

# Emerging Systems for Operating Data Centers

Dr. Sven Graupner Hewlett Packard Laboratories Palo Alto, California, USA

GI FG Betriebssysteme Herbsttreffen TU Dresden, Dec 16, 2004

© 2004 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice





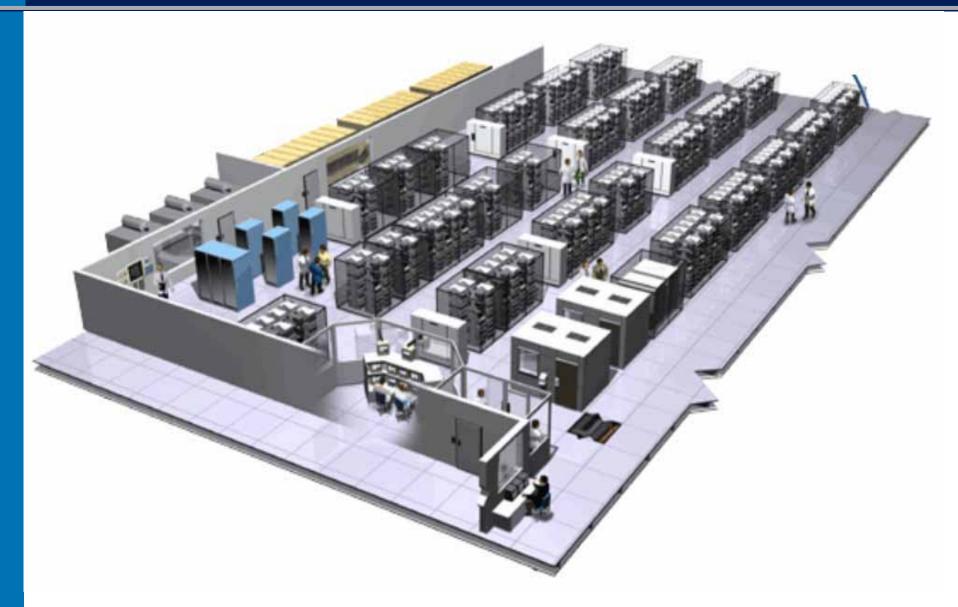


- New domain for OS: Enterprise IT environments
- Research Prototype from HP Labs
  - Formal Basis: Layered Information Model
  - Resource Grounding
  - Resource Allocation
- Summarizing remarks

San Francisco, CA	O S OPERATING SYSTEMS D D D D D D D D D D D D D
December 6-8, 2004	Sponsored by USENIX in cooperation with ACM SIGOPS
11:00 a.m12:30 p.m.	
AUTOMATED MANAGEMENT I	
Automated Worm Fingerprinting Sumeet Singh, Cristian Estan, George Varghese, and Stefar University of California, San Diego	n Savage,
Understanding and Dealing with Operator Mistakes in In Kiran Nagaraja, Fabio Oliveira, Ricardo Bianchini, Richard F Thu D. Nguyen, Rutgers University	
System Administration as Search: Finding the Needle in Andrew Whitaker, Richard S. Cox, and Steven D. Gribble, G	
Washington	
Microreboot—A Technique for Cheap Recovery George Candea, Shinichi Kawamoto, Yuichi Fujiki, Greg Fi Armando Fox, Stanford University	AUTOMATED MANAGEMENT II Correlating Instrumentation Data to System States: A Building Block for Automated Diagnosis and Control Ira Cohen, Hewlett-Packard Laboratories; Jeff Chase, Duke University; Moises Goldszmidt, Terence Kelly, and Julie Symons, Hewlett-Packard Laboratories
	Automatic Misconfiguration Troubleshooting with PeerPressure Helen J. Wang, John Platt, Yu Chen, Ruyun Zhang, and Yi-min Wang, Microsoft Research
December 30, 2004	Using Magpie for Request Extraction and Workload Modelling Paul Barham, Austin Donnelly, Rebecca Isaacs, and Richard Mortier, Microsoft Research, Cambridge, UK

# **Enterprise Data Center**

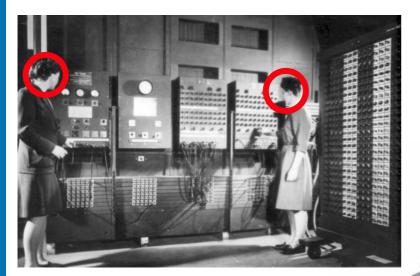




### Ratio: Operators / Machines



#### 1954: **20-to-1**...



# ...50 years later: **1-to-20** avg. (Google: 1-to-500)

# Single Computer Environment

- has pool of local resources shared among applications:
  - CPUs, Memory, Disks, Busses
  - IO devices
- purpose: host data and perform applications.
  - editor, compiler, browser
  - text processor



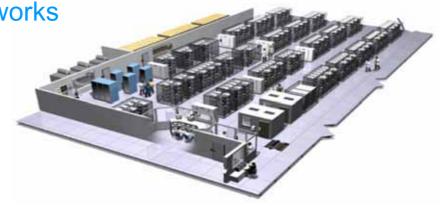
- **Operating System:** resource management, application control, monitoring and accounting, access control.
  - Linux, HP-UX, Windows



# Data Center Environment



- has pool of local resources shared among applications:
  - Servers, Storage Arrays, Networks
  - Devices (LB, FW)
- purpose: host data and perform applications.
  - SAP, CRM, ERP
  - Oracle

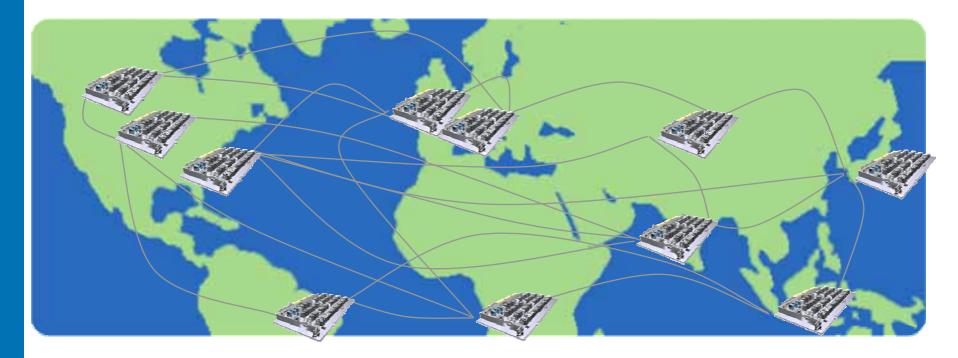


- Operating System: resource management, application control, monitoring and accounting, access control.
  - HP OpenView, Tivoli: monitoring assistance

# **Globally Networked Enterprise IT**



- distributed enterprise IT systems (e.g. supply-chain, procurement),
- global scale, e.g. HP 140 data centers world wide, 7,000 applications
- Enterprise Grid (leverage Grid technology)







### operates Systems.



- a device, a machine;
- a resource environment;
- a data center;





### operates Systems.



- a device, a machine;
- a resource environment;
- a data center;

also:

- an application;
- an IT solution;





### operates Systems.



- a device, a machine;
- a resource environment;
- a data center;
- an application;
- an IT solution;
- local to
- large-scale, distributed.





### operates Systems.

Open or self-operating enclosure					
	Operating System	*	Operated System		

- a device, a machine;
- a resource environment;
- a data center;
- an application;
- an IT solution;
- local to
- large-scale, distributed.

### Hypothesis

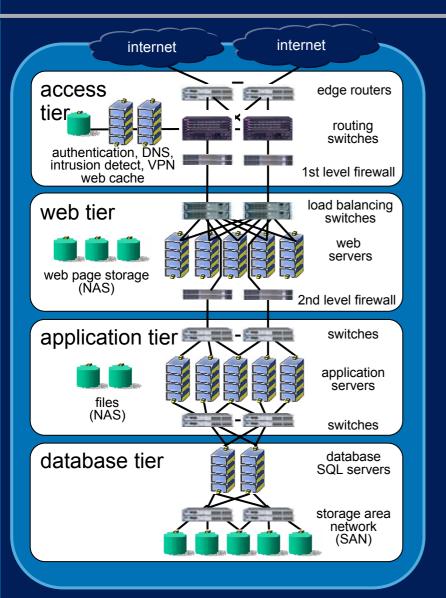


### **New OS Technologies will:**

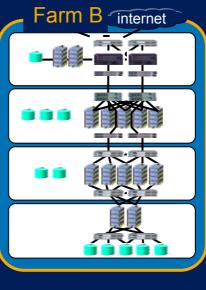
- largely automate the operation of data centers and IT solutions
  - Resource management, configuration management, application deployment, HW&SW life cycle management, monitoring, accounting;
- merge into commercial IT software and solutions for automating their operation ("self-")
  - HP's Adaptive and IBM's Autonomic Computing, SAP's Utility Initiative, Oracle's 10g;
- fundamentally change the way IT solutions are planned, designed, implemented, operated
  - using more formal methods and models that also enable computer-assistance and design tools;

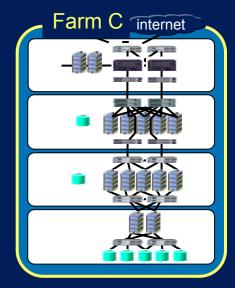
### **Enterprise Applications**





### Application Container: "Farm"



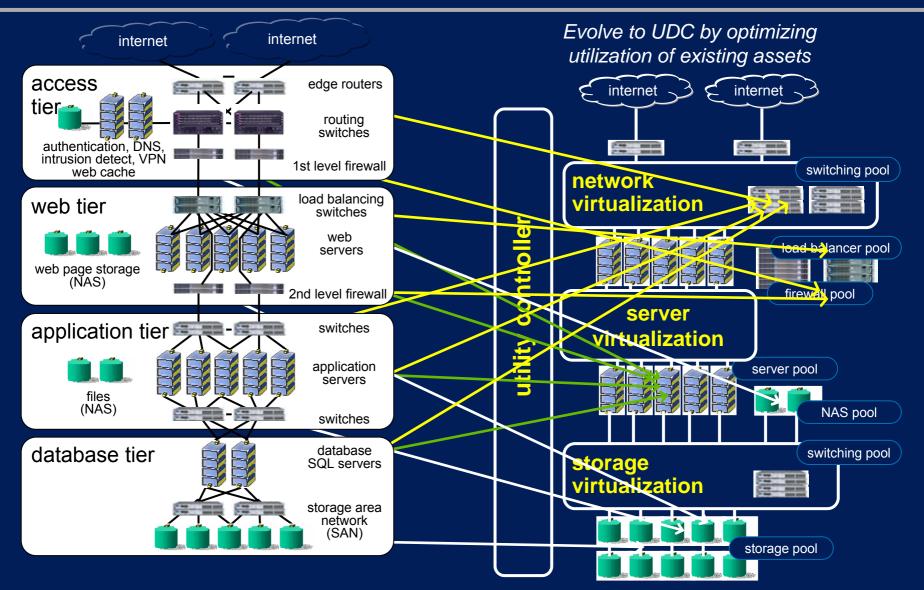


Customer "B"

Customer "C"

### Programmable Resource Infrastructure





# Creating an Application



#### 1. Architect application:

- business case
- service growth projection
- SLO requirement
- availability
- security needs
- time to implement

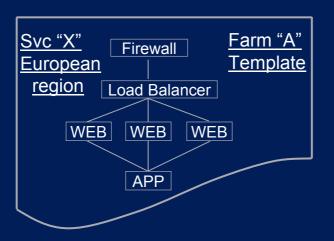


#### 3. Activate the Application

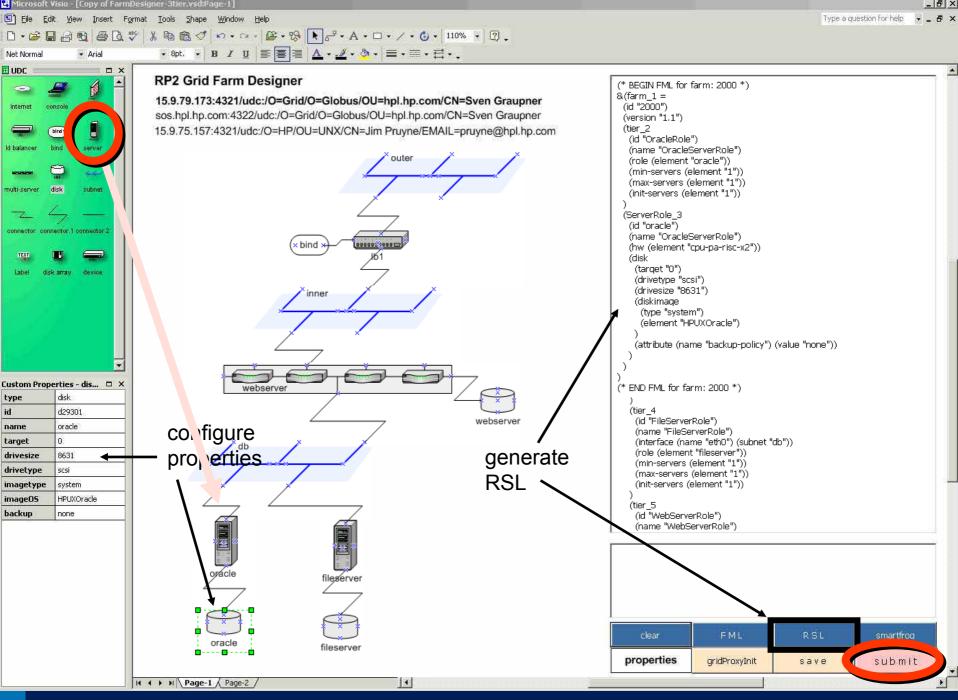
- SUBMIT
  - automatically locate and allocate resources
  - auto-configure network and storage
  - auto-configure firewall & load balancers
  - auto-configure & boot servers

#### 2. Design a resource topology :

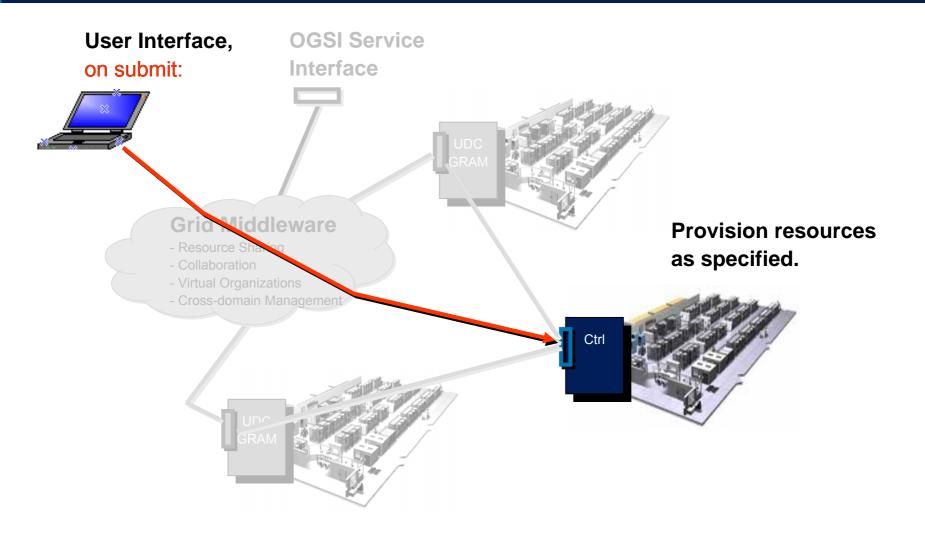






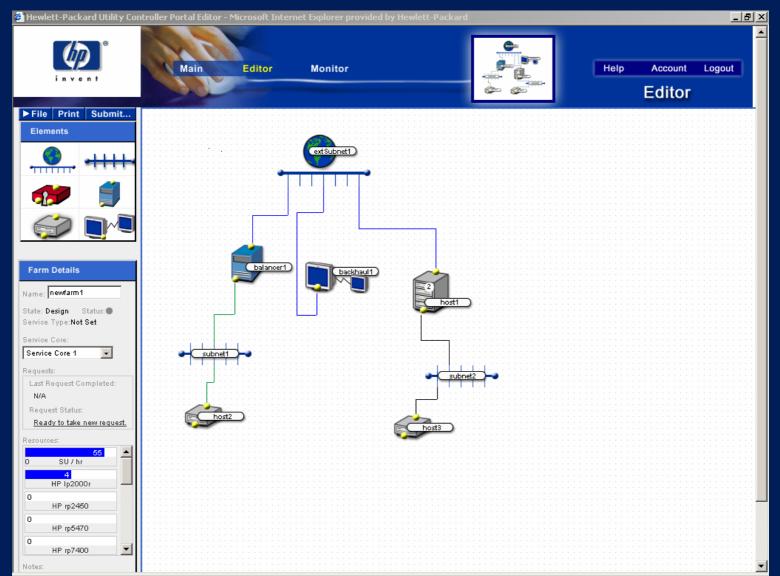


# Submit Design for Deployment



### Farm Editor









- New domain for OS: Enterprise IT environments
- Research Prototype from HP Labs
  - Formal Basis: Layered Information Model
  - Resource Grounding
  - Resource Allocation
- Summarizing remarks

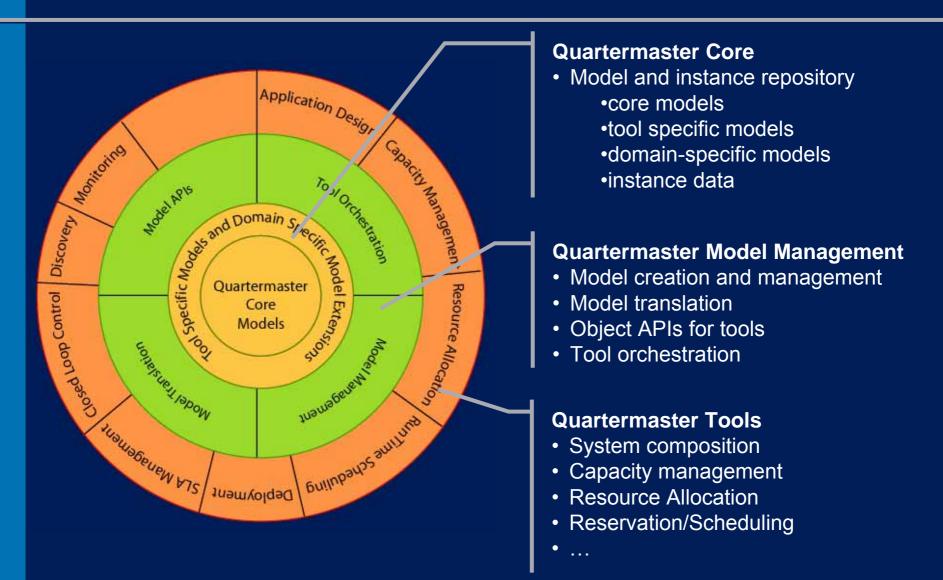
# **QM** Contributions



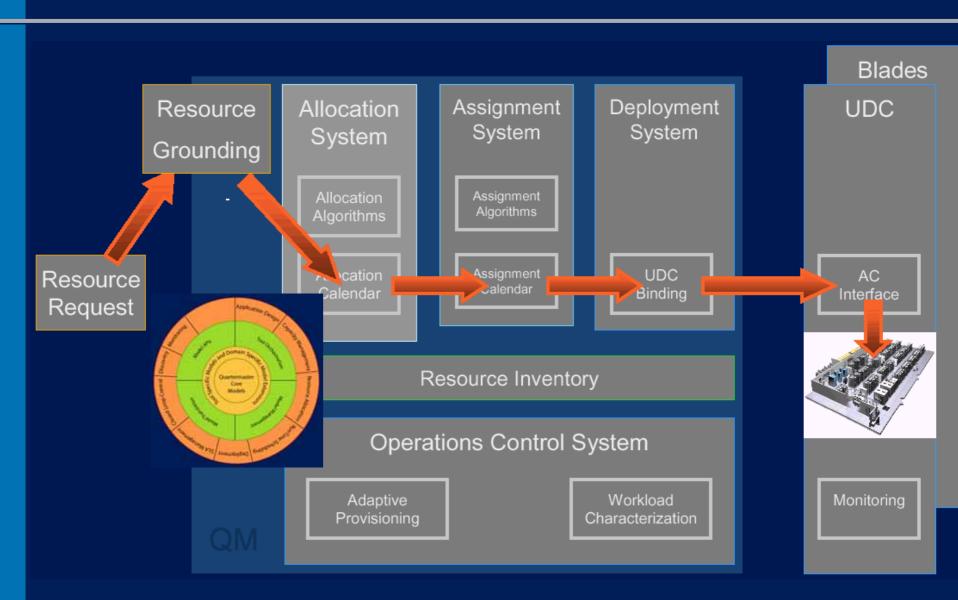
- Design of Resource Topologies for enterprise applications,
- *Grounding* resource topologies from application specifications into allocatable resources in a data center,
- requesting and making Allocation decisions for resource topologies including a complex Resource Request Format,
- ability to deal with Resource Constructions,
- Assignment of allocated resource capacity to resource instances in a data center,
- Deployment of resources (creation, configuration of resources) using actuators,
- Run-time Control functions in form of dynamic adjustment of resource use, which is also referred to as *Flexing*.

### QM Research Prototype



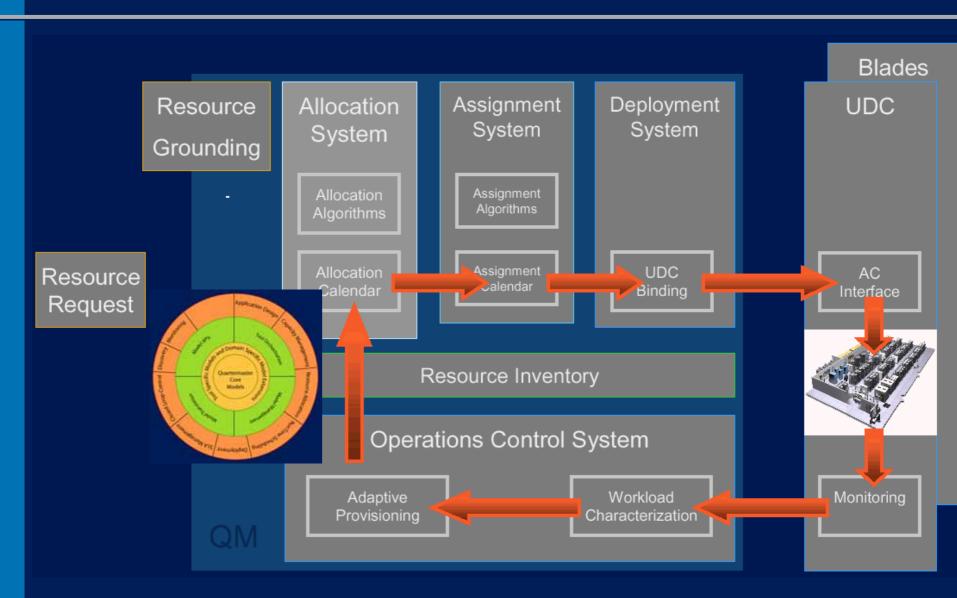


### **Resource Request Flow**





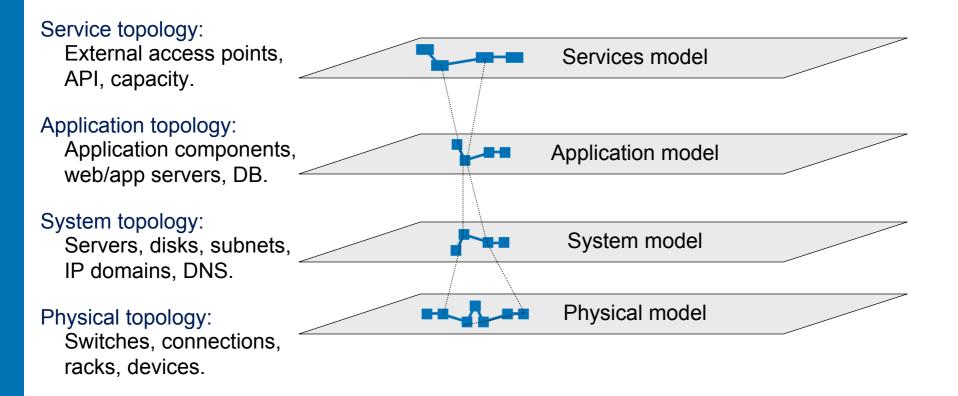
## **Run-time Resource Adjustment**





### Apply same abstractions to each layer.

Maintain relationships within and across layers.



### **QM Core Model Entities**

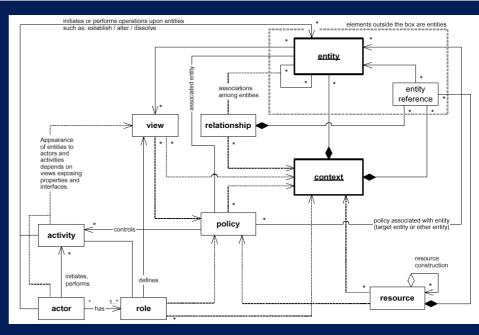


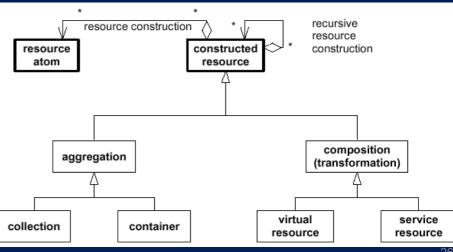
#### Purpose:

 Identify and structure all information maintained in QM according to a uniform model.

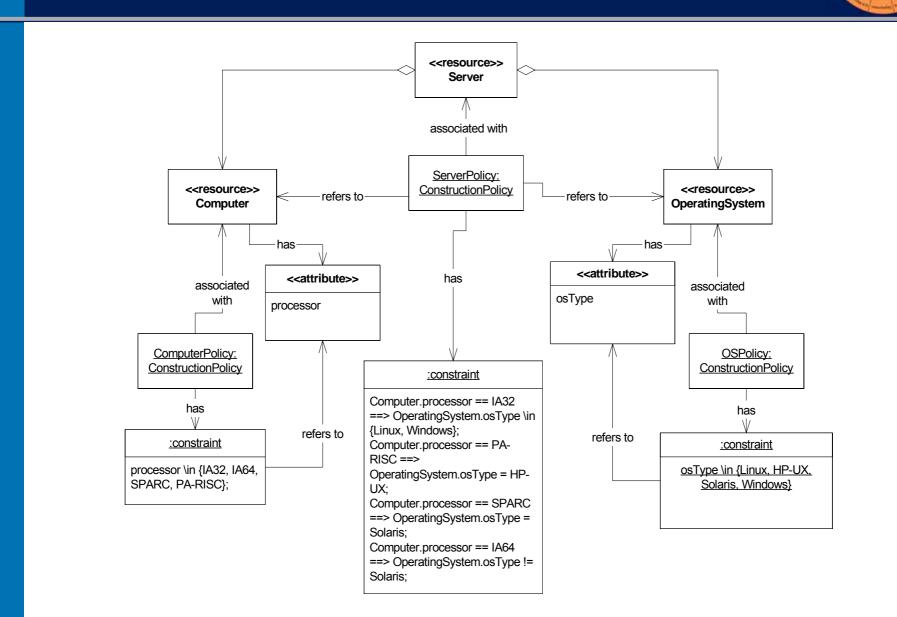
#### First-class Entities:

- Actor,
- Role,
- Activity,
- Resource (atom, construction),
- Relationship,
- Context,
- View,
- Policy.





### Model Representation in CIM



# **Example: Resource Grounding**



#### Purpose:

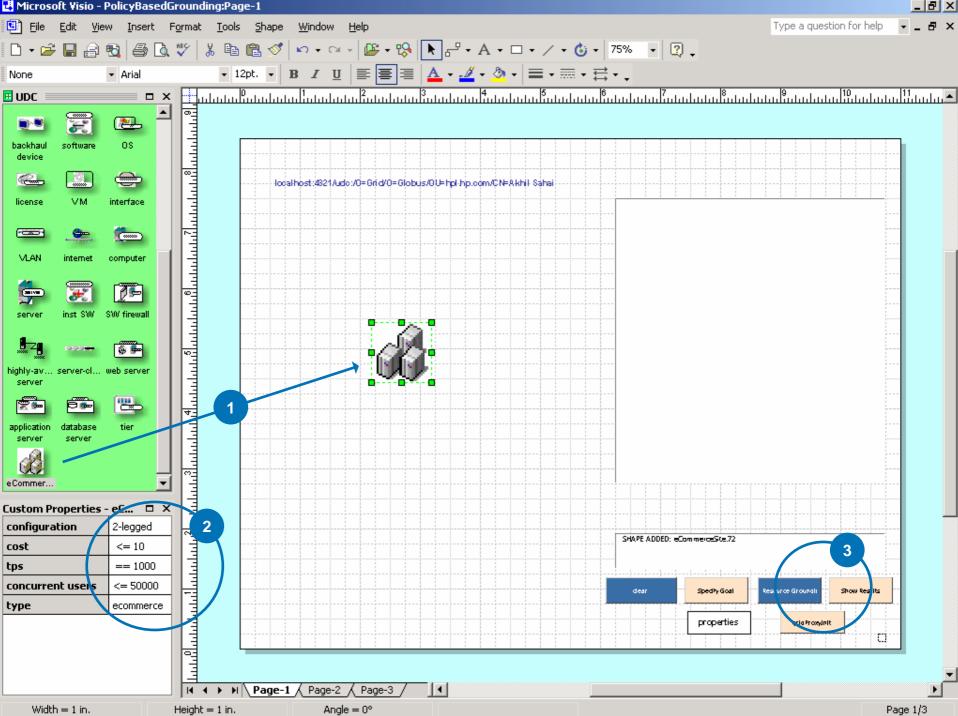
- Describe resource needs in higher terms such as "e-commerce site that can handle up to 300,000 queries per minute".
- Many ways exist to construct this solution. Resource grounding finds a solution that meets all specified criteria and, recursively, and obeys all constraints.

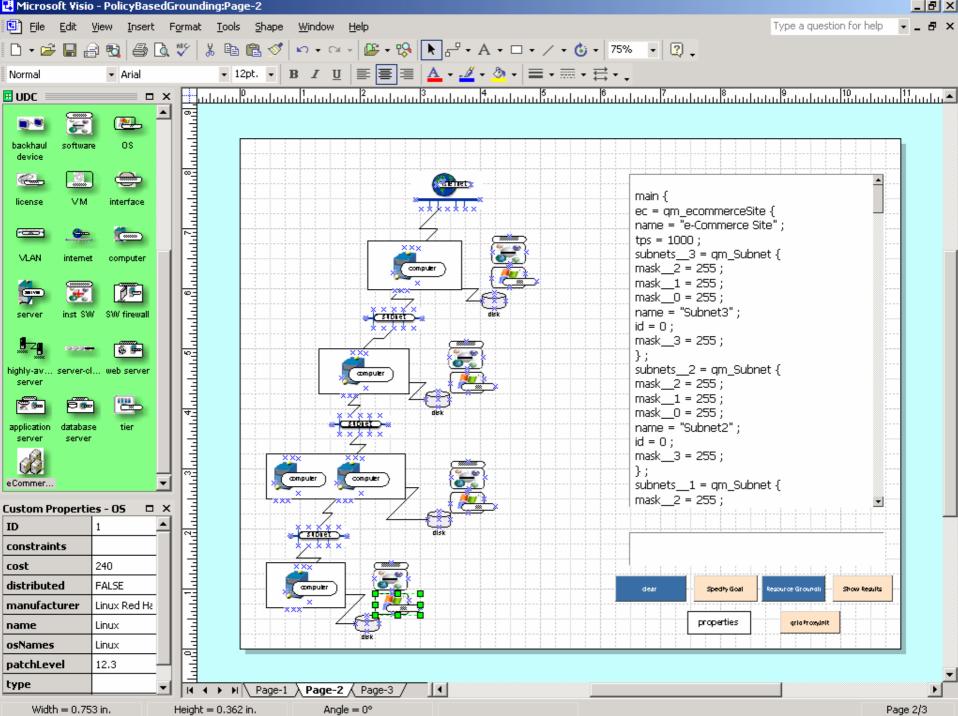
#### Definition:

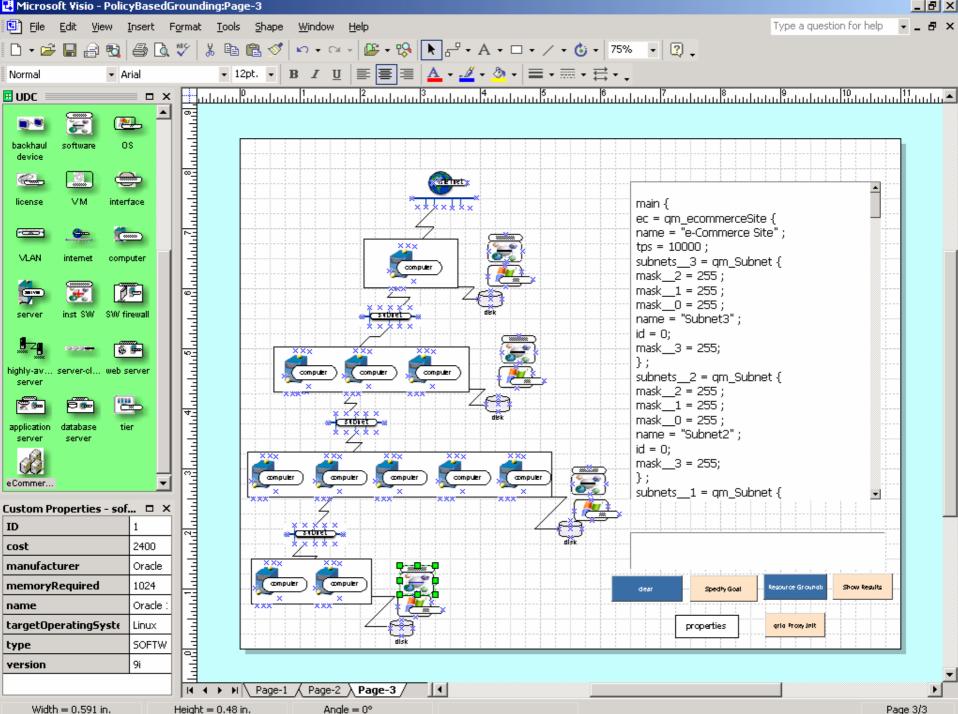
- A "grounded" resource is a resource that can be allocated.
- Resource grounding is the process of resolving non-grounded resource specifications into grounded, allocatable resource specifications.
- Resource grounding is based on pre-configured templates using a constraint satisfaction engine to recursively identify matching templates and binding all parameters.

#### OS Analogy:

– None.







Width = 0.591 in.

Height = 0.48 in.

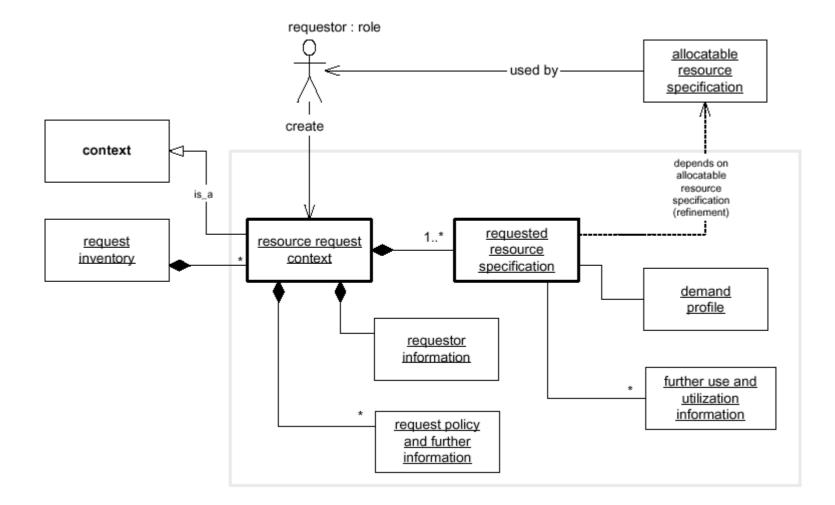
Page 3/3

### **Example: Resource Allocation**



- = Commitment to provide specified <u>quantities</u> of requested resource types for the specified <u>times</u>.
- Purpose:
  - Specifies resource need for complex resource arrangement:
    - Multiple resource types (atoms or constructions),
    - · Per resource type: Demand Profile (quantity over time),
    - Topology information.
    - · Policy information.
    - Requestor information.
- OS Analogy:
  - malloc( int nbytes);
  - sbrk( int n);
  - OS syscalls do not incorporate demand over time ("profile").

### **Resource Request Format**

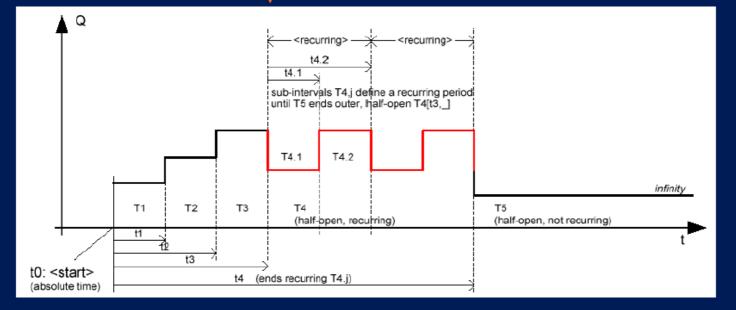


### **Resource Profile**

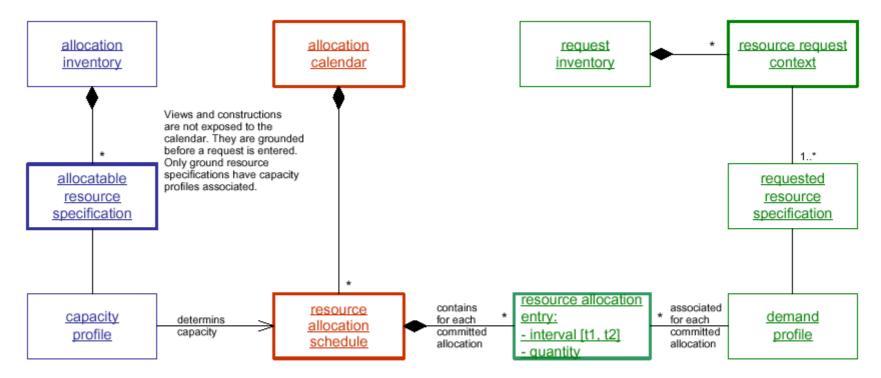


- Purpose:
  - Specifies resource quantity.
    - over <u>time</u>.
  - Used for both, demand and capacity.

Quantity Specification Q3	Example
Constant value: n	5 machines
Range flexing: [n,m] initially k, $n \le k \le m$	5 - 10 machines, initially 8
Discrete statistical distribution of <i>n</i> resources as tuples: [ <i>n</i> , <i>p</i> ( <i>n</i> )], <i>n</i> =0, 1, 2N, $\Sigma p(n)=1$	1 machine requested with probability 0.3, 2 machines with probability 0.6, 3 machines with probability 0.1
Statistical distribution as probability function	Gaussian normal distribution with expectancy and standard deviation of a demand



# Allocation Subsystem Components



Color coding:

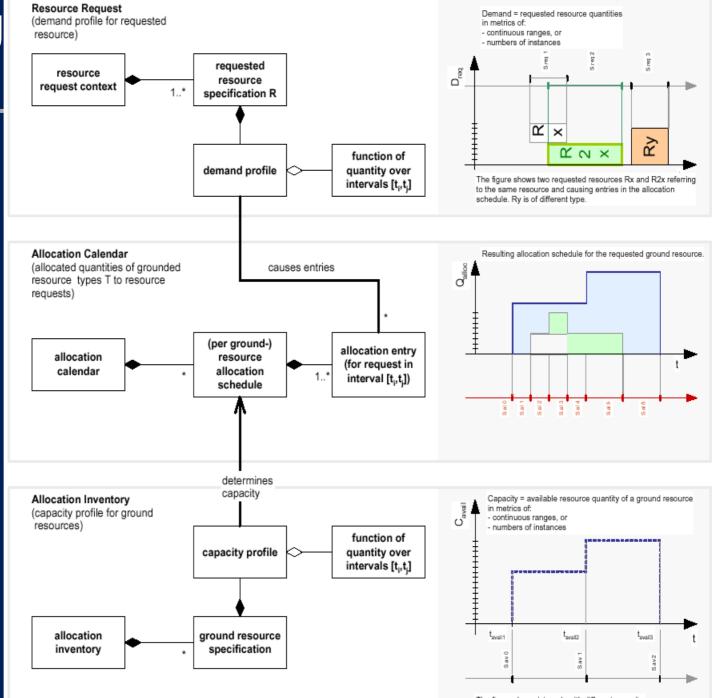
- blue: allocation inventory-related

- red: calendar-related

- green - request-related

information.

### Matching Profiles



The figure shows intervals with different capacity.

#### Summary



- New domain for OS Technology is emerging: systems for operating enterprise IT environments.
- Highly relevant area for the coming decade. Change will affect HW+SW vendors as well as system integrators and service providers in how IT systems are built and operated, significantly more based on formalisms and tools supporting design and validation of designs.
- CS graduates will need skills in architecture, design, specification and operation of complex IT environments.



#### More detail...



#### Automating the Operation of Enterprise IT Environments

#### **Habilitation Thesis**

(Preliminary Draft)

#### Dr.-Ing. Sven Graupner

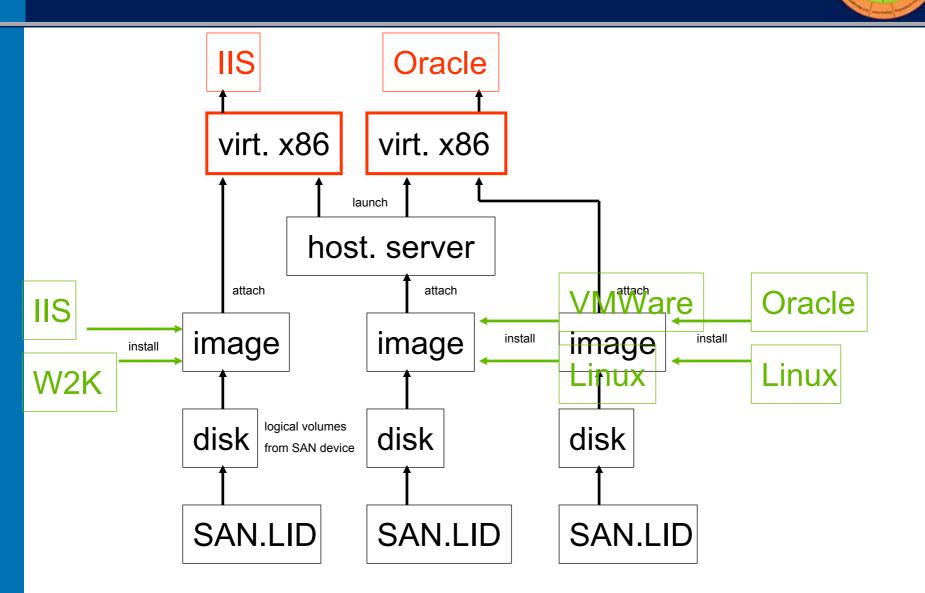
Hewlett-Packard Laboratories 1501 Page Mill Rd Palo Alto, CA 94304 USA

to be submitted to the Faculty of Computer Science Chemnitz University of Technology Germany

Palo Alto, November 12, 2004.

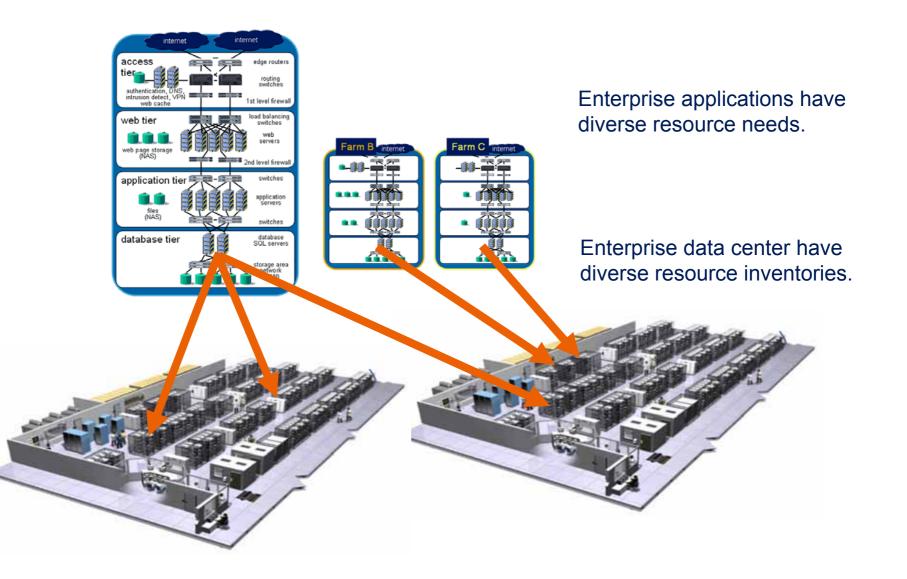
December 30, 2004

### **Example: Resource Construction**



# **Resource Allocation for Topologies**





# Comparing with



- Distributed OS → different granularity of resources, ability to deal with resource constructions.
- Management Systems → fully integrated, operates autonomously, operator role only sets policy.
- Compute Grids → heterogeneous resource topologies targeted for enterprise applications, longer life time, transactional, resource planning, aimed at by Enterprise Grid.
- Autonomous Systems implies self-operation and (operating) systems for this purpose.

## Aspects of Operating Systems



System Structure	Releasing Resources
External Interface	Process / Application Management
Underlying Control Points	Protection and Isolation
Base Abstractions	Self-Operation and Impact of Failure
Resources and Resource Management	Virtualization and Working Set
Requesting Resources	Interrupt Handling and Signaling
Grounding Resource Requests	Middleware and Interconnect Layer
Resource Scheduling	Information Base and Model

1 Brown		HOME HILFE		Bitte wählen Sie ⊻
D Springer	L'ALT	SEARCH	Alle	EV Los ERWEITERTE SUCHE
	ming, SWE	Zeits	chriften   Lehrbücher	Reihen   Deutsches Programm
Weitere Fachgeb	iete 💌	Zud	en Fachbereichen	×
> Home / Informatik / Pr	ogramming, SWE & Operating Sys	stems		
	Web Services in the Enter Concepts, Standards, Solu Reihe: Network and Syster Sahai, Akhil, Graupner, Sw 2004, Approx. 310 p., Geb. ISBN: 0-387-23374-1 Erscheinungstermin: Deze	utions and Manage ms Management, en	ment	Druckversion Diesen Titel empfehlen Alle Bücher dieser Autoren Sahai, Akhil Graupner, Sven  Verwandte Fachbereiche Database Management & Info Retrieval
Über dieses Buch	Inhaltsverzeichnis			Programming, SWE & Operating Systems
Über dies	es Buch			Wirtschaftsinformatik
The emergence of v industry hype surrou and implications to helping software ard behind web service they fit into the enter enterprise, as well a professional referer	veb services is transforming to unding these technologies ob- enterprises. Here the authors chitects, developers and IT ma technologies: the challenges prise stack, how they relate to as the existing and emerging s nee is a guide for business ma es, and how to use them in m	scures the underst take the "big pictur anagers understan and opportunities the business and standards and thei anagers and analy	anding of their impac e" perspective, d the concepts they present, how IT layers of the r relevance. This	:t

#### Geschrieben für:

Students and Academics, Software Architects, Developers and IT Managers who develop and maintain enterprise applications and web services.



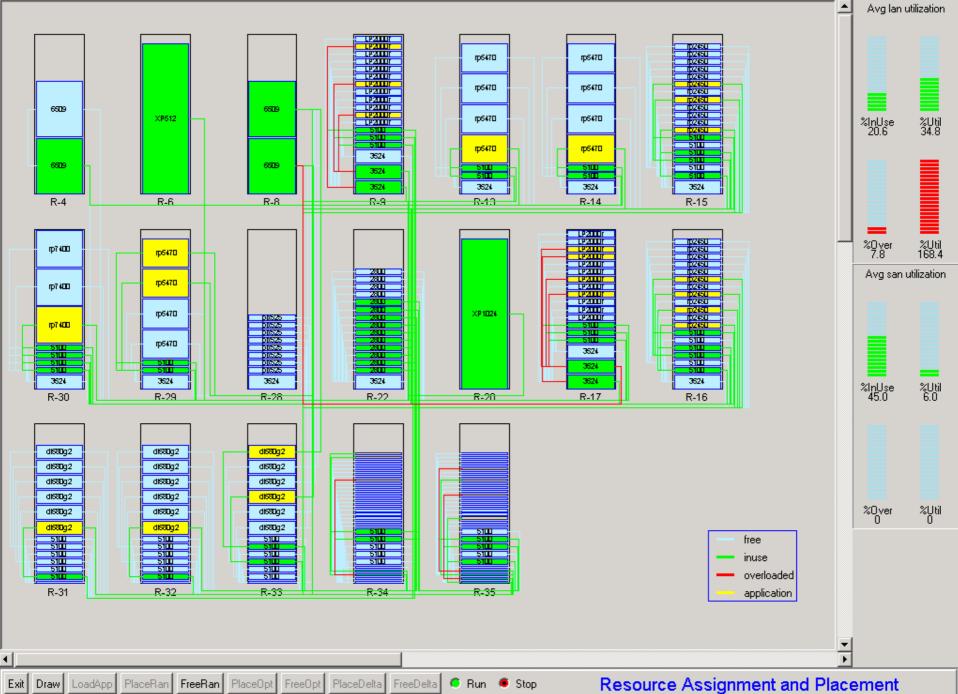
#### Automate dynamic resource assignment

- Quickly solves complex assignment problems that are beyond the scope of a human operator
- Creates optimal assignments to avoid bottlenecks
- Responds to changes in user requirement in "real-time"

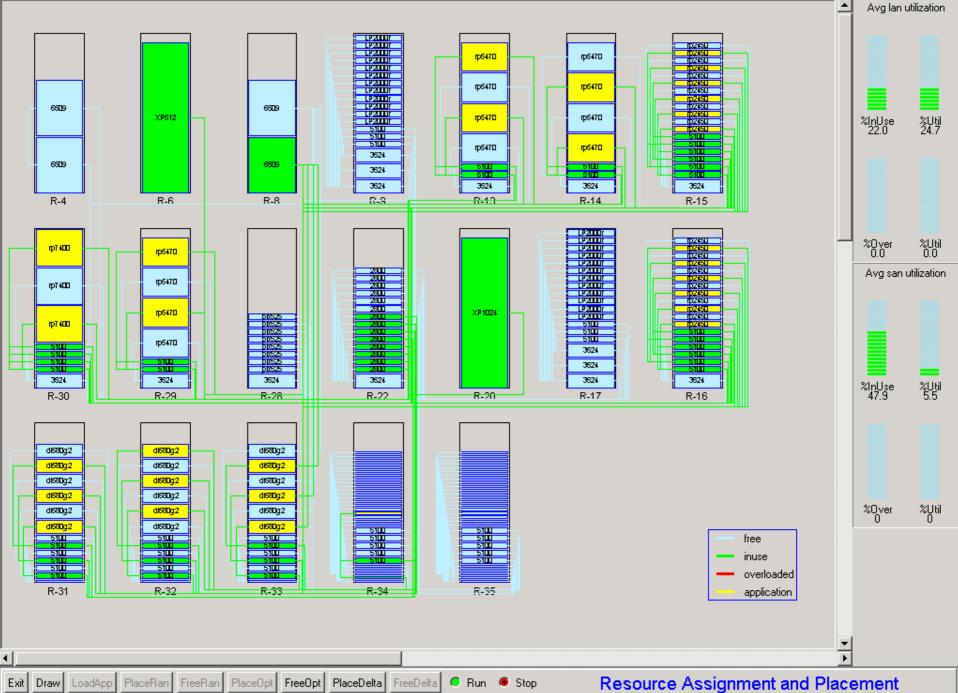
#### Core Innovation

- Solved a very complex combinatorial non-linear NPhard problem by transforming it into a linear one that can be handled efficiently by commercial solvers
- Advantage of approach: Clearly indicates if a feasible optimal solution exists and if not, why not.









### Some Related Work

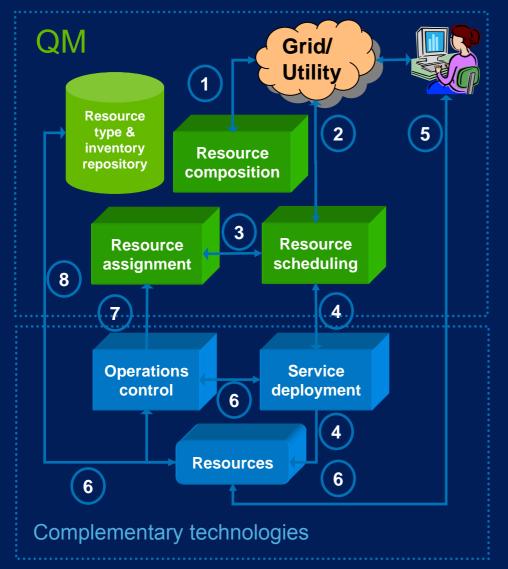


- Grid standards for Grid/utility computing
- Imaging / provisioning systems: Altiris, Rembo, Novadigm (+ many others)
- MS Dynamic Systems Initiative, MS Provisioning System
- Adaptive systems: ThinkDynamics, Corosoft
- IBM eLiza, autonomic computing

 Proprietary (product-specific) solutions in place today for installation / configuration / activation

### QM Research Prototype





- User uses a resource composition service to design a custom environment (or selects a preconfigured template).
- 2. User schedules deployment of application.
- 3. Resources needed for the deployment are assigned.
- 4. Service is deployed, and
- 5. Resources are made available to user.
- 6. On-line monitoring is used to adjust resources as necessary.
- Resource availability & utilization is used to improve future decisions.
- 8. The type/inventory repository tracks any changes in resources.

#### HP Utility Data Center (UDC)





# Configuring Resource Topology

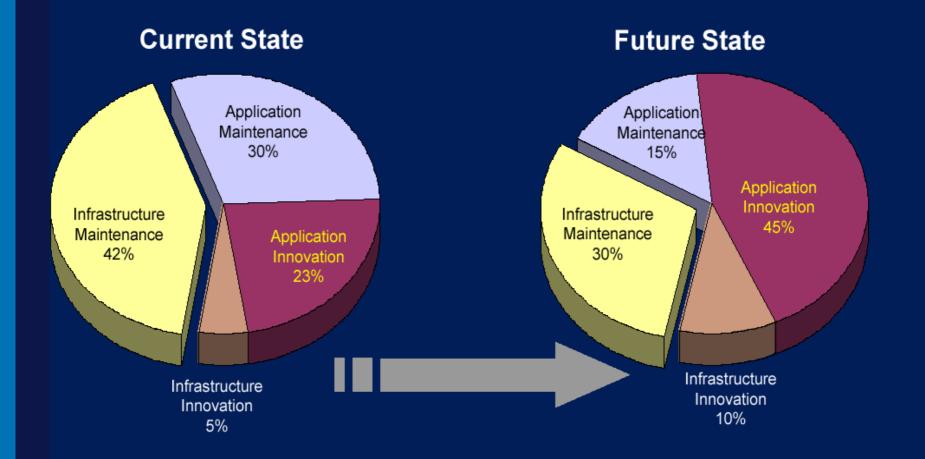


	Configure Server Web Page Dialog	? 🗙 lp Account Logout
File Print Submit	Configure: Serve	Editor
	Name Direct Attached Storage   Type HP Ip2000r   ef. Gateway None	
Farm Details	Add Disk Delete Disk	
ame: SSS2 tate: Design Status: • envice Type: Not Set envice Core:	Server Group Server Count: 4	
Pseudo Service Core 💌 equests: Last Request Completed: N/A Current Status: <u>Ready to take new request</u>	Nameeth0: IP Addresseth1: IP AddresswebSNot AssignedNot ConfiguredwebS-1Not AssignedNot ConfiguredwebS-2Not AssignedNot ConfiguredwebS-3Not AssignedNot Configured	
	OK Cancel Help	
HP 1p2000r 0 HP 1t6000r 0 HP j8000	webs	



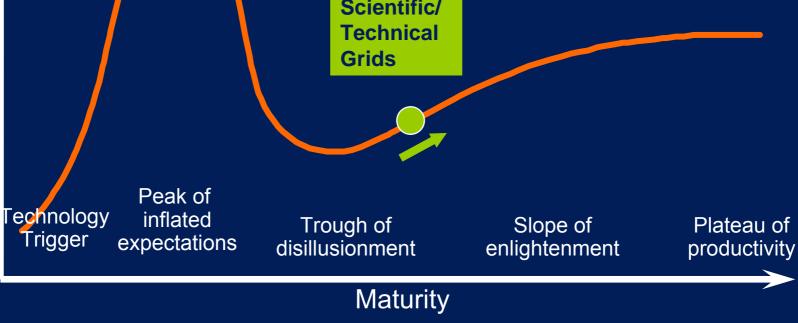
# Addressing the Pain Points in the Enterprise.

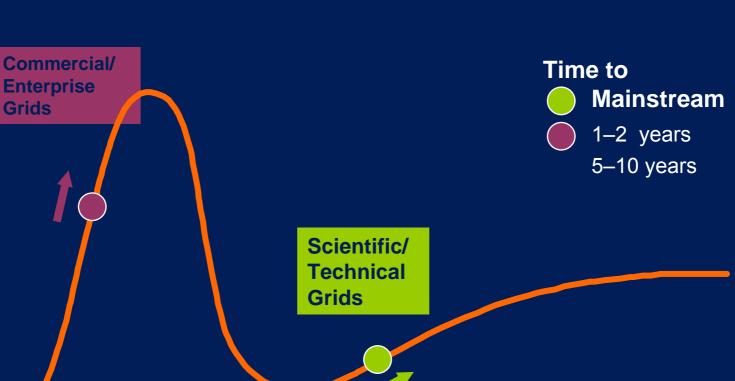




Source: HP

Visibility



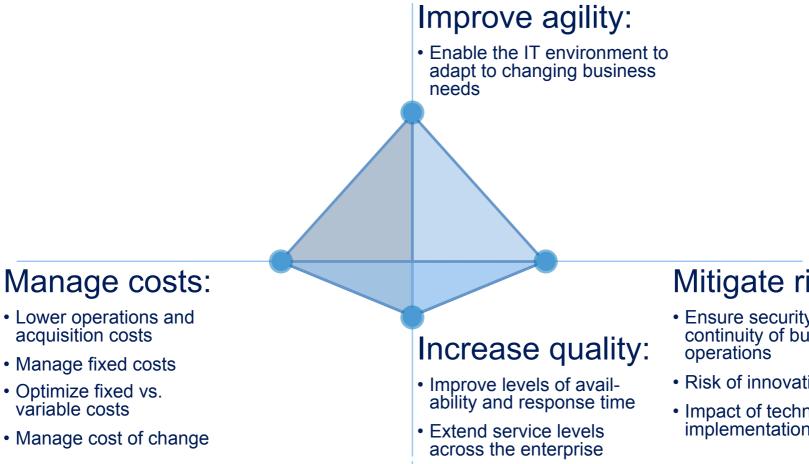


# Hype cycle



# The Opportunity: Increased Agility





- Ability to change the service as well as deliver it
- Agility as service level

#### Mitigate risk:

- Ensure security and continuity of business
- Risk of innovation
- Impact of technology implementation



	Scientific Grid	Enterprise Grid
Workload	compute jobs	mostly transactional
Lifetime	duration of job, range of hours, sometimes days	duration of application deployment, range of months and years
Resources	homogeneous, parallel machines or clusters	diverse resource, environment including servers, storage, networks (LAN, SAN), devices
Schedulable Unit	job	Resource Topology

#### Service-Oriented Architecture What kinds of features does an SOA have?



Feature	Principles
<b>Security</b> – Three level security mechanism (transport, message and authentication/authorization) utilizing standards (http/s, WS- Security, etc) and security integration (Select Access, Netegrity SiteMinder).	Standards-based, Architecturally Layered, Identifiable Participants
<b>Discovery</b> – Ability for discovering and addressing all services available in the SOA.	Identifiable Participants, Discoverable Participants
<b>Registry</b> – Ability to register all services discovered in the SOA with a standard UDDI registries. Registry stays synchronized with discovery mechanism.	Standards-based, Discoverable Participants, Model Oriented
<b>Metrics</b> – Measure various statistics about service interactions (response time, errors, etc.) and aggregate them.	Architecturally Durable, Model Oriented
<b>SLO</b> – Define service level objectives for metrics at different aggregation levels with breach transition notification.	Architecturally Durable, Model Oriented
<b>Audit</b> – Archive any service interaction. Can define the criteria for audit and interaction context to archive.	Architecturally Layered, Identifiable Participants
<b>Routing</b> – Can be deployed to have the ability to control the flow of interactions. QoS and customer routing available.	Architecturally Durable, Architecturally Adaptive, Flexibly Deployed
<b>Management Integration</b> – SOA itself is built from set of interacting services and described as a model that forms the core for integration and adaptive change.	Extensible Platform, Model Oriented, Architecturally Adaptive



### QM: Capacity Profile

Capacity – Capacity is the current or anticipated quantitative availability of a resource (type or instance). Capacity may be expressed in metrics of continuous ranges or numbers of instances.

Examples of capacity expressed in terms of numbers of instances are: 100 machines of type A or 20 devices of type B. Examples of capacity expressed in terms if continuous ranges are: 1GB/s bandwidth or 10 TB storage.

► Capacity Profile – Capacity profile of a ground resource is a function of anticipated capacity of a resource over time.

Only ground resources have capacity provisioned. Capacity of a resource is a function over time about what quantity of a resource type is available (or are expected to be available) over a time period. Capacity profile is provided by the inventory maintainer role.

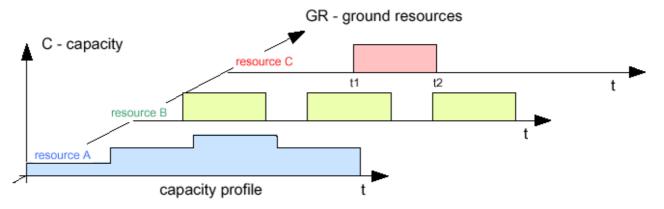


Figure 4: Capacity profiles of ground resources (are part of ground resource specification).



#### **QM: Demand Profile**

Demand – Demand is the current or anticipated quantitative use of a resource. Demand may be expressed in metrics of continuous ranges or numbers of instances.

Examples of demand expressed in terms of numbers of instances are: 87 machines of type A or 12 devices of type B. Examples of demand expressed in terms if continuous ranges are: 0.86 GB/s bandwidth or 8.4 TB storage.

▶ **Demand Profile** – Demand profile of a requested resource is a function of anticipated demand of a resource over time.

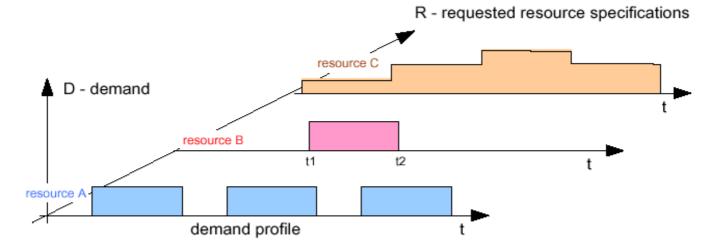


Figure 5: Resource demand profiles as part of a resource request.

#### **Information about HP Labs**

hp

### HP Labs Worldwide





#### HP Labs web site





#### http://www.hpl.hp.com

#### **Research Centers in HP Labs**



