Red Hat Enterprise Virtualisation and IBM Flex Systems Fabrics

Bill Bauman
Global Alliances

VMworld Europe 2012

Agenda

- Review legacy (traditional) computing infrastructure
- Cloud computing over(re)view and discussion
- Introduction to fabric-based computing
- Well-designed fabrics

Legacy Computing Infrastructure

Web Servers

Print Servers

File Servers

Database, CRM, ERP Servers

- Servers sized for specific task
  - Every server has different CPU, Memory and I/O capabilities and storage capacities

Legacy Network & Data Infrastructure

2 x 1Gb Ethernet

2 x 2Gb FC Data (optional)

2 or 4 x 1Gb Ethernet

2 x 4Gb FC Data (File Servers)

2 x 10Gb Ethernet

2 x 1Gb Ethernet

2 x 8Gb FC Data (DB Servers)

- Network and I/O differs by server task
  - Not all servers have high speed, shared I/O (Data)

Legacy Infrastructure Connected Overview

Print Servers

1GbE

1GbE

Ethernet

1GbE

10GbE

Web Servers

1GbFC

SAN

8GbFC

Database, CRM, ERP Servers

1GbE

4GbFC

File Servers

Legacy Virtualization Infrastructure

- Most customers virtualized using existing, legacy infrastructure
  - Future deployments followed legacy model

Blade Virtualization Host Servers

1GbE

Ethernet

1GbE

10GbE

1GbFC

SAN

8GbFC

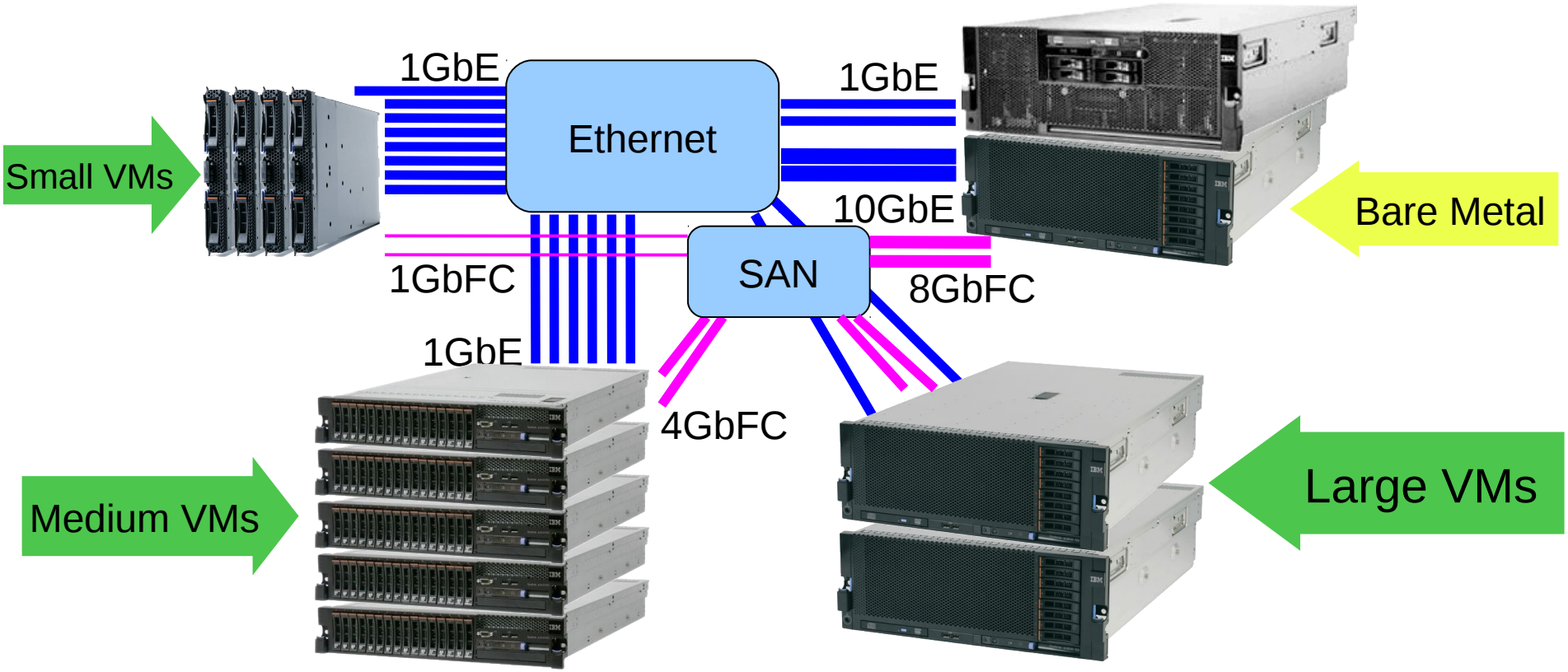
1GbE

4GbFC

Database, CRM, ERP Servers (Bare Metal)

Rack Virtualization Host Servers

Legacy Virtualization Infrastructure Continued

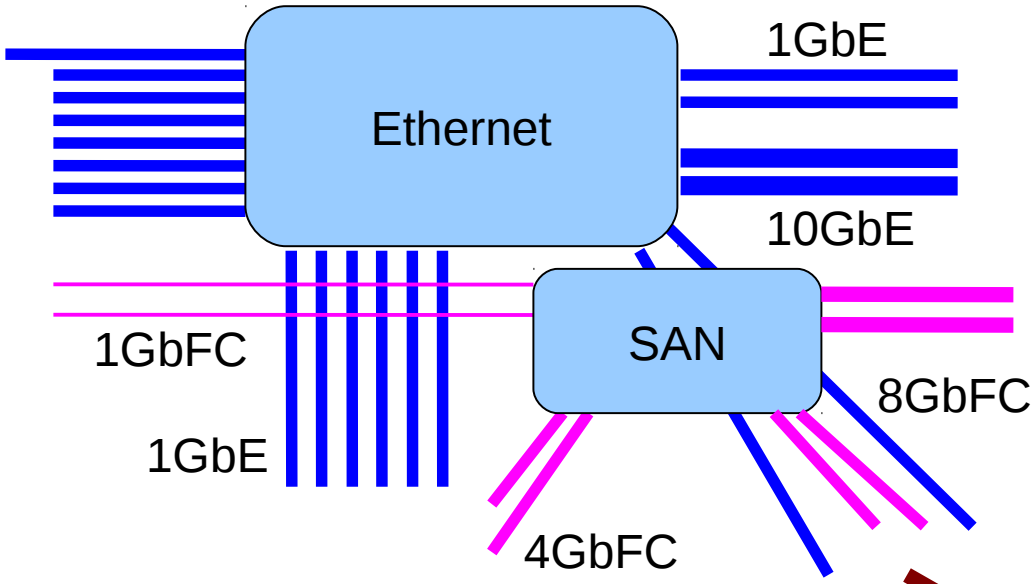


- Virtual Machines manually placed or pooled by VM resource requirements
  - Some or many applications still not virtualized

Agenda

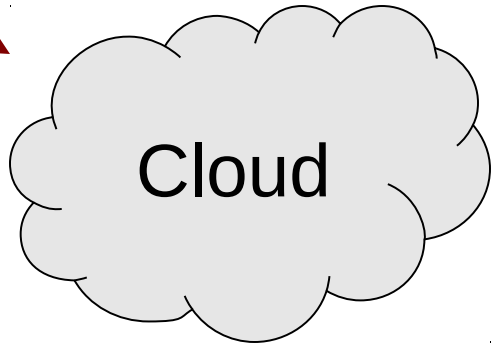
- Review legacy (traditional) computing infrastructure
- Cloud computing over(re)view and discussion
- Introduction to fabric-based computing
- Well-designed fabrics

Reaching The Clouds



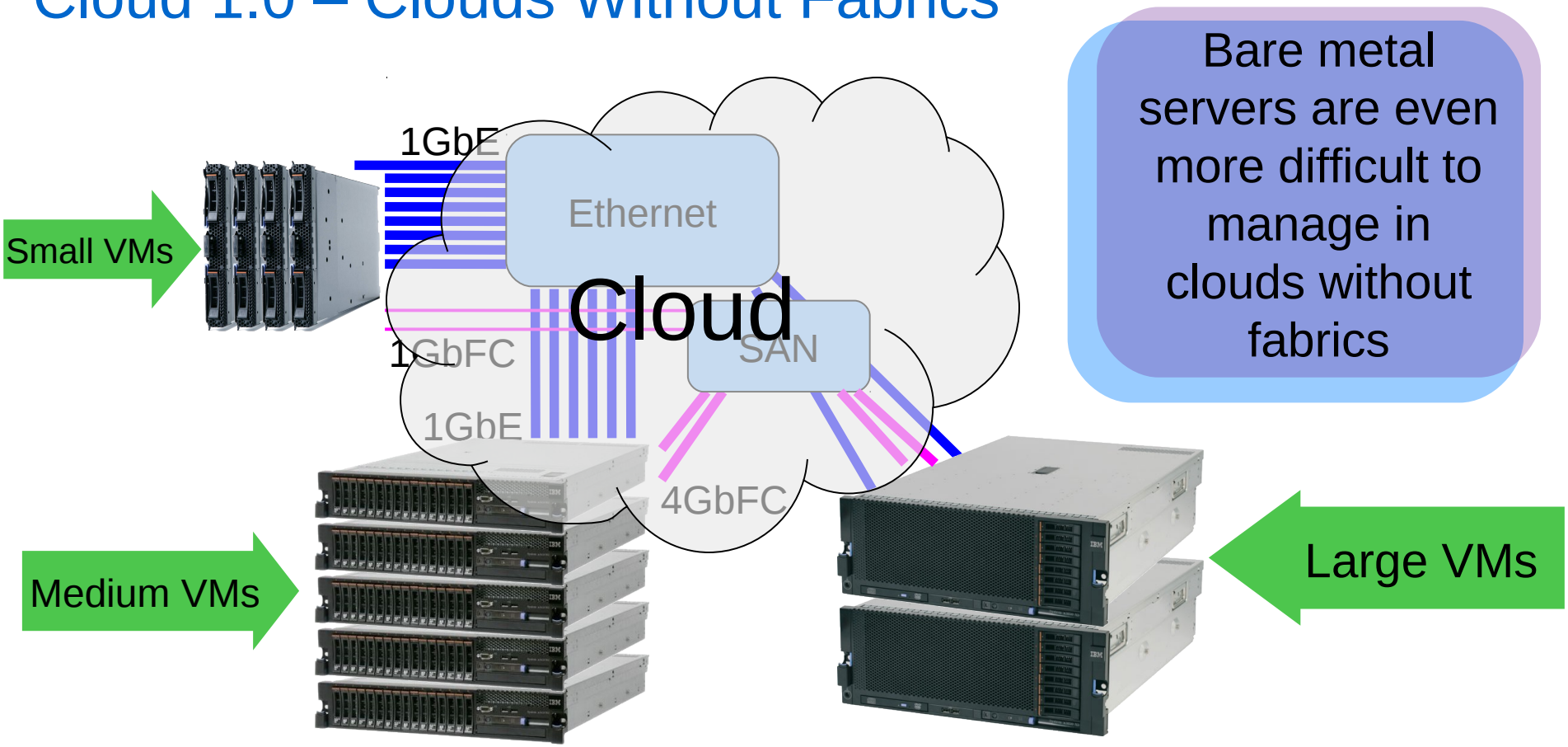
Virtualization and automated workloads become synonymous with computing in the cloud

Becomes



- Connected infrastructure continues to become increasingly complex to design and manage
  - Architectural diagrams no longer show all details of connectivity

Cloud 1.0 – Clouds Without Fabrics



- Cloud computing 1.0 dramatically reduces time to deployment of new workloads
- To deliver workloads to appropriate hosts, administrators must:
  - Implement and write complex rules-based delivery systems or
  - Create isolated pools of resources, decreasing total system utilization

Cloud Computing (still 1.0)

- From a logical, or software perspective, cloud computing addresses the issues of (and more)
  - Automation
  - Elasticity & Scaling (up and down)
  - Self-service

- As discussed, most clouds have not been designed with a for-purpose, fabric-based infrastructure

Agenda

- Review legacy (traditional) computing infrastructure
- Cloud computing over(re)view and discussion
- Introduction to fabric-based computing
- Well-designed fabrics

Some Principles of a Fabric-Based Infrastructure

- All physical resources should be equal
- All workloads can run anywhere
- Data localization no longer inhibits resource placement
- Spend time working on things that matter, let the fabric service the cloud

- Fabric-based computing should generally be a Highest Common Denominator (HCD) design
  - The largest workload should run adequately on all systems

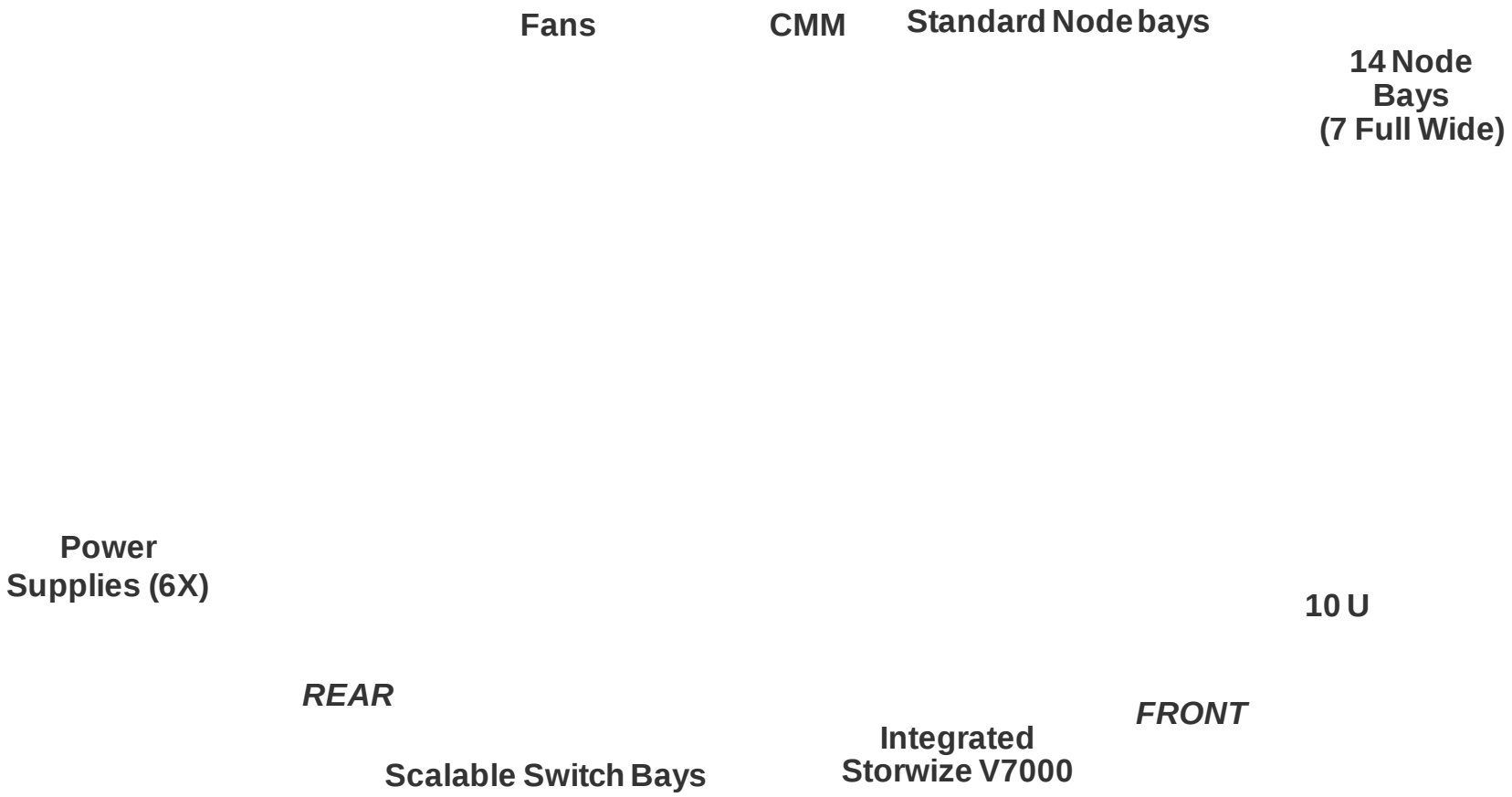
Agenda

- Review legacy (traditional) computing infrastructure
- Cloud computing over(re)view and discussion
- Introduction to fabric-based computing
- Well-designed fabrics

Highest Common Denominator & Equal Access

- Fabric nodes (servers) should be designed to support all workloads
  - What is the largest VM to be deployed?
    - In-memory databases, CRM, ERP, etc.
      - 64 vCPU? 512GB, 1TB virtual RAM?
  - What are the highest bandwidth and lowest latencies required for the most demanding workloads?
    - Accessibility to the highest speed I/O should be available on all nodes, regardless of the current workload that node is supporting
- All nodes in the fabric should have equal access to all fabric resources
  - High Performance Computing (HPC) clusters have been designed this way for years
    - Cost has been prohibitive for mainstream, commercial computing until recently
      - IBM System Networking and integrated 10Gb+ switching in PureSystems changes the economics of equal access

Flex Systems – Designed for Fabrics



- Integrated storage reduces time to data and end user wait periods
- Up to 220Gb uplink (external) bandwidth and <1microsec latency
  - Reduces congestion for non-localized data accesses
  - 14 nodes x 10Gb = non-blocked data access to all nodes inside *or* outside the chassis

Network Switching and Why it Matters

The Problem:

Blade to blade **communication** flows north-south through the TOR, causing **latency** from request / response traffic. Added network latency will impact the overall **workload** the servers can support.

Communication between blades requires traffic to TOR

The Flex System Difference:

Do more with your servers and **reduce** network delays. Node to Node communication happens **within** the chassis.

Communication can be contained within chassis running at 10Gb converged I/O

Why this matters:

- Reduces switch latency
- Additional servers needed to overcome performance loss in network delays
- Low latency, web-serving, and database apps create significant server to server chatter and stress on the network
- Non-blocked external I/O

Optimized, Automated and Integrated network architecture

Fits within your existing and future environment

The Problem:

Today's networking offerings lack the flexibility to meet the **demands** of the next decade of I/O. Clients are often burdened **now** with the costs of technology for **tomorrow**.

Extreme Flexibility

- Designed to meet port and bandwidth requirements for next decade
- Pay for what you need today with Features on Demand (FoD)

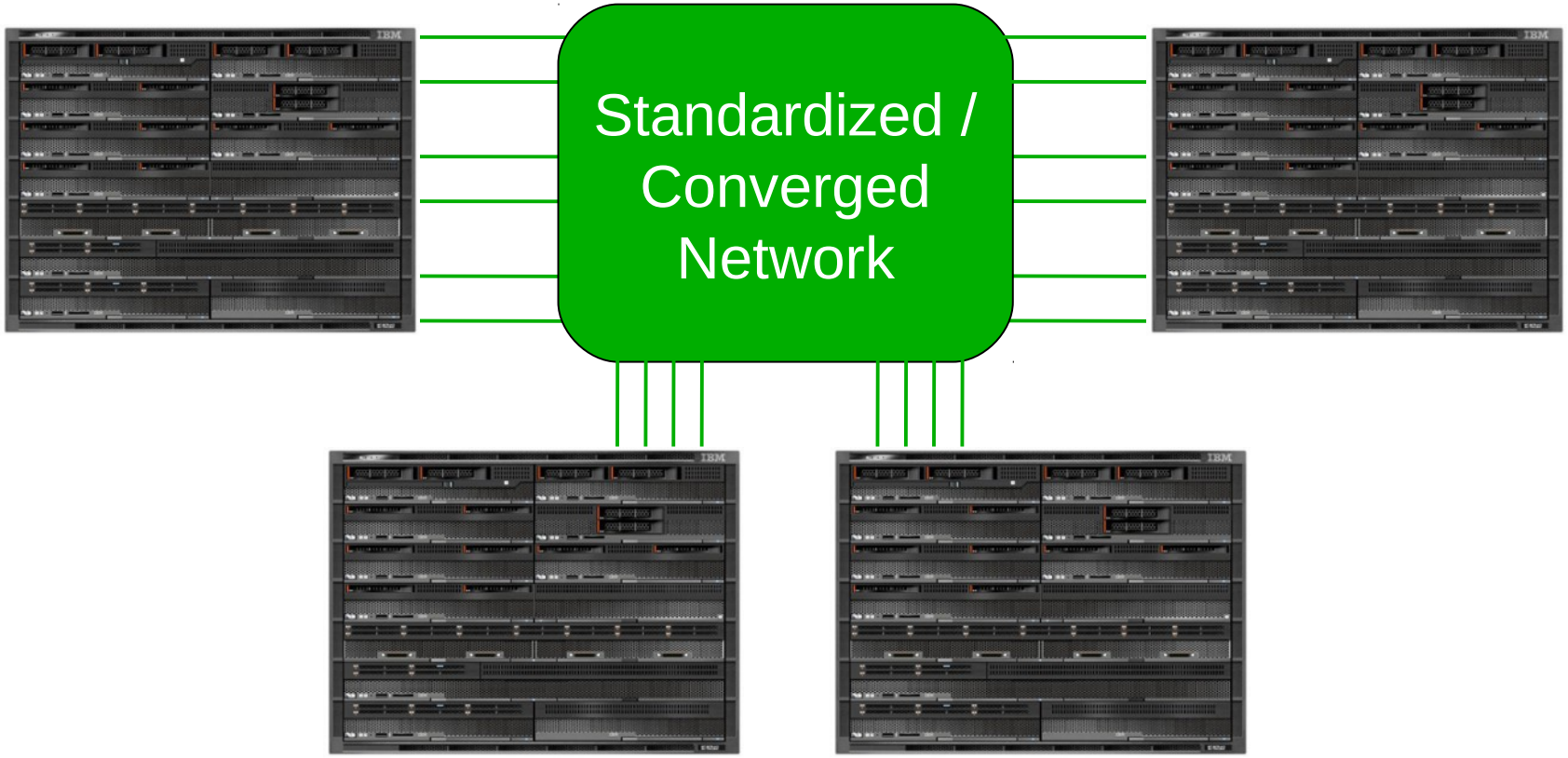
Highest Performance

- First 40Gb capable Ethernet Switch
- First 16Gb capable SAN Switch
- First 56Gb capable Infiniband FDR switch
- Up to 220Gb uplink BW and <1microsec latency

Standards based Convergence

- 10Gb iSCSI and FCoE offering
- First 40Gb end to end FCoE offering (post GA)
- Standard based for seamless integration

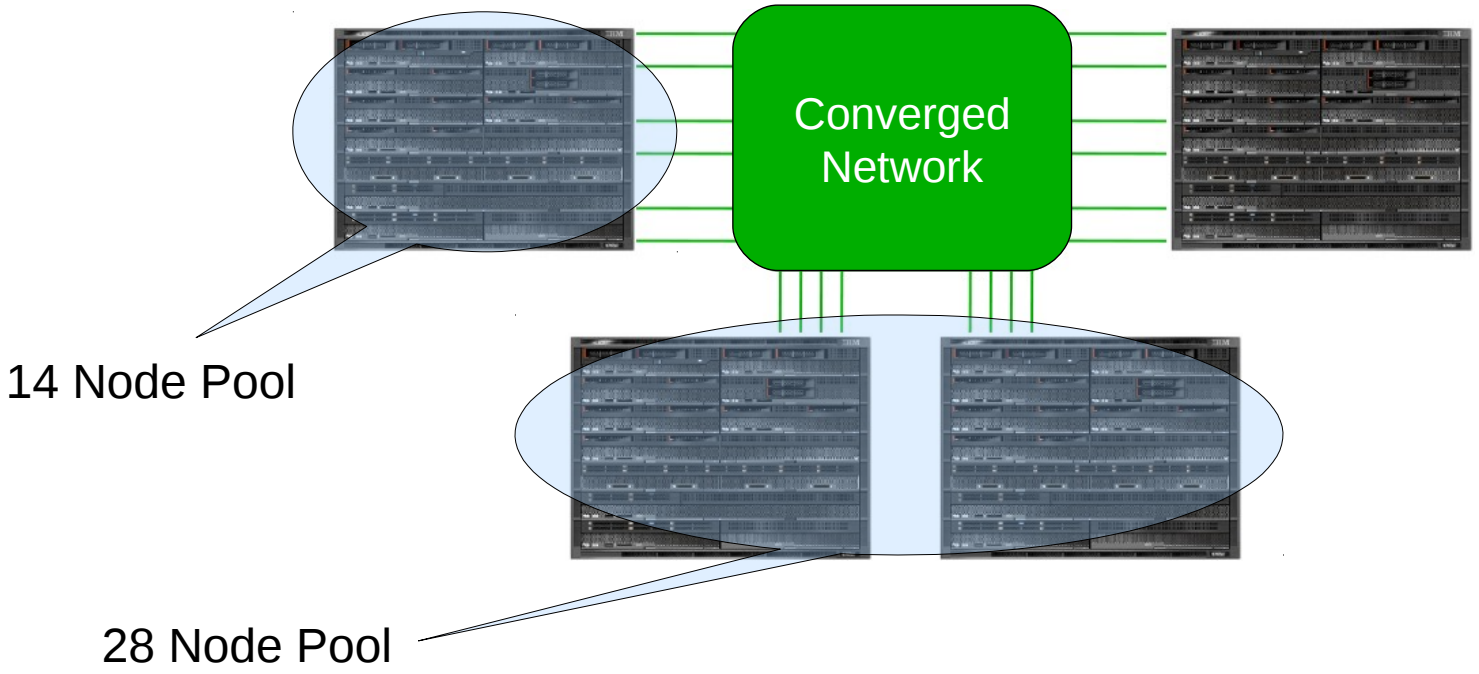
Flex Systems Fabric



- In-Chassis architecture
  - All systems share equal access to all resources
  - All systems are equal in capability
- Out-of-chassis is nearly identical in access time

Resource Pools

- Resource pools must be large enough to accommodate design points of the fabric-based infrastructure
  - Flex Systems shows that the hardware can scale, the key for resource pools is software scalability

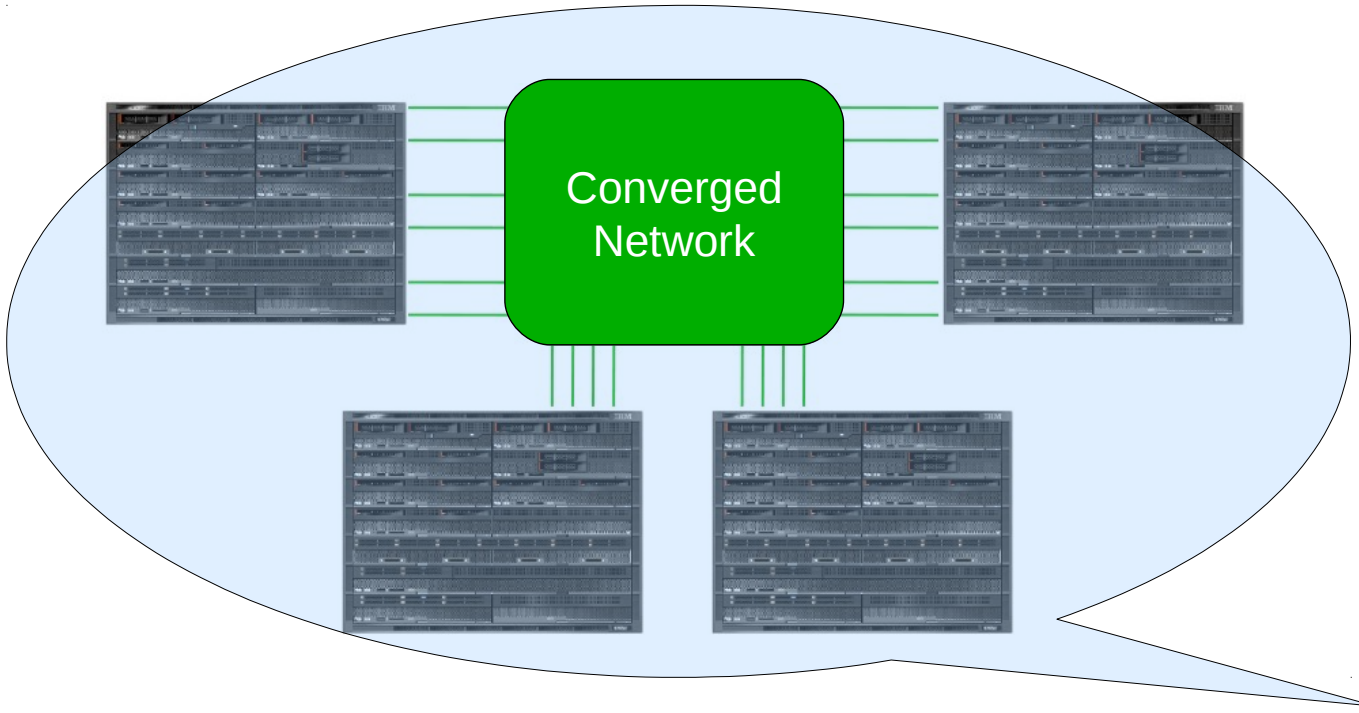


Resource Pool Sizing Guidelines

- The maximum size of a resource pool determines the maximum size of one fabric or one cloud
  - Difficulties in automation outside the resource pool
    - Workload placement generally limited to one pool
      - placing workloads outside that pool generally requires manual administration
    - Workload mobility (automatic load-balancing, etc.) is generally limited to the individual pool
    - Maintenance (live migrating to other nodes) is generally limited to a single pool (especially automated maintenance)
  - The larger the pool, the generally higher the utilization of that pool
    - Increased flexibility of workload placement
    - Increased flexibility for maintenance tasks
    - More even distribution of workload – better response times to end users

Resource Pools

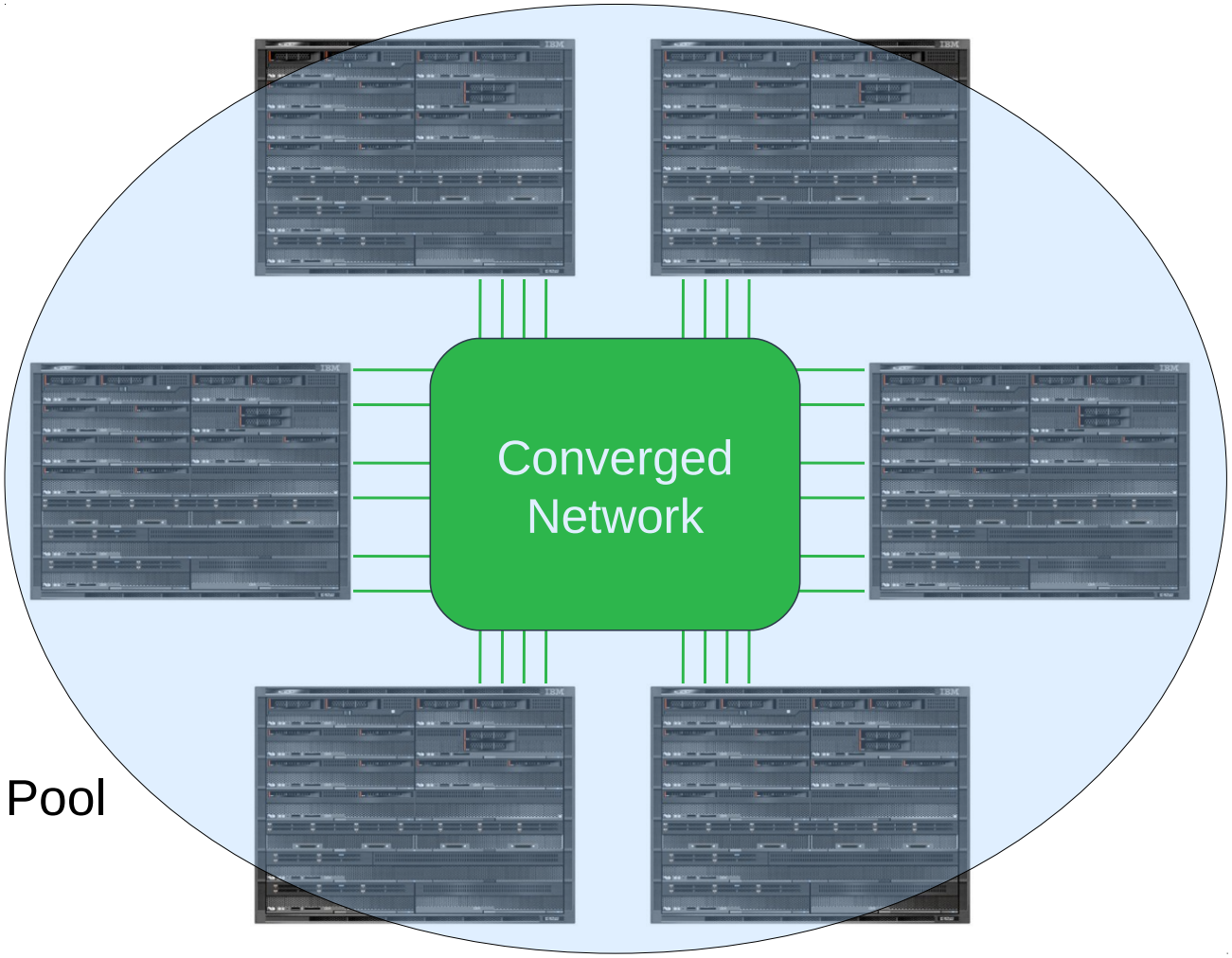
- Resource pools are generally limited by the largest cluster size to which the virtualization solution can adequately scale
  - RHEV-M – Officially: 200 / Best Practice: Add memory to management database as appropriate



56+ Nodes in a pool

RHEV 3.0 Resource Pools

- Maximum flexibility in workload deployment & placement



84 Node Pool

- RHEV-Manager scales Clusters to hundreds of nodes
  - Resource pools of 100 to 200 nodes is reasonable in RHEV 3.0

IBM Systems Director 6.3

Simplified management to Optimize your Datacenter

- **Supports larger Data Centers**
  - Faster inventory, discovery and start-up times
  - 2X number of endpoints managed
- **Faster Time to Value**
  - Initial install and setup in 75% less steps
  - Simplified discovery, installation, service, and the update process
- **Integration of the Hardware and Virtualized realms**
  - Provides a single and consistent solution with the widest range of flexibility in the industry supporting the x86, Power and z systems allowing hypervisor of choice.
  - Workload automation based on server / storage / network health
  - Workload management based on server / storage / network capacity
- **Deployment of Virtual Storage 2X faster with new integrated features**
  - Define pools of storage with differing qualities of service (RAID levels)
- **Systemic Reliability**
  - Automated server restart – “Always up” console
  - Automated Agent restart – “Always up” agent
- **Unified security management – Single pane of glass**
  - For managing users, access, certificate/ keys, policies.
- **Enterprise Ready Database and Licensing**
  - An enterprise ready Database (DB2) now automatically installed / embedded
  - No Database experience required
  - Simplified license management with user-defined activation
- **Embedded serviceability for speedy problem resolution**
  - Service and Support Manager comes standard
  - Improved messages so the users know what actions to take
  - Collection of all logs for services gets problems solved quicker

With Systems Director 6.3 -

Manage up to 20,000 endpoints,

2X beyond what you could before

Single console for the integrated management of server (x85 to RISC), storage (IBM and non-IBM) and network resources.

Improve end-user and operations productivity with faster time to value

Enable faster, more flexible management of new application / workloads with virtualization and integrated hardware health

IBM Systems Director Scalability

- Manage up to 20,000 endpoints
  - Systems management scalability is critical to a well-designed fabric and cloud infrastructure
- Every aspect of a fabric should be managed, including the operating systems of each individual virtual machine
  - IBM Systems Director can enumerate the hardware and the software of any fabric
    - VMControl provides integrated management of vCenter and KVM on RHEL-based nodes
    - RHEV-Manager VMs are managed as invididual entities as of today (2012-05-14)
      - Nodes are hardware manageable via level 2 h/w mgmt.

Flex Systems & Fabric-based Infrastructure

- Fabrics built on IBM Flex Systems do not suffer performance degradation when accessing data internal to or external from the chassis
- Flex Systems scale to hundreds or thousands of nodes but don't *require* hundreds or thousands of nodes to be cost-effective fabric-based building blocks
- Red Hat Enterprise Virtualization (RHEV) scales to hundreds of nodes in single resource domains (clusters)
- IBM Systems Director scales to meet the needs of a well-managed Red Hat & IBM Fabric

RED HAT ENTERPRISE VIRTUALISATION
RHEV 3.1 Storage Enhancements Overview

	Storage
Hotplug Disk	- Hot plug/unplug virtual machine disk image
Direct LUN	- UI support for configuring direct LUN access for virtual machine
Shared Disk	- Allow VM disk to be shared e.g., Shared disk for database
Multiple Storage Domains	- Allow VM to use disks from multiple storage domains
Live Snapshots	- Live snapshots of Virtual Machine – including guest agent
'Floating' Disks	- Create and manage disks that are not associated to a VM
'Pluggable' Filesystems	- Infrastructure to support other shared file systems e.g., Gluster, GPFS, etc
Resize Storage	- Dynamically resize guest and host storage
Storage	- Storage Live Migration

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States, other countries, or both.

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.

Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

*, AS/400®, e business(logo)®, DBE, ESCO, eServer, FICON, IBM®, IBM (logo)®, iSeries®, MVS, OS/390®, pSeries®, RS/6000®, S/30, VM/ESA®, VSE/ESA, WebSphere®, xSeries®, z/OS®, zSeries®, z/VM®, System i, System i5, System p, System p5, System x, System z, System z9®, BladeCenter®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.