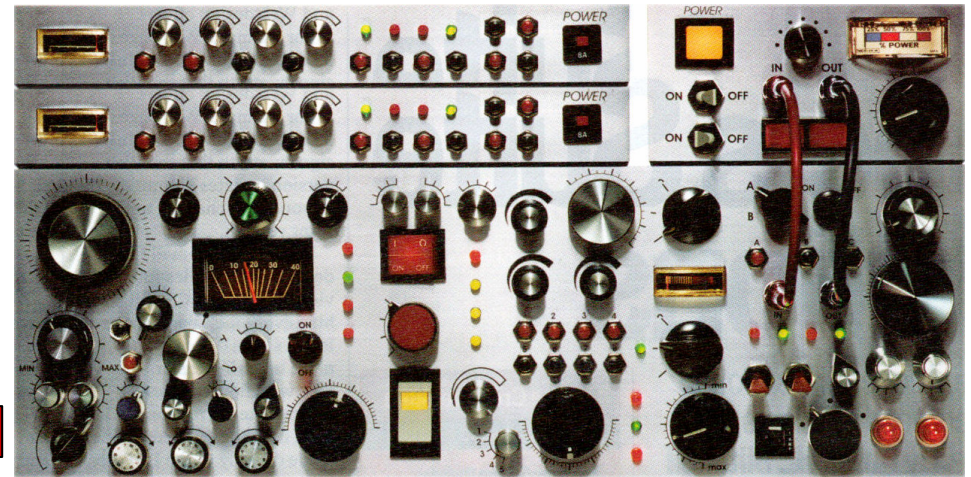


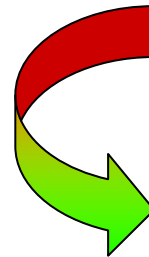


BladeCenter Simplification

Watson / Haifa Research Projects



Basil Smith
4/3/2005



Example: Integrated Server / Storage Management

Current effort focused on simple provisioning of BladeCenter with external storage (diskless blades)



IBM Director

Server & Storage Provisioning Task (SSPT)

Servers & Adapter Configuration



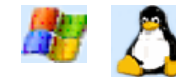
xSeries BladeCenter

Fibre Switch Configuration



- Qlogic
- Brocade

OS & Image Clone & Deployment



Storage Configuration



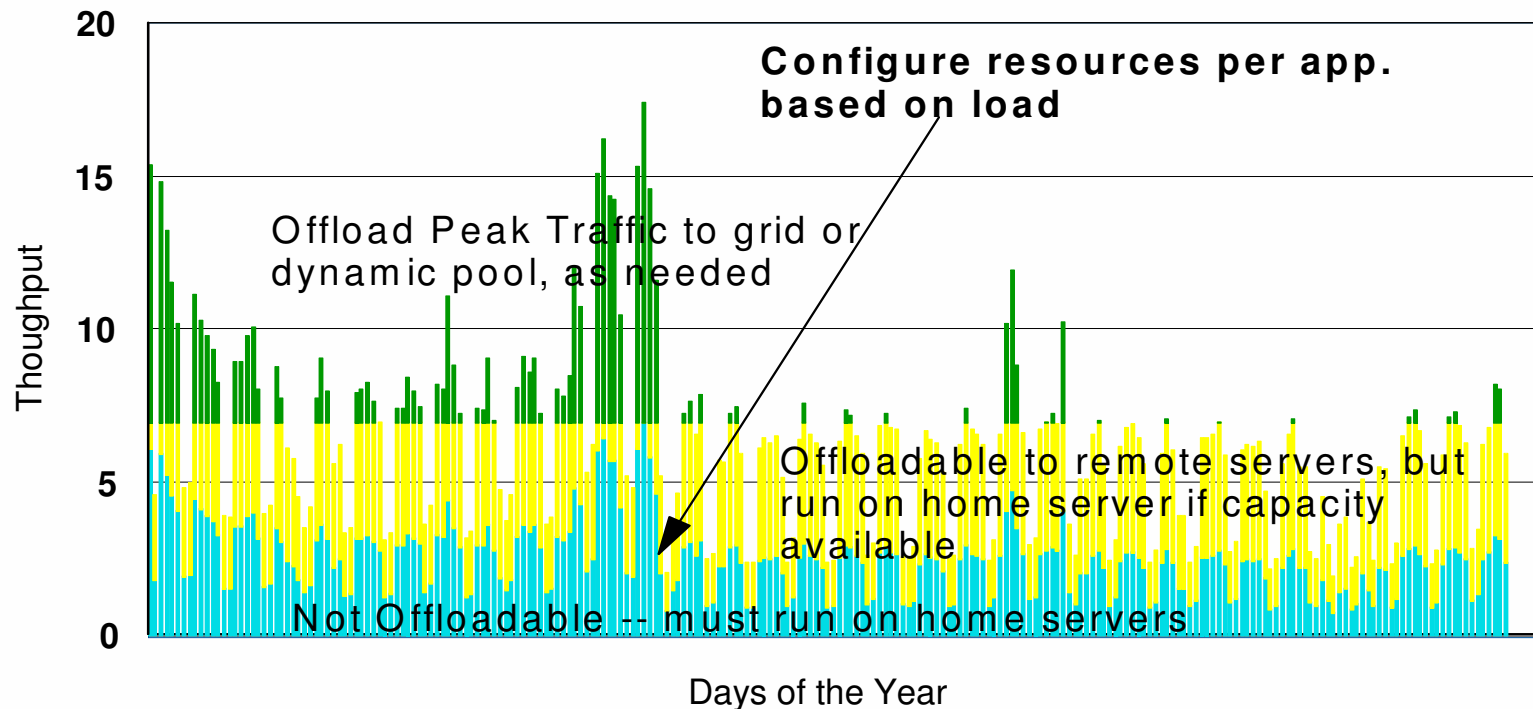
- FASTt
- Adaptec FC
- Adaptec iSCSI Blade
- Adaptec iSCSI box
- SMI-S storage

First a digression

Why is this important?

What does it have to do with Grid Computing?

Dynamic On-Demand Provisioning and WorkLoad Balancing: Motivation

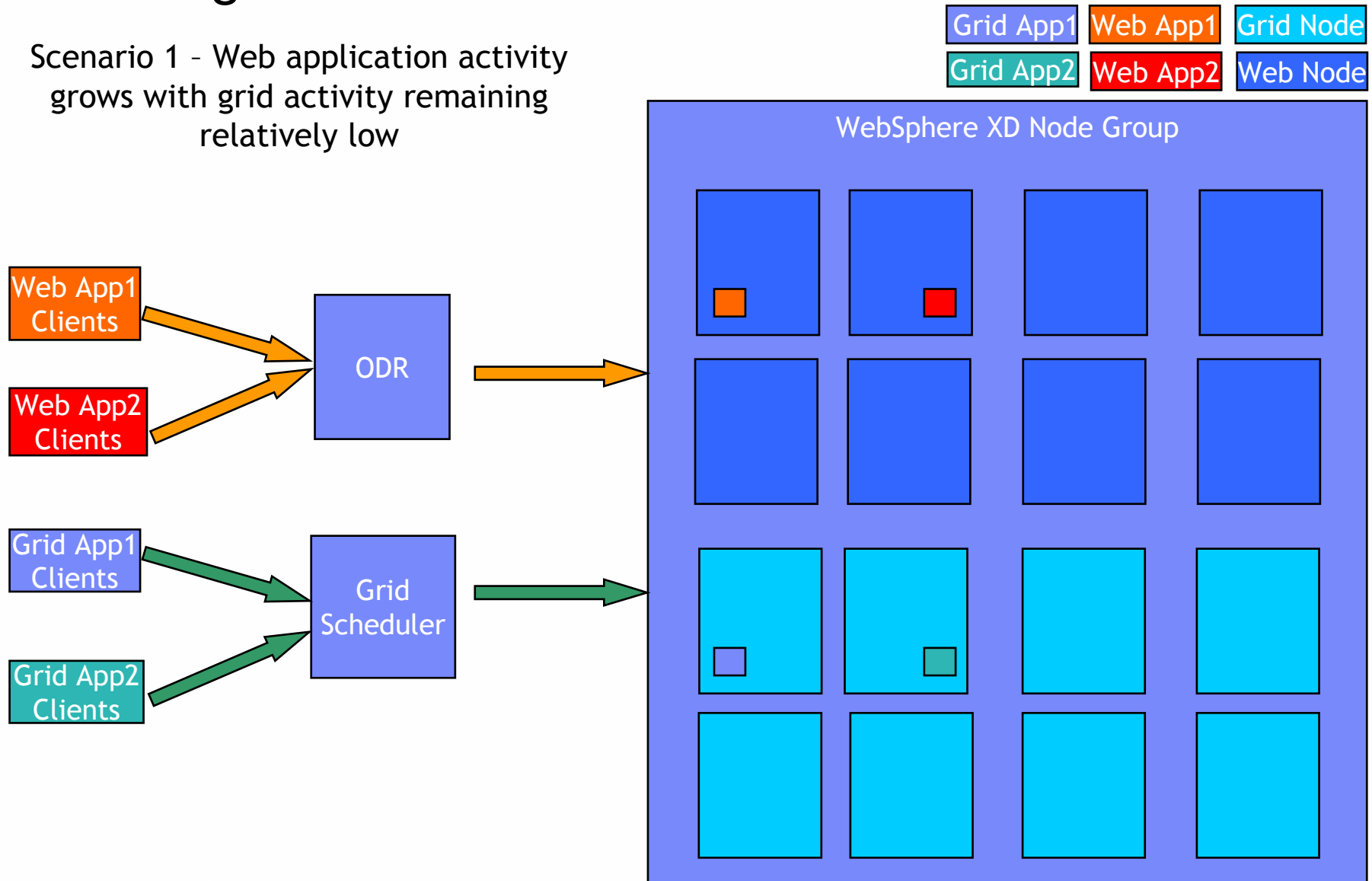


#1 and #2 items from on demand customer survey:

- (Provide) flexibility of increasing/decreasing capacity (51%)
- (Provide) choice of service levels (49%)

Balancing Grid and Web Workloads

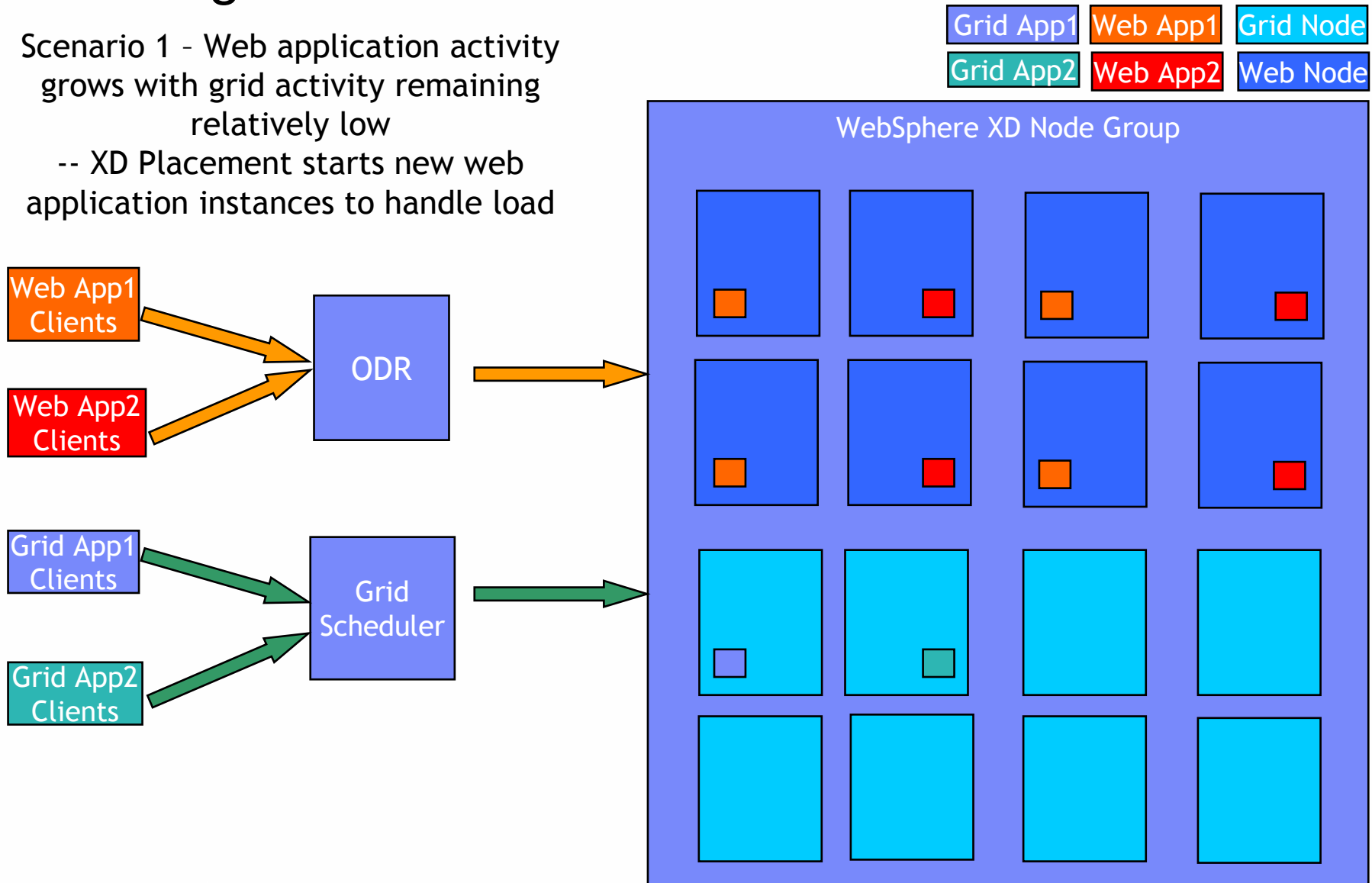
Scenario 1 - Web application activity grows with grid activity remaining relatively low



Balancing Grid and Web Workloads

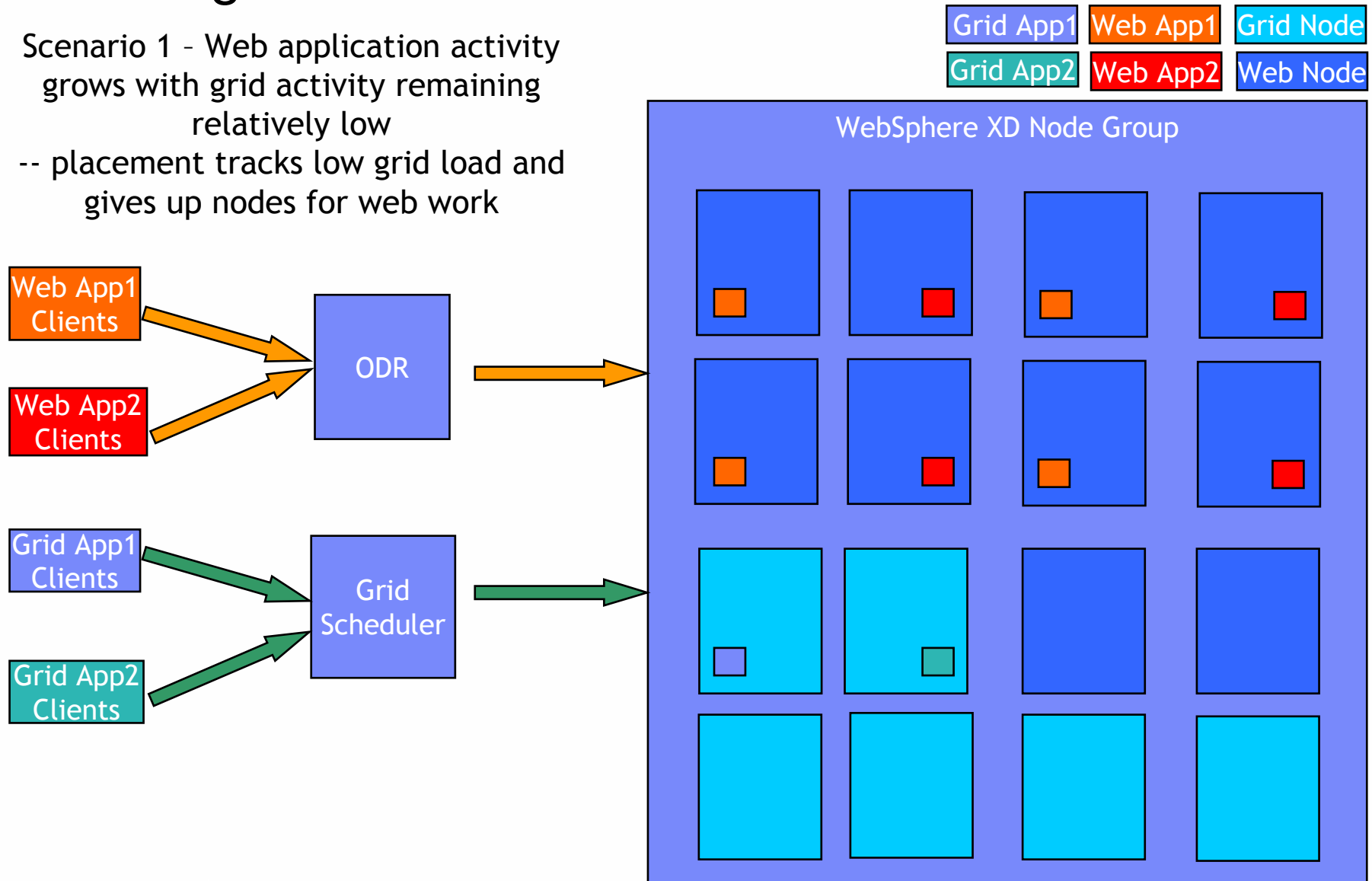
Scenario 1 - Web application activity grows with grid activity remaining relatively low

-- XD Placement starts new web application instances to handle load



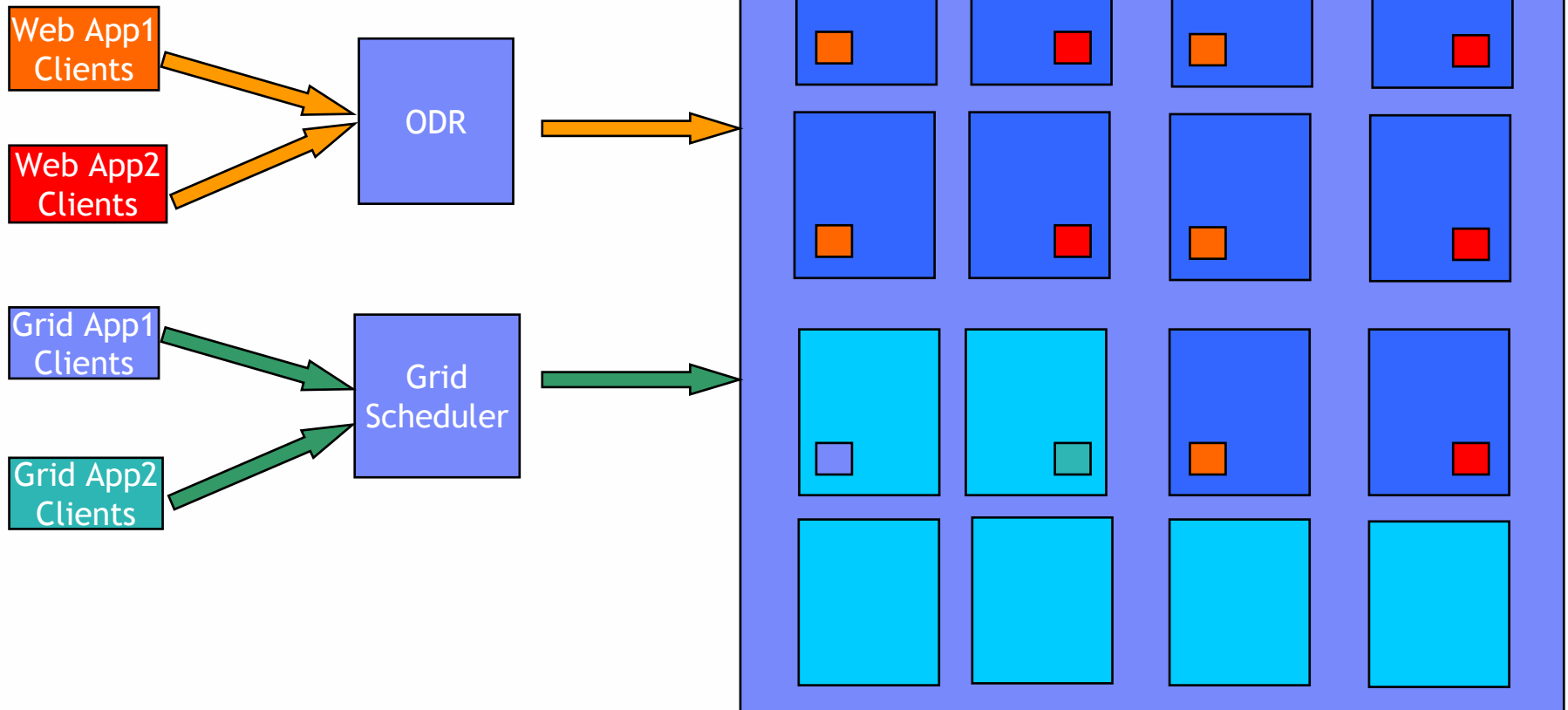
Balancing Grid and Web Workloads

Scenario 1 - Web application activity grows with grid activity remaining relatively low
-- placement tracks low grid load and gives up nodes for web work



Balancing Grid and Web Workloads

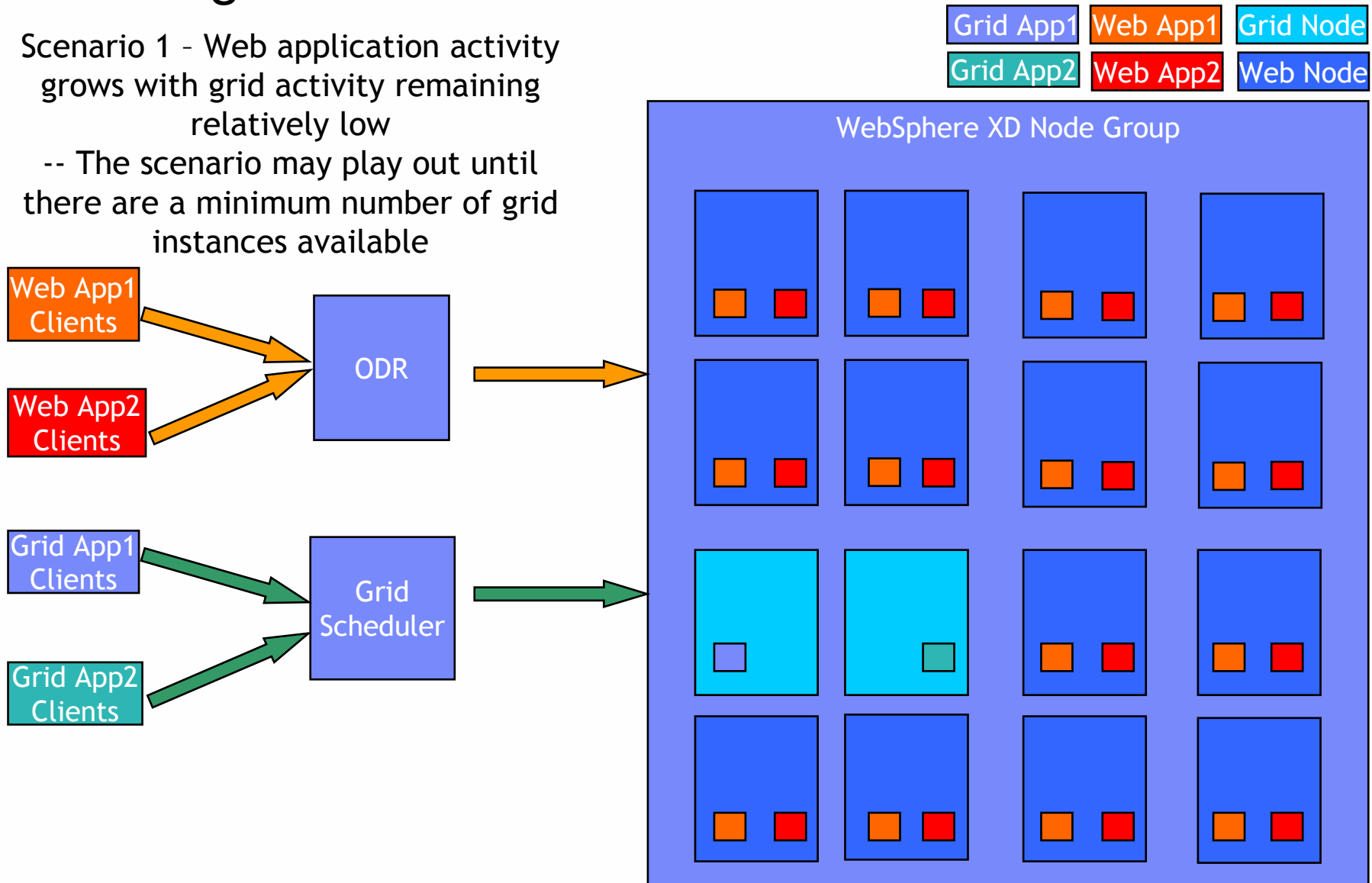
Scenario 1 - Web application activity grows with grid activity remaining relatively low
-- XD Placement continues by utilizing new nodes for web work



Balancing Grid and Web Workloads

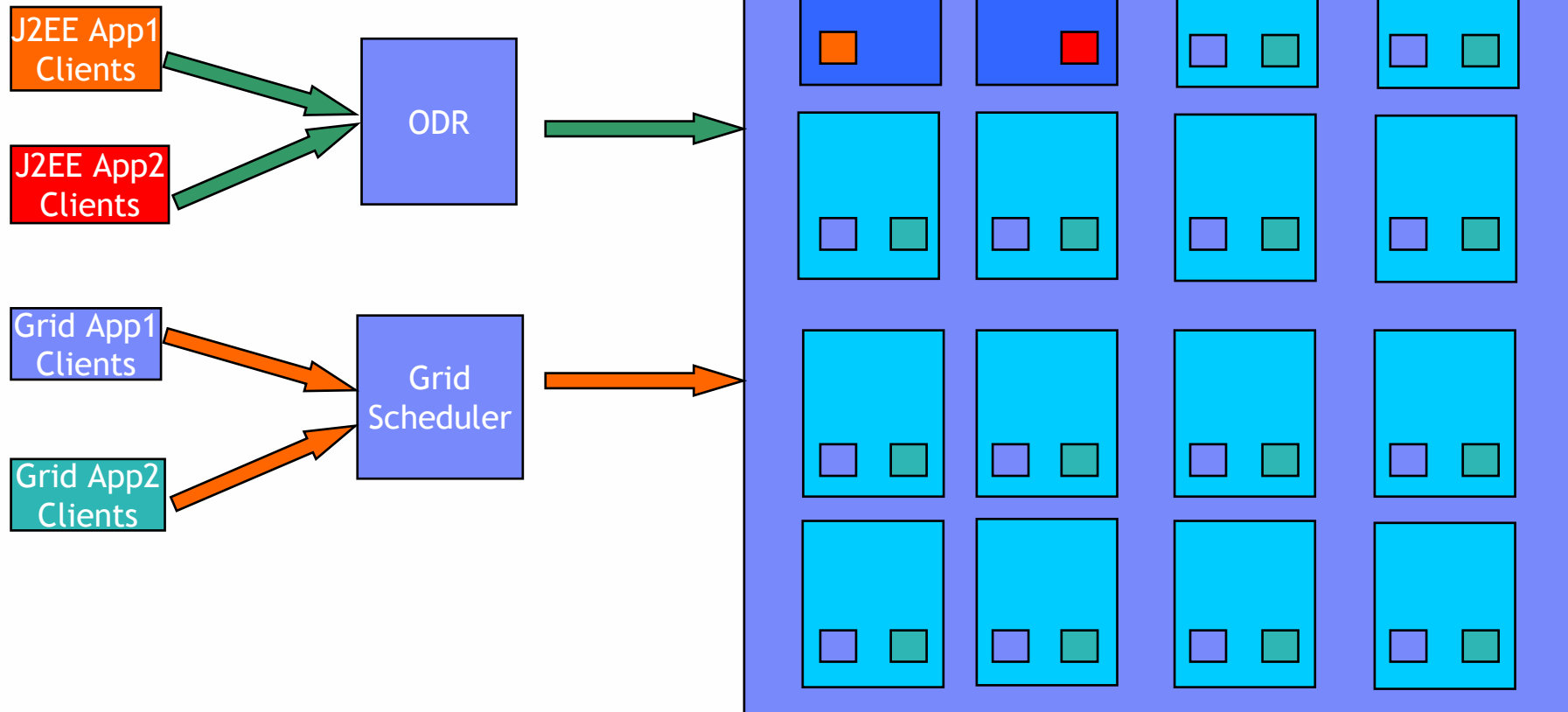
Scenario 1 - Web application activity grows with grid activity remaining relatively low

-- The scenario may play out until there are a minimum number of grid instances available



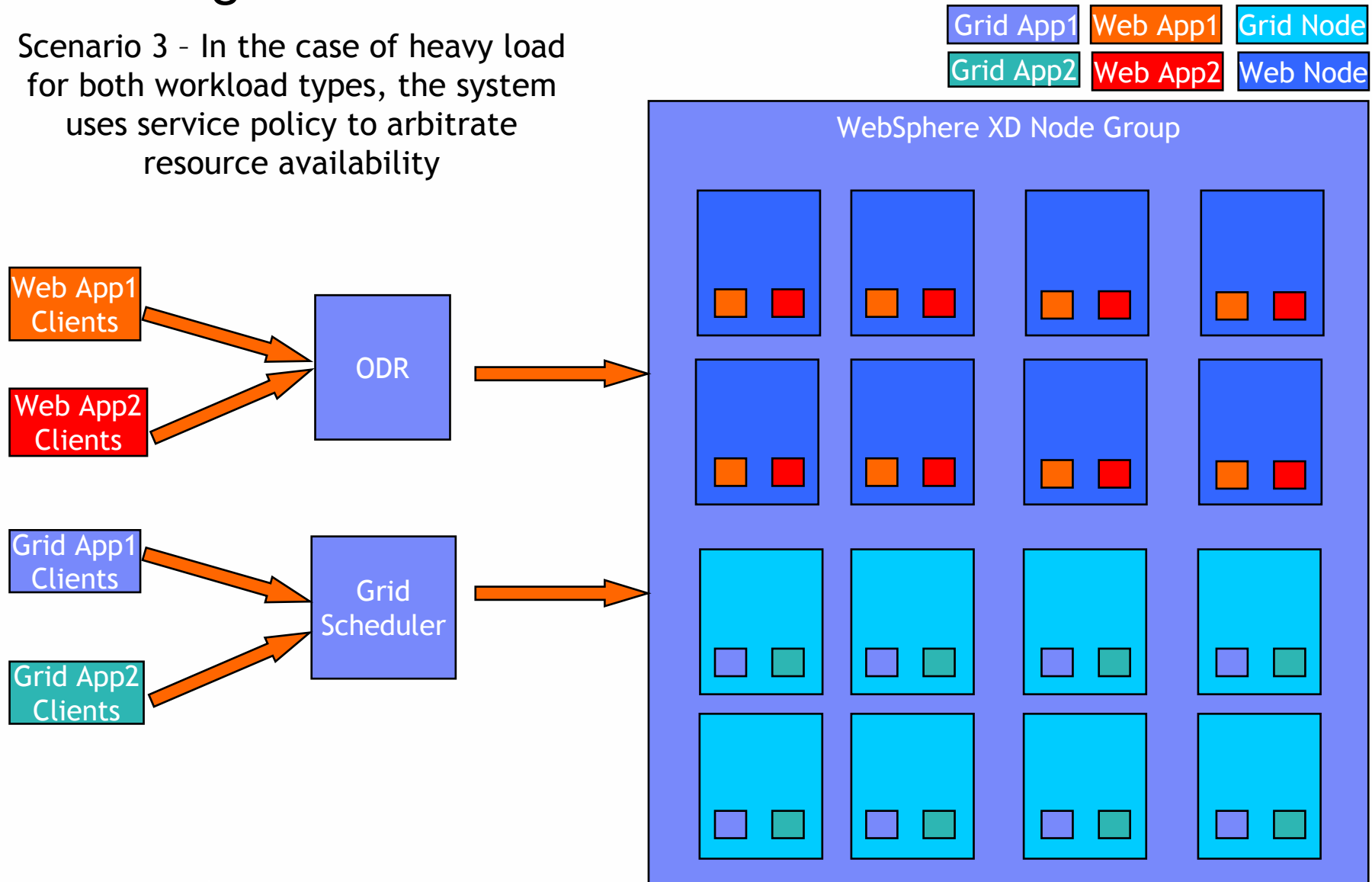
Balancing Grid and Web Workloads

Scenario 2 - Grid application activity grows with web activity remaining relatively low
-- The scenario plays out in the opposite way...



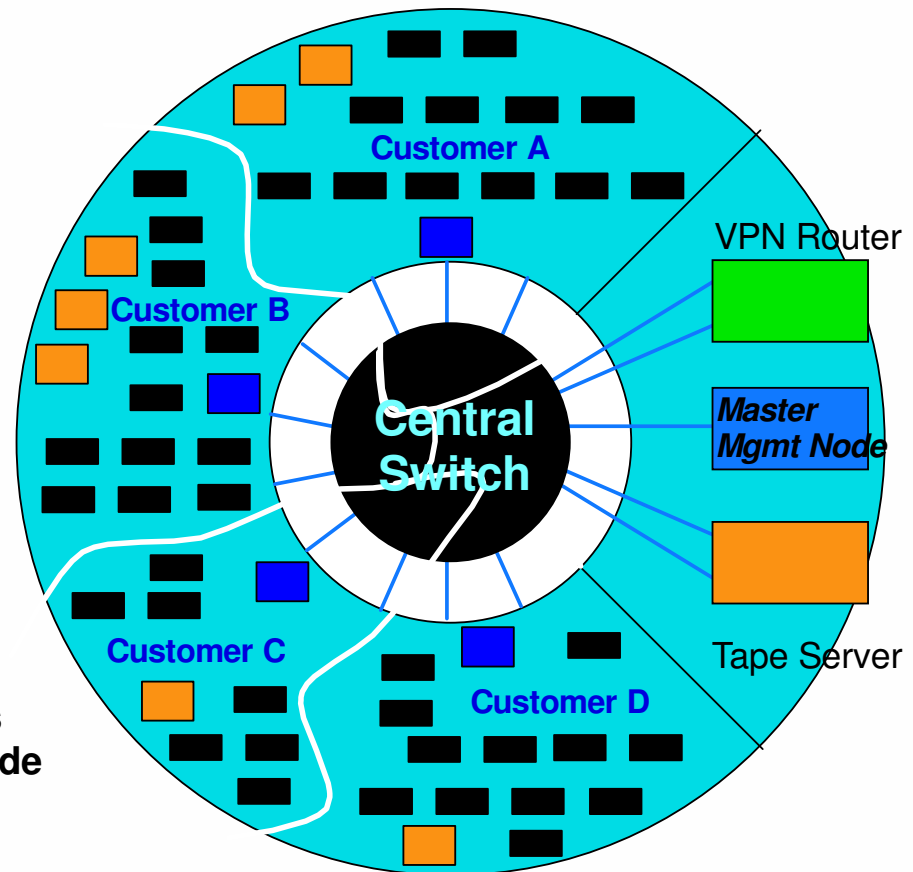
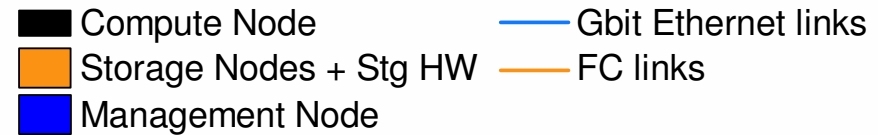
Balancing Grid and Web Workloads

Scenario 3 - In the case of heavy load for both workload types, the system uses service policy to arbitrate resource availability



HPC Utility Resources

One or more Customer partitions, each of which is a Virtual Cluster carved out of the resources available to the utility



Number of Management Servers and/or Storage Servers will decide total number of supported customers

Recapping the digression

Why is this important?

What does it have to do with Grid Computing?

It's the underlying infrastructure that enables Grid computing

The ability to quickly (seconds to few minutes) create and provision servers (real or virtual) and

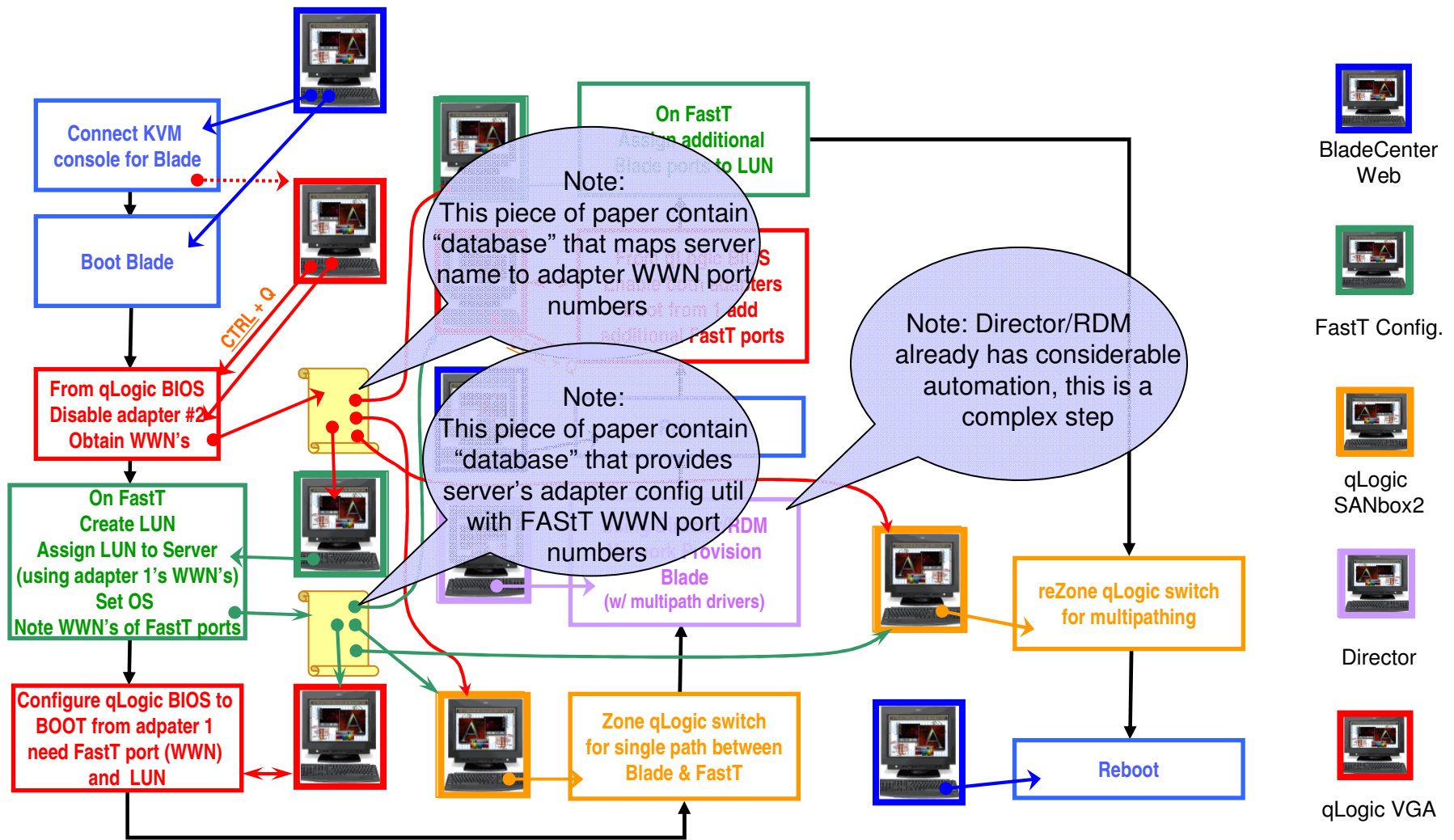
The ability to quickly mothball unneeded servers

Now

Why is this hard?

Integrated Server / Storage Management – Why is this hard?

e.g., Provision WAS image on Blade 5 with 10 GB C: drive from FASTT

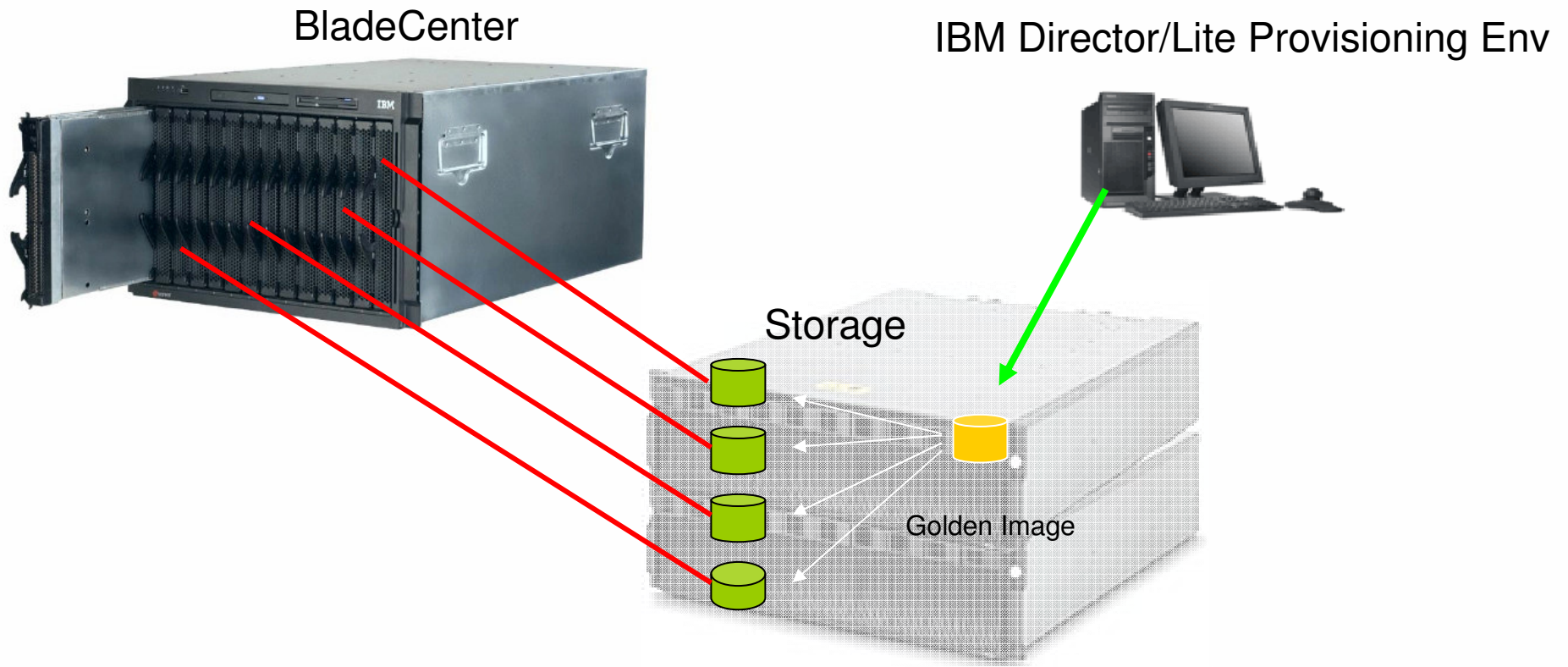


Integrated Server / Storage Management - Target

- Simple goal oriented management
 - Goals should be in easily understood customer context
 - e.g., Deploy WAS image on Blade 5 with 10 GB C: (boot) drive from storage box 2
 - Configuration and deployment decisions choices are still manual, but in familiar easily understood customer terms
 - Most frequently from a server perspective
 - Avoids manual mapping of goals to obscure complex set of steps
 - Provides simple automation of repetitive procedures
 - e.g., Deploy WAS images on Blades 1-14 each with 10 GB C: drives form storage box 2
 - e.g., Divide iSCSI Blade storage equally among 12 diskless Blades, provisioning each blade with a boot volume, and load disk image onto each volume
- Provide interfaces so that policy driven tools can invoke goal driven automation tools (Grid Schedulers for example)
- Leverage server storage synergies to improve underlying process
 - Examples:
 - Using storage subsystem copy functions to improve clone deployment process
 - Using LUN reassignment in moving images from one server to another

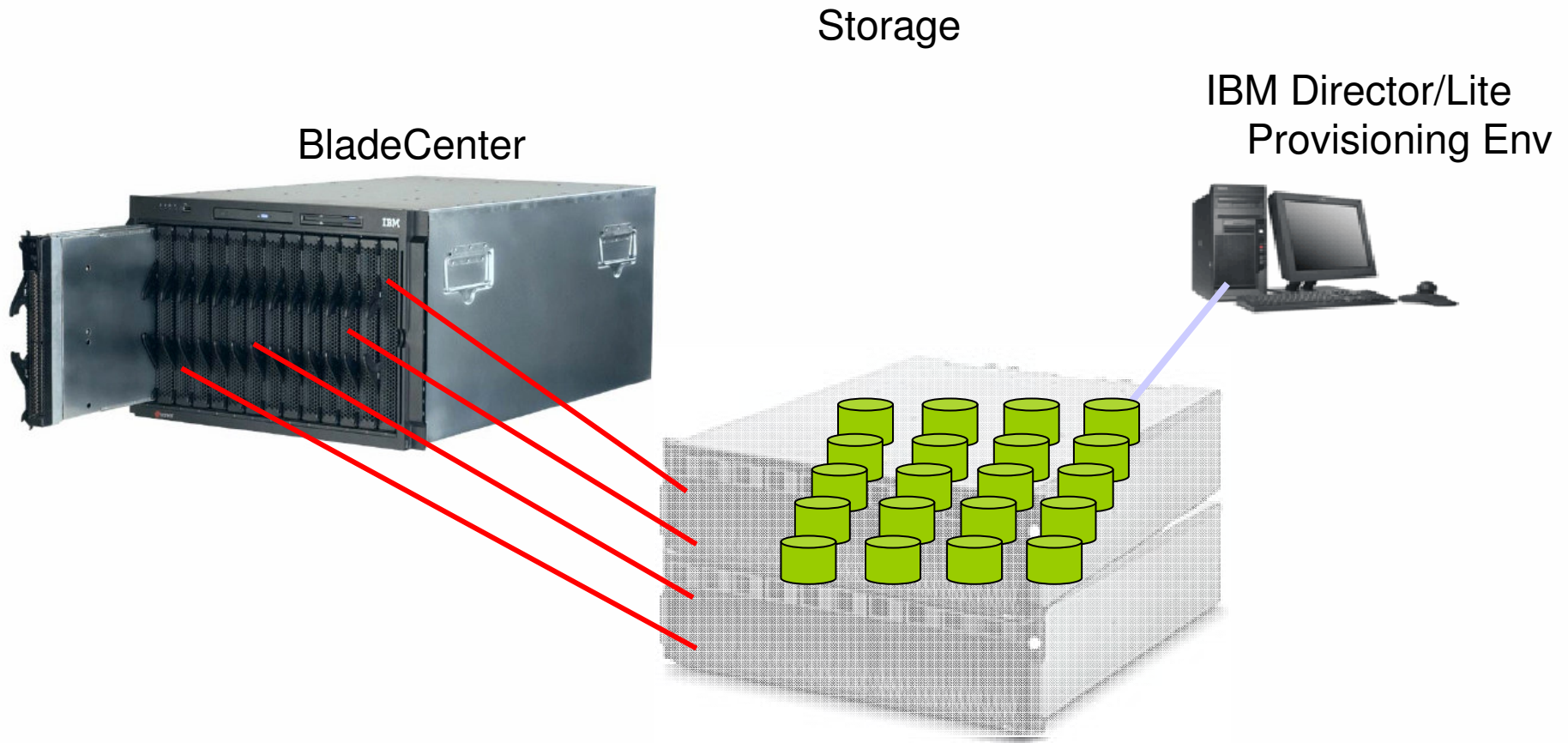
Rapid Provisioning with Flash Deploy

- Install a server instantaneously from a master image with the flash-copy storage function (<5 minutes)
- Volume copy
- Tie in to Director/Lite Provisioning environment



Rapid Provisioning with Ready to Boot Images

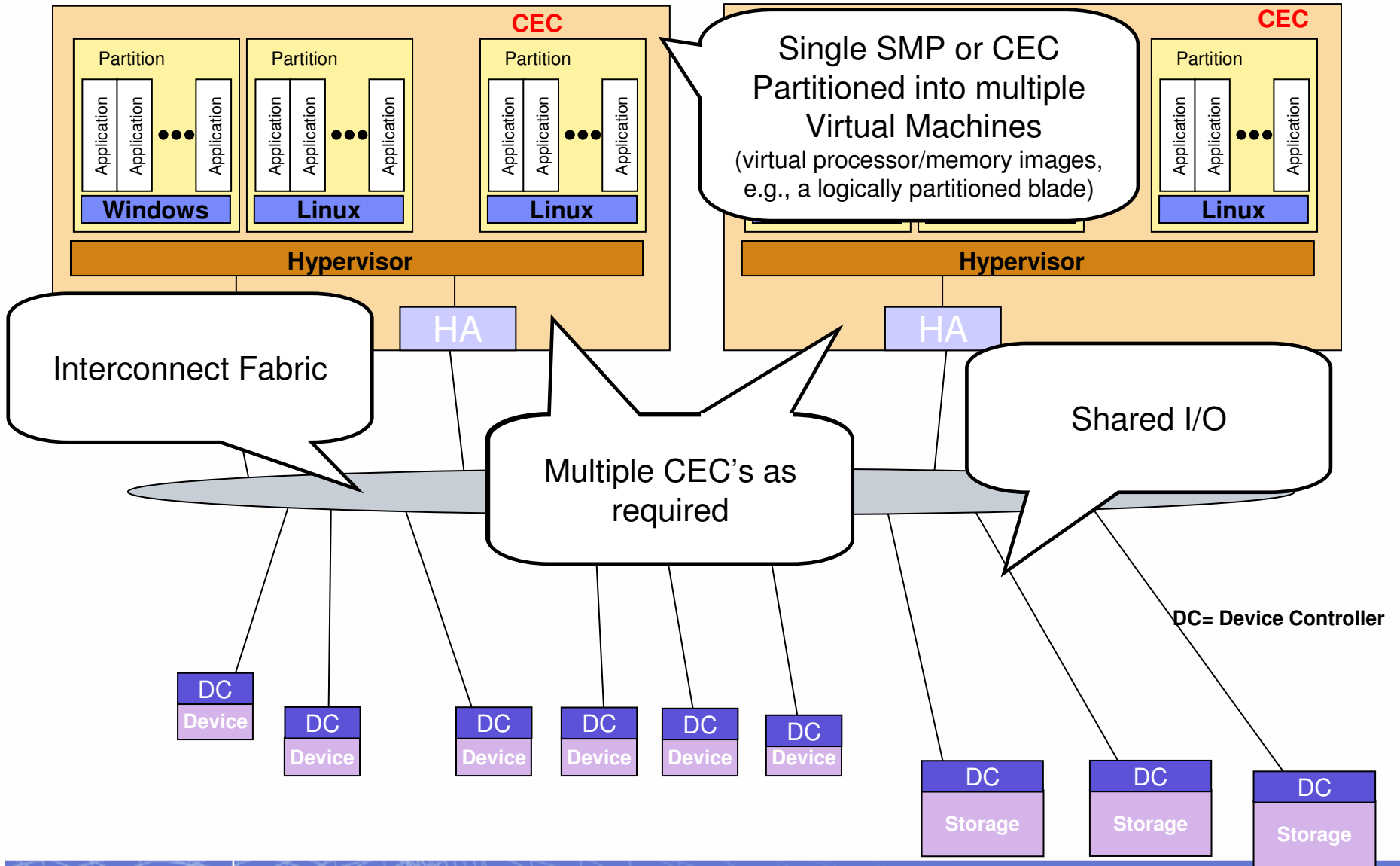
- Preinstalled images ready to boot.
- On-Demand mapping of images to blades
- Num. Images > Num. Blades
- Linux at night, Windows during the day
- Tie in to Director/Lite Provisioning environment



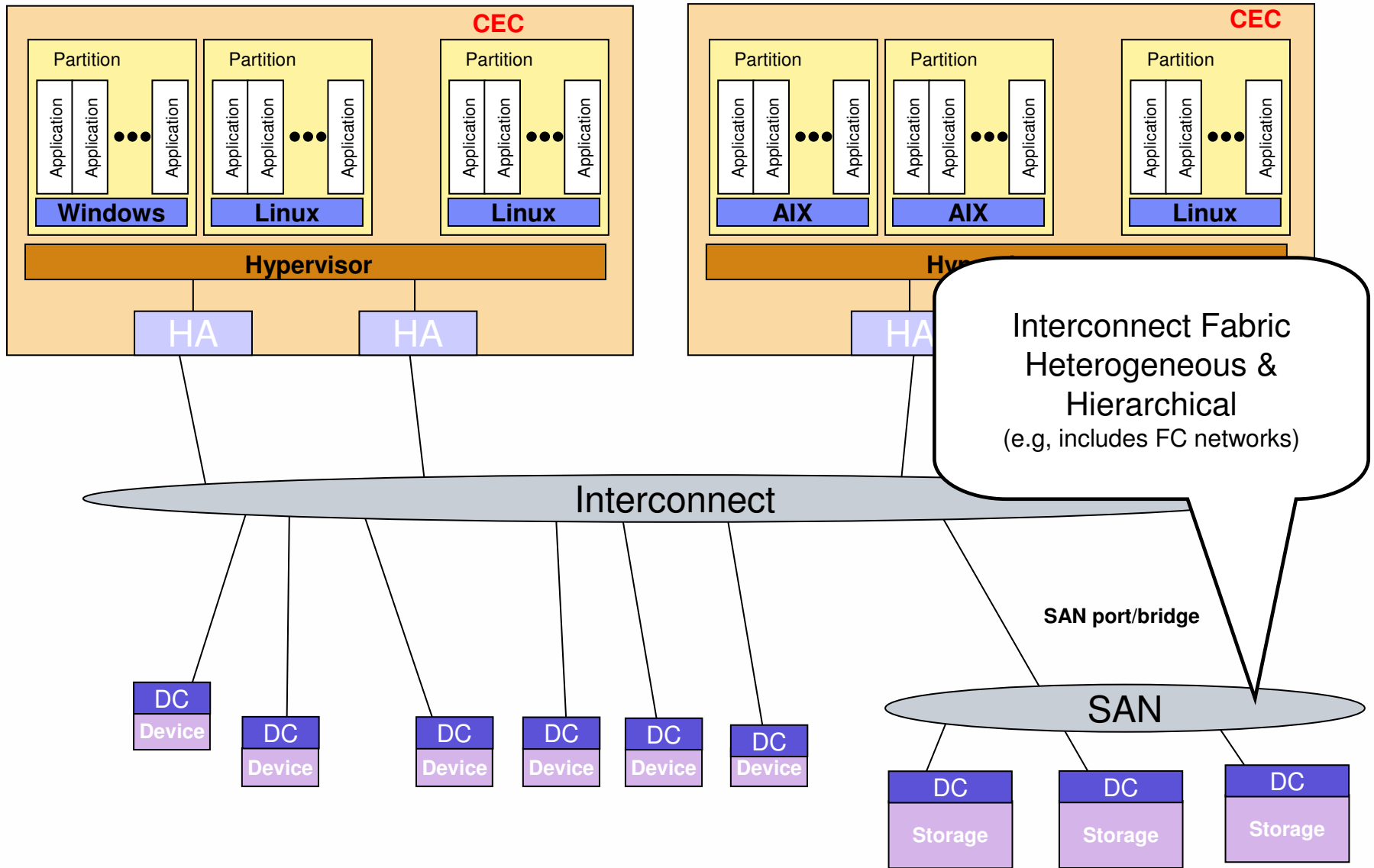
Forward looking assumption

- ✓ Compute (CPU/Memory), I/O (e.g, disk) and Network (i.e., TCP/IP) will be virtualized
 - OS images will universally move to running on virtualized machines
 - I/O resources will be virtualized and shared (i.e, on demand allocated to) virtual servers
 - Virtual (TCP/IP) networks will provide the logical structure (e.g, as seen in three and four tier web application)

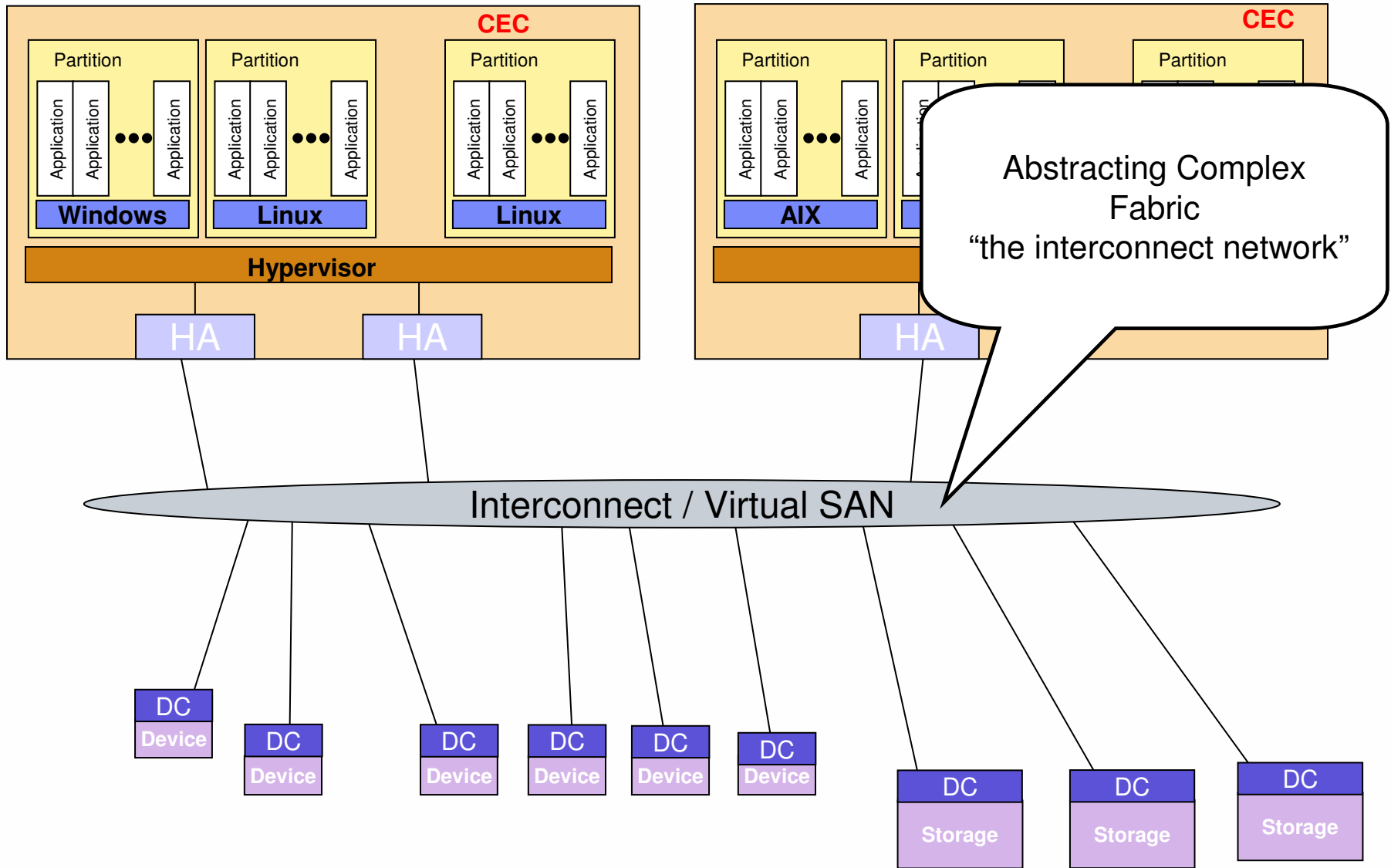
Virtualized System View



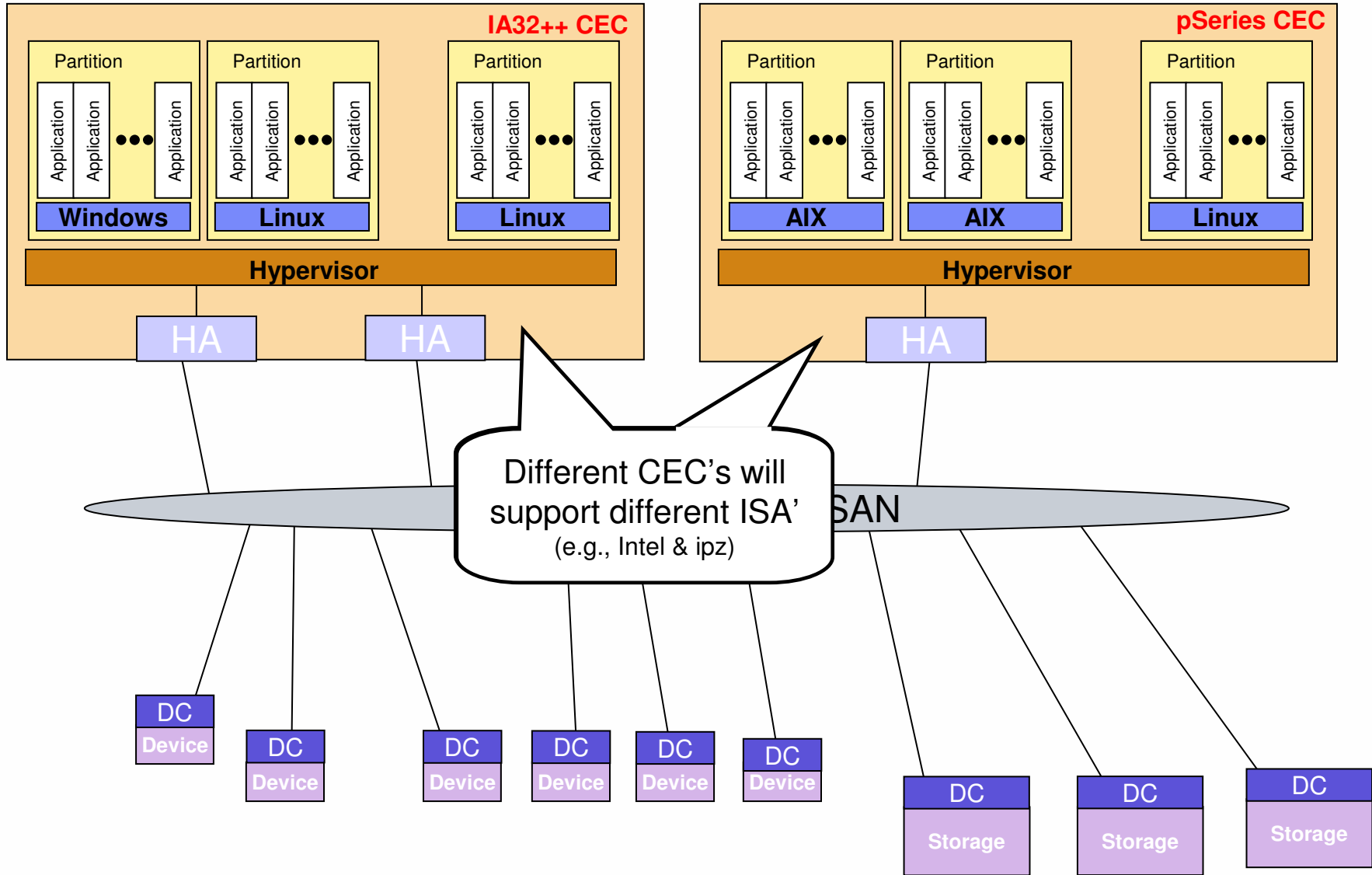
Virtualized System View



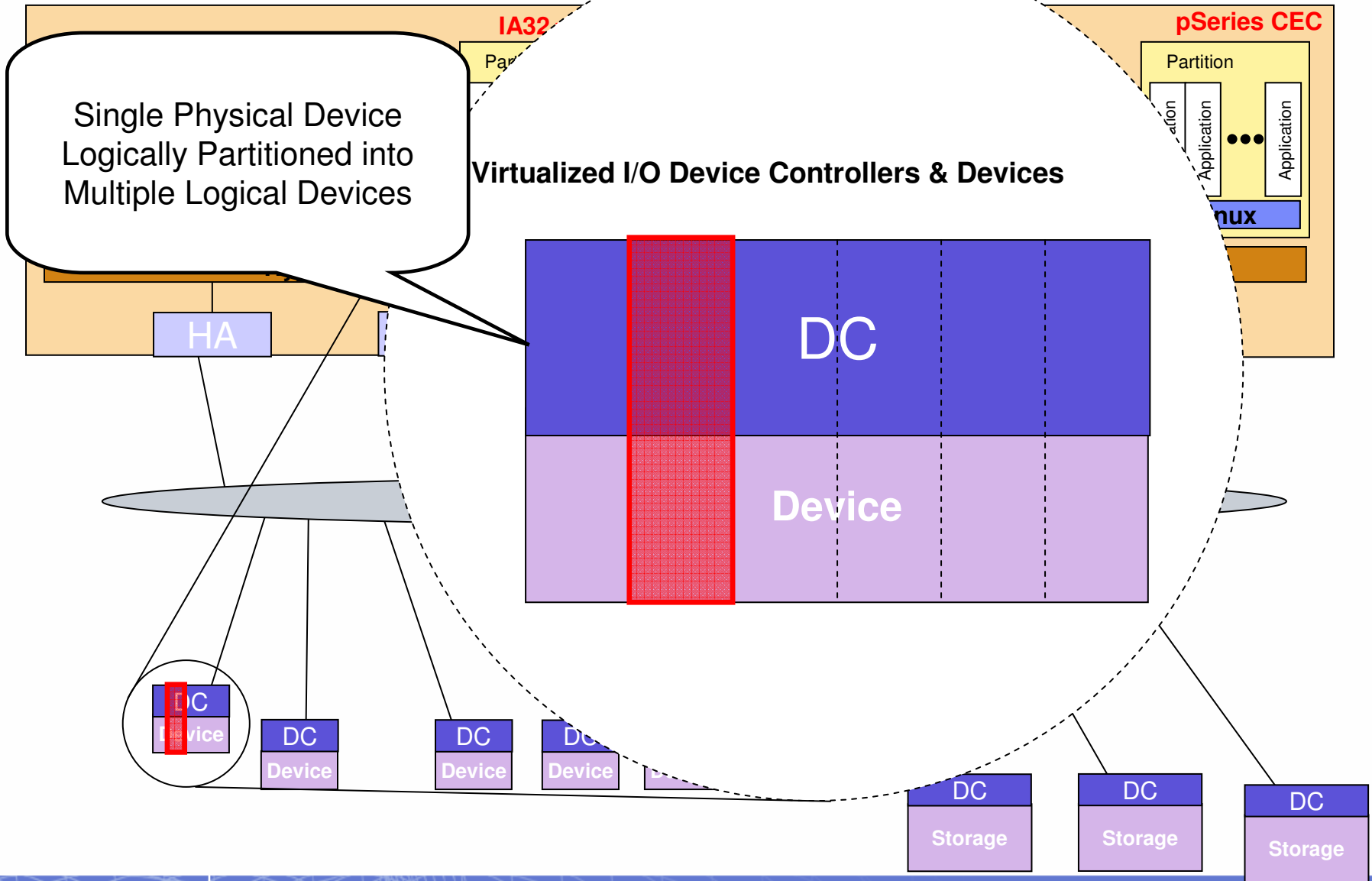
Virtualized System View



Virtualized System View

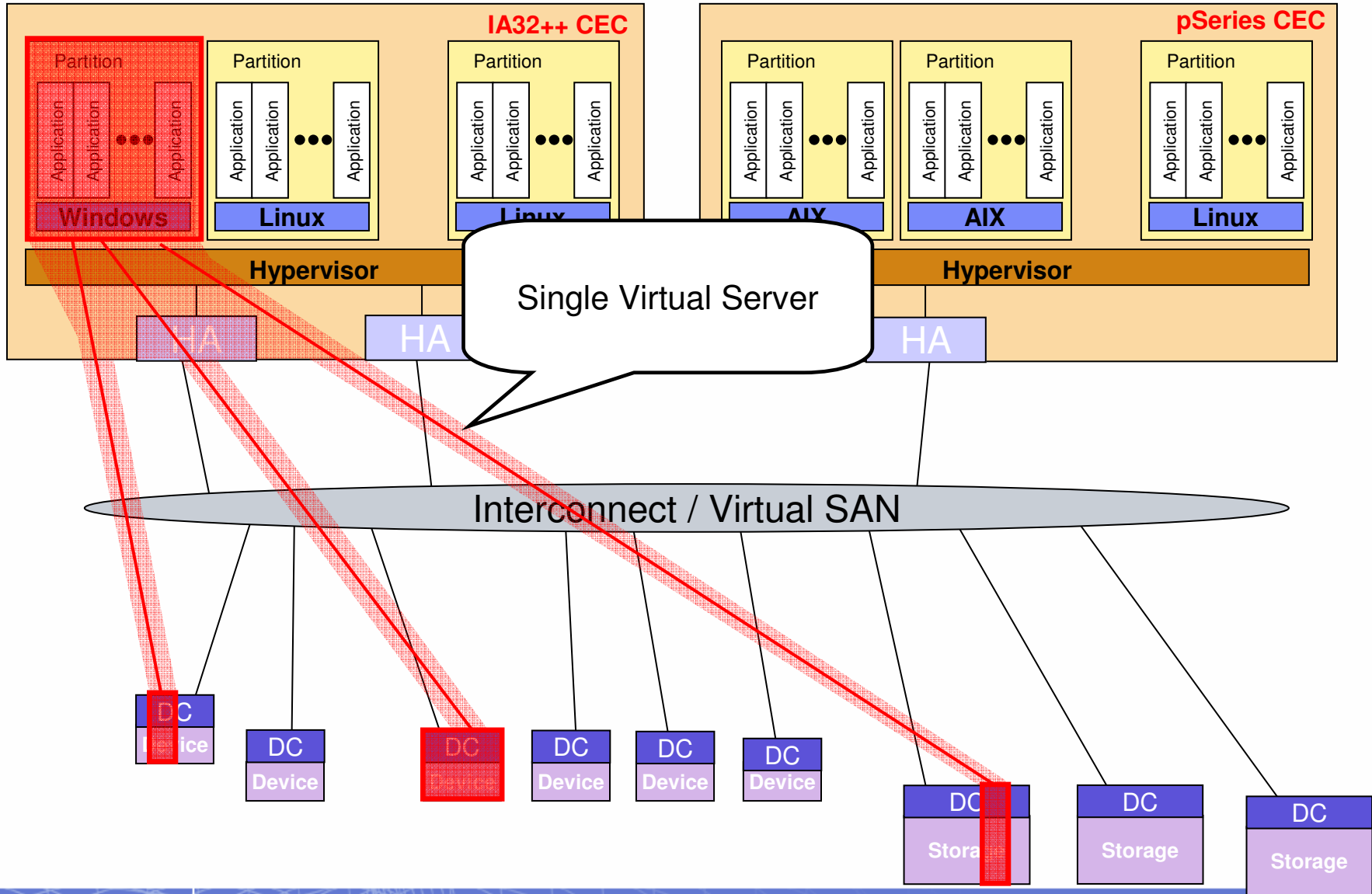


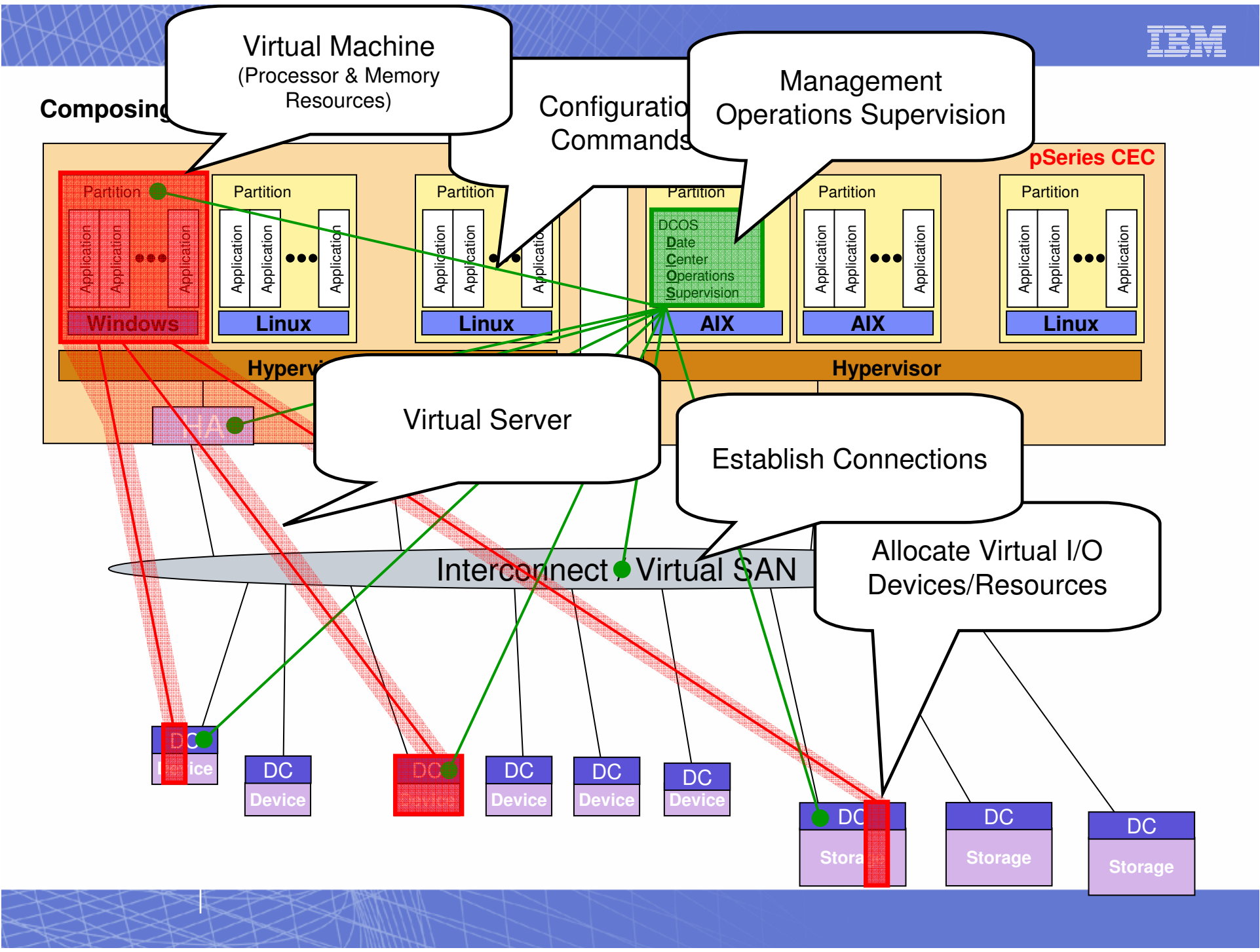
Virtualized System View



Composing a server ...

Virtual Server = Virtual Machine + Virtual I/O and connections through Interconnect





Step 1: ... Class

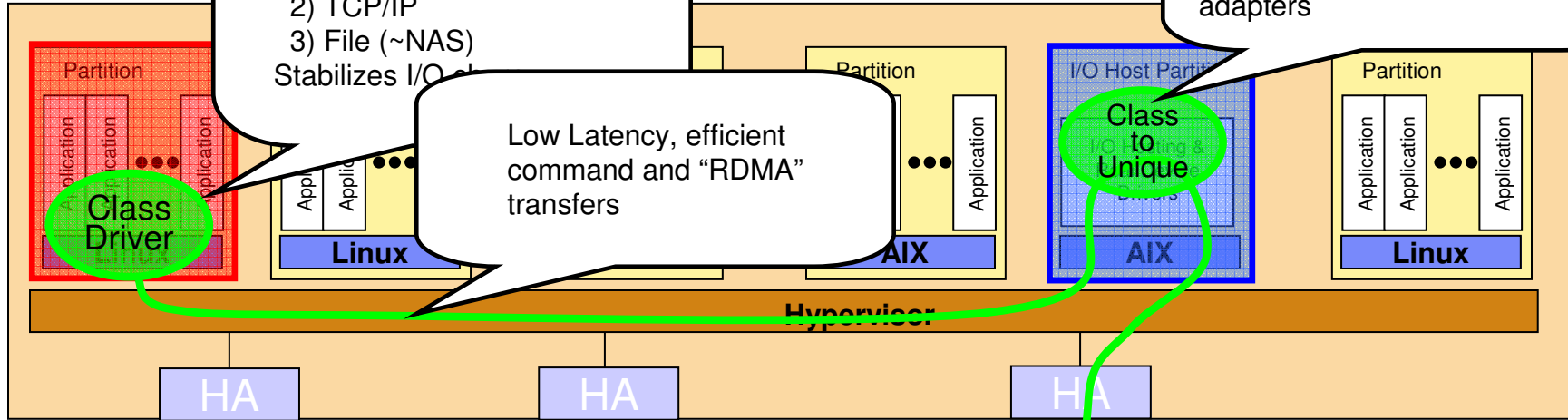
Standard stub one per each class of I/O; Classes:

- 1) Block Storage (~SCSI)
- 2) TCP/IP
- 3) File (~NAS)

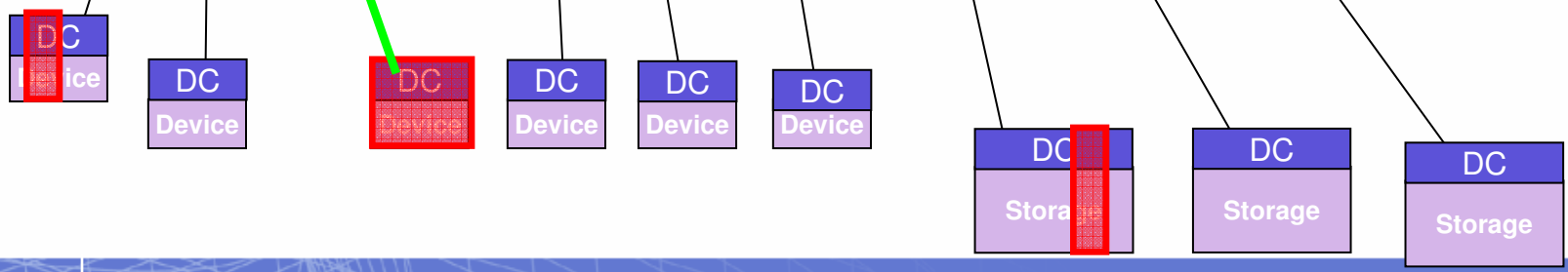
Stabilizes I/O

Low Latency, efficient command and "RDMA" transfers

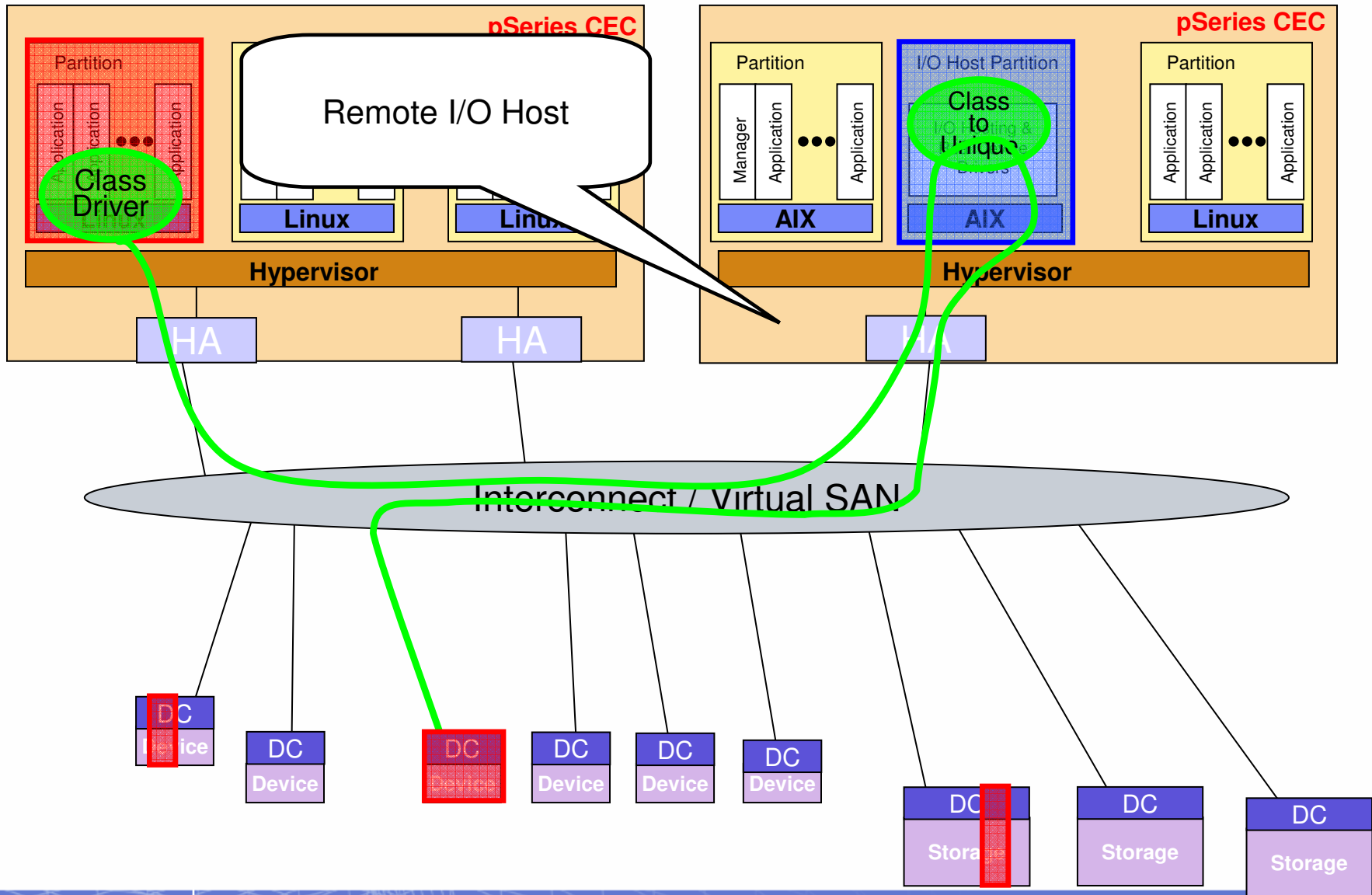
Terminates class I/O, Uses standard device drivers to drive I/O to legacy/standard I/O adapters



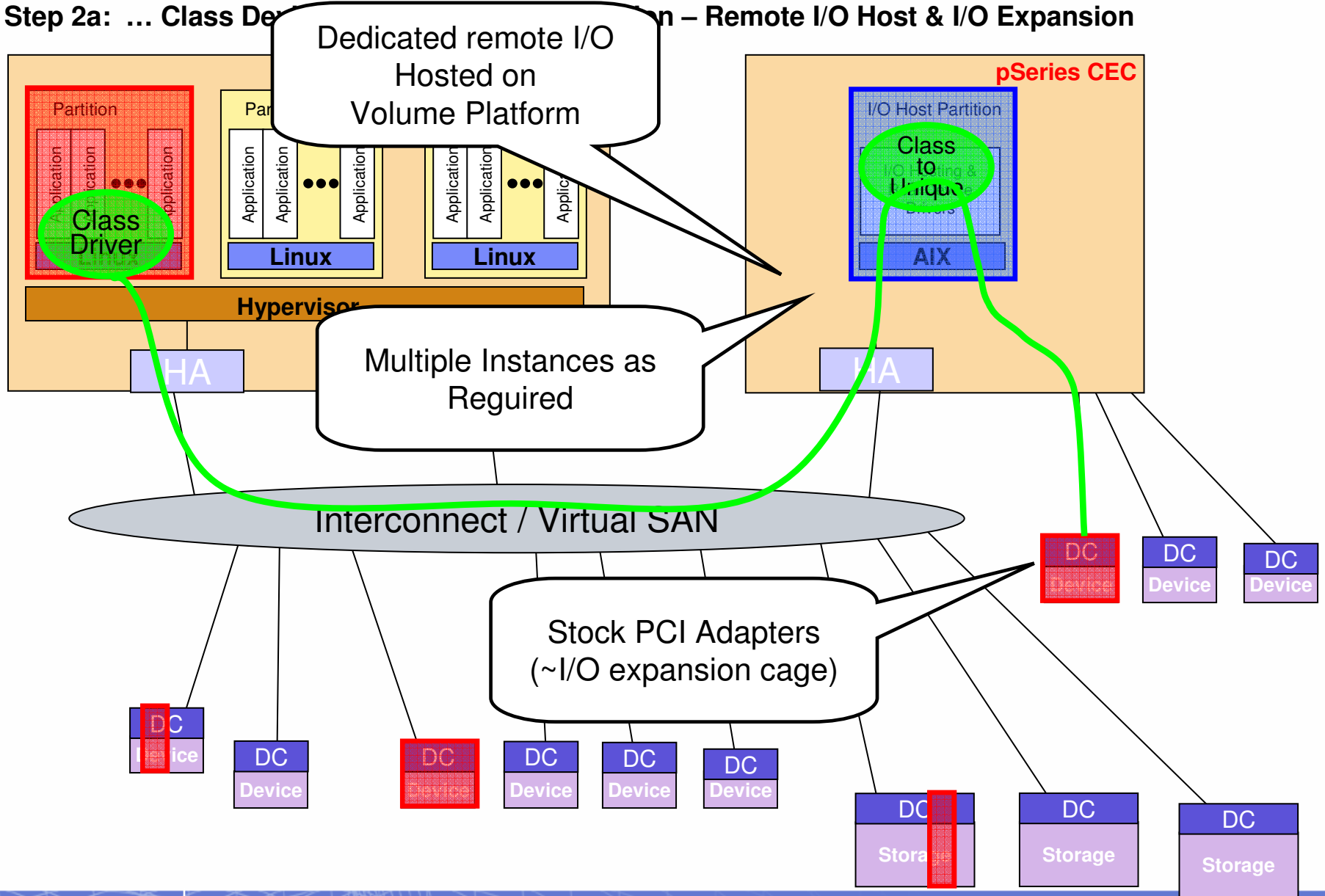
Interconnect / Virtual SAN



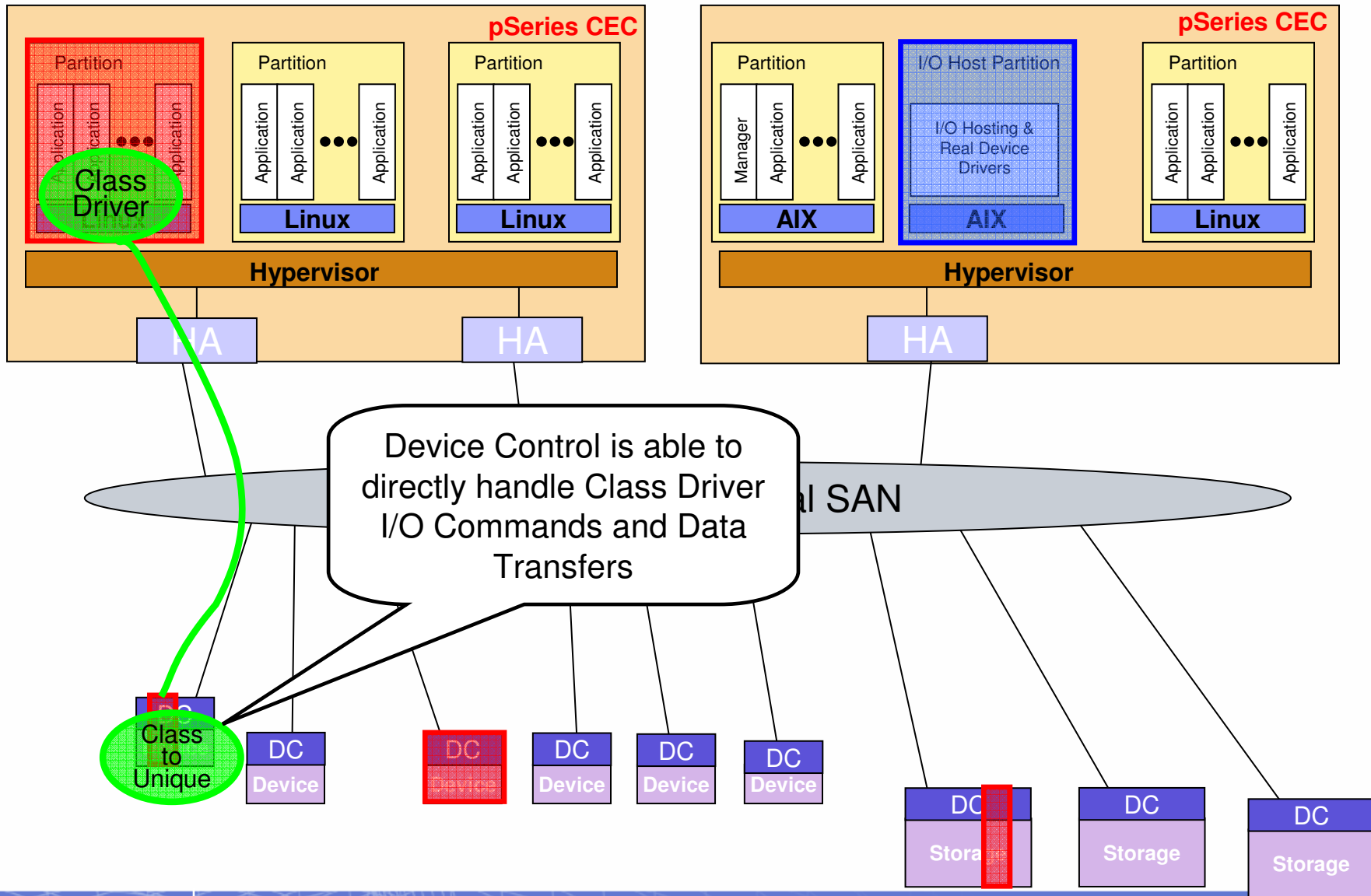
Step 2: ... Class Device Driver & I/O Host Partition – Remote I/O Host



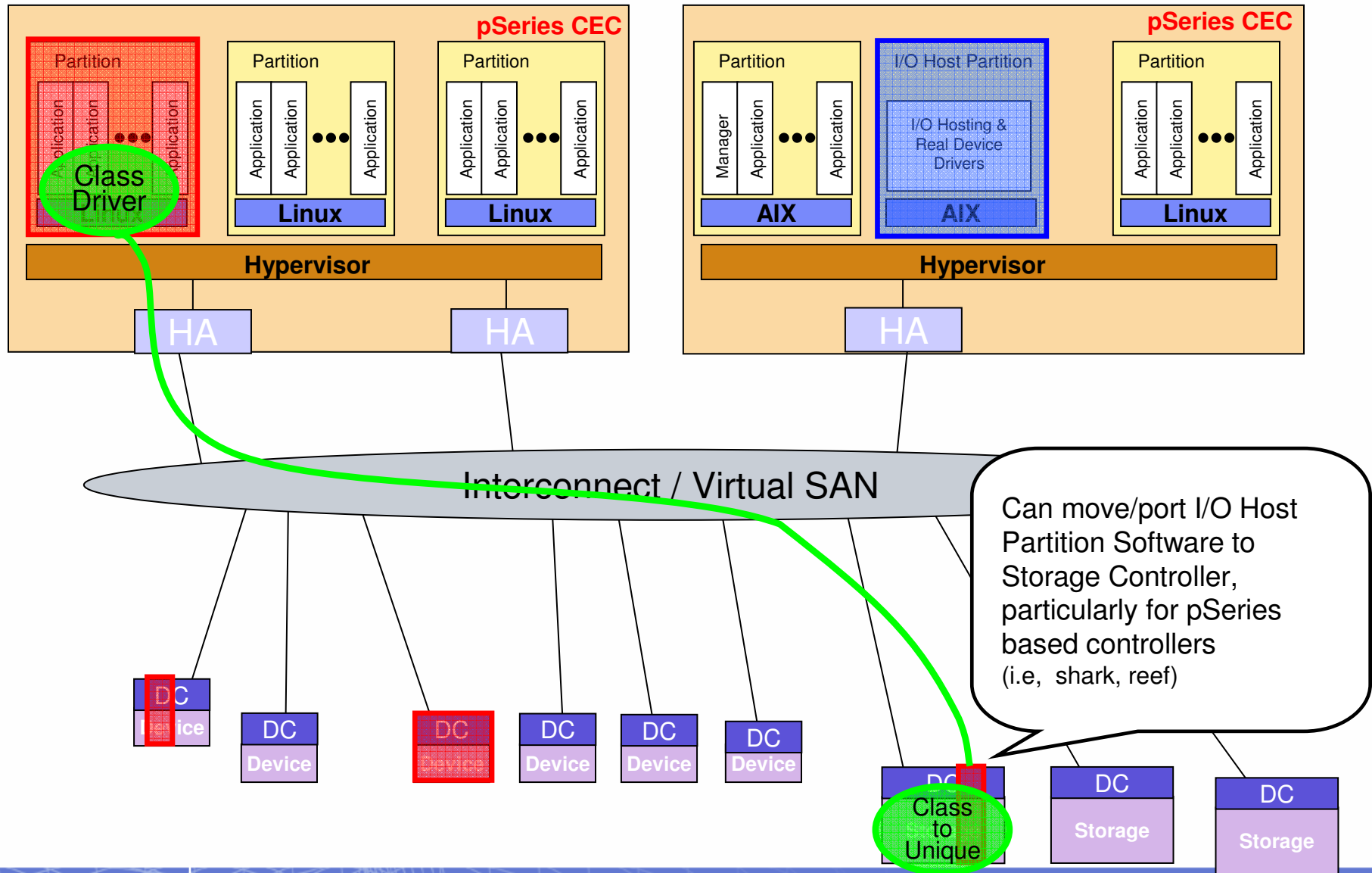
Step 2a: ... Class Driver Partition - Remote I/O Host & I/O Expansion



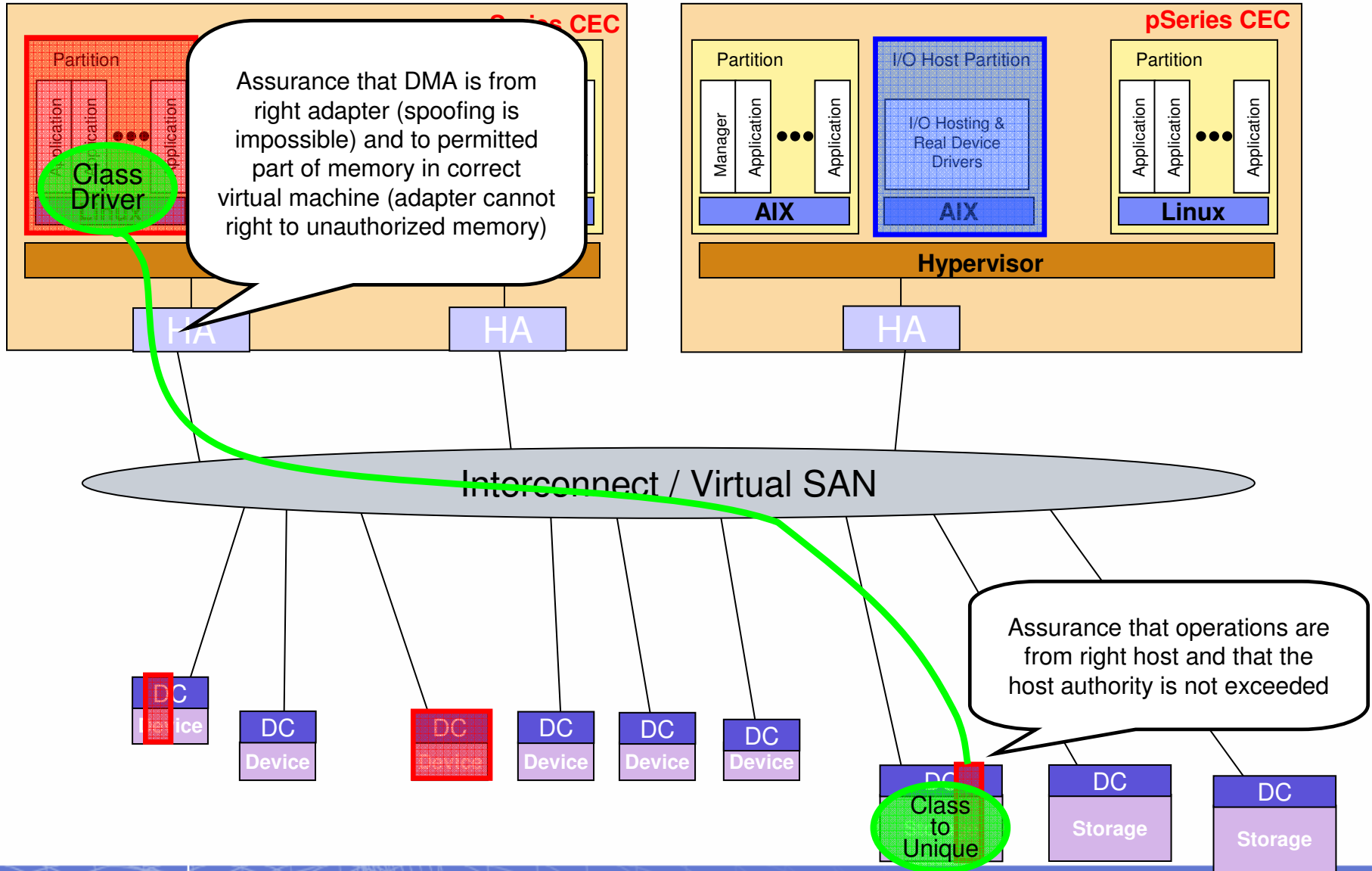
Step 3: ... Class Device Driver & Class Devices - Native Devices



Step 4x: ... Class Device Driver & Storage Devices (leveraging our ability to integrate)



Step 5: ... Class Device Driver, Addressing Security and Protection



Needed improvements in IBM Director GUI (improve intuitiveness)

The screenshot shows the IBM Director Console interface with the following components:

- Groups Pane (Left):** A tree view containing categories such as 'All Groups', 'All Systems and Devices', 'Chassis and Chassis Members', 'Clusters and Cluster Members', 'Hardware Status Critical', 'Hardware Status Information', 'Hardware Status Warning', 'Physical Platforms', 'Platforms', 'Platforms and Platform Members', 'Scalable Systems and Members', 'Systems with ASF', 'Systems with ASF Secure Remote Mand', 'Systems with Asset ID', 'Systems with CIM', and 'Systems with SNMP Agent'.
- Group Contents (1) Pane (Center):** Displays a single system named 'SLUGGO' with a warning icon.
- Tasks Pane (Right):** A list of tasks including 'Microsoft Cluster Browser', 'Network Configuration', 'Process Management', 'Remote Control', 'Remote Deployment Manager', 'CMOS Update', 'Custom', 'Donor Image', 'Linux Native Install', 'Power Restore', 'RAID Clone Configuration', 'RAID Custom Configuration', 'Remote Storage', 'Configure Fibre HBA', 'Delete Mapping', 'Delete Mapping and LUNs', 'Get Mapping', 'Put Mapping', 'Scan', and 'Script'.

Annotations on the screenshot:

- A red arrow points from the text 'Groups' to the 'Group Contents (1)' pane.
- A red arrow points from the text 'Things' to the 'Groups' pane.
- A red arrow points from the text 'Tasks' to the 'Tasks' pane.

At the bottom of the console, the status bar shows: IBM Ready, Host: Iriedle, User ID: LRIEDLE\riedle, and 7:36 AM EDT.

Conclusion / Comments

- Simplification and abstraction of BladeCenter infrastructure management is essential component for generalized scale out computing
- Dependable agile Web computing will need to be built on the fundamental management abstractions of the underlying infrastructure.
- The hardwired “server” resource of the Classic Grid vision is disappearing – replace by agile allocation of resource components to compose servers on demand.
 - ▶ Fundamentally a different model for exploiting underutilized resources.
 - ▶ Virtualization and separating of compute, I/O and Network resources into allocatable resource pools makes things both more complex in practice but simpler abstractly
- Staggering amount of work to be done