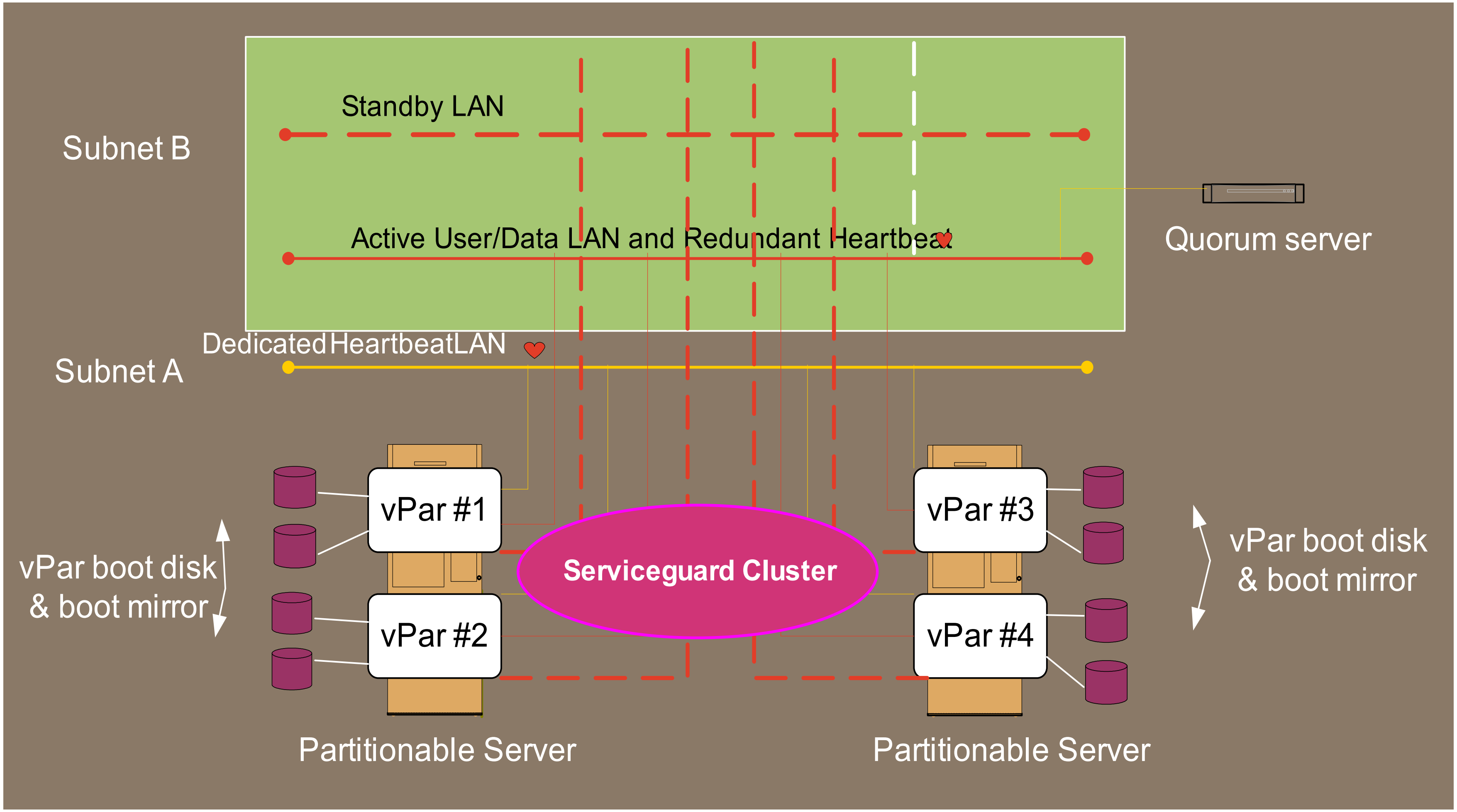
Design Principles – vPars

- Configured with independent boot disks
- I/O cards with multiple ports are not shared by vPars
- Only one vPar owns all ports on any multi-port I/O card
- Majority of nPar configuration guidelines apply to vPars

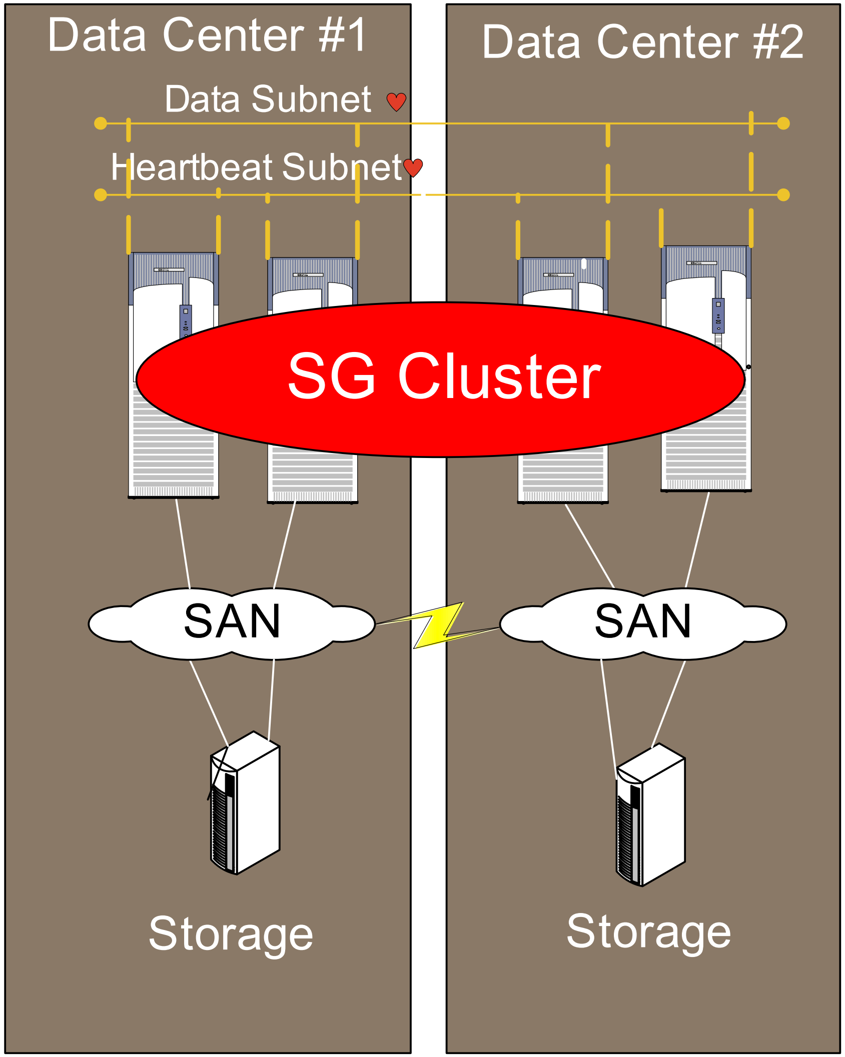
Design Principles – vPars as Cluster Nodes

- A vPar can be thought of as a node in an HA cluster
- Configure a vPar with at least two CPUs
- Add the following types of I/O cards for each vPar:
  - I/O cards for primary boot and alternate boot mirror
  - LAN cards for dedicated HB, active LAN, standby LAN
  - I/O cards for shared disk primary path and alternate path
  - I/O card for removable media (DVD-ROM/DDS-DAT)
- Combo card support can ease vPar I/O requirements
- Requirements will determine best partitioning solution

Example: vPar Cluster LAN Architecture



Disaster Tolerant Architecture Guidelines



- Minimize single-site SPOF’s
- Configure multiple servers per site, if possible
- Redundant physical paths for site-to-site cabling, such as networking and storage
- Three-site architecture is preferable to two-site architecture

Summary –
HA Design Principles

General Design Principles:

- Minimize potential single system SPOF’s:
  - Multiple cell boards per nPar
  - Redundant I/O chassis per nPar
  - Redundant I/O cards for boot storage, data storage, and networking
  - Redundant power inputs to server
  - Path failover software for storage and networking

Cluster Design Principles:

- Redundant cluster heartbeat paths
- Configure cluster nodes across independent Superdome cabinets
- Cluster arbitrator should be powered independently of the cluster nodes
- Configure an arbitrator if a single partitionable server contains 50% of the cluster nodes
- A single server should never be configured with a majority of the cluster nodes

Reference Resources

Superdome hardware:

- http://www.docs.hp.com/hpux/hw/index.html

Cluster Design/Documentation:

- http://www.docs.hp.com/hpux/ha/index.html

Virtual Partitioning:

- http://www.docs.hp.com/hpux/11i/index.html

HP WORLD 2003

Solutions and Technology Conference & Expo

Interex, Encompass and HP bring you a powerful new HP World.

interex

shared knowledge • shared power

encompass

AN HP USER GROUP

hp

invent