Integration of Cloud Computing and Cloud Storage

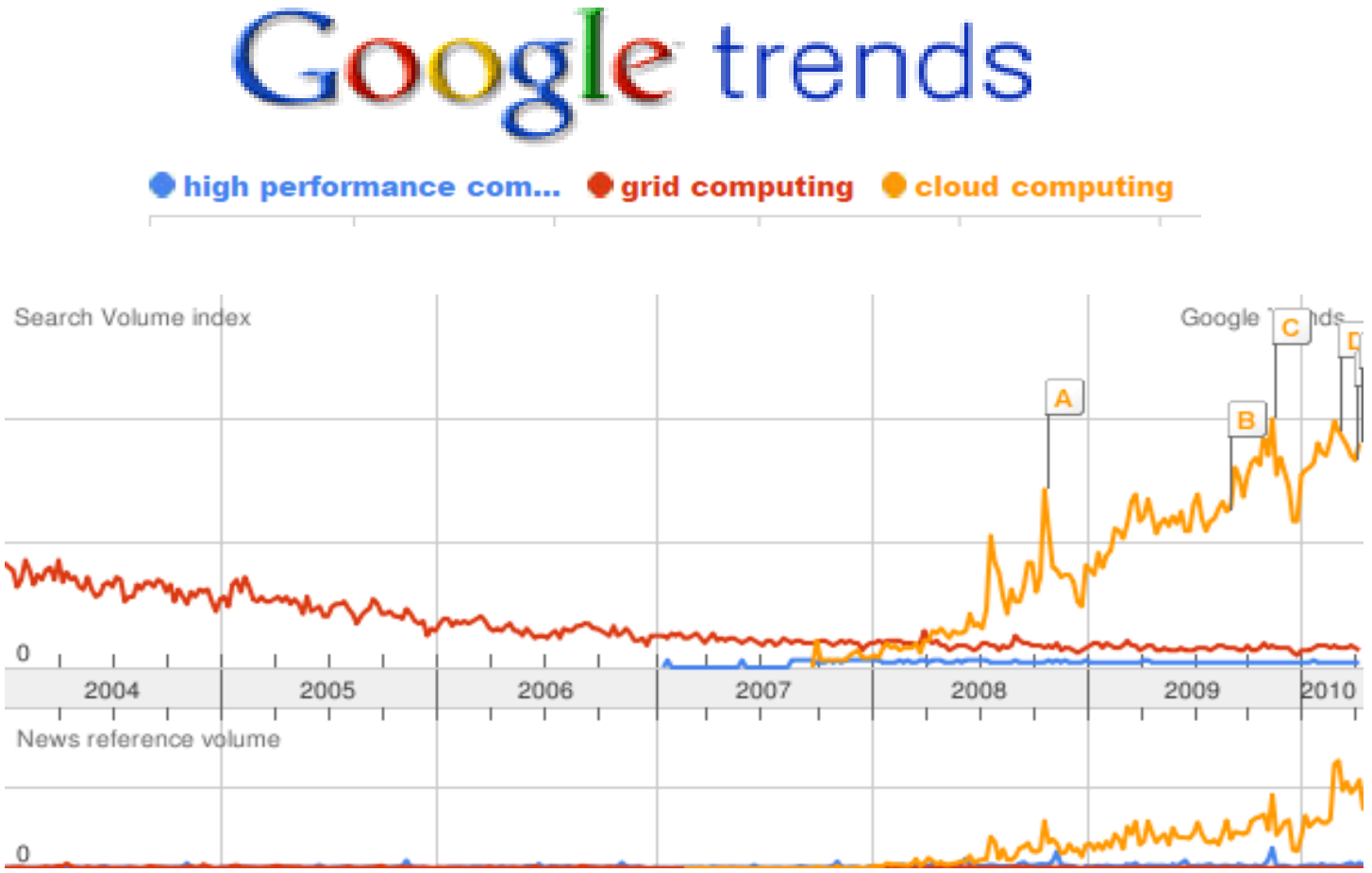
Chief Scientist ,Renaissance Computing Institute
Adjunct Prof. Computer Science, NC State University

IEEE Mass Storage Conference Tutorial
May 3, 2010

Outline

- Introduction to Cloud Computing
- Building a cloud computing architecture
  - Properties for a cloud
  - Types of clouds
  - Building images in a cloud
  - Designing a cloud system
- Large data sets and analysis requirements
- Cloud Computing Demo
- Closing remarks

Level of General Interest in Cloud Computing



So many choices…

“The interesting thing about cloud computing is that we've redefined cloud computing to include everything that we already do…. I don't understand what we would do differently in the light of cloud computing other than change the wording of some of our ads.”

Larry Ellison
CEO Oracle

Some Important Characteristics in Building Cloud Computing Architectures

Start with Basic General Design Goals For Any System

- Rapid efficient delivery of IT services
- Reliable, secure, and fault-tolerant
- Data and process aware services
- Secure dropped-session recovery
- More efficient delivery to remote users
- System is cost-effective to operate and maintain
- On-demand or batch processing capabilities
- Removal of conflicts among supported software applications
  - Incompatible versions
  - Inconsistent upgrade patterns among different locations
  - User issues with obtaining and using the software

Add Basic General Design Goals For A Cloud Computing System

- Ability for users to request specific HW platforms to build, save, modify, run virtual computing environments and applications that are:
  - Reusable
  - Sustainable
  - Scalable
  - Customizable
- Timely inclusion of new software images
- Root privileges *(as required/authorized)*
- Time and place independent access
- Full functionality via consumer devices and platforms

A List of Some Important Characteristics For a Cloud Computing System

- An operational paradigm allowing the users to seamlessly and securely provision and/or combine
  - Computer hardware
  - Operating systems
  - Application software
  - Storage
  - Rich set of customizable services
- As a system that is scalable up, down, in, and out
  - With resources accessible over a network
  - Based on a service-oriented architecture
- With users controlling the options to purchase, lease or reserve each equipment and service capability on a mix-n-match component or unit basis

A List of Some Important Characteristics For a Cloud Computing System

- An operational paradigm allowing the **users** to **seamlessly and securely provision and/or combine**
  - Computer hardware
  - Operating systems
  - Application software
  - Storage
  - Rich set of customizable services
- As a system that is **scalable up, down, in, and out**
  - With resources accessible over a network
  - Based on a service-oriented architecture
- With **users controlling the options** to purchase, lease or reserve each equipment and service capability on a **mix-n-match component or unit basis**

Some Definitions -- Cloud Types

- **Public Cloud** is a system where hardware, software and/or application services are accessible to the general public over the Internet with access usually purchased on some type of a pay-as-you-go per component usage basis
- **Private Cloud** is a system where computational hardware, software, and/or applications and software services are only accessible internally within an organization or educational institution
- **Hybrid Cloud** is a private cloud that scales up service availability by externally provisioning resources from a public cloud when there are rapid workload fluctuations or hardware failures.

Cloud Computing Images - A Key Concept

- The combination of the operating system, applications and customizable software forms an **“image”**
- An “**image”** is a tangible abstraction of a software stack
- Ingredients needed to create an image
  - Any base-line operating system
  - If virtualization is needed for scalability, a hypervisor layer
  - Any desired middleware or application that runs on that operating system
  - Any end-user access solution that is applicable

Some Metadata Describing an Image

- Identifier
- Location
- Name
- Owner
- Memory footprint
- Speed of access information
- Licensing information
- Hardware requirements
- Loading time
- Access permissions

Some Image Properties

- Images can be associated into logical image groups
  - Mapped onto a particular logical resource (“hardware”) groups
  - Mapped onto individual computers
- Association of the meta-data with an image is made primarily via a data-base

Types of Images

- Simple Virtual or Bare Metal Images
  - Simple Virtual Image: image loaded into an operating system/application virtual environment of choice
  - Bare-metal image: operating system and application stack loaded straight onto the hardware without any other software layer between the image and that target hardware.
- Composite Images
  - Composite images are aggregates of two or more images that are loaded synchronously (called environments)
  - Composite images can construct “virtual clouds” and cloud services

Traditional Versus Cloud Computing
Some Comparisons of Technical Characteristics

Traditional	Cloud
• Control of resource use managed by the site	• Control of resource use managed by the user
• Site-specific environment	• User defines site environment
• Site defines modes of access	• User defines modes of access
• Site-driven prioritization	• Explicit user choices for service level options & prioritization
• Ease of deployment for provider, constraints on user	• More difficult deployment for provider, ease of use for user
• Site controls the technology	• Users control the technology

Clouds Grouped by Services

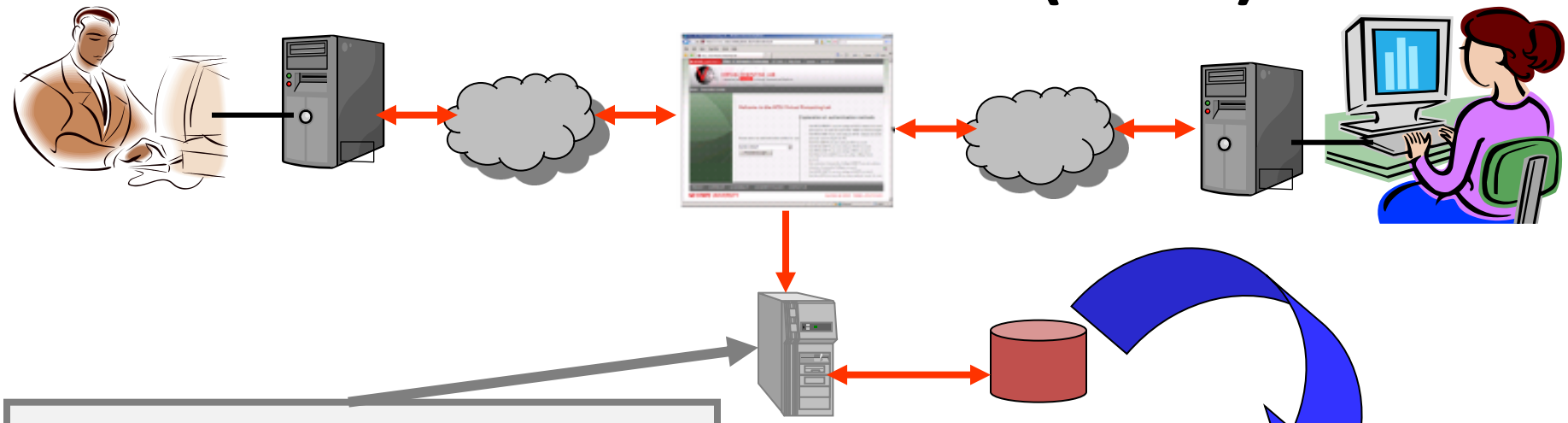
- **Hardware as a Service (HaaS)** – On demand access to a specific equipment configuration possibly at a particular site
- **Infrastructure as a service (IaaS)** – On demand access to user specified hardware capabilities, performance and services which may run on a variety on hardware products
- **Platform as a Service (PaaS)** - On-demand access to user specified combination of hypervisors, operating systems and middleware that enables applications and services
- **Application as a Service (AaaS)** - On-demand access to user specified application(s)
- **Software as a Service (SaaS)** - may encompass anything from PaaS through AaaS
- **Cloud as a Service –** On demand ability to construct a local cloud within an overall cloud service
- **Security as a Service –** On-demand use of cloud configuration for security of applications and systems

Cloud Services and Cloud Architectures

- Cloud architecture abstracts resources at several levels
  - Application and operating system level via images and hypervisors
  - Hardware location level via compute manager and compute nodes
  - Network level (via virtual networks, VLANs, VPNs)
- Each cloud service type needs an architecture that will optimize that type of service delivery


IaaS
amazon web services™
MOSSO the hosting cloud
GO GRID
IronScale
PaaS
Google App Engine
RIGHT SCALE
Windows Azure
Join Now
developer.force.com
Get a Free Developer Edition Account
SaaS
NETSUITE ONE SYSTEM. NO LIMITS.
Salesforce
SAP Business ByDesign

Infrastructure as a Service (IaaS)
Hardware as a Service (HaaS)



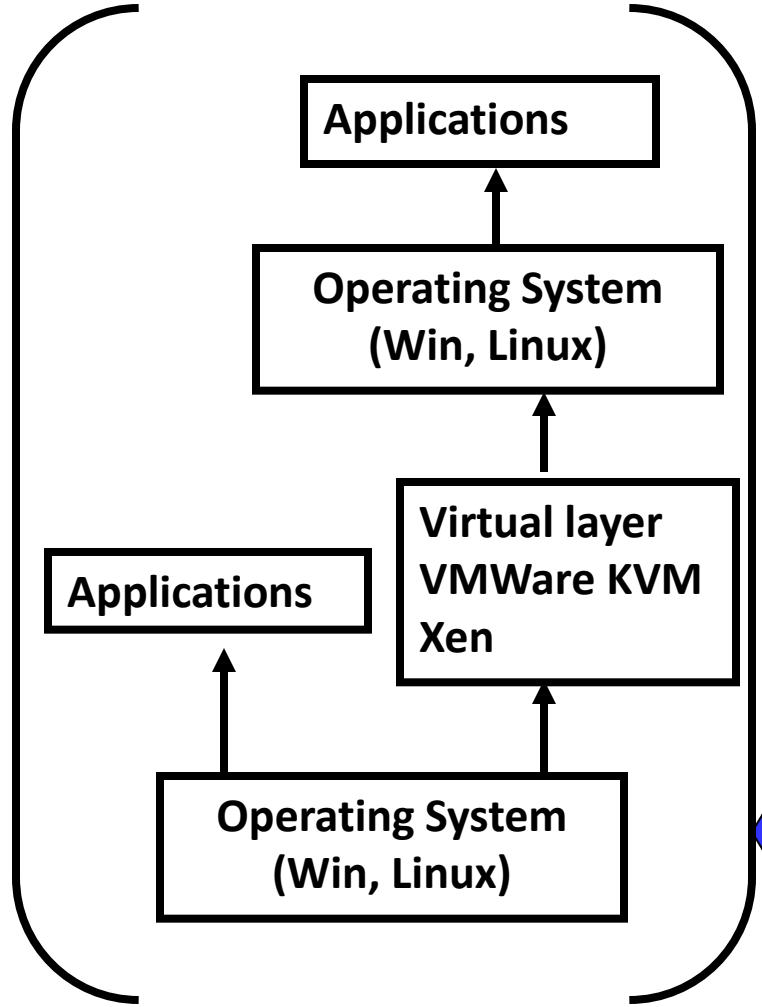
Managing server handles

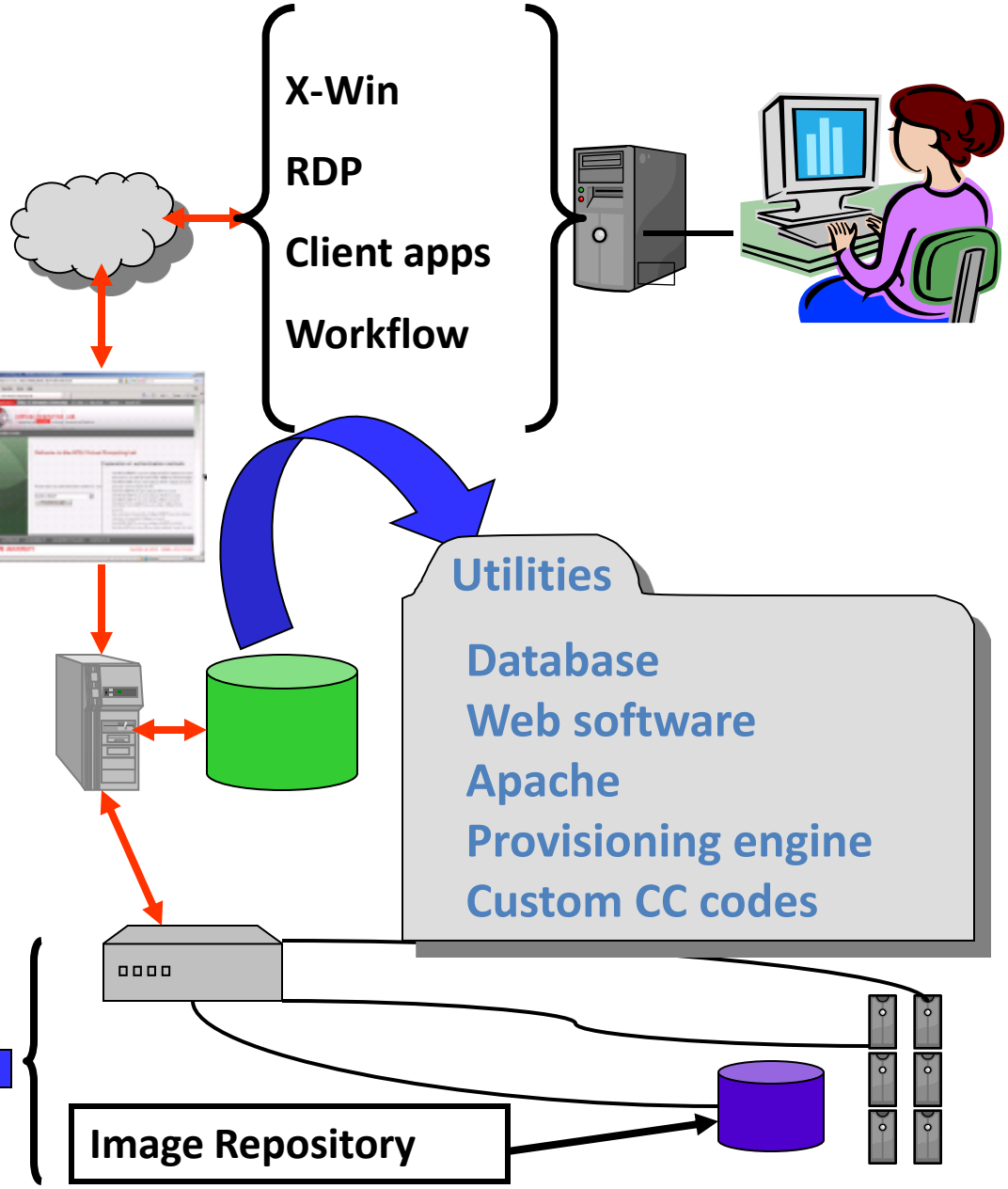
- User requests
- Resource scheduling
- Authorization
- Security
- Multi-site coordination
- Performance monitoring
- Virtual network mgt
- Software licensing etc.,

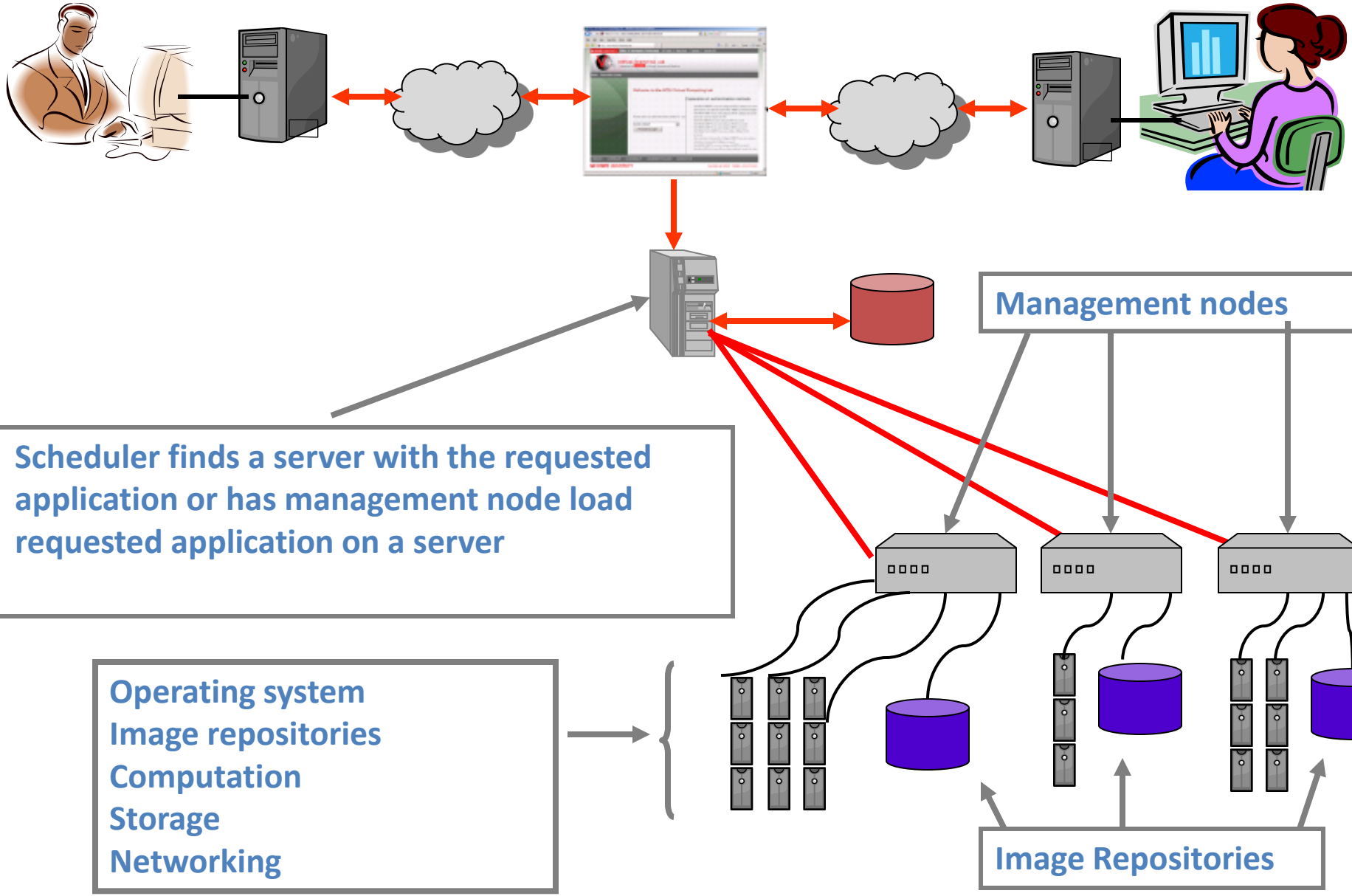
Utilities

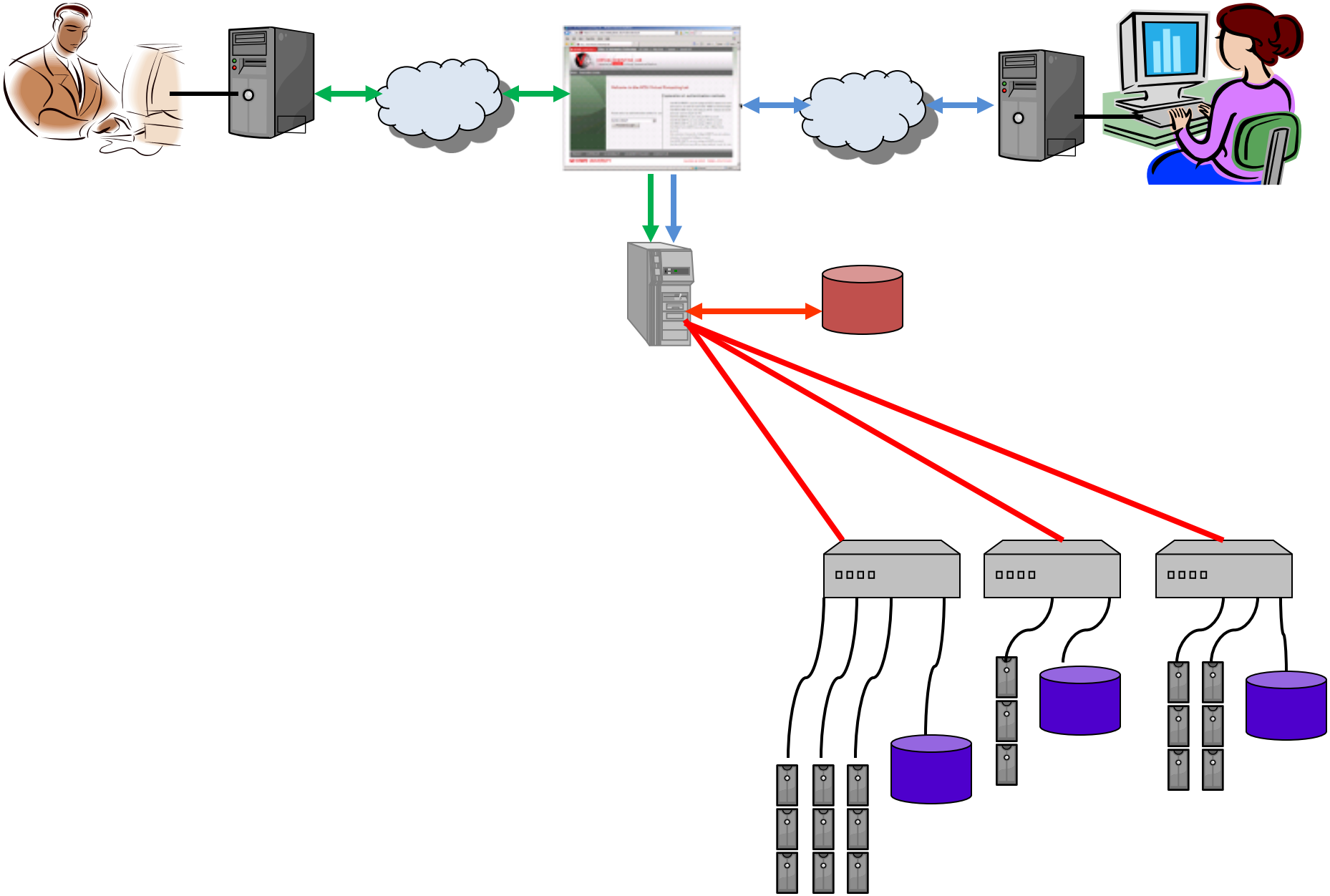
Database
Web software
Apache
Provisioning engine
Custom CC codes

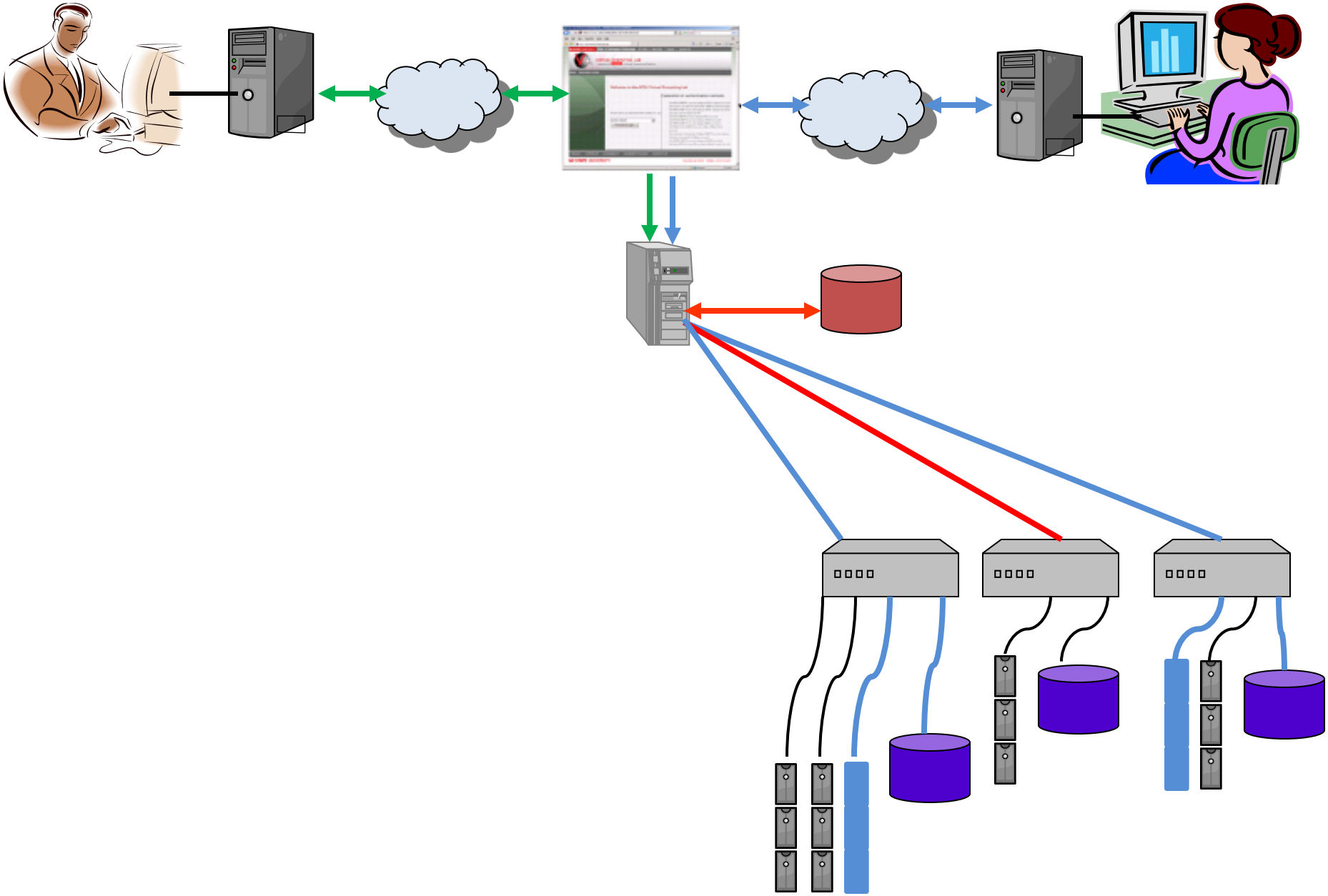
IaaS/HaaS Logical Architecture

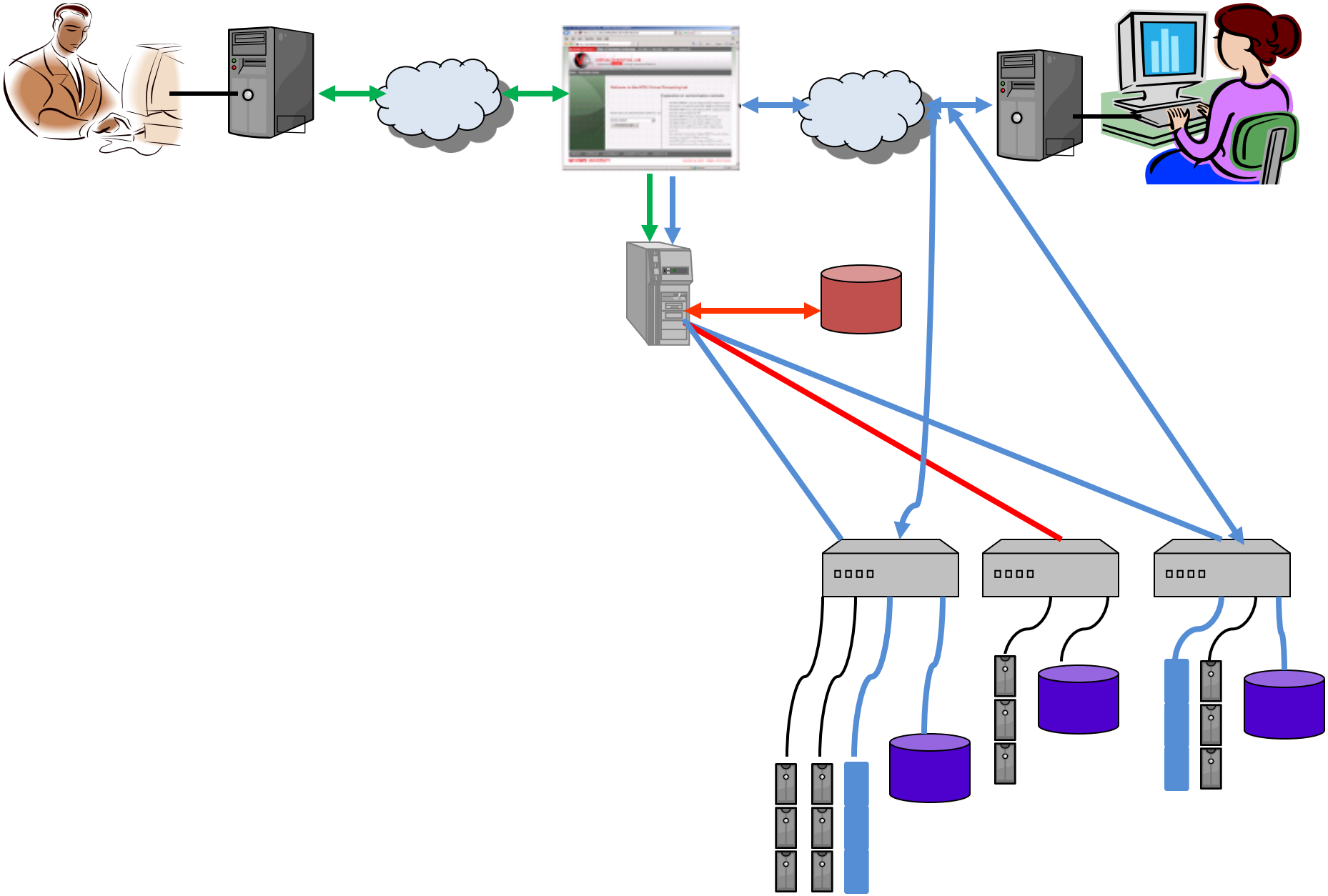


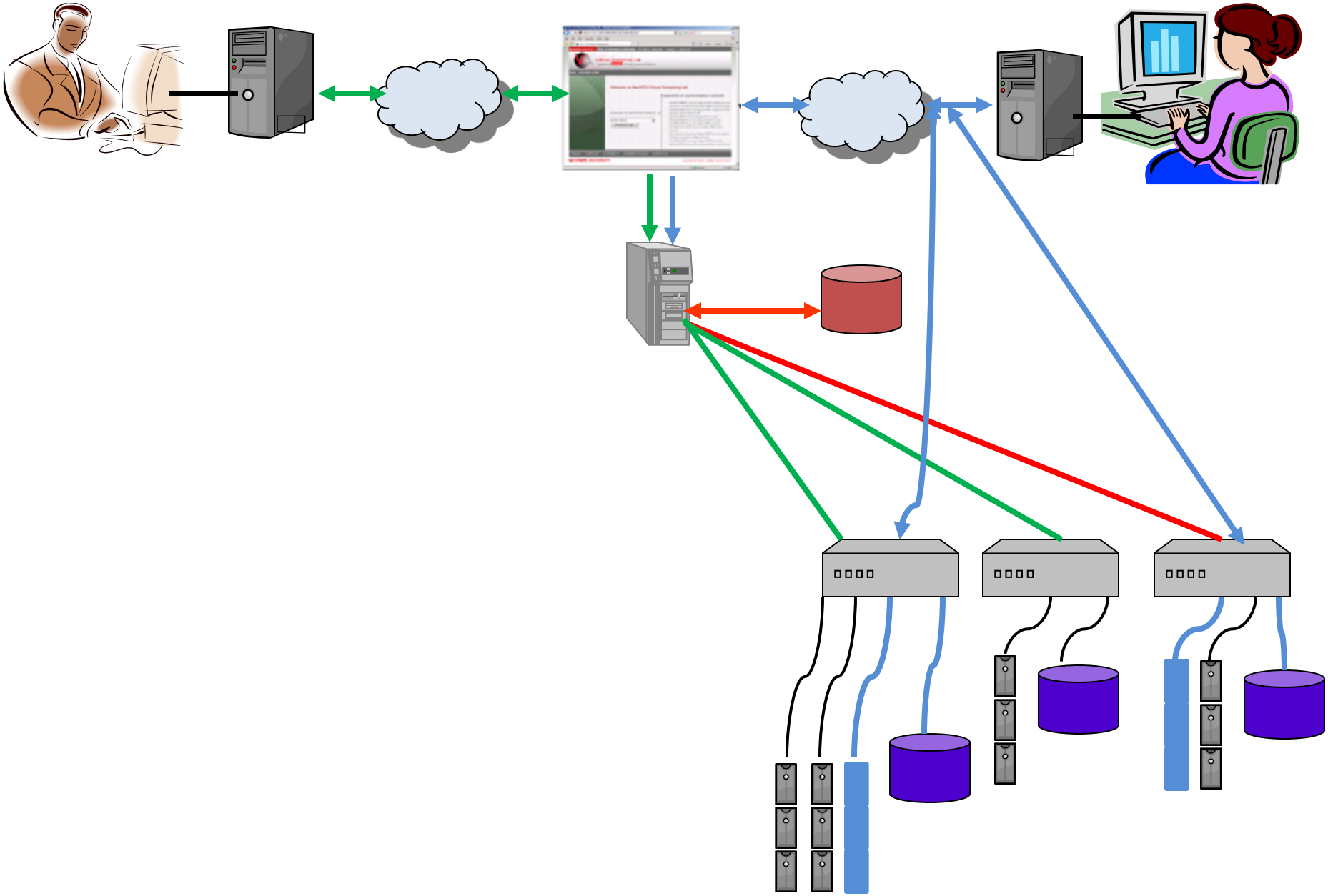


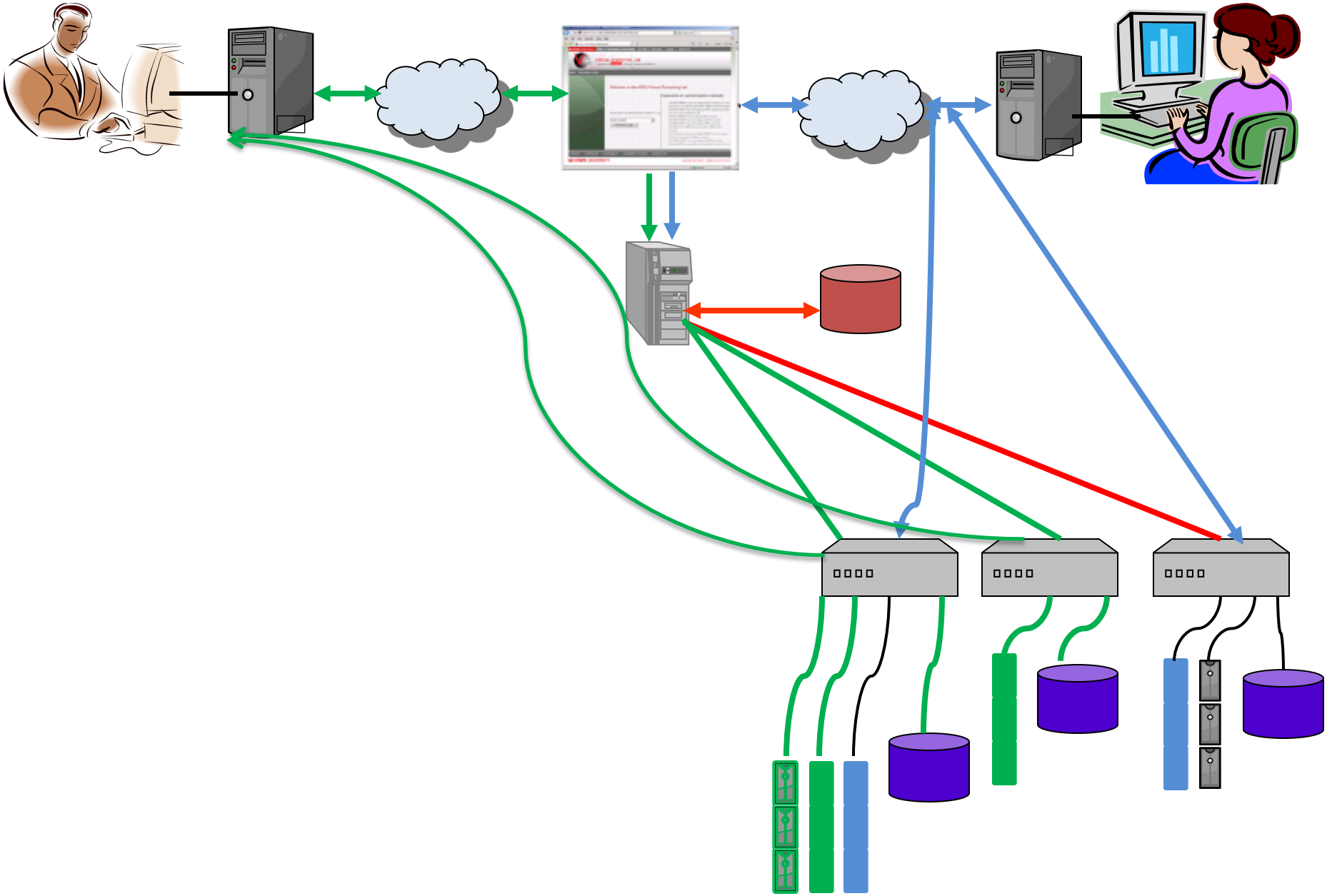

Management nodes
Scheduler finds a server with the requested application or has management node load requested application on a server
Operating system
Image repositories
Computation
Storage
Networking
Image Repositories







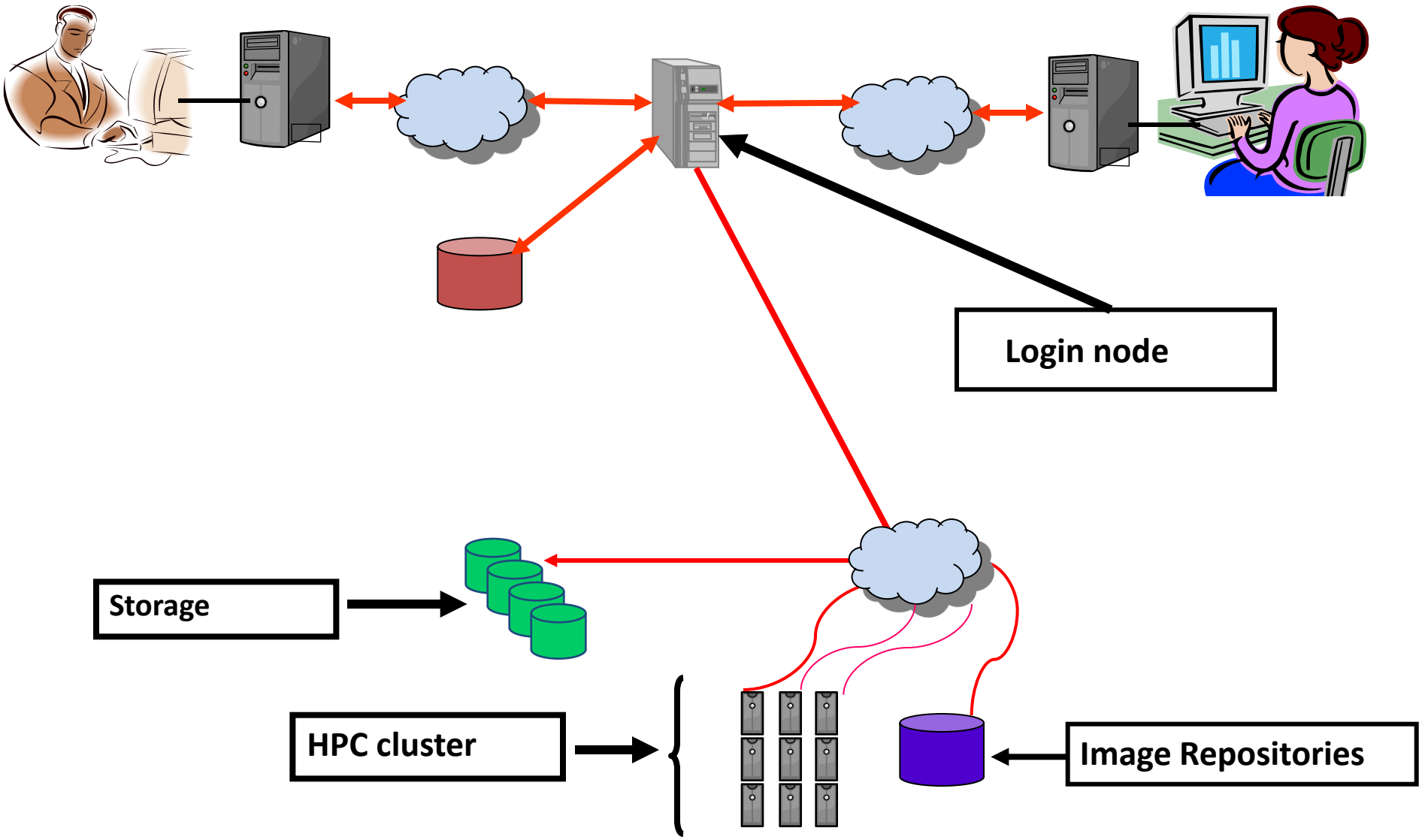


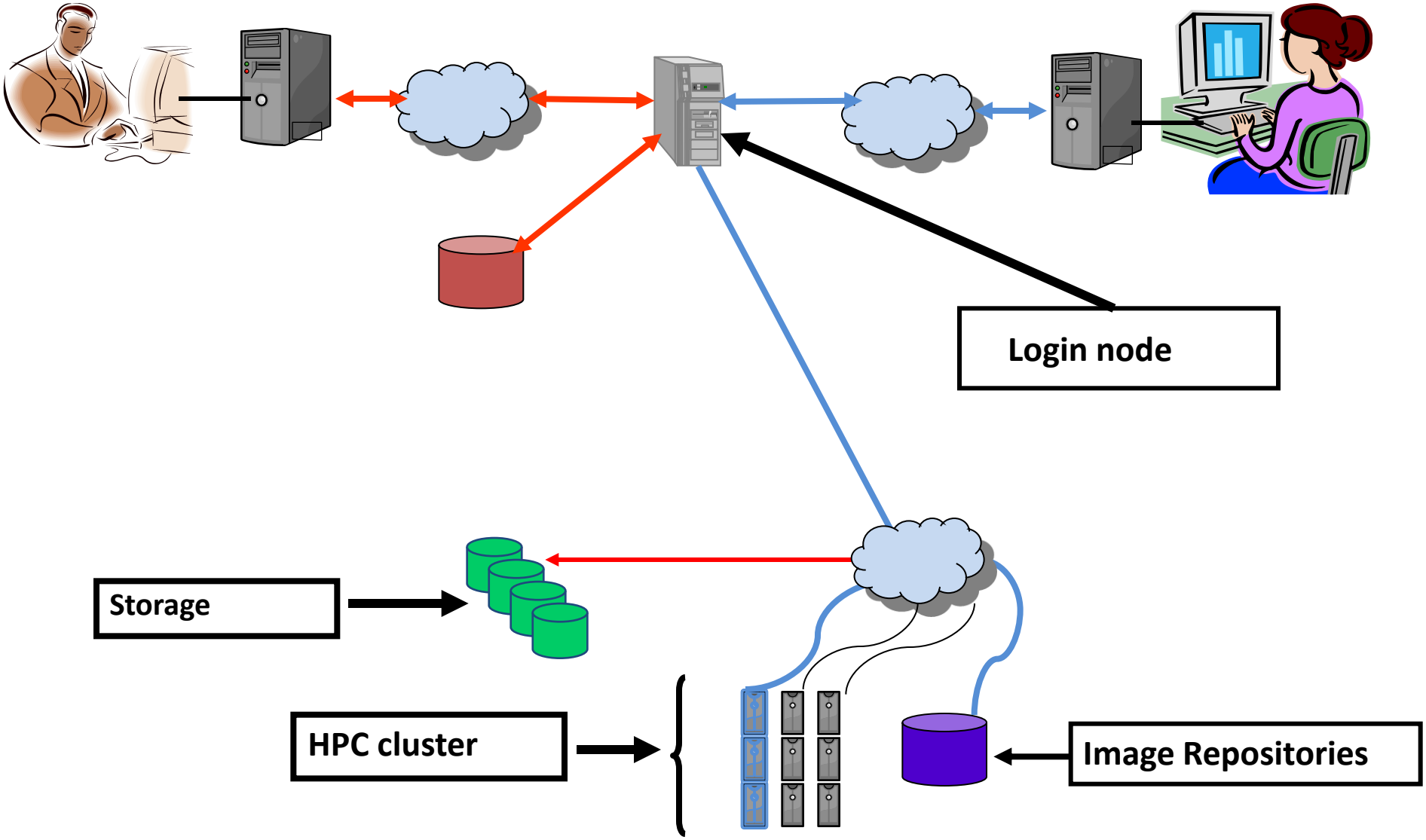


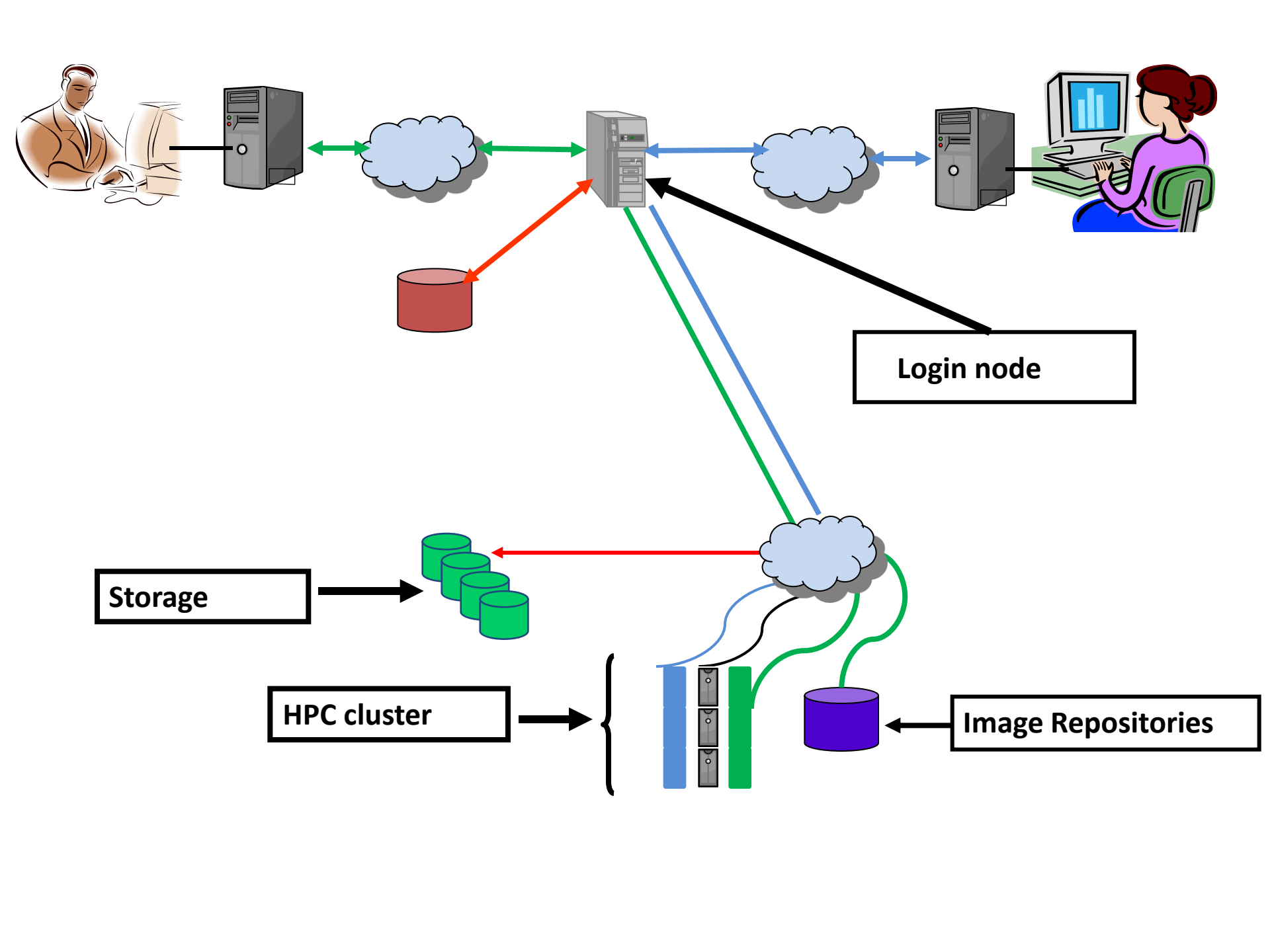
High Performance Computing Option

HPC Configuration

- In the education and research space it is possible to design dynamic hardware reconfiguration of a cloud computing system
- Identify sporadic usage patterns and re-purpose cloud hardware between distributed and HPC usage


Login node
Storage
HPC cluster
Image Repositories

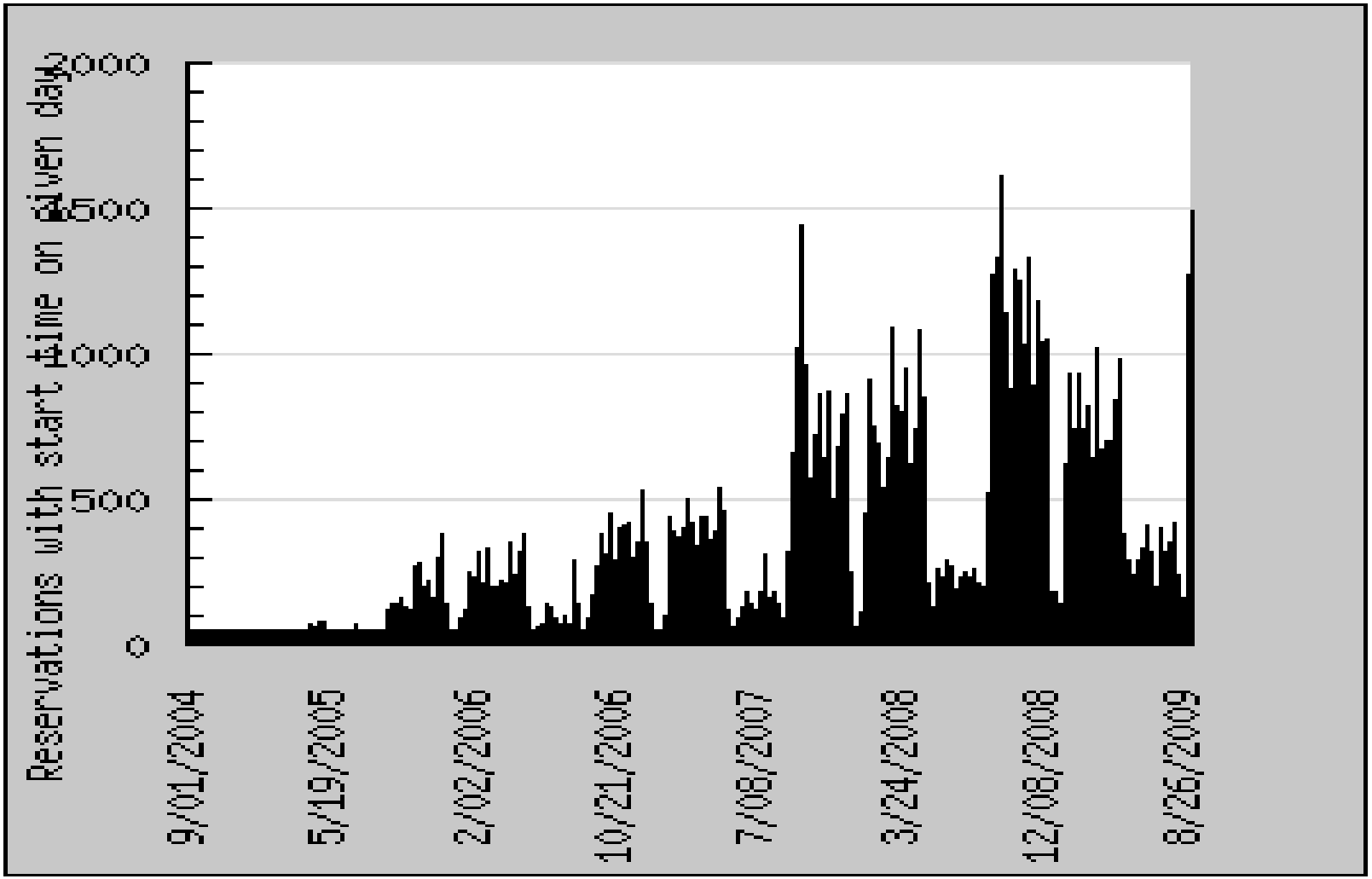

Login node
Storage
HPC cluster
Image Repositories


Login node
Storage
HPC cluster
Image Repositories

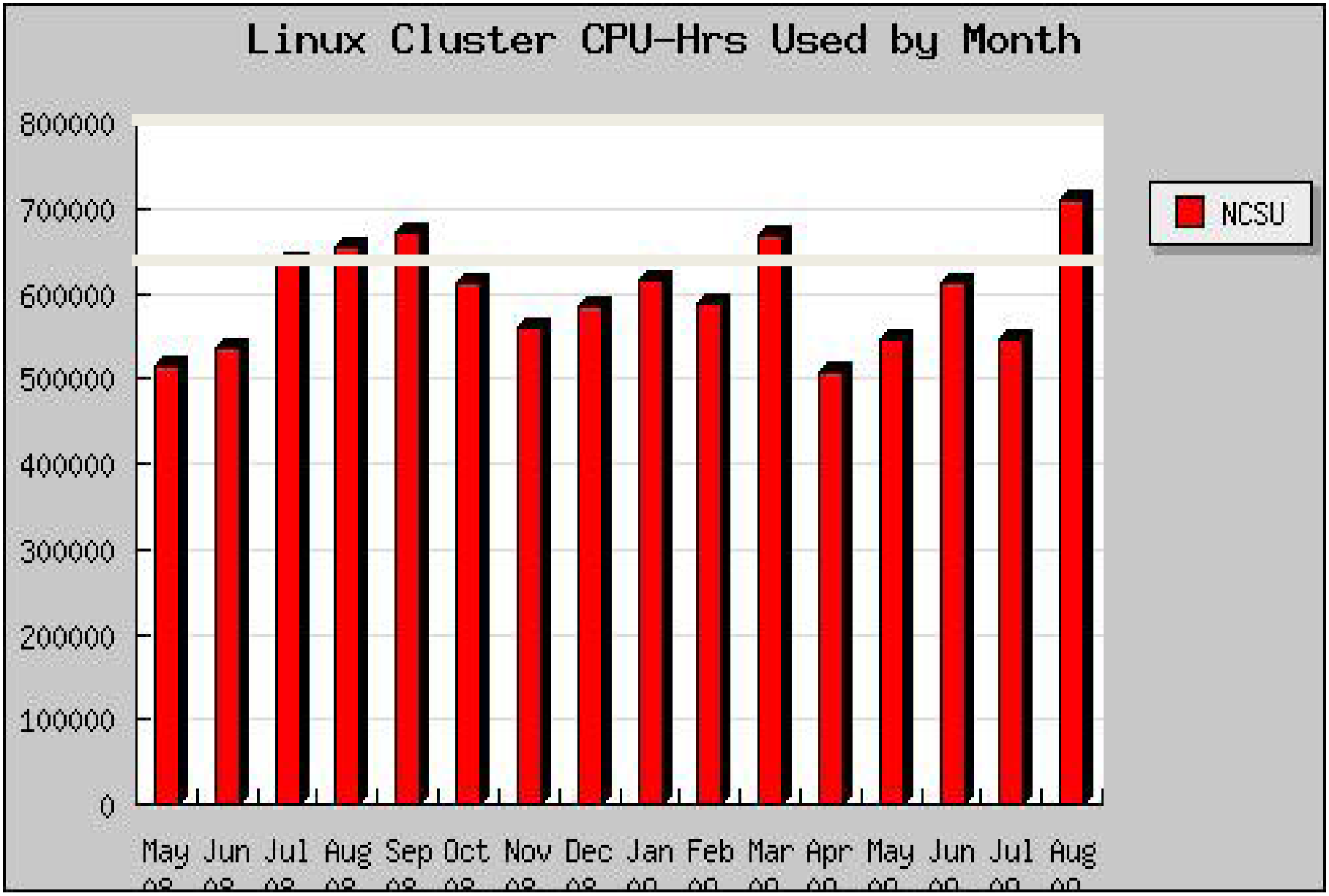
Integrating Cloud Computing and HPC

Periodic Sporadic Usage

North Carolina State – Student system Usage Patterns 9/04 – 9/09



NC State HPC Usage patterns 5/08 – 5/09



Dynamic Blade – Server Re-Configuration

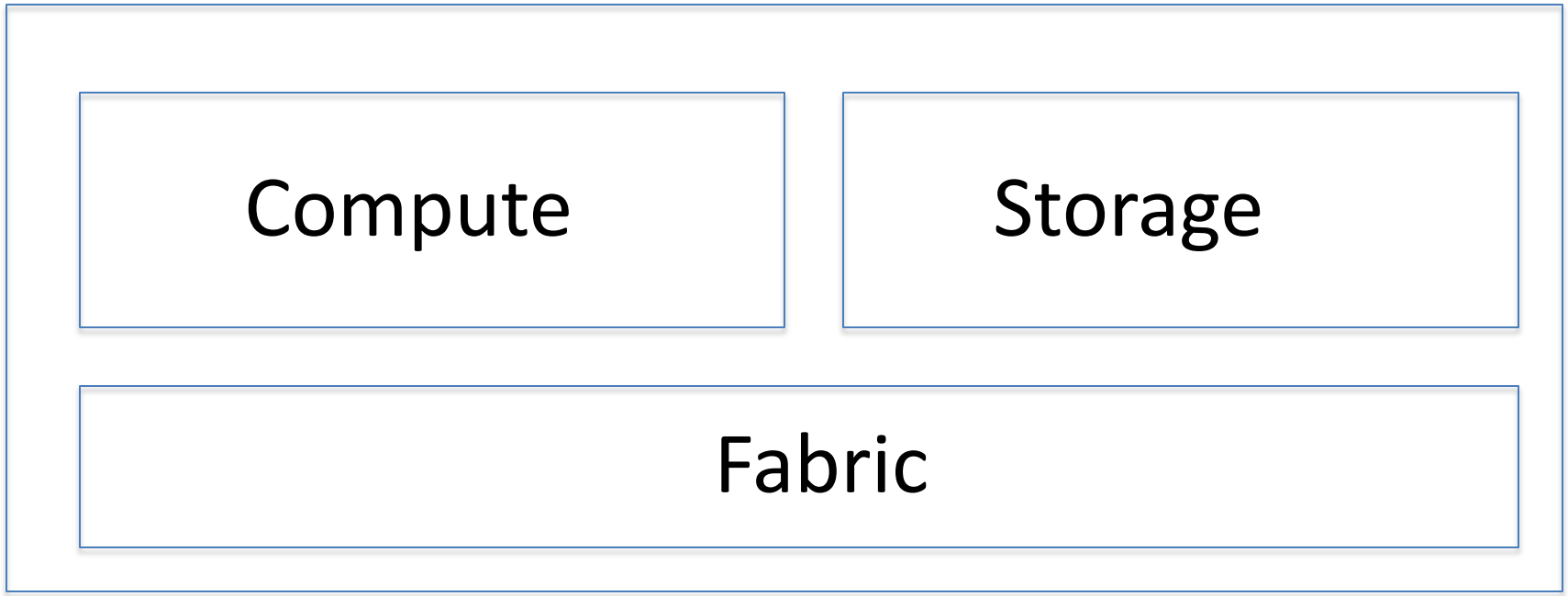
- Insert multiple NICs in the blades or servers
- Build multiple VLANs (non-routing IPs) to
  - Control the out of band management network
  - Access/load images from the image repository
- Create either
  - A public (routable) IP connection
  - VLAN together a cluster of blades / servers
  - Block reserve group of machines for HPC (depending on IB, Myrinet, or 10 GigE config)

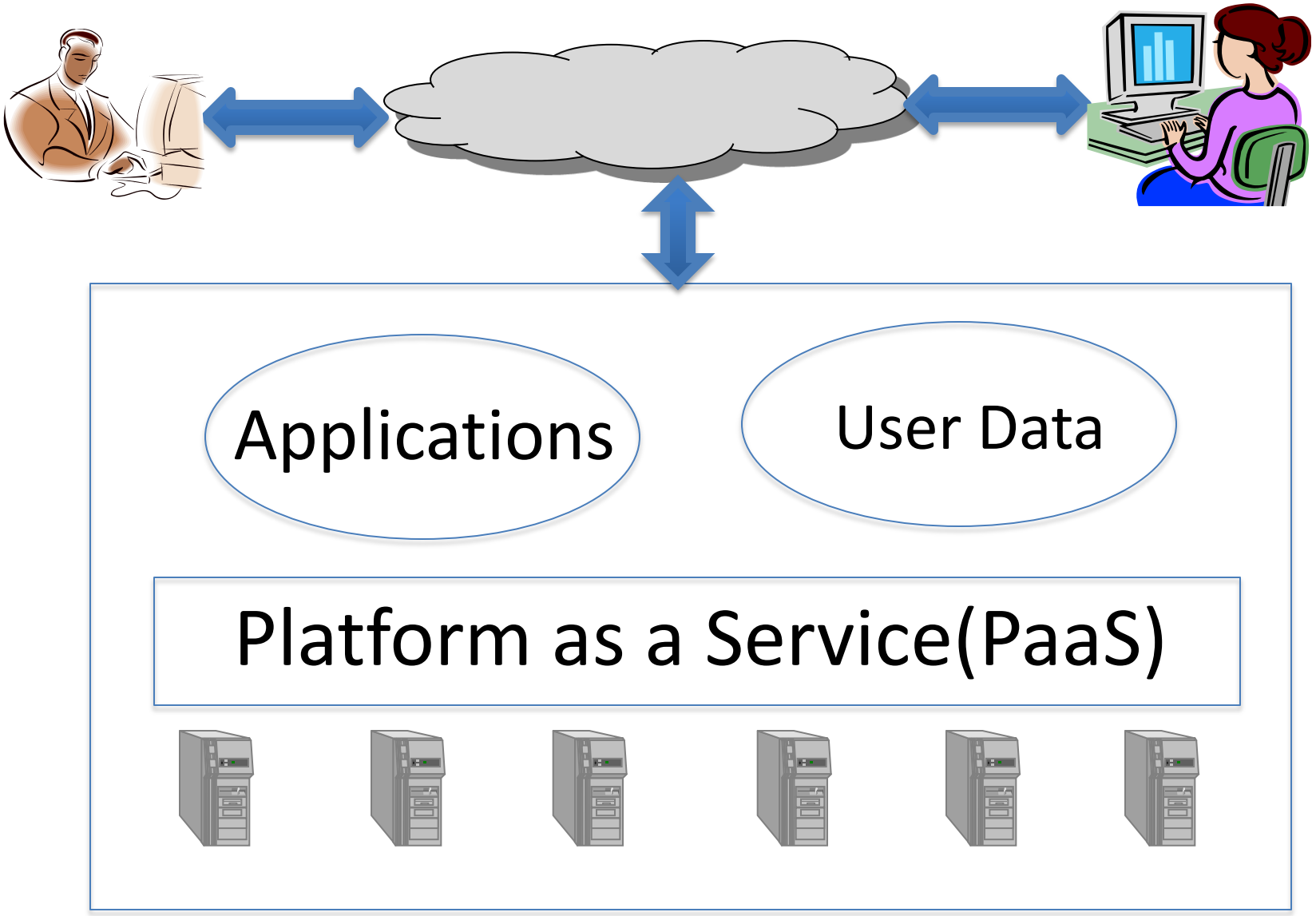
Platform as a Service (PaaS) Characteristics

- Computing platform and solution stack as a service
- Program the cloud, but at a relatively high level, such as Web Services and delivery of web-based applications
- development platforms with the development tool itself hosted in the cloud and accessed through a browser

Platform as a Service (PaaS)
Software as a Service (SaaS)
Application as a Service (AaaS)

Platform as a Service (PaaS) Architecture




Applications
User Data
Platform as a Service(PaaS)
Data Center

Other Cloud Services

- Application as a Service and Software as a Service
  - Both specific cloud services focused on particular software application(s)
  - Extension of existing application stack already running on the local site (more toward business applications)
- Cloud as a Service – construction of entire cloud architecture within a larger cloud computing system

Data Sets and Repositories

- Computational capability needed to process and analyze the data
- Transform data into useful information
- Two options
  - Push the data to the computational system
  - Bringing the computational system to the data

Large Data Sets and Analysis Requirements

- Data set itself
  - Location
  - Size
- Multiple dimensions to consider
  - Technical
  - Operational
  - Economic

Location
Balancing Storage and Computation

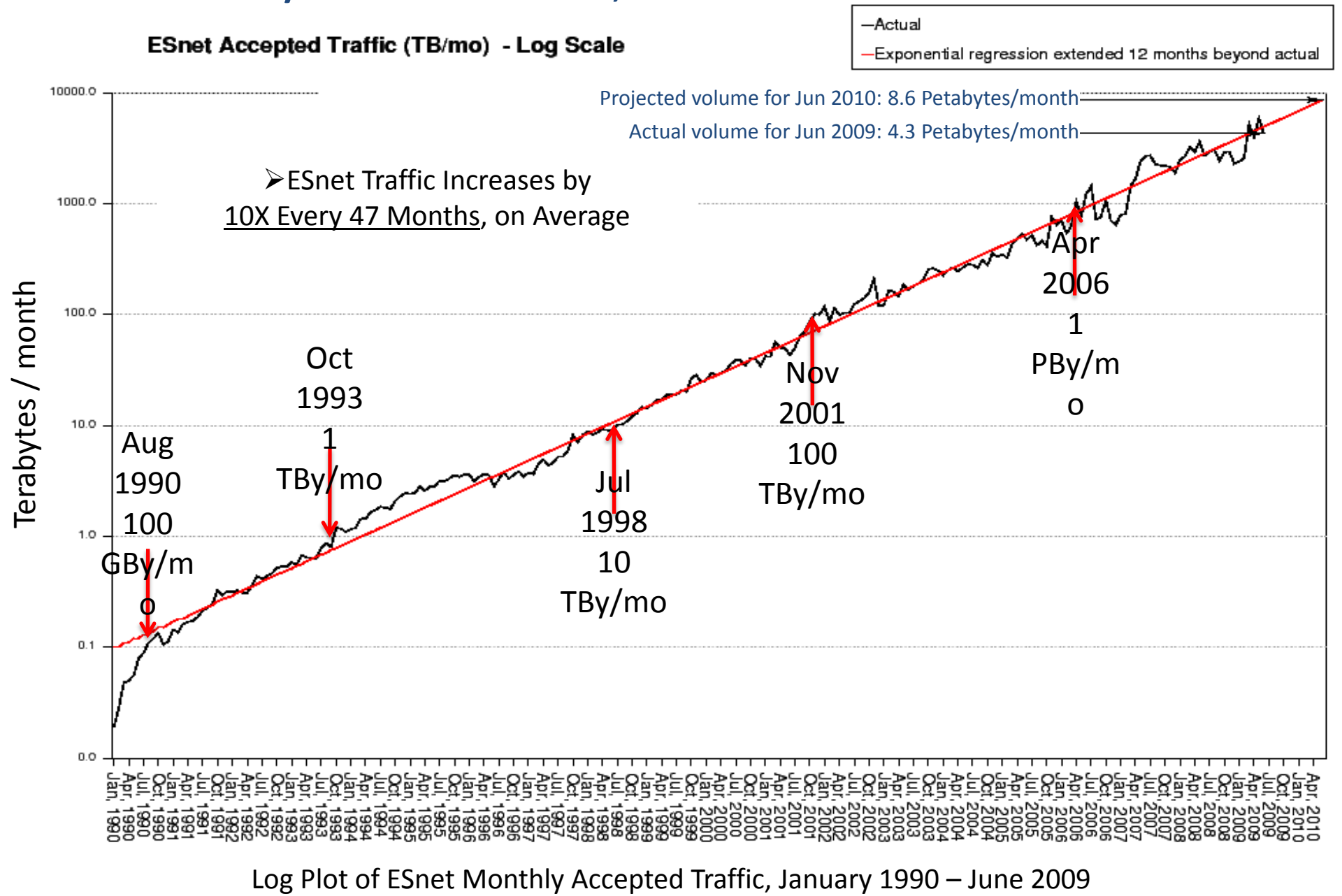
- Where should the data be located and where should the computations be performed?
- Two potential options
  - Option 1 - traditional approach is to move all/part of the data to the computational system
  - Option 2 –move the computation near the data repository

Size of the Data Sets

- Number and size of data repositories are expanding
  - Rate of data collection increasing
  - Aggregate expansion of the total size of the data
- Compare the network bandwidth growth to the growth in the size of data repositories

Observation of Current and Historical ESnet Traffic Patterns

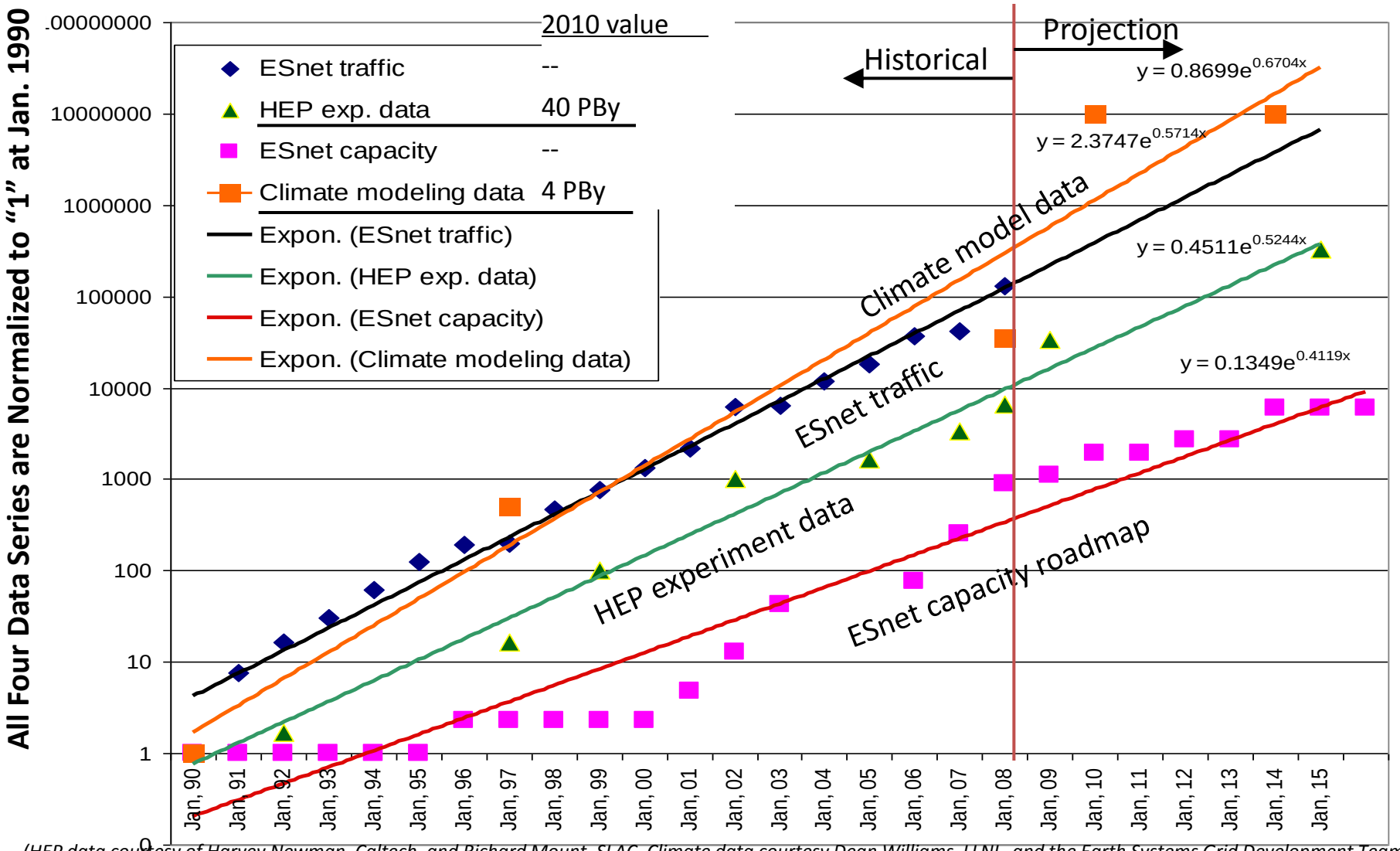
Courtesy of William Johnston, LBL



Log Plot of ESnet Monthly Accepted Traffic, January 1990 – June 2009

Network Traffic, Science Data, and Network Capacity Long-term trends

Courtesy of William Johnston, LBL



(HEP data courtesy of Harvey Newman, Caltech, and Richard Mount, SLAC. Climate data courtesy Dean Williams, LLNL, and the Earth Systems Grid Development Team.)

Size of the Data Sets

- Best performance from the current ESnet traffic patterns is approximately 8.6 Pbytes/month (total throughput)
- Single user may get 20% sustained bandwidth or approximately 1.75 Pbytes/month
- Moving a full data repository of 100 Pbytes may take on the order of 57 months!
- Moving only 1% may take several weeks!
- Size of a fully constructed cloud computing image may be approximately several Gbytes and take fractions of a second to transfer between locations

“Instead of pushing the data to the cloud
Bring the cloud to the data

What is needed to make this happen?

Desirable Cloud Computing and Storage Technical Properties

- **Portability** – ability to move cloud computing job among different computational systems
- **Interoperability** – ability to run cloud job among different computational systems
- **Scalability**
  - Up and down – ability to have cloud computing hob access different generations of similar hardware architecture
  - In and out – ability to expand or contract the size of a cloud computing job on a specific hardware generation architecture

Cloud Computing and Storage
Some Operational Issues/Observations

- Software licensing agreements at each cloud computing location
  - Multi-institution
  - Multi-site cloud installations
- Service Level Agreements
- Security - privacy and sensitive data
- Data and application audits
- Terms and conditions (Master Service Agreement, click-wrap EULAs, distribution of risks)

Demo

- Cloud computing demonstration
  - User interface
  - Access to the image repository
  - Computation options
  - Example session

Some Closing Remarks
Cloud Computing and Storage

- Cloud computing (CC) is a disruptive paradigm in several dimensions
  - Technical
  - Economic
  - Education
  - Research
- There is an intertwining among these dimensions that must be observed when constructing cloud computing
- CC is actually having the largest impact in the business and commercial sector
- The majority of cloud services developed are directed to business applications – not STEM projects and research

Some Closing Remarks
Cloud Computing and Storage

- Not everything is best optimized by migrating a cloud computing image to the data repository
- What are the proper decision criteria when and where to locate computation and storage
- Which leads to the second part of the tutorial

Policy-based Data Management – Reagan Moore

Questions