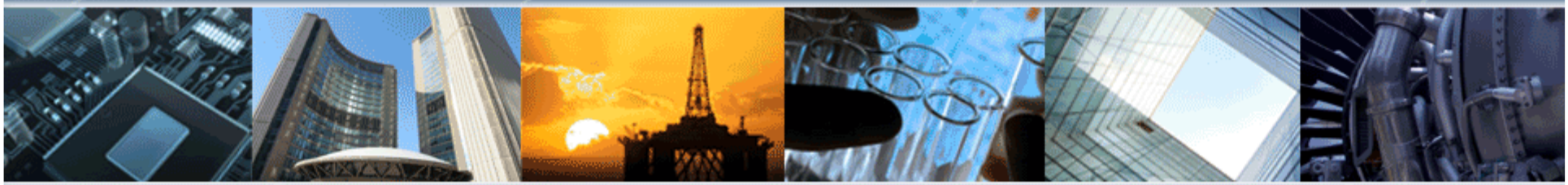
Platform™
The Power of Sharing

Dynamic HPC Datacenter - the best of Grid & Cloud HPDC 2009 Munich



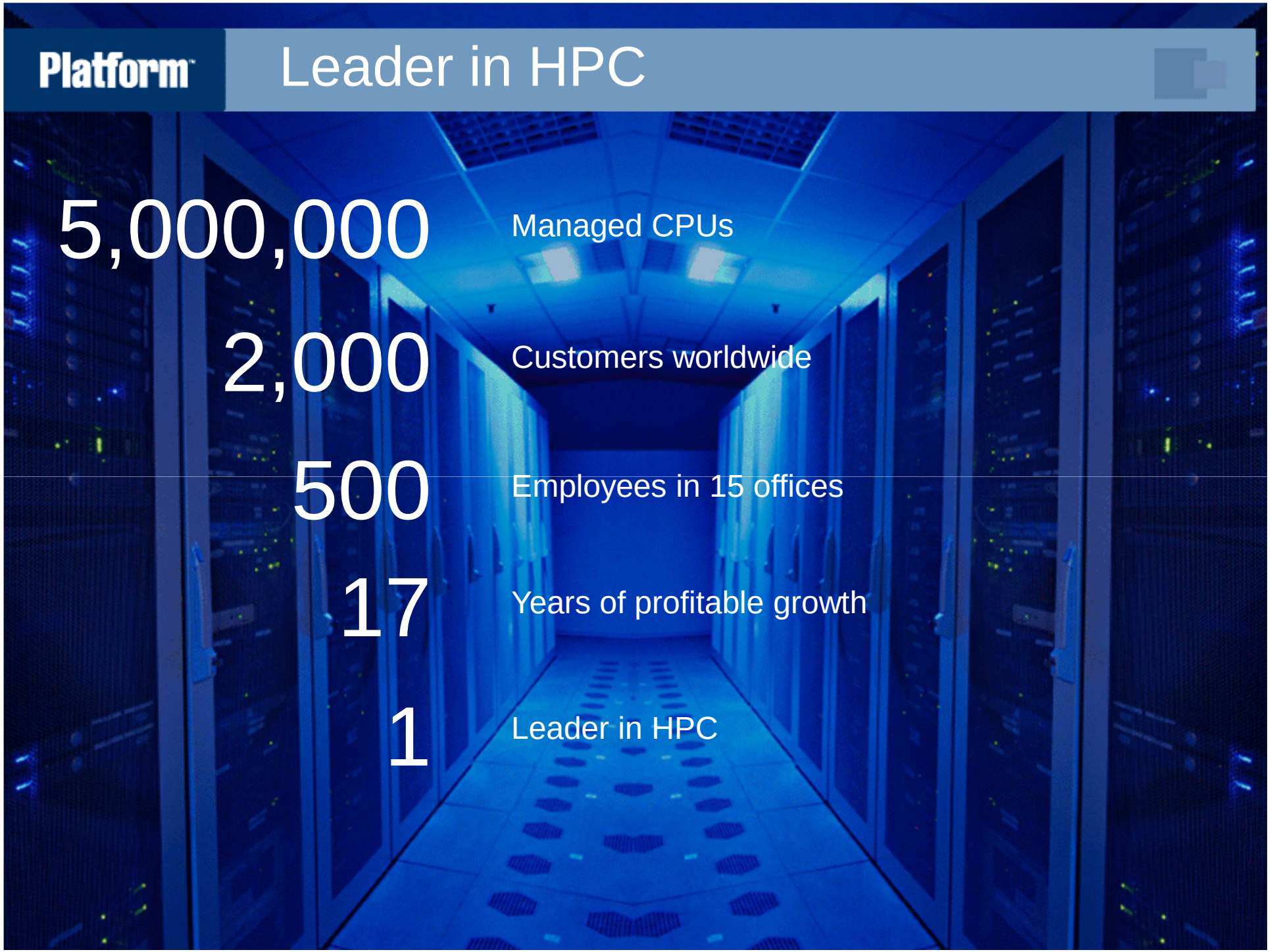
Bernhard Schott
EU-Research Program Manager bschott@platform.com Platform Computing GmbH

Date: June 2009

Agenda

- Introduction on Platform Computing
  - Platform solutions for HPC in research and industry
- The best of Grids and Clouds
  - Managing the dynamic Datacenter
- Introducing Platform DDC
  - Dynamics from the Cloud – Performance from the Grid
- Summary
- Sorry, not today: Green datacenter

http://www.ogf.org/OGF25/materials/1654/Energy+Optimization+of+Existing+Datacenters+-+Bernhard+Schott+-+Platform.pdf


Platform
Leader in HPC
5,000,000
Managed CPUs
2,000
Customers worldwide
500
Employees in 15 offices
17
Years of profitable growth
1
Leader in HPC

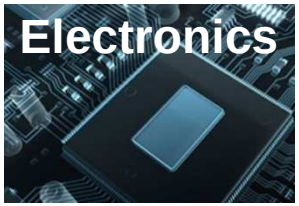
Platform Computing

- **Recognized leader and pioneer in Grid computing and HPC**
  - 17 years solving the most challenging enterprise distributed computing problems
  - Global offices, resellers and partners
  - 24x7 worldwide service, support, and consulting
  - Continual innovation in new product development & open standards
  - Growing and profitable since its inception

Winner of Gartner's "Cool Vendor, 2006"
Platform Computing awarded "Cool Vendor in IT Operations Mgmt, 2006"

Platform Computing in Top 10 Virtualization Vendor to Watch in 2008:
http://www.cio.com/article/160951

Platform — Industries Served by Platform



- AMD
- ARM
- Broadcom
- Cadence
- Cisco
- Infineon
- MediaTek
- Motorola
- NVidia
- Qualcomm
- Samsung
- Sony
- ST Micro
- Synopsys
- TI
- Toshiba

- BNP
- Citigroup
- Fortis
- HSBC
- KBC Financial
- JPMC
- Lehman Brothers
- LBBW
- Mass Mutual
- MUFG
- Nomura
- Prudential
- Sal. Oppenheim
- Société Générale

- Airbus
- BAE Systems
- Boeing
- Bombardier
- Deere & Company
- Ericsson
- Honda
- General Electric
- General Motors
- Goodrich
- Lockheed Martin
- Nissan
- Northrop Grumman
- Pratt & Whitney
- Toyota
- Volkswagen

- Agip
- BP
- British Gas
- China Petroleum
- ConocoPhillips
- EMGS
- Gaz de France
- Hess
- Kuwait Oil
- PetroBras
- Petro Canada
- PetroChina
- Shell
- StatoilHydro
- Total
- Woodside

- CERN
- DoD, US
- DoE, US
- ENEA
- AWE
- Georgia Tech
- Harvard Medical School
- Japan Atomic Energy Inst.
- MaxPlanck Inst.
- MIT
- SSC, China
- Stanford Medical
- TACC
- U. Tokyo
- Washington U.

- Abott Labs
- AstraZeneca
- Celera
- DuPont
- Eli Lilly
- Johnson & Johnson
- Merck
- National Institutes of Health
- Novartis
- Partners Health Network
- Pharsight
- Pfizer
- Sanger Institute

Other Industries

AT&T Bell Canada DreamWorks Animation SKG GE
IRI Telecom Italia Telefonica Walt Disney Co.

Solutions with Partners

Platform OCS 5 and Platform Manager integrated in Dell cluster systems

Platform LSF, Platform Manager form key parts of Unified Cluster Portfolio

Platform enterprise solutions support a wide range of IBM HPC systems

Platform delivers first certified Intel® Cluster Ready solution, Platform OCS 5

Integrates Platform LSF and Platform Symphony in grid solutions

Platform OCS 5 powers the Red Hat® HPC Solution

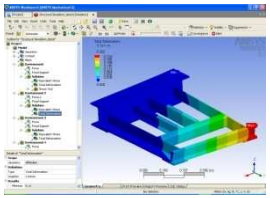
OEMs Platform's core technology in SAS® applications

Platform™

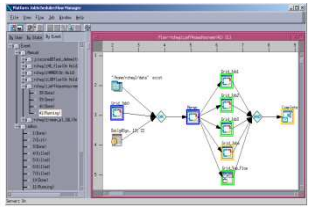
Platform Solutions for Research & Education

Delivering HPC Resources to Users

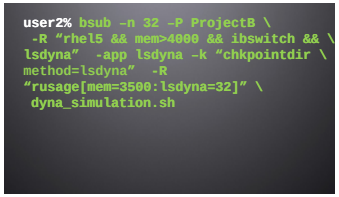
User Interfaces



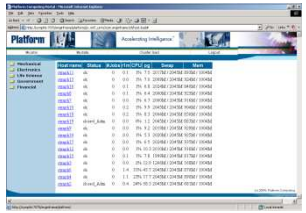
Integrated App



Flow manager



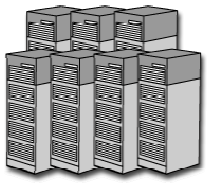
CLI



Web portal

Workload Management

Scheduling Master



HPC clusters

Desktops

Application Execution

Platform Accelerate provides:

- Ease of use and secure UI
- High performance workload scheduler
- Data management via Web portal or flow manager
- Fault tolerance of hardware failure

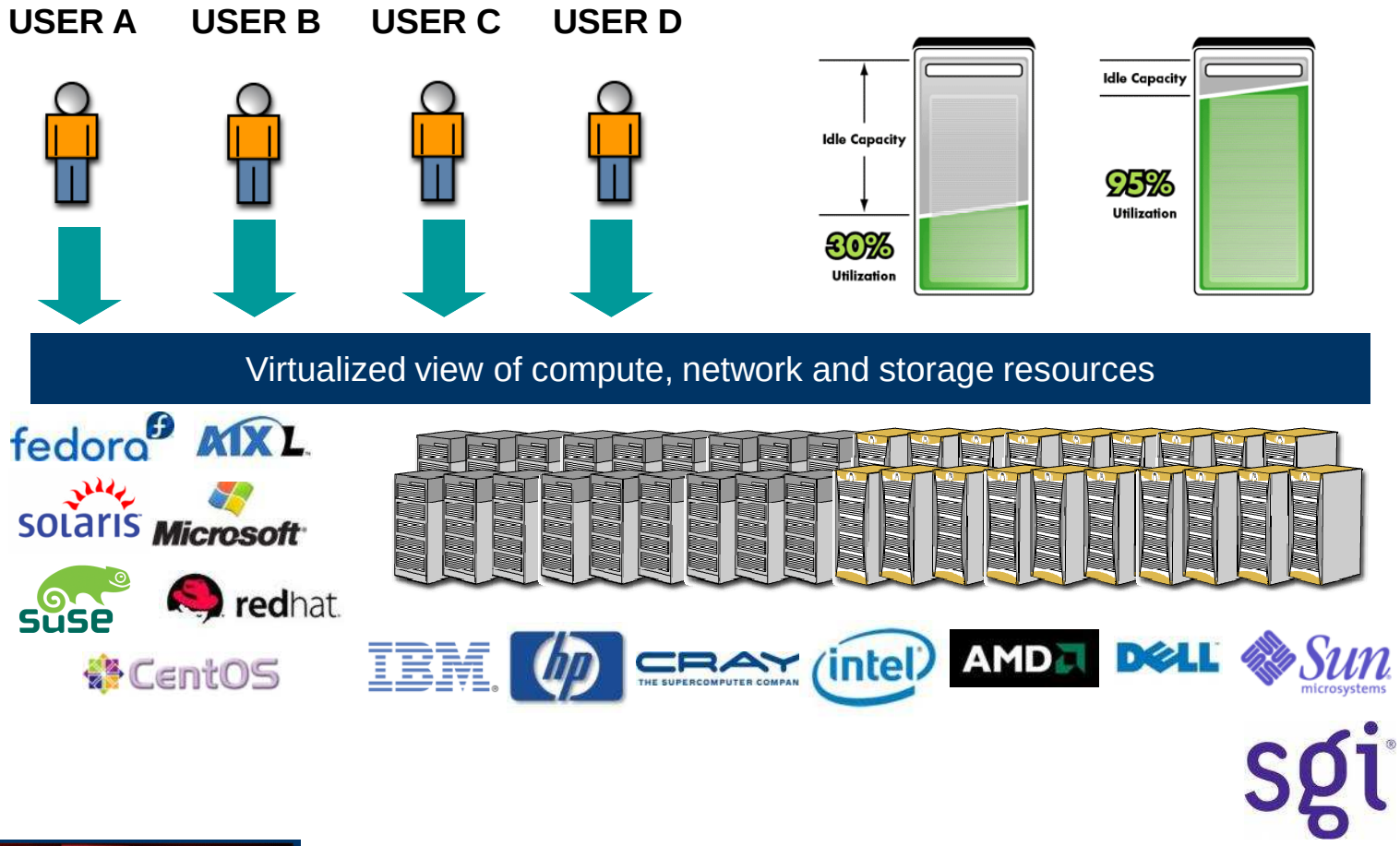
Solution components:

- Platform LSF
  - Platform LSF Desktop
- EngineFrame
- Platform Process Manager

Platform Accelerate

Faster Apps, Higher Asset Utilization

When we grid enable services or applications, applications run more quickly, asset utilization is higher and reliability comes "**for free**"

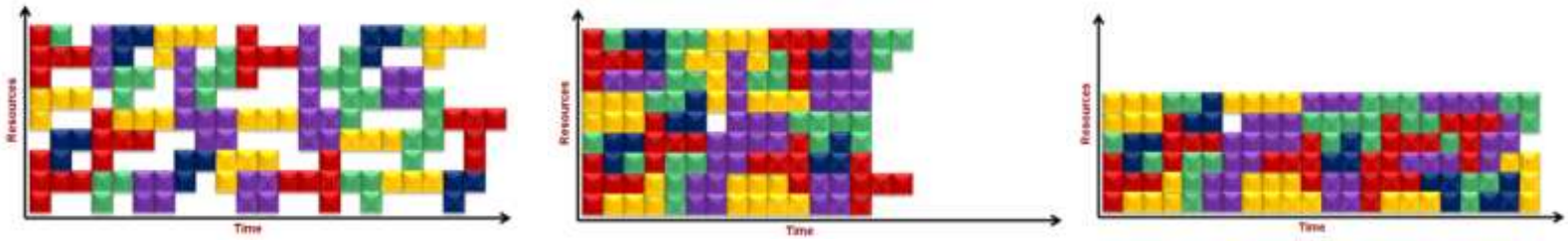


Platform Accelerate

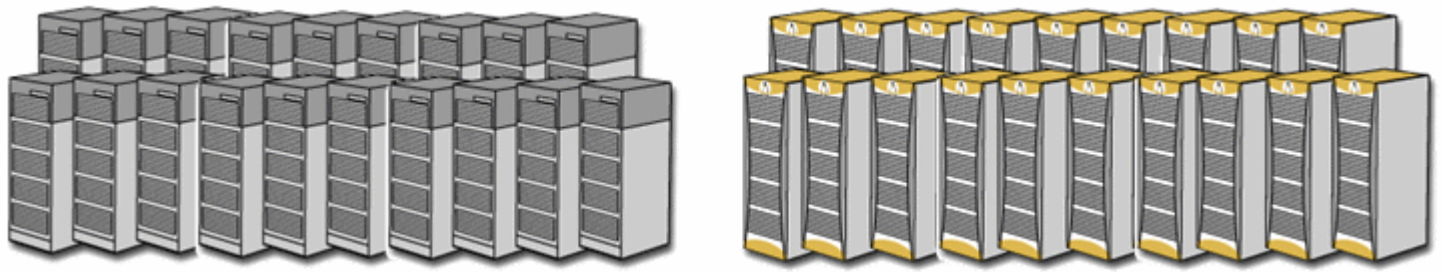
Platform LSF

Platform Faster Apps, Higher Asset Utilization

By scheduling workloads intelligently according to policy, Platform LSF reduces application run-times and optimizes resource use.



Virtualized view of compute, network and storage resources



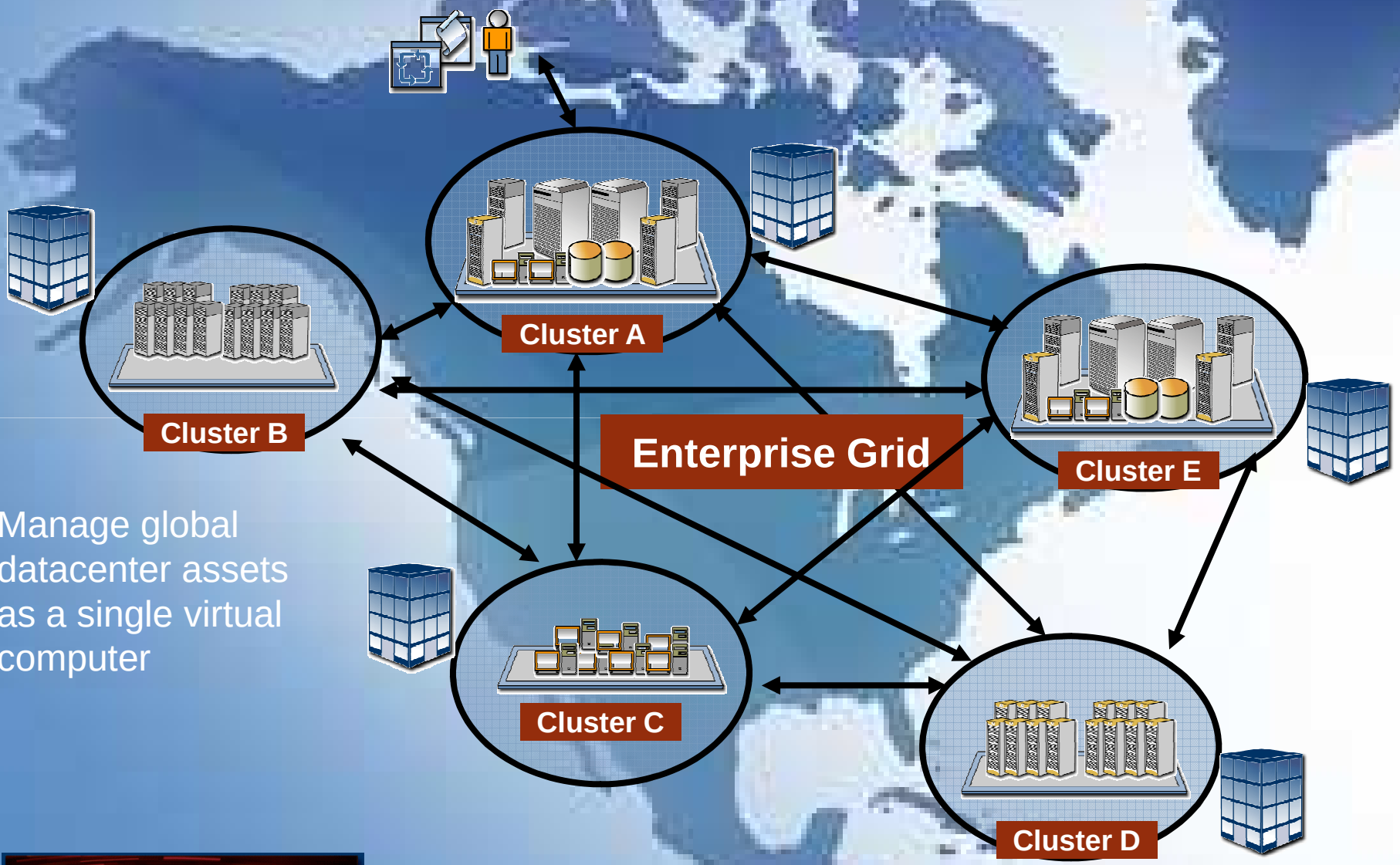
Platform Accelerate

Platform LSF

Multi-site Resource Sharing

- Unmatchable **resource sharing**
  - Combine clusters into one resource pool
  - Centralized control across sites
  - Increase productivity
    - Faster turnaround
    - Resource prioritization
    - Better job throughput
  - Defer future hardware purchases
  - Lower operational cost through standardization
  - Reduce costs of downtime with high fault tolerance

Platform Multi-site Resource Sharing



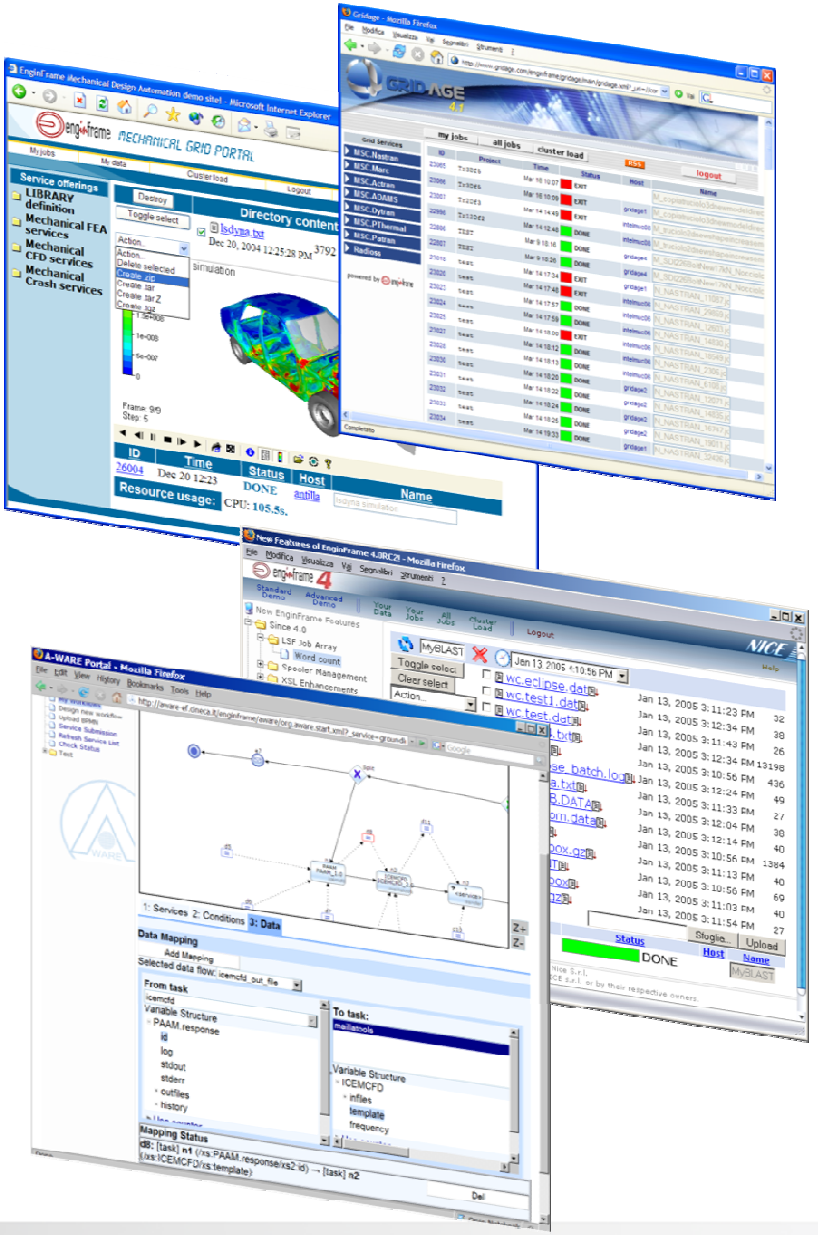
Manage global datacenter assets as a single virtual computer

Platform Accelerate

Platform LSF MultiCluster

Simplifying User Access to HPC

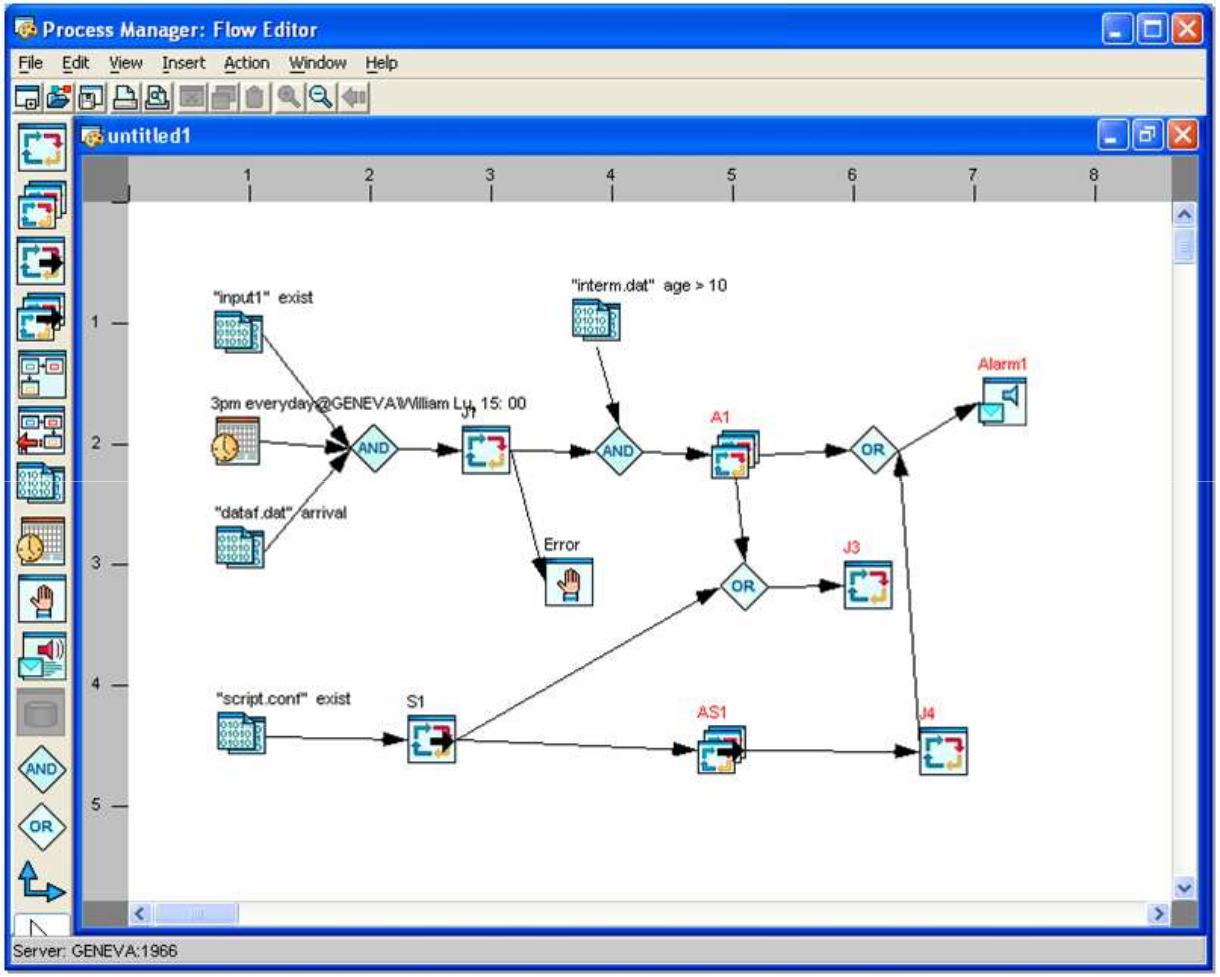
- Highly customizable job submission
- Flexible, cost-effective integration of CAE & Windows desktops
- Intuitive job monitoring and management



Platform Accelerate

EngineFrame

Platform Manage, Automate Complex Workflows

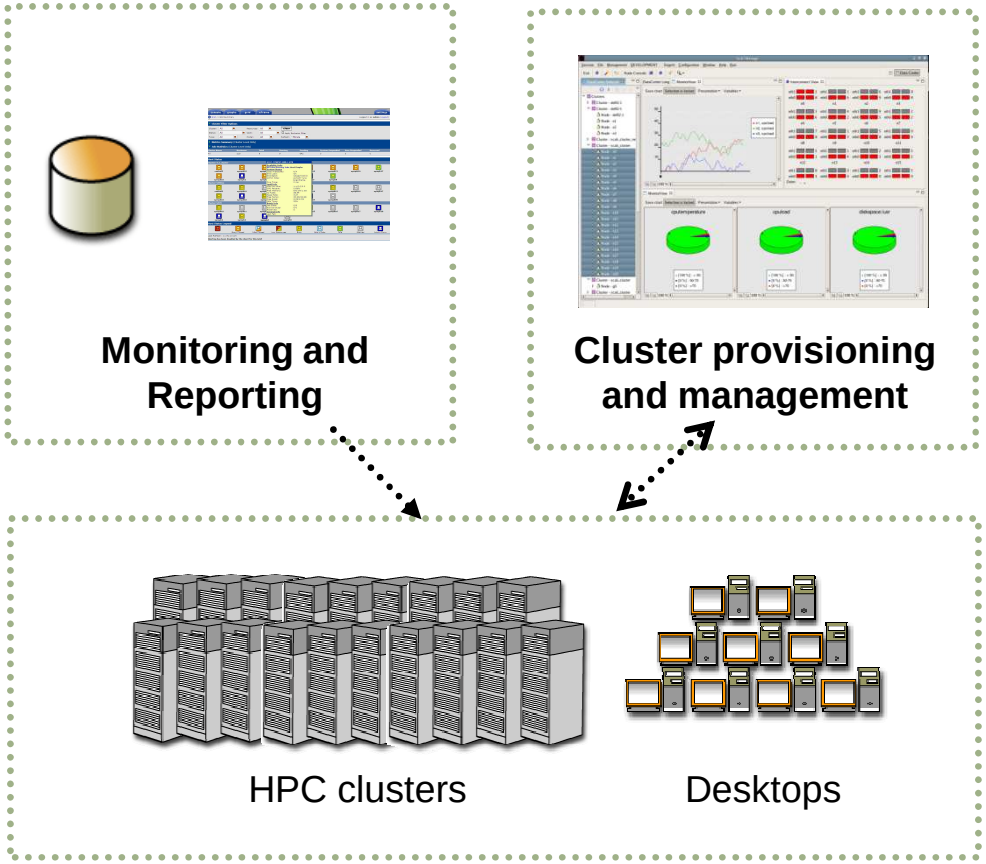


- Manages data and job flow in one interface
- Automatically parallelizes the flow and executes in distribute environment
- Documents the flow to retain IP

Platform Accelerate

Platform Process Manager

Operating HPC Systems Economically



Platform Manage provides:

- Monitoring of multiple clusters for hosts, workload, & floating software licenses
- Host provisioning
- Remote administration

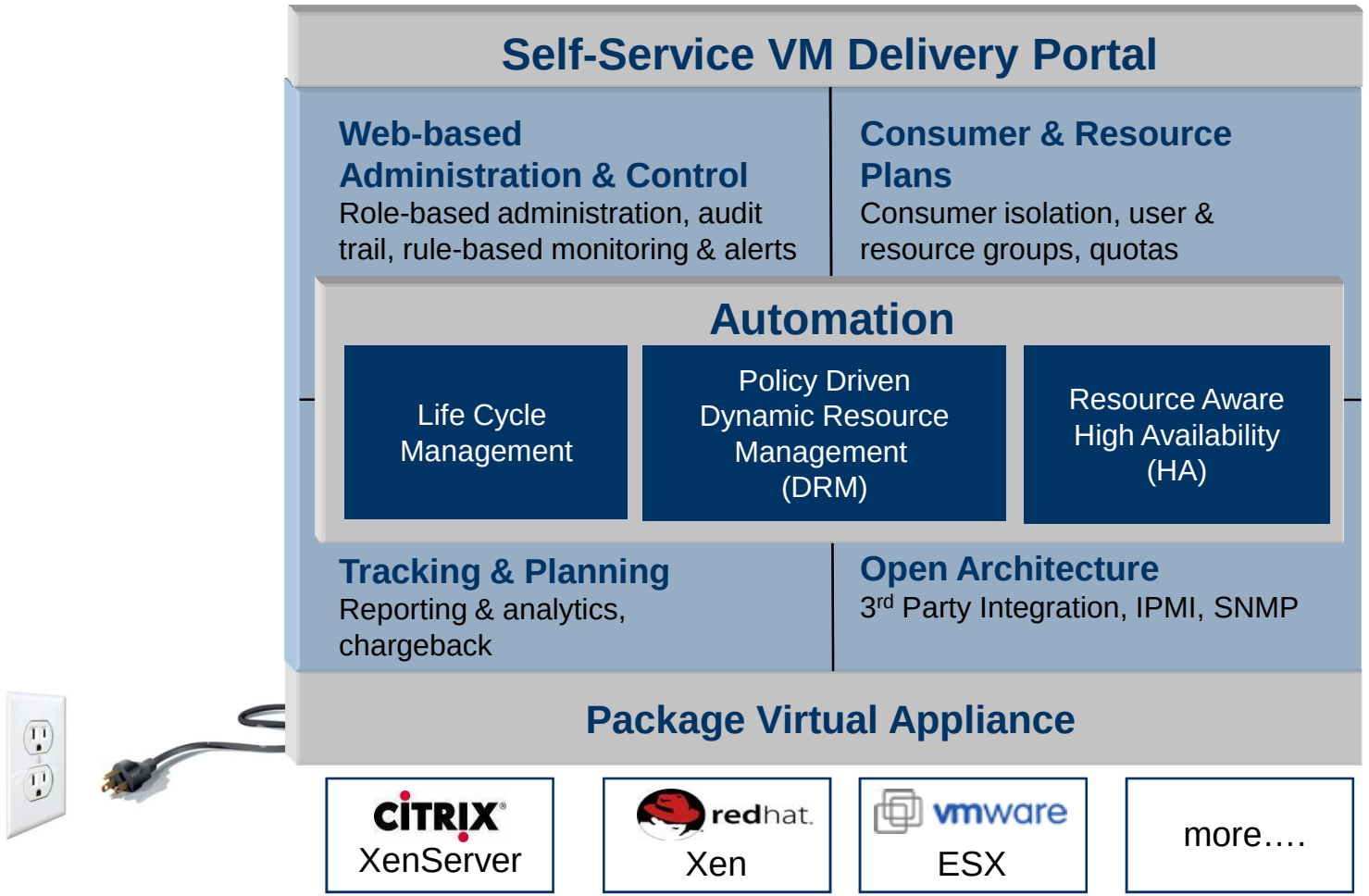
Solution components:

- Platform RTM
- Platform Manager
- OCS

Platform Manage

Automated VM Delivery and Management

Platform VMO



VMO Maximizes ROI from Virtualization

Platform Reporting, Tracking, Monitoring

- Dashboard for multiple workload management clusters
- Job profiling
- Historical performance reporting
- Floating license monitoring
- Job accounting and charge back

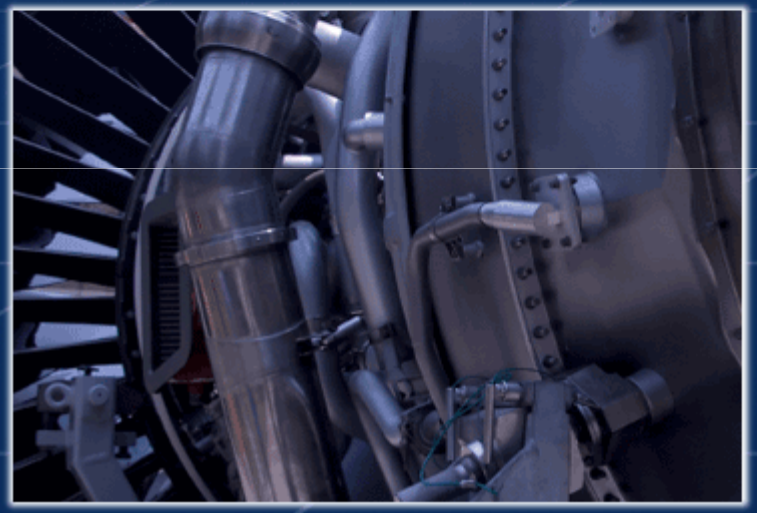
Platform Manage

Platform RTM

www.HPCCommunity.org

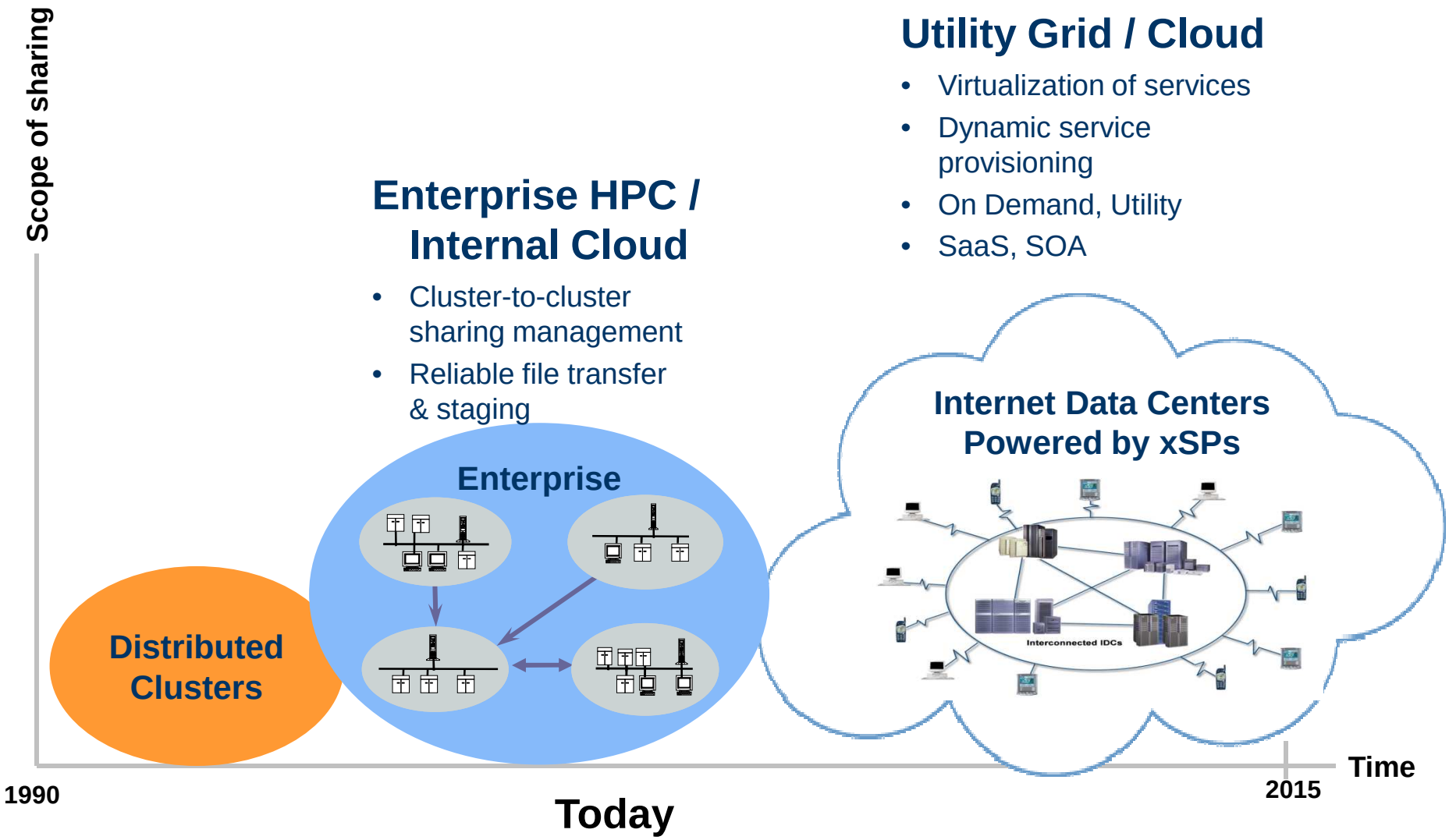
- Open source components Platform LSF
- LSF Perl extensions
- Free access to KUSU & LAVA
- EGO-SDK developer kit for direct SOI access

Platform™



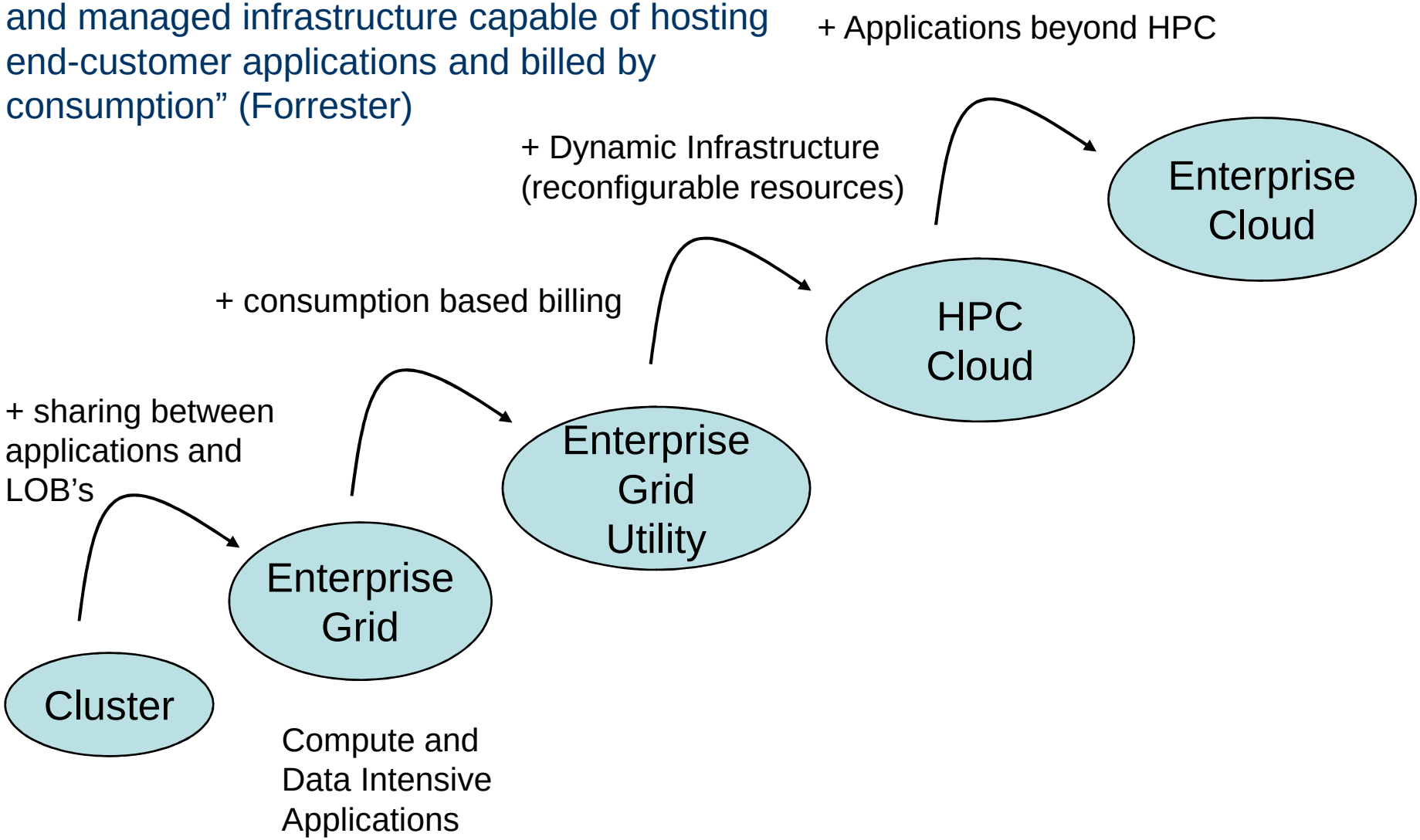
Taking the best from Grids & Clouds for HPC

Evolution of HPC Adoption



Grids to Clouds

Cloud – “A pool of abstracted, highly scalable, and managed infrastructure capable of hosting end-customer applications and billed by consumption” (Forrester)



HPC systems, application clusters

Common Practice:
HPC resources are acquired for specific purpose. They are typically dedicated for single type of work

Capacity limit

- The total capacity is limited by the size of the system or cluster

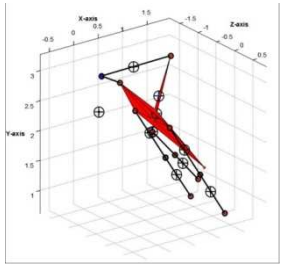
Utilization

- Provisioned for peak load
- Even if it is not fully utilized, it can't be repurposed for other applications

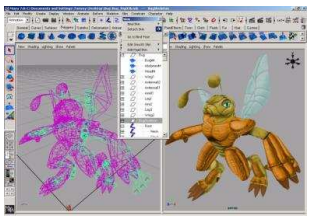
Quick Resource Provisioning

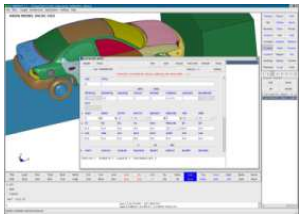
- Users compare their own HPC resource with external "cloud"

The Concept of Cloud



o Unlimited application resources
o Instant resource availability
o Ease of use





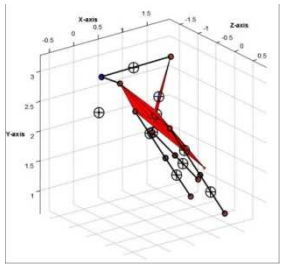
Providing application or compute resource as a service

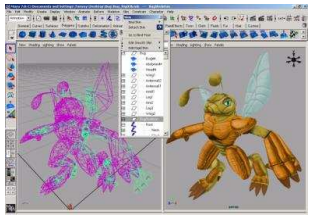
HPC: Matching Supply & Demand

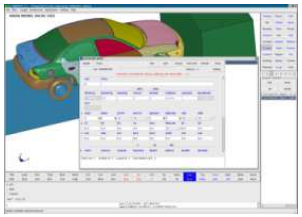
DEMAND

SUPPLY

End Users







Modeling

Rendering

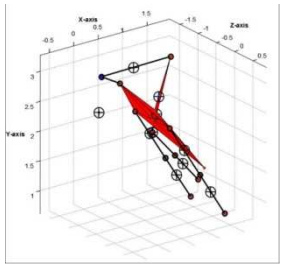
Analysis

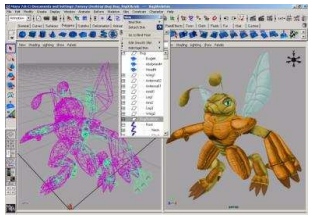
HPC: Matching Supply & Demand

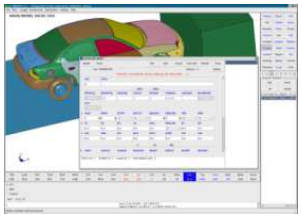
DEMAND

SUPPLY

End Users

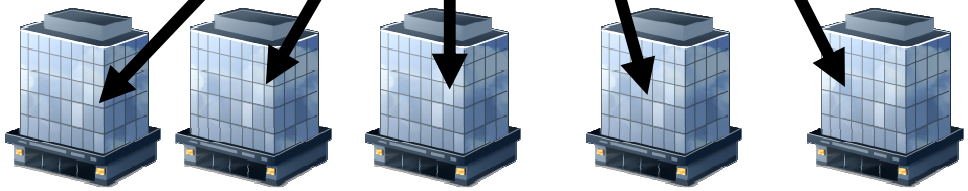






Cloud Environment

Dynamic resource management



HPC: Matching Supply & Demand

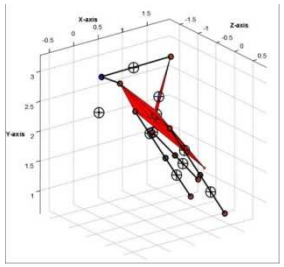
DEMAND

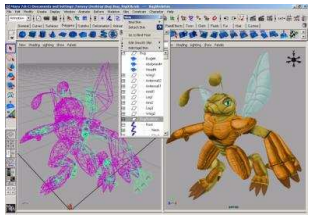
SUPPLY

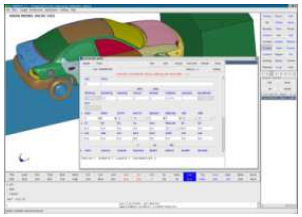
Mixing grid & cloud:

- Workload management
- Cluster management
- Dynamic VM and OS management
- Accounting & chargeback

End Users

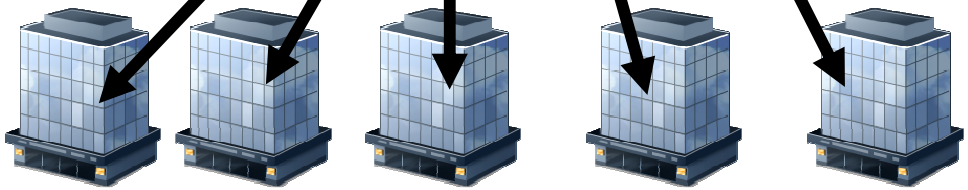






Cloud Environment

Dynamic resource management



Matching Supply & Demand

A note on heterogeneity:

- Applications require different resources:
  - OS: RHEL3.x, RHEL4.x, RHEL5, SL3.5, Windows2003, ….
  - Configurations, patches: mounts, permissions
  - Pre-installed tools and libraries: Java, compilers, …
- Users require applications in different version and configurations

- Legacy issue:
  - “Old” applications are valuable!
  - Legacy applications need older OS, tools, config, …
  - How old are your applications?

- The real world is heterogeneous – also in HPC!

Internal and External Cloud

External Cloud by Service Providers

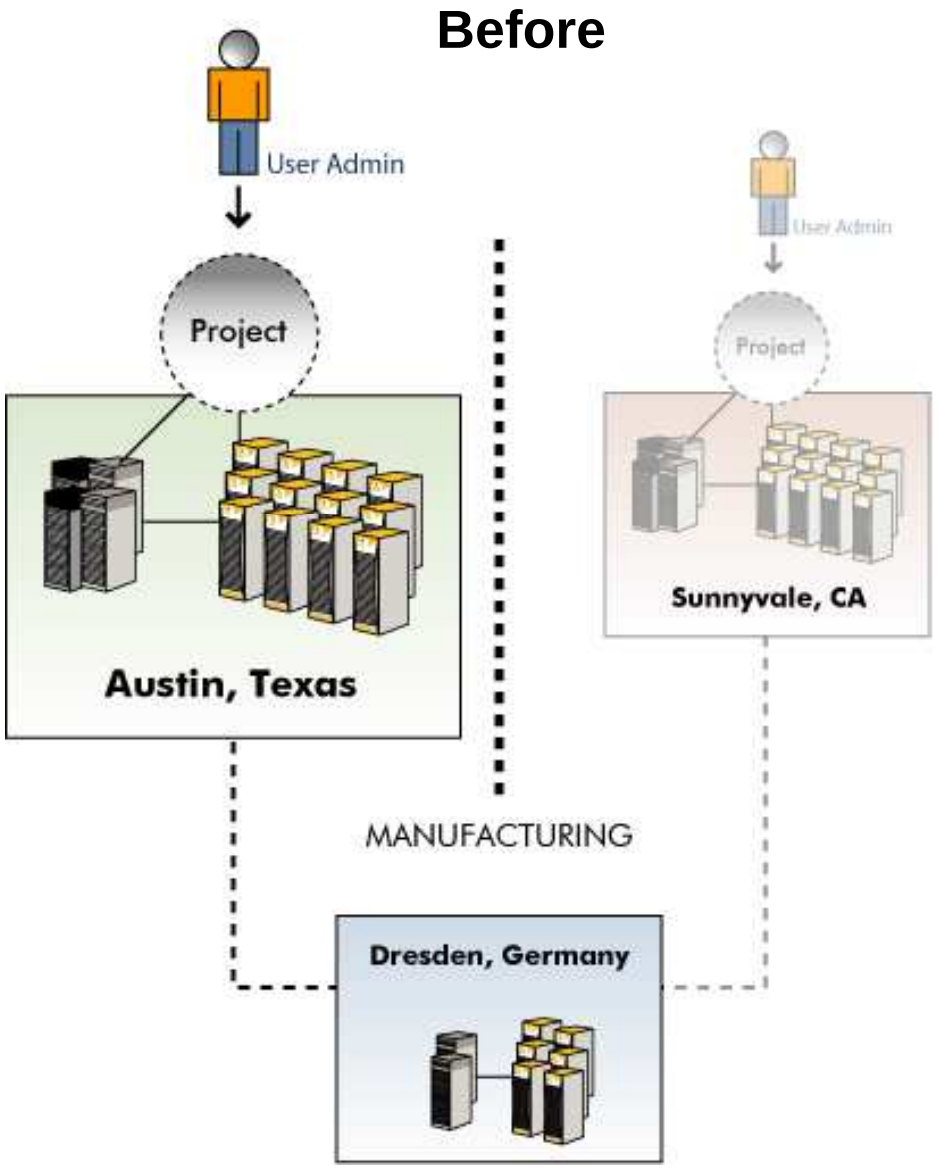
- CapEx reduction
- Non-mission critical SLAs
- In-house IT has limited scale, scope or expertise

Internal Cloud by HPC Center

- CapEx and OpEx reduction
- Maximize value of underutilized resources
- Mission critical SLAs
- High security requirements
- Enterprise-specific services
- Less legal issue for application licenses

AMD HPC environment

- More design, simulation & verification – faster
- Better utilization of resources in an always-available computing environment
- Better products to market faster and at lower cost



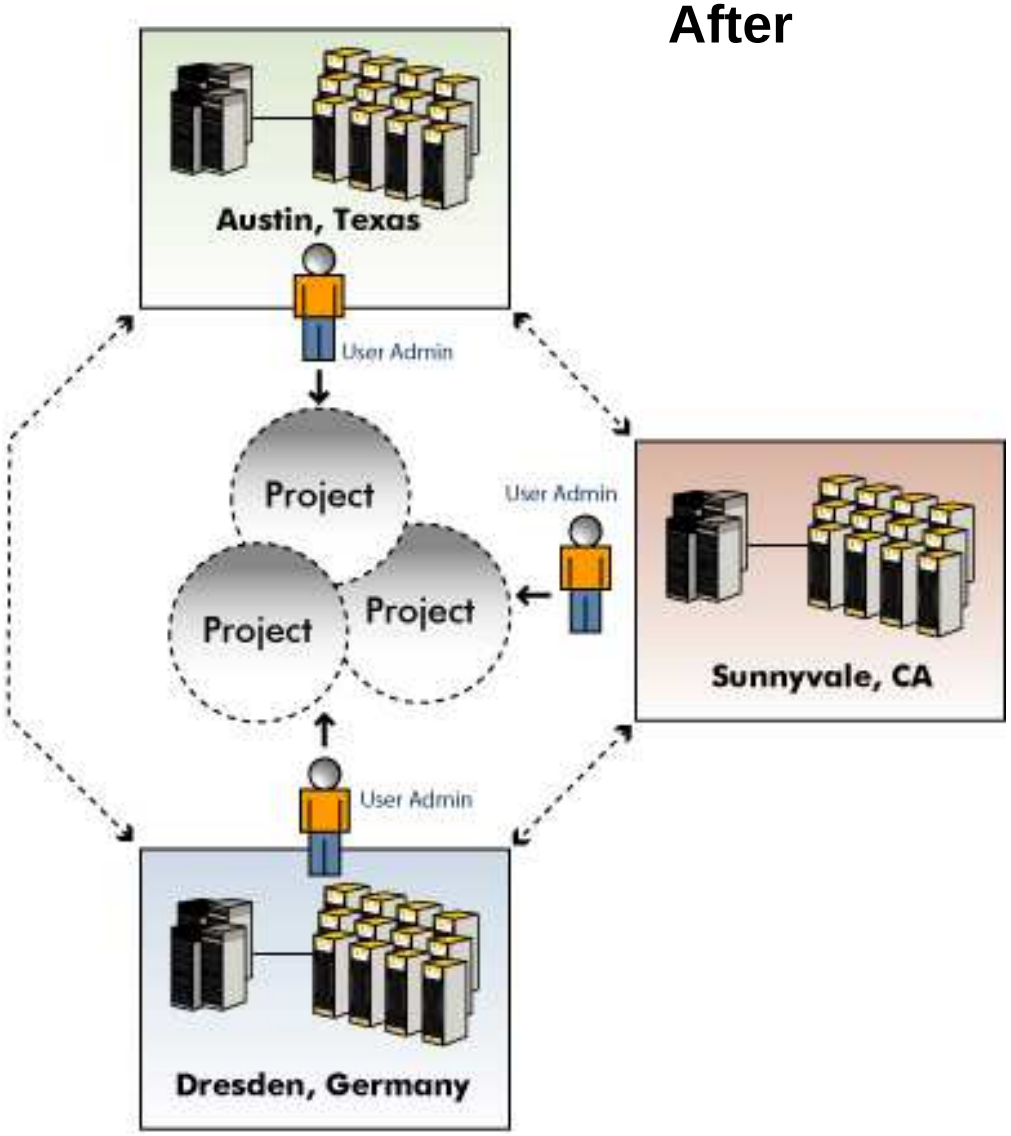
Powered by

AMD HPC environment

AMD

- More design, simulation & verification – faster
- Better utilization of resources in an always-available computing environment
- Better products to market faster and at lower cost

After



Powered by

Platform™

Citi – Corporate Shared HPC Services

FX derives Pricing & Hedging

Converts Pricing & Hedging

Enterprise Mkt Risk

Credit Derivs, Pricing/Hedging

Counterparty Credit Risk

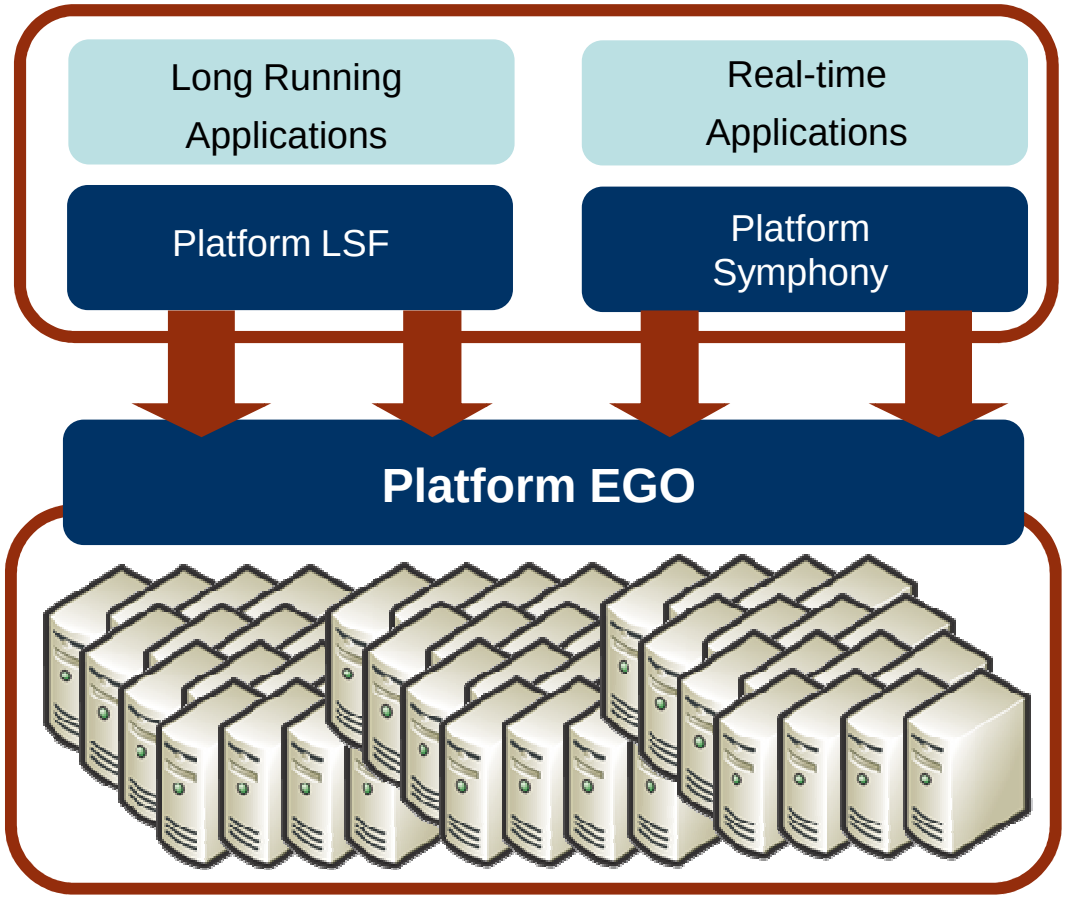
Fraud, Anti-Laundering

Acc'ting, Actuarial Analysis

Operational Risk

CRM, Data Mining, Credit Scoring

More & more apps from LOB silos



Powered by

Platform Dev Test Environment

Software build and QA environment

- A Dozen Products
- 5 dev centers distributed globally
- Products need to support 30 different x86/64 OS

Internal test cloud for x86/64 OS

- Engineers request OS through web portal
  - Define environment
  - Define schedule
  - Define size
  - Define physical machine or VM
- Resources are provisioned automatically
- Next step: Extending the solution for technical support and field engineers

Resources ready in minutes vs. 2 days

Cloud Infrastructure Requirements

Isolated application run time environment

- Different applications can run concurrently on a multi-core node/server
- Problem in one application does not affect the others
- Create personal/group cluster

Change node/server personality

- Re-domain a server/node
- Switch OS, particularly between Windows and Linux
- Running a legacy OS on the latest hardware

Reduce resource fragmentation

- Application migration

Capacity Planning

- “What if” analysis

DDC for HPC Cloud

Workload scheduler

Cloud Portal

Dynamic provisioning scheduler

- Provision OS/VM
- Migrate VM

HPC Systems

OS or VM Image database

DDC extends to deploy into multiple virtual clusters

DDC for HPC Cloud

Workload scheduler

Schedule jobs

Cloud Portal

• Provision OS/VM
• Migrate VM

Dynamic provisioning scheduler

HPC Systems

OS or VM Image database

DDC extends to deploy into multiple virtual clusters

PROs & CONs of VM vs RMs

	VM	RM
PROs	- Reliability • Isolated from hardware • Checkpointing - SLA • Quick provisioning - Resource utilization • Job migration	- Application Performance - No need to have special infrastructure
CONs	- Performance cost (=application cost) Getting better - Infrastructure cost	- Application reliability - Slow provisioning - Resource utilization

Conclusion: use both VMs and RMs (real machines)

Platform™

Platform DDC
– dynamics from the Cloud
– performance from the Grid

DDC: personality for compute hosts

- Multipurpose hosts – feature *personality*
  - Hosts can be setup with multiple personalities (OS, OS+patch+pre-installed apps, …)
  - Dynamically provisioning hosts' personalities based on workload demand
  - Allow admin to manually change hosts' personalities
- Control Virtual Machines (VM)
  - Integration with VM server (RedHat Xen, …)
  - Dynamically provision VMs based on workload demand (start/stop VM on a physical machine)

Provisioning environment supported

Dualboot/Multiboot: (Supports Windows / Linux)

Machines are pre-installed with multiple operating systems. DDC uses network boot to select which partition to boot

Diskless booting: (Linux only)

Machines do not use local disk, instead copy an OS image from a server at boot time

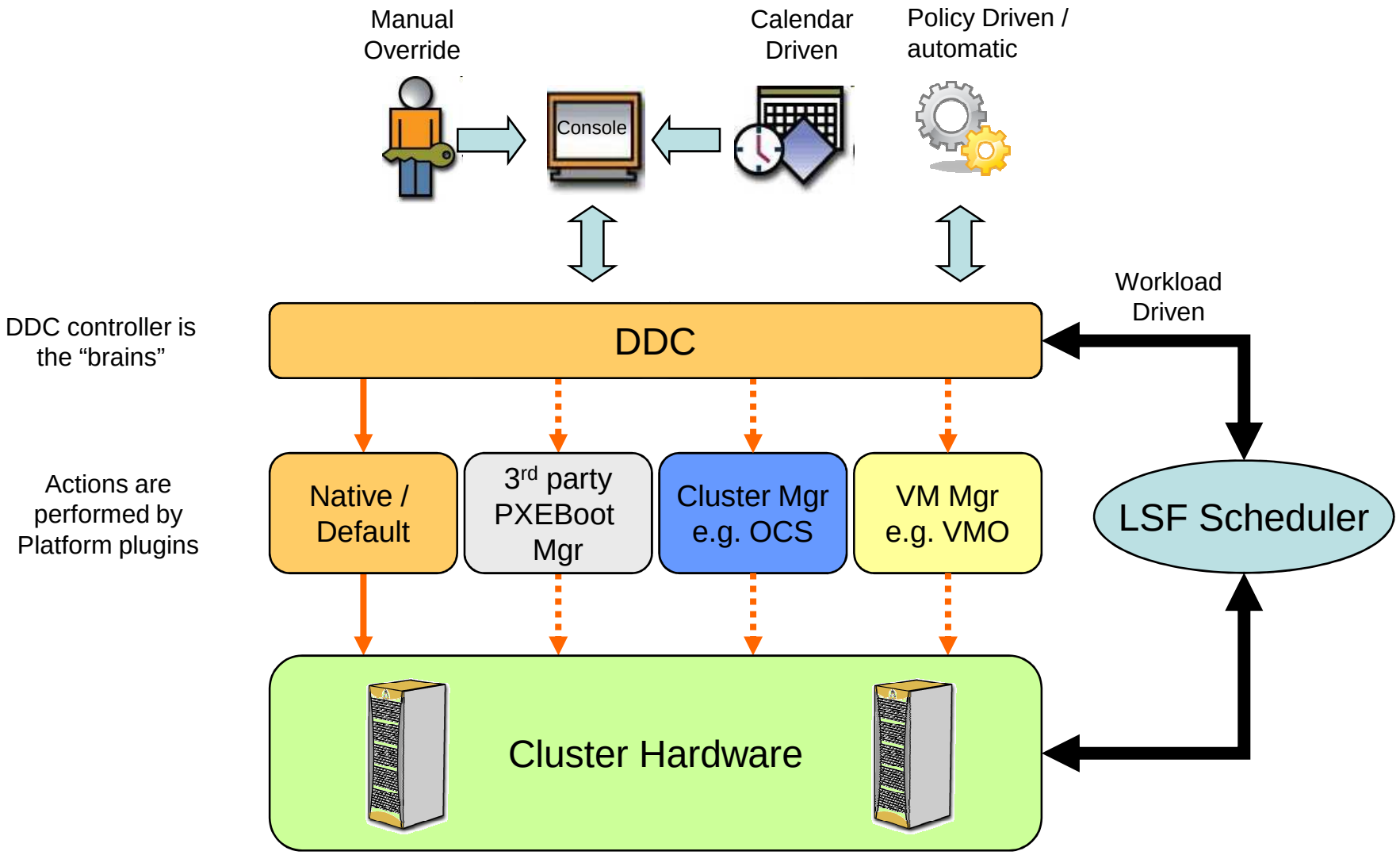
“Image” based installation: (Linux only)

Machines are re-installed each time when they are re-purposed

Controlling VMs:

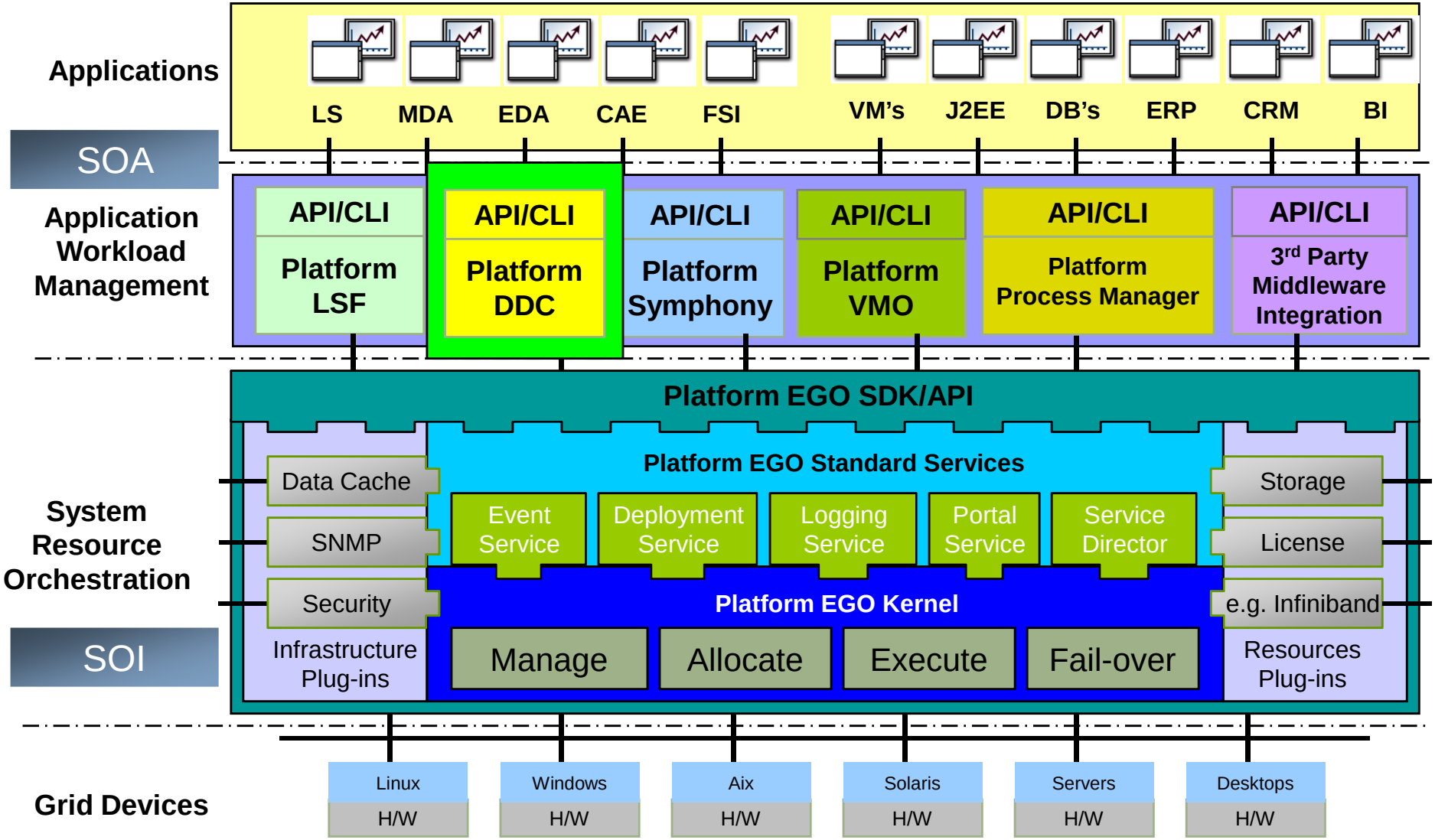
VMs are pre-installed, they can be dynamically started/stopped based on the resource usage/demand

DDC Architecture concept



Platform Enterprise Grid Orchestrator

Open & Decoupled SOA / SOI Architecture

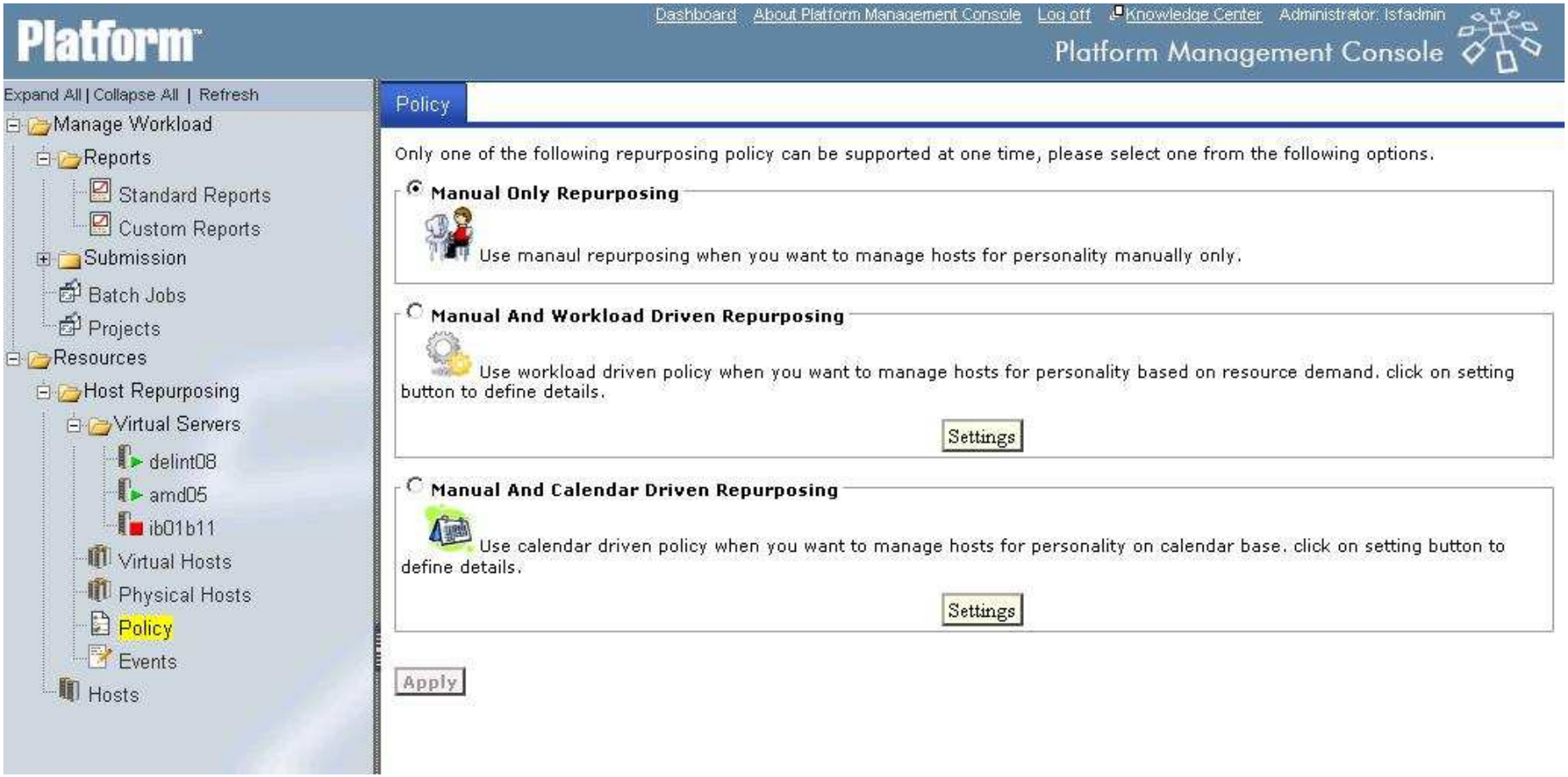


Platform For LSF Administrator

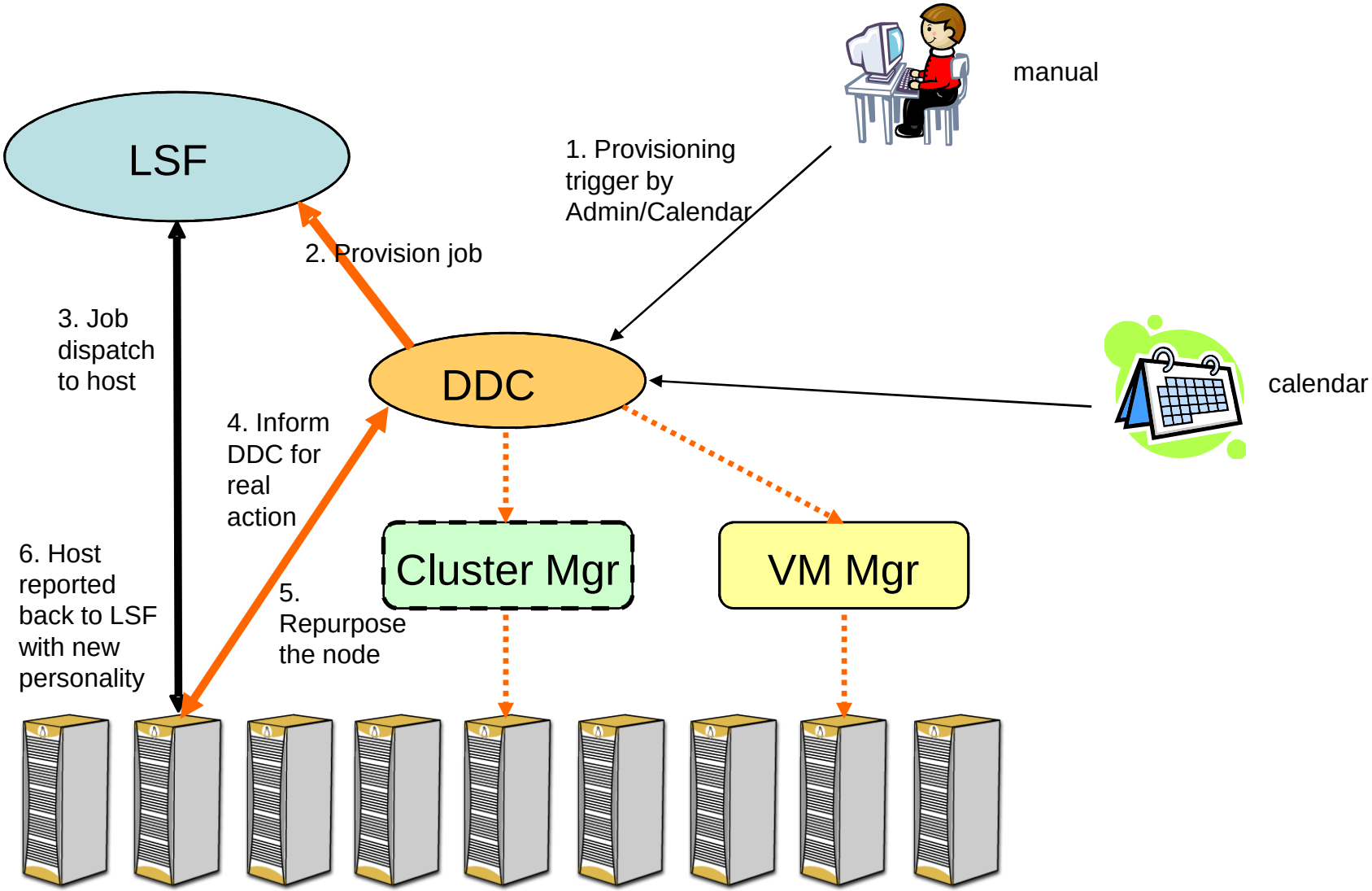
1) Hosts' environment setup
   - Hosts' OS setup, dual boot, multi-boot, diskless, image based provisioning, VMs etc…… all according to DDC's guidance
   - Install applications, LSF etc…
2) Install DDC
   - Install the package
   - Configure the hosts information, I.e. name, IP, ipmi, personalities, into DDC
   - Configure mappings between host personalities and LSF host types, OS types – possibly with other Boolean resources
   - Configure DDC provision mode, policy and control parameters
3) LSF
   - All hosts, all different personalities, need to be LSF hosts
   - Define special queue for DDC to use, provisioning jobs
   - User submit jobs requiring certain personality, can use the "-R [ostype=="personality"]", or by default we use the host types

Three modes of provisioning

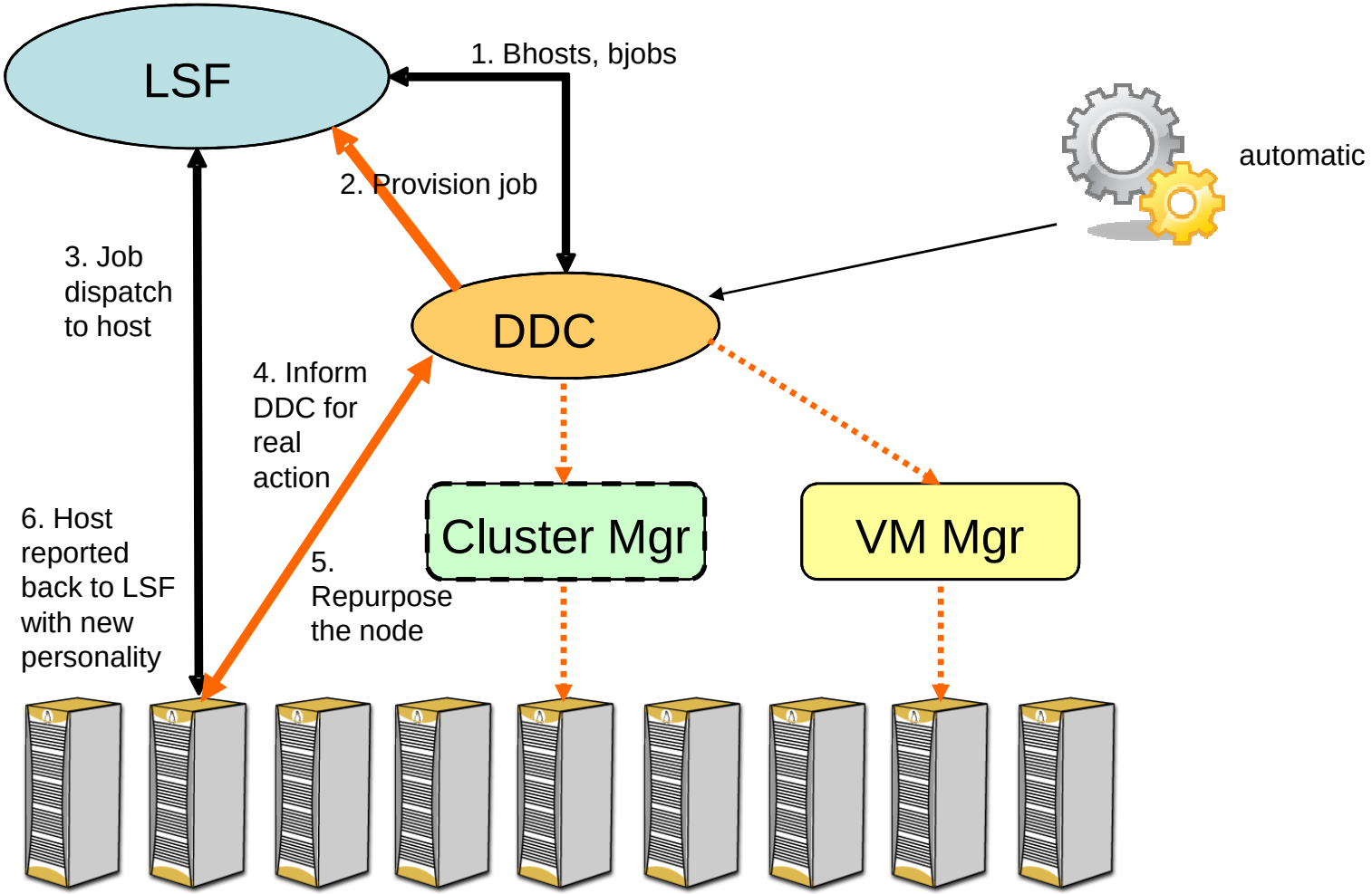
- Workload policy driven provisioning
- Calendar policy driven provisioning
- Admin can take manual actions any time



Workflow – Manual/Calendar



Workflow -- Automatic

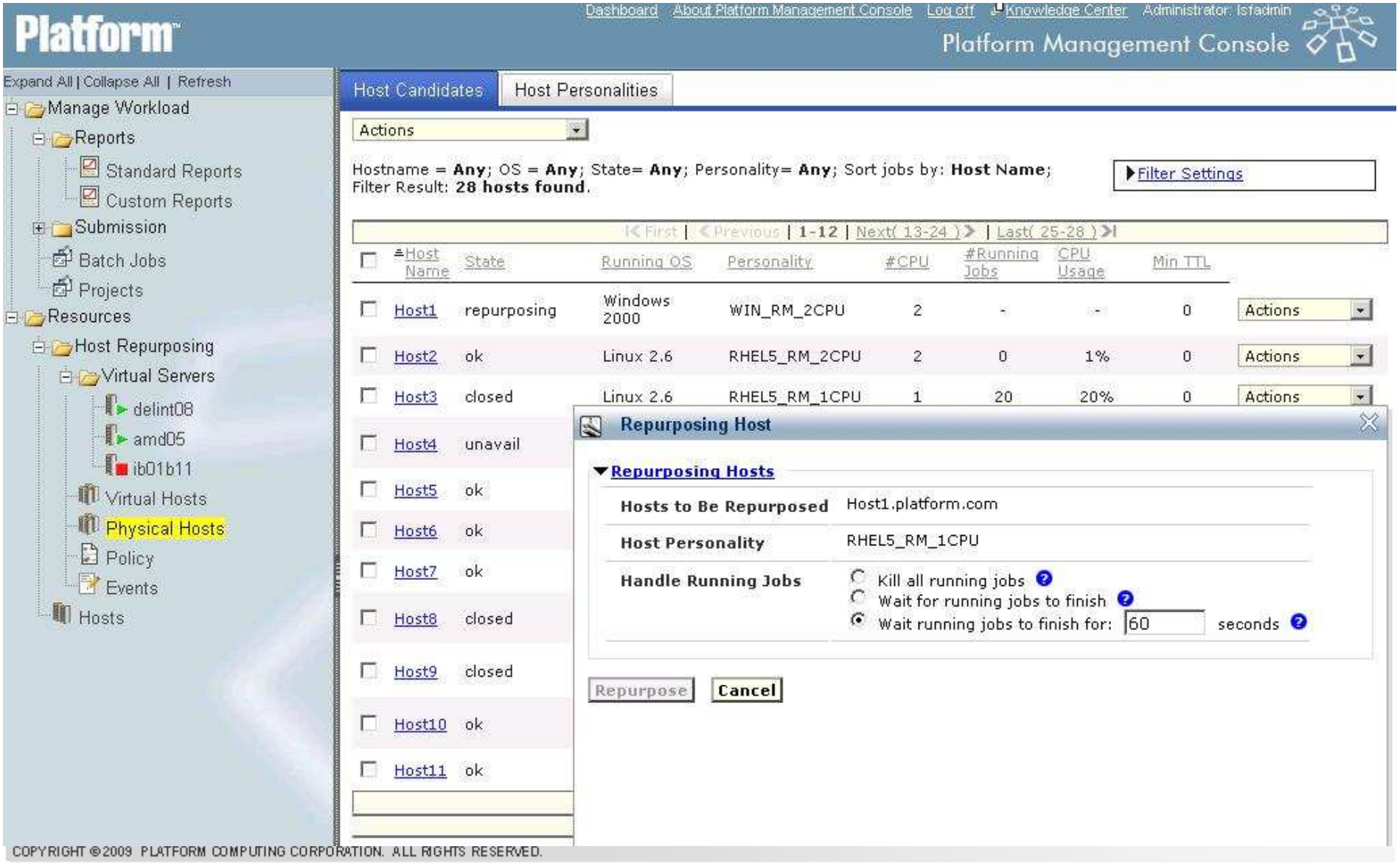


Use case 1: Admin manually switching resources

1) Description:
   - Resource switching by admin (manual only)
   - User asks admin requiring more Linux hosts to serve their workload, and Admin takes manual action to add 70 Linux hosts
2) Example below:
   - Before: Linux:20, Windows:80
   - After: Linux:90, Windows:10

Manual mode provision

Admin switching one host, Host1, to RHEL5_RM_1CPU

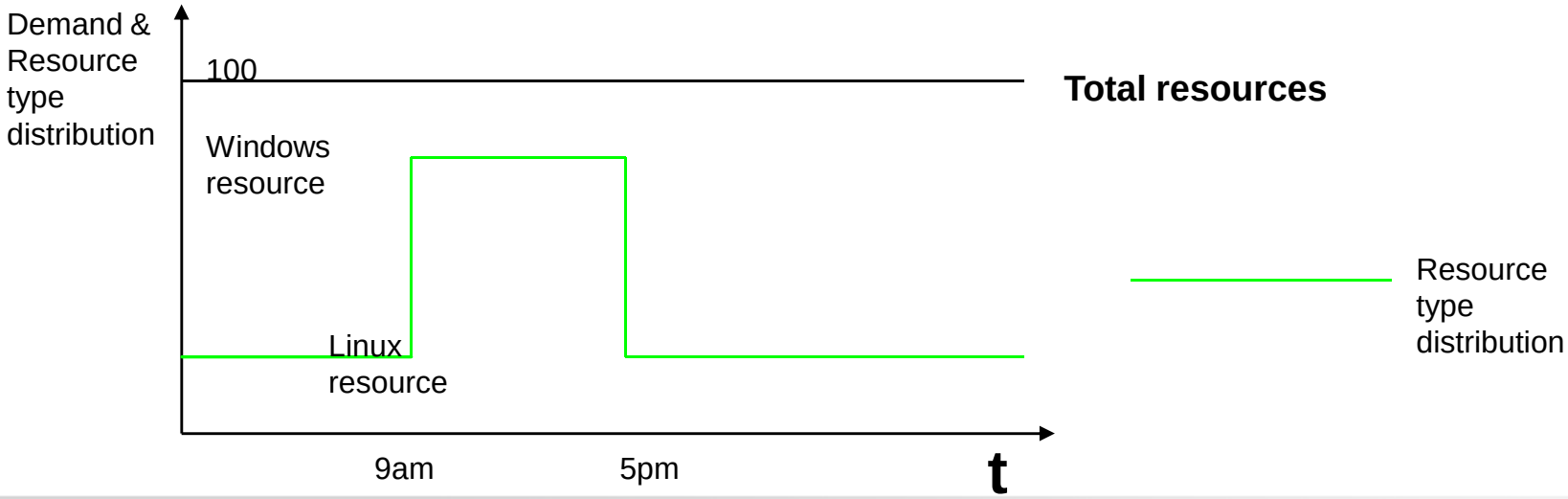


Manual mode provision (cont)

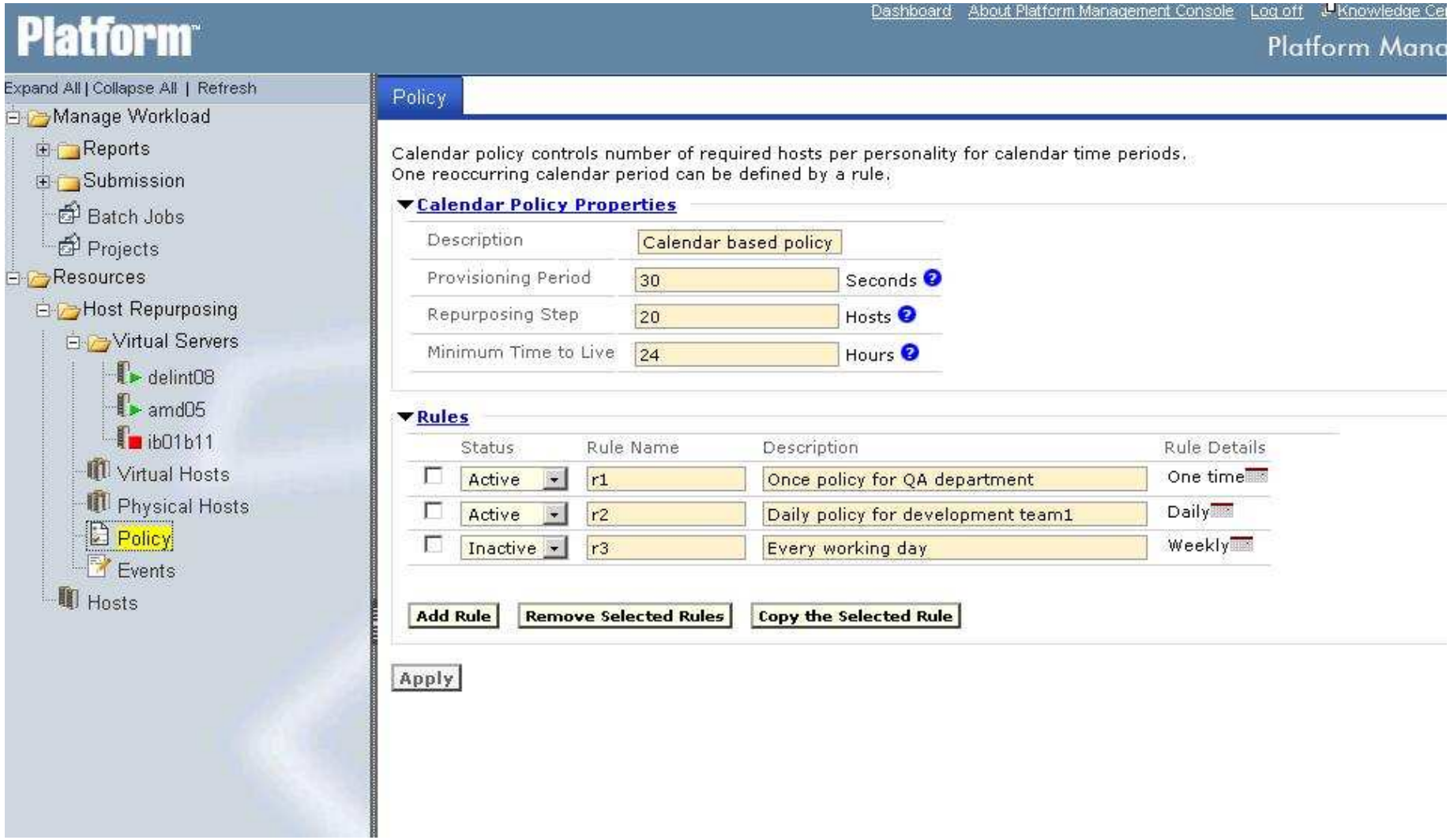
Admin manually switching hosts personality by specifying number of hosts required, Admin can also define the resource requirement string to further refine the scope. The workload manager will select the best hosts, earliest available hosts, for repurposing.

Use case 2: Switching resource based on calendar time

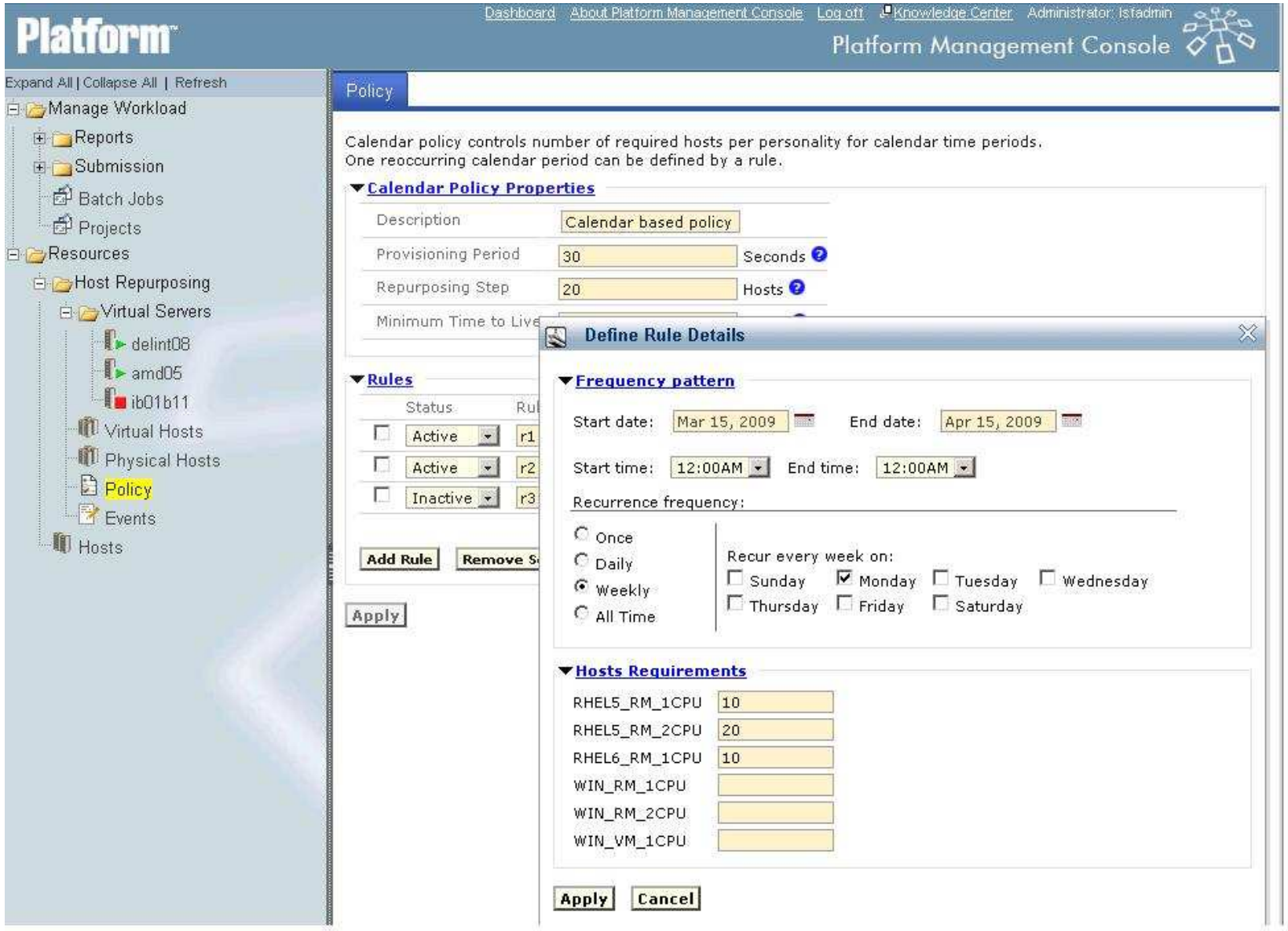
1) Description:
   - Workload comes at fixed pattern, I.e. Mainly regression test, using Linux, after hours and interactive Windows work during day time.
   - Calendar rules defined for this action
2) Example below:
   - Two rules defined for 9am and 5pm.
   - 9am: Linux:80, Windows:20
   - 5pm: Linux:20, Windows:80



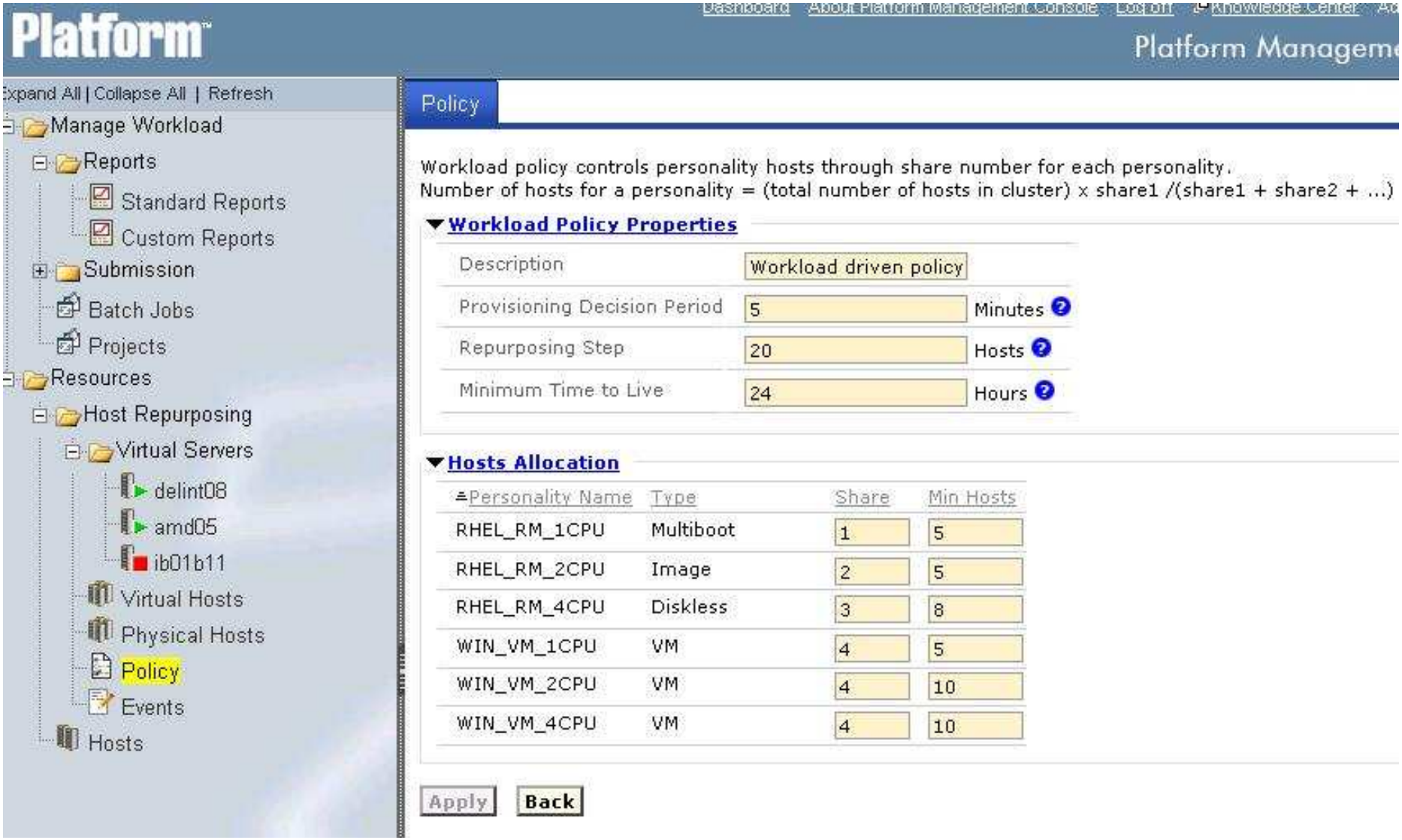
Calendar driven provision



Calendar driven provision (cont)

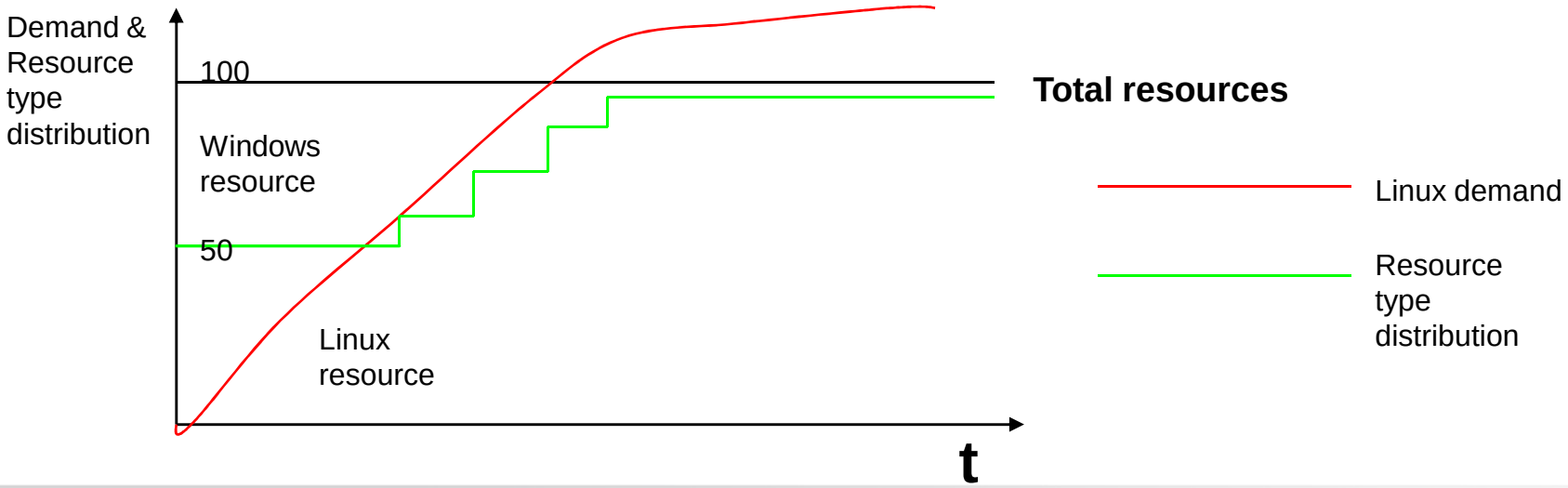


Automatic provision – Resource based policy



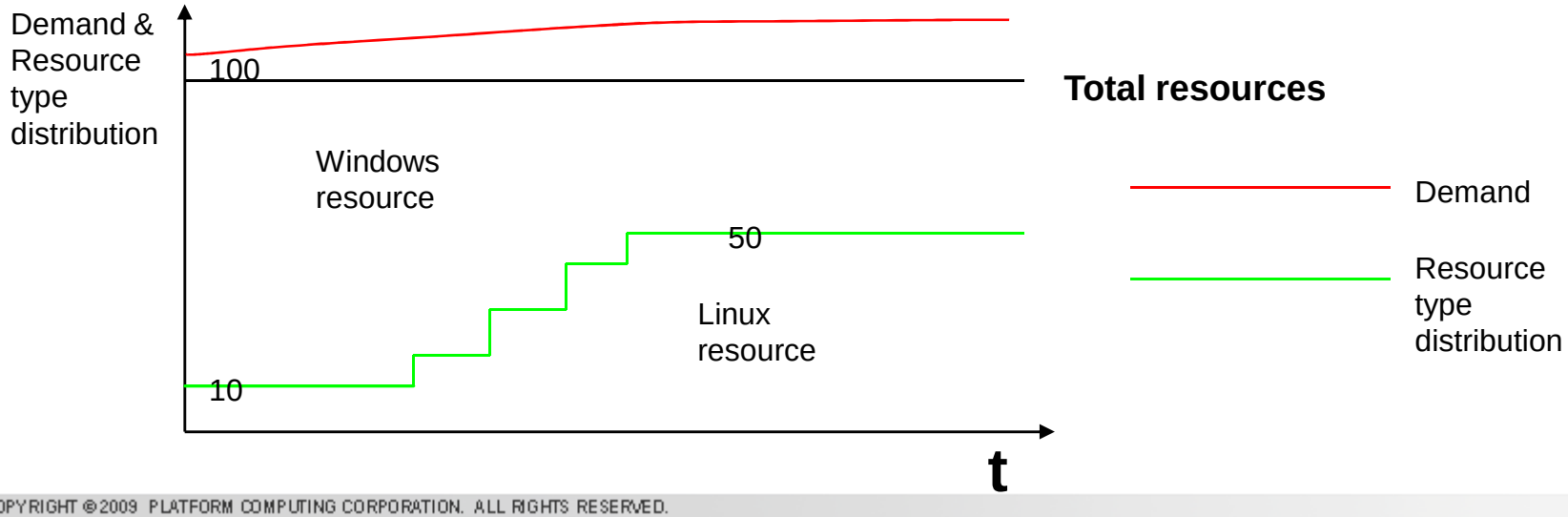
Use case 3: Workload policy. High workload comes for one particular type of resource

1) Description:
   - High workload comes demand one resource type when the cluster resources are evenly distributed, and no demand for the other type.
   - DDC will sense the shrinking availability of that resource type and will try to move other types to the demanded type, with each decision cycle and number of hosts constraint.
2) Example below:
   - Two types of resources, each with a share value of 50, with 10 maximum hosts switching step and a 5 minutes cycle period.



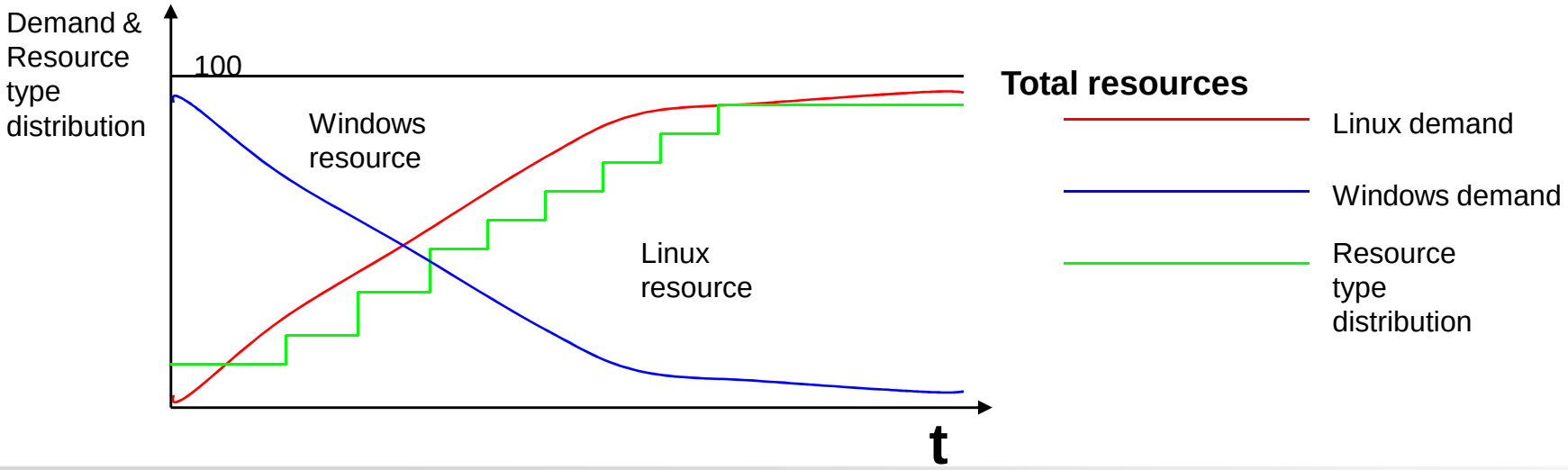
Use case 4: Workload Policy: Cluster busy with all resources used

1) Description:
   - High workload comes demand all resource types when the cluster resources are not evenly distributed.
   - DDC will sense the shrinking availability of all resource types and will try to move the resource supply to the desired balance level, based on the each resource types' weight
2) Example below:
   - Two types of resources, each with a share value of 50, with 10 maximum hosts switching each time and a 5 minutes cycle period.



Use case 5: Workload Policy: Workload demand swing from one type of resource to another type

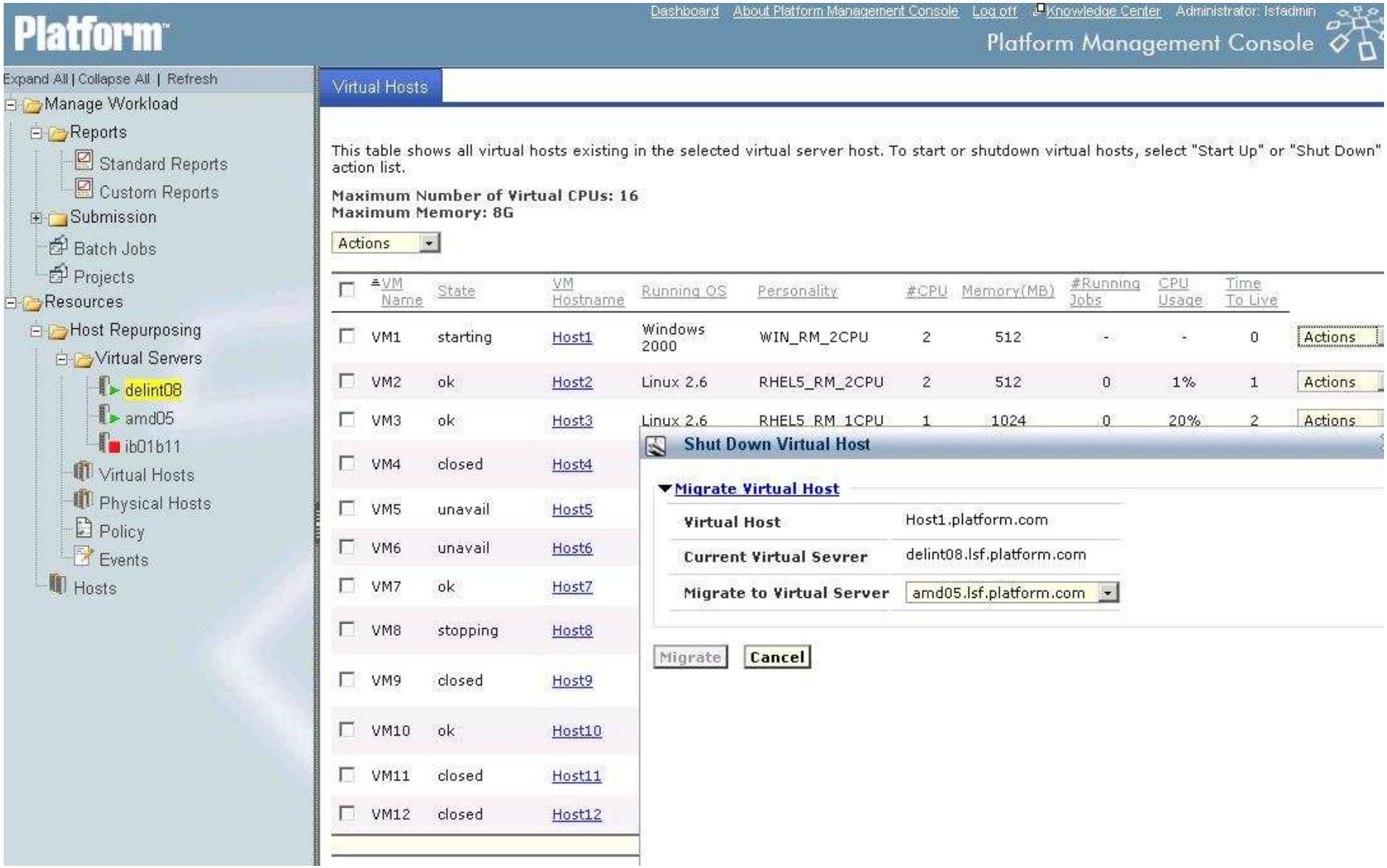
1) Description:
   - Ø High workload comes demand changes from one resource type to another resource type.
   - Ø DDC will gradually change the personality from one type to another type
2) Example below:
   - Ø Two types of resources, with 20 maximum hosts switching each time and a 5 minutes cycle period.



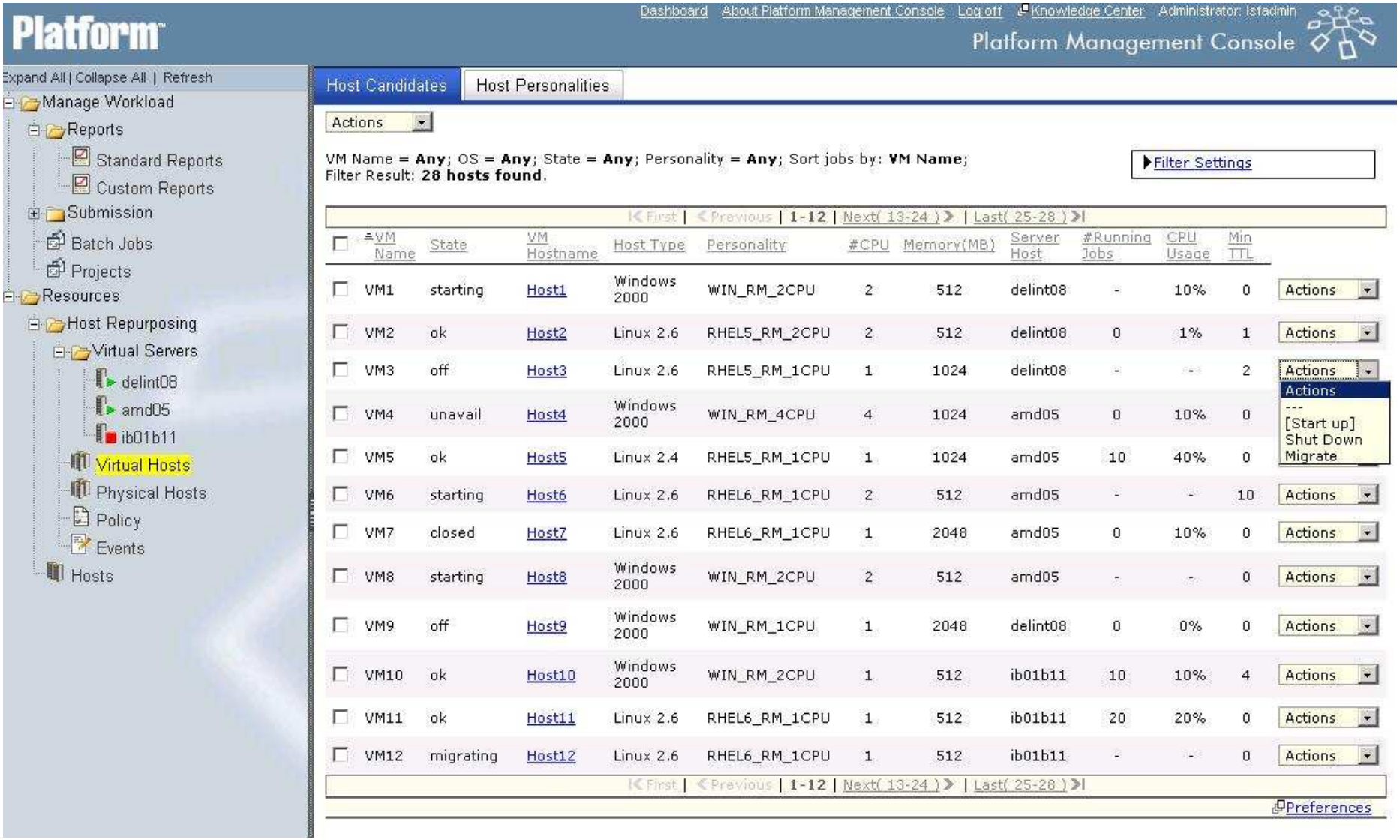
Controlling VMs

1) Similar rules can apply to VMs too:
   - Maintain the minimum amount of VMs started and start more when the VMs are used.
   - Only start a limited number each time
   - Decision time period would be the same
   - Each VM type would also get a share so we can keep a ratio of them when system is busy
   - The total number of VMs running on a server is limited by the server's number of CPUs and memory size
2) Use Boolean resource to indicate VMs
   - User can submit jobs specifically asking for VMs, or vise versa

VM server GUI

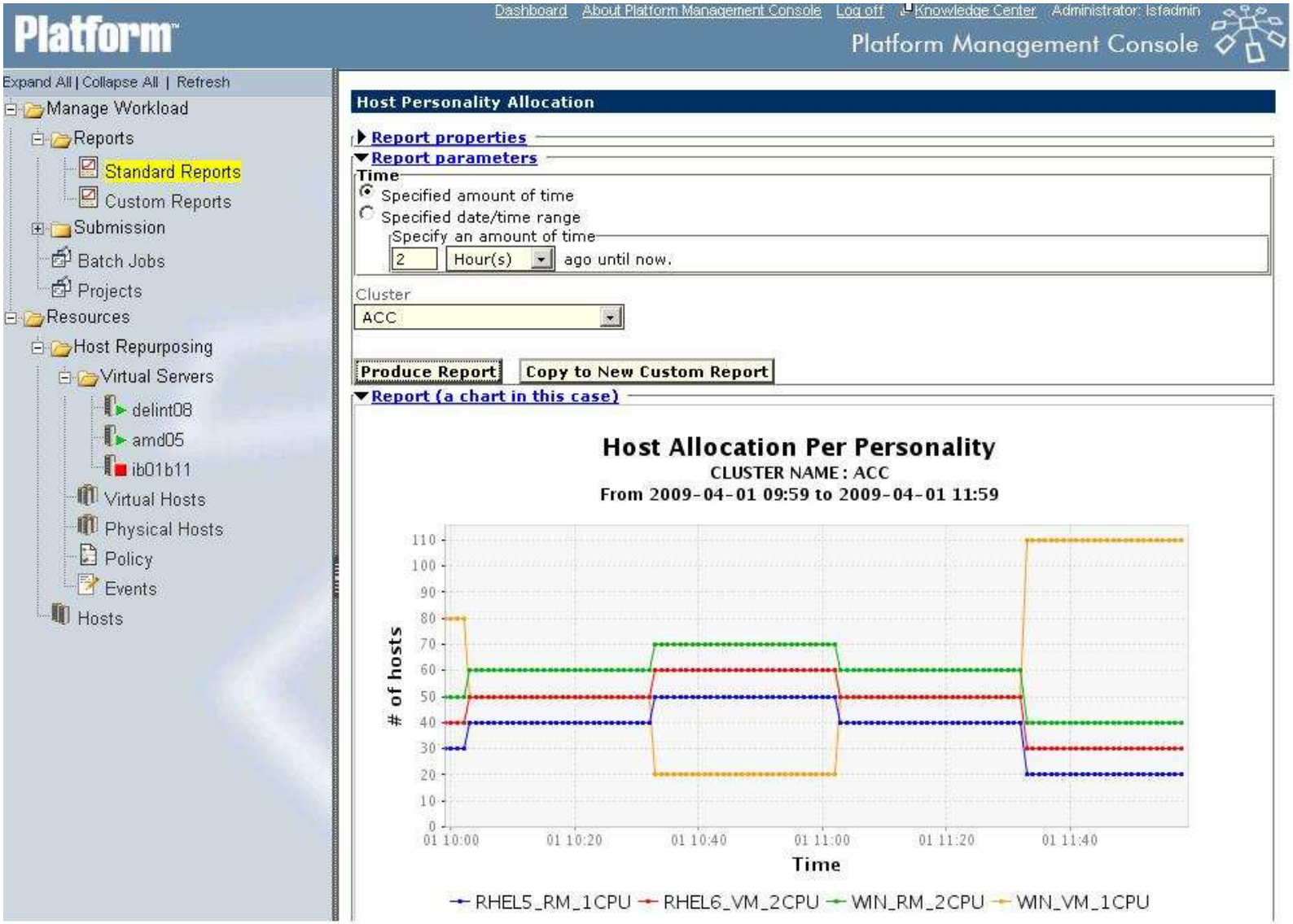


VM hosts GUI

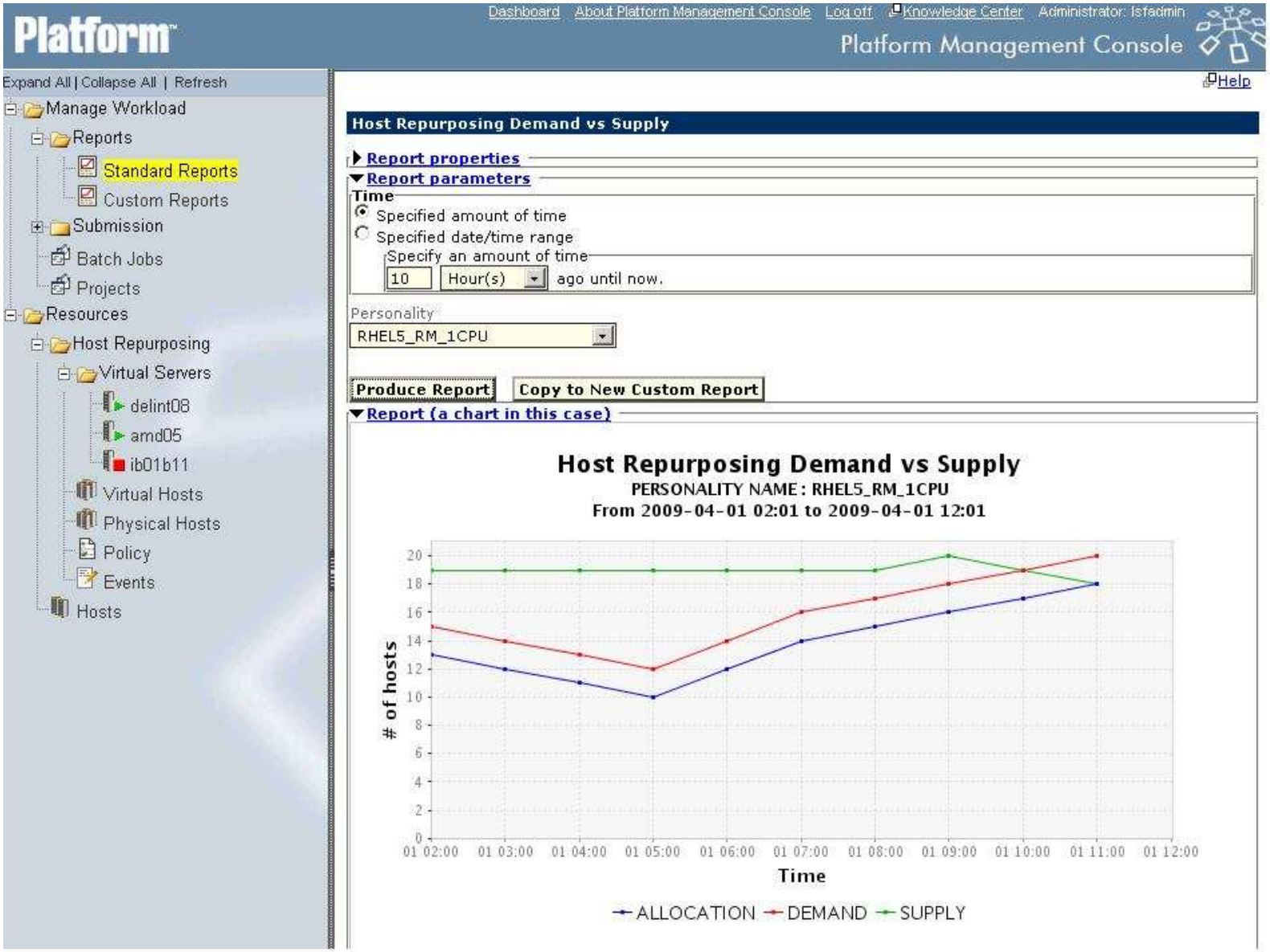


Events

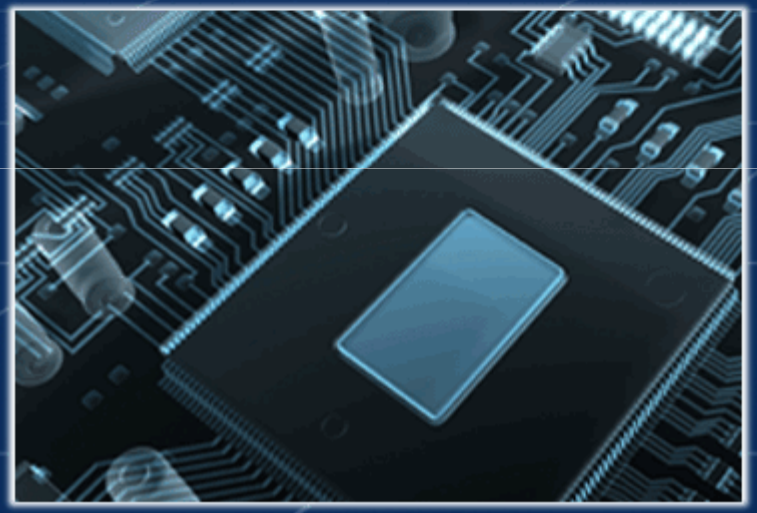
Report: Host Personalities



Report: Demand vs. Supply



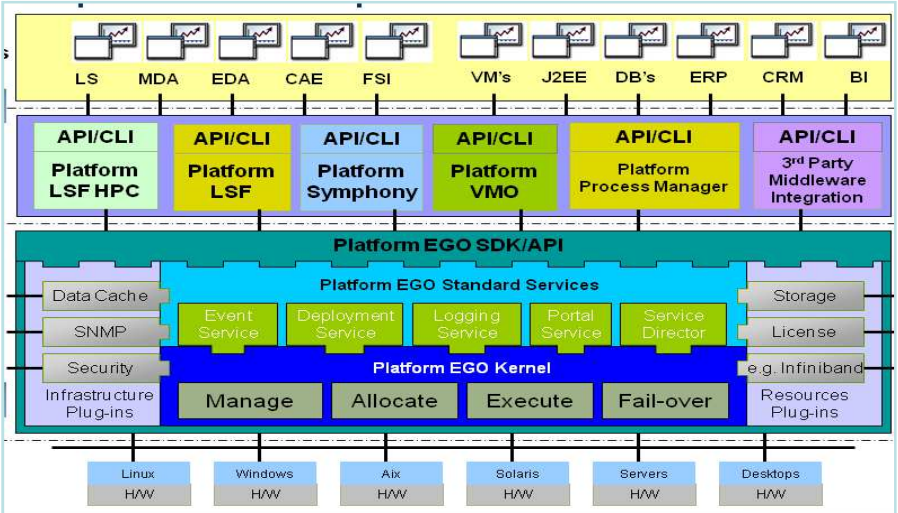
Platform™



Summary

The Dynamic HPC Datacenter

- Can the HPC Datacenter please the user (=customer)?
  - Provide Cloud-style agility towards both users and resources
  - Beat the legacy issue
- Taking the best of both Grid and Cloud to enforce business resp. scientific process execution – effectively & efficiently
- Looking forward to discuss and share with you!



Bernhard Schott
Dipl. Phys.
EU-Research Program Manager

Platform Computing GmbH
Mobile +49 (0) 171 6915 405
Email: bschott@platform.com
Skype: bernhard_schott
Web: http://www.platform.com/

Platform™

Powering High Performance

Thank you

www.platform.com