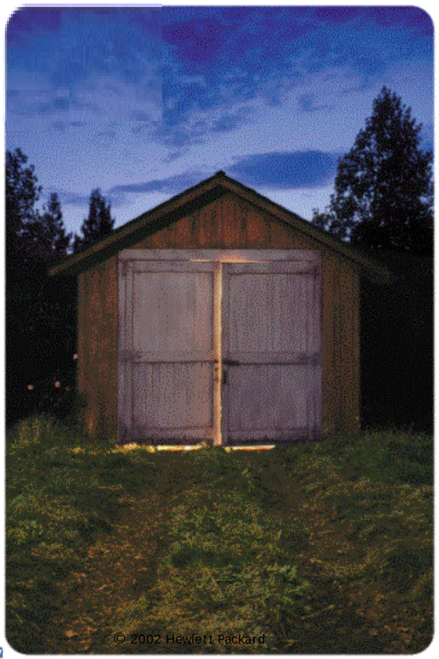


Planetary-Scale Computing

infrastructure
&
services

hp laboratories
May 2002

Aad van Moorsel
using slides from
Patrick Scaglia
Rich Friedrich
John Wilkes
Giuliano DiVitantonio

Service Centric Computing

HP and planetary scale computing

- In the mid-80's Joel Birnbaum of HP Labs first talked about pervasive/utility computing
- In the second half of the 90's HP Labs began work on "e-services computing"
  - Anyone remember e-speak?
- In the late 90's HP Labs began work on planetary scale computing
  - Invent a utility that allocates globally connected resources on demand, anytime, anywhere
  - Realize service centric computing

Service Centric Computing

Planetary scale computing

- A new computing model that allocates IT resources on demand, anywhere
  - Static or dynamic
  - Economical
    - Commodity computing
    - Self-aware with proactive control
  - Programmatically configured
    - Rather than re-cabling…
    - Federated "bricks" of server and storage
  - Federating on a planetary (geographic) scale
    - $O(10^5)$ element data centers

Service Centric Computing

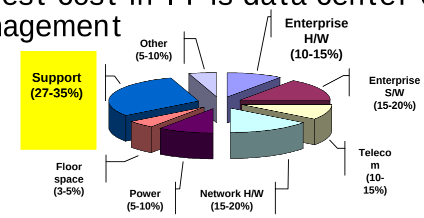
Planetary scale computing

- IT infrastructure becomes a virtual resource service
  - Description, specification
  - Provisioning
  - Management
  - Billing
  - Trust
- Conceptual target:
  - thousands of resources per service, thousand of services per data center, thousands of data centers

Service Centric Computing

Resulting challenge

- "pay as you grow" is attractive
- horizontal scaling enables it
- managing the sheer number of devices that results (>10k node data centers) is a problem
  - the largest cost in IT is data center operations and management

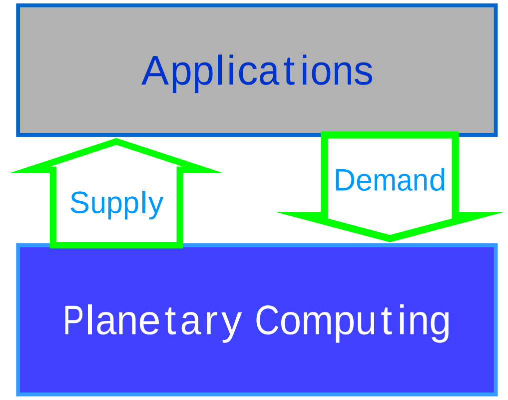


Source: Giga May 2001

Service Centric Computing

basic idea: consumer/supplier

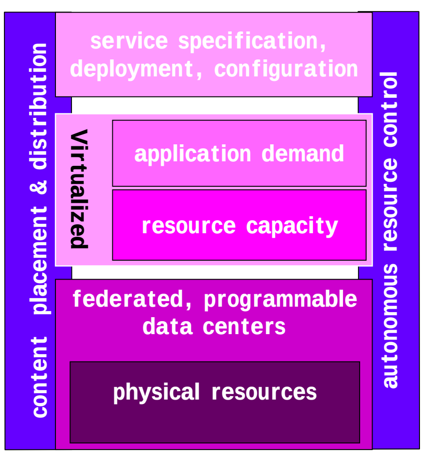
intelligent provisioning: effective use of physical resources



intelligent brokering: match service demand with resource capacity

Service Centric Computing

Basic idea (cont.)



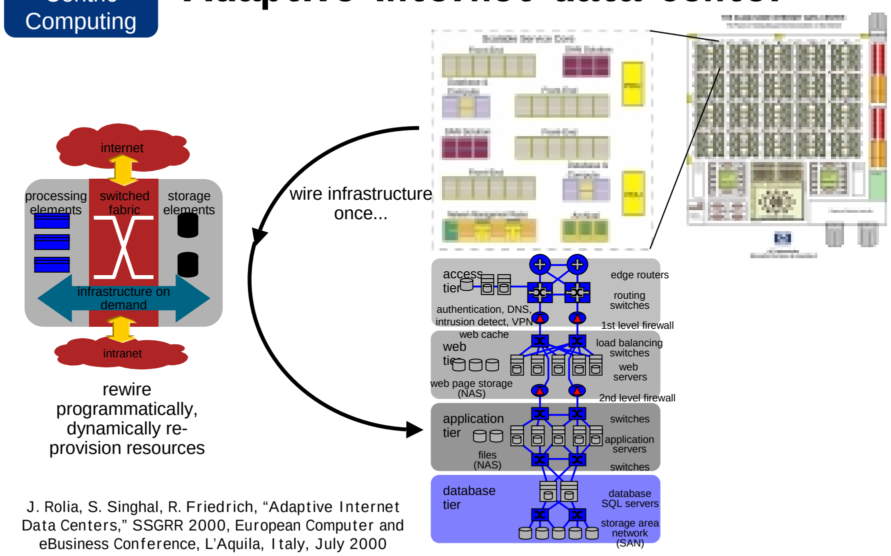
Layer 3 Services
- Services consist of application components on virtual resources
- Service semantics / SLAs

Layer 2 Virtual Resources
- Intelligent provisioning of virtual resources to physical service cores
- Resource planning based on QoS/power/cost/geography/administration

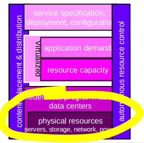
Layer 1 Physical Resources
- Servers, storage and switches
- Software images, licenses
- Data center resources: power, floor space, economics

Service Centric Computing

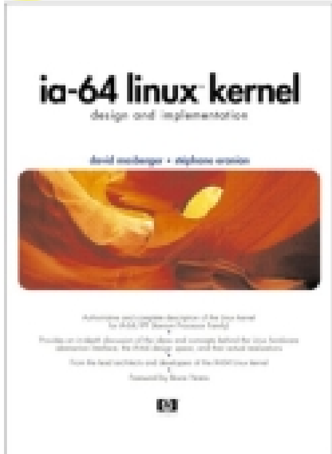
Adaptive internet data center

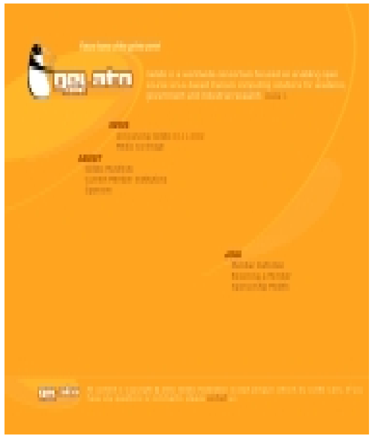


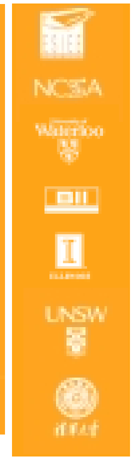
J. Rolia, S. Singhal, R. Friedrich, "Adaptive Internet Data Centers," SSGRR 2000, European Computer and eBusiness Conference, L'Aquila, Italy, July 2000



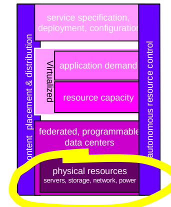
Scalable commodity open source platform



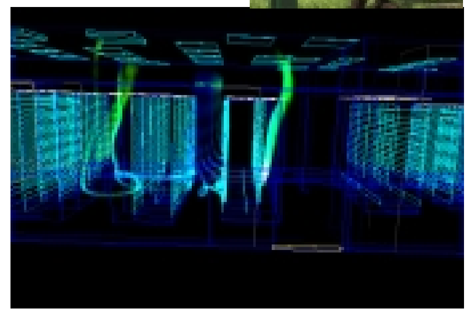




http://www.gelato.org

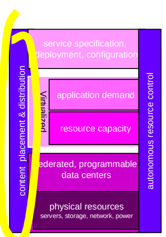


Dynamic thermal management in large scale data centers

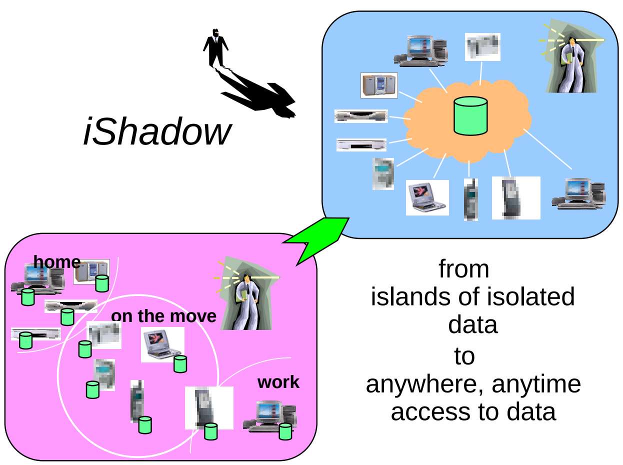


- Power Density -
  - Microprocessor: 200 W/cm^2 **(by 2003, today 60 W/cm^2)**
  - System – 300 W, thin 1U form factor 10 to 15 KW per EIA Rack foot print
  - Room- 2700 W/m^2 (~300 W/ft^2)
- Use 3D modeling to understand thermal characteristics of data centers
- Exploit this for dynamic resource allocation

Patel, C.D., Sharma, R.K, Bash, C.E., Beitelmal, A, Thermal Considerations in Cooling Large Scale High Compute Density Data Centers, ITherm 2002 – 8th Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems" May 2002, San Diego, California

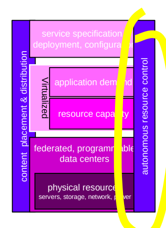


Planetary scale storage

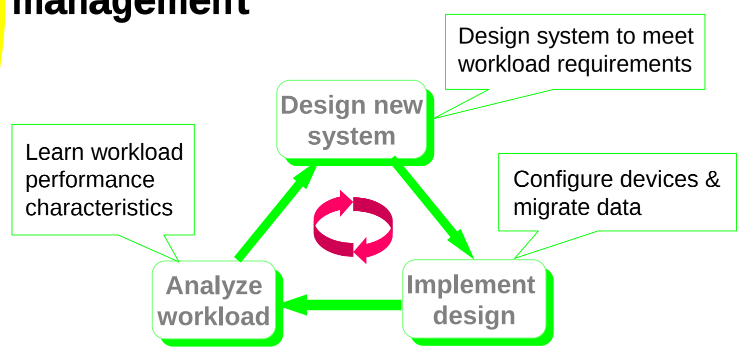


from
islands of isolated data
to
anywhere, anytime access to data

Alistair Veitch, Erik Riedel, Simon Towers, John Wilkes. Towards global storage management and data placement, 8th Workshop on Hot Topics in Operating Systems (HotOS-VIII, 20-23 May 2001, Schloss Elmau, Germany)

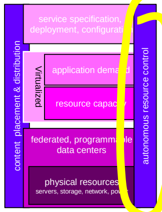


Hippodrome: automatic storage management



benefit: "autonomic" storage

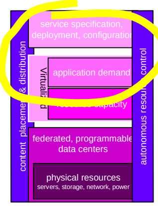
Hippodrome: running circles around storage administration
Eric Anderson, Michael Hobbs, Kimberly Keeton, Susan Spence, Mustafa Uysal, and Alistair Veitch. Conference on File and Storage Technology (FAST'02) January 2002



Self aware services

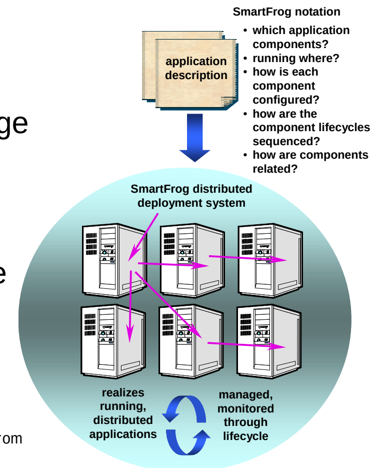
- How to manage 50,000 servers, 1 million objects ?
  - centralized management, human-centered operation, polling architectures don't scale
- services monitor own health and the health of local dependents to determine the root cause of failures
  - based upon statistical measures and bayesian network reasoning

"Self-Aware Services: Using Bayesian Networks for Detecting Anomalies in Internet-based Services"; Bronstein, Alexandre; Cohen, Ira; Das, Joydip; Duro, Marsha; Friedrich, Richard; Kleyner, Gary; Mueller, Martin; Singhal, Sharad in Proceedings of Integrated Network Management VII (IM-2001), 14-18 May2001, Seattle, IEEE/IFIP



SmartFrog: service description and deployment

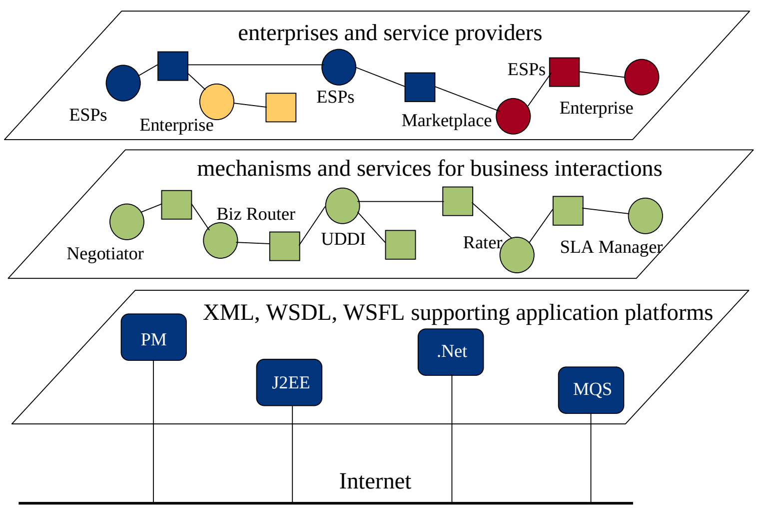
- Configuration description language
  - precise, desired configuration of applications composed of sets of components running across a distributed system
- Service deployment architecture for massive systems
  - realize application description
  - monitor and manage the resulting applications through their lifecycles



Patrick Goldsack, "SmartFrog: A framework for configuration", from the Workshop on Large-Scale System Configuration, Edinburgh, November 2001. (Online proceedings available at www.dcs.ed.ac.uk/home/paul/wshop)

Service Centric Computing

web services everywhere



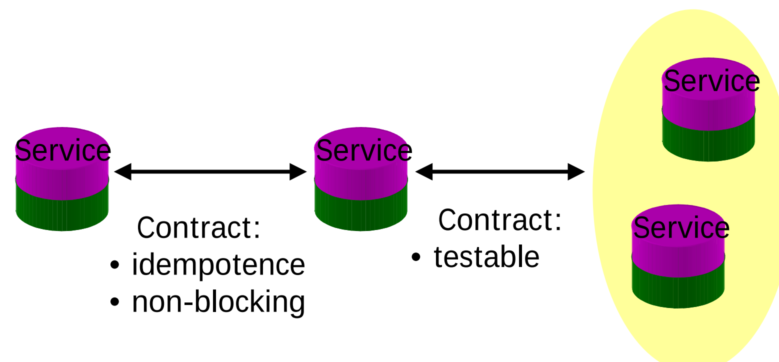
business verticals:
- procurement
- HR
- utility computing

reusable, core service blocks:
- services
- proxies
- app APIs

Service Centric Computing

exactly-once transactions

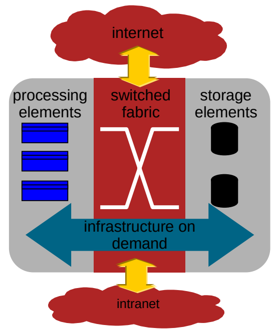
Client	↔	Service	↔	Environment
(Travel Agent)		(Car Broker)		(AVIS/Hertz/Alamo)



Service Centric Computing

From research to reality

- HP announced the Utility Data Center (UDC) Nov 2001
- Based on HP Labs research on adaptive internet data center
  - ability to direct resources to any application dynamically
  - self healing, policy driven.
  - Open system: Windows, Linux, HP-UX, Sun Solaris

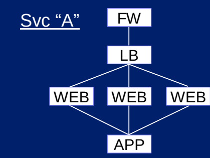


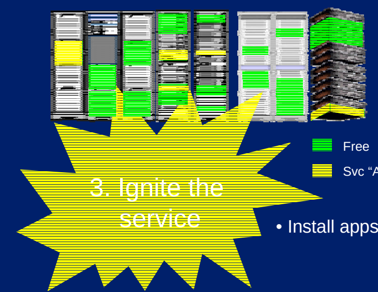
… to create a dynamically configurable utility fabric that can be programmed per service or customer, based on SLAs and demand…

Service Centric Computing

creating a service with the UDC

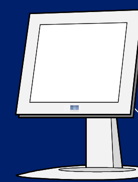
1. Architect new service:

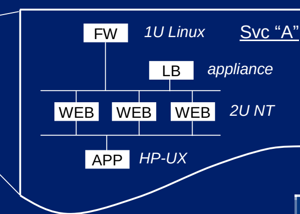




- Install apps

2. Build a service template:





Discover and apply free resources

- Specify connectivity
- Auto-configure network and storage
- Auto-load OSes

Service Centric Computing

conclusion

- HP focus on service-centric (utility) computing
- self-management research at all layers (for all the known reasons, but also to deal with new dynamism):
  - storage self-management
  - utility data-center resource allocation
  - self-aware services
  - service lifecycle management
  - exactly-once multi-party web service conversations
  - …