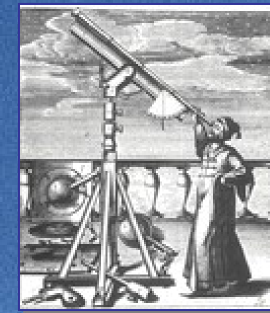
# From Research Infrastructures to Industry Participation: A Long Way to Go?

**Tony Hey**
**Corporate VP for Technical Computing**
**Microsoft Corporation**

# A New Science Paradigm

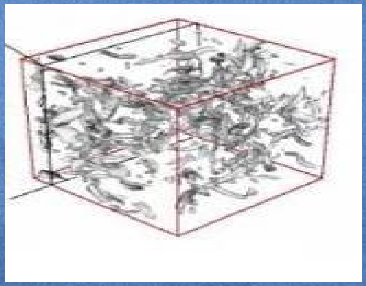
- **Thousand years ago:**
  Experimental Science
  - description of natural phenomena



- **Last few hundred years:**
  Theoretical Science
  - Newton's Laws, Maxwell's Equations …

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - \mathrm{K}\frac{c^2}{a^2}$$

- **Last few decades:**
  Computational Science
  - simulation of complex phenomena



- **Today:**
  e-Science or Data-centric Science
  - unify theory, experiment, and simulation
  - using data exploration and data mining
    - Data captured by instruments
    - Data generated by simulations
    - Data generated by sensor networks
    - Scientist analyzes databases/files

(With thanks to Jim Gray)

# e-Science and Cyberinfrastructure?

- In 2003, the NSF published the ‘Atkins Report’ on ‘Revolutionizing Science and Engineering through Cyberinfrastructure’

- Report defined Cyberinfrastructure as:
  - Grids of computational centers
  - Comprehensive libraries of digital objects
  - Well-curated collections of scientific data
  - Online instruments and vast sensor arrays
  - Convenient software toolkits

# e-Science and e-Infrastructure?

## Malcolm Read of JISC (e-IRG 2005)

**e-Infrastructure includes:**

- Networks (internet, light paths……..)
- Computers (workstations, servers, HPC………….)
- Access controls (security, AAA…………)
- Middleware (metadata…………..)
- Finding tools (portals, search engines………)
- Digital libraries (bibliographic, text, images, sound………)
- Research data (national and scientific databases, individual data……..)

**“But little of this is universally available”**

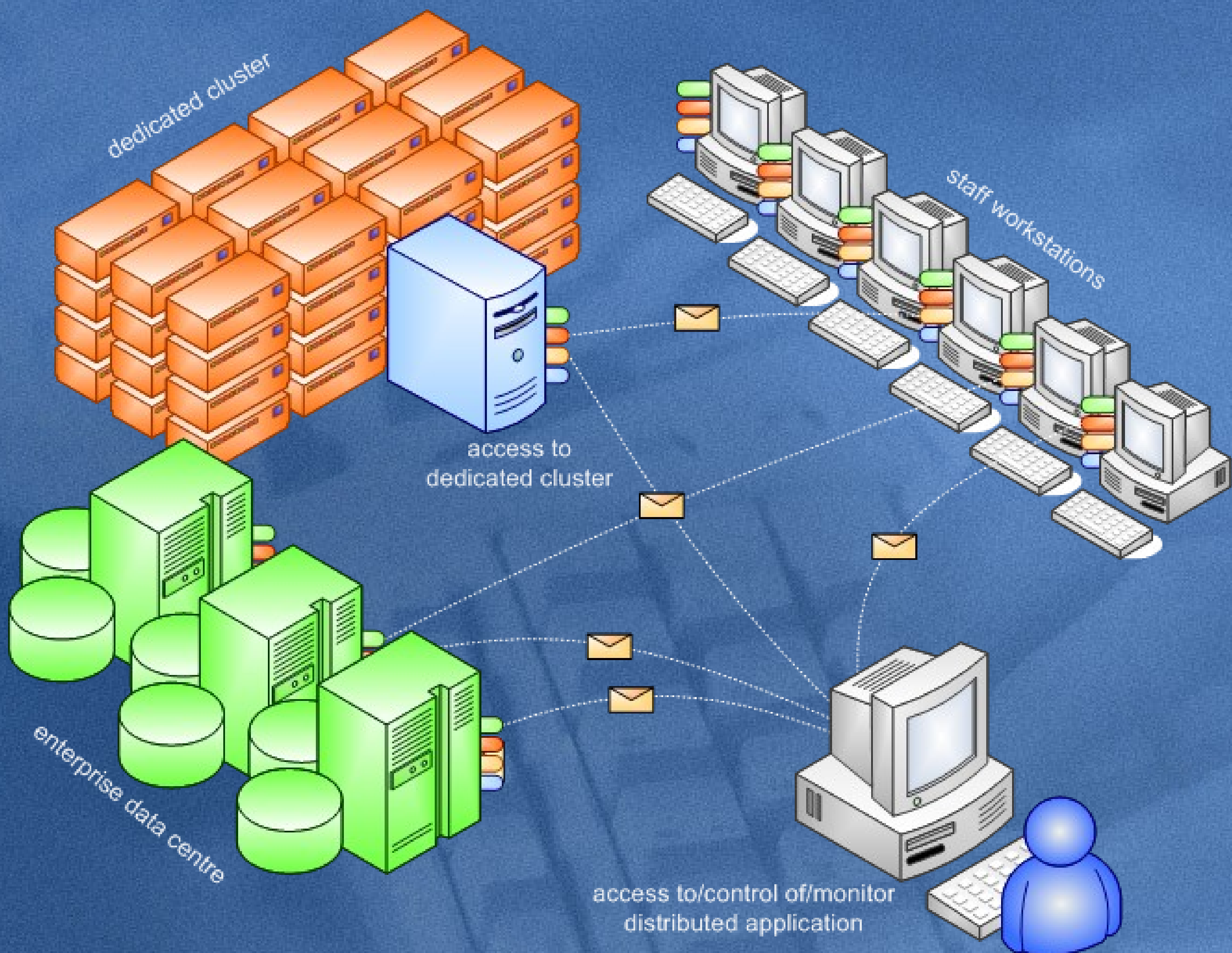
# Research Infrastructure

- In the US, Europe and Asia there is a common vision for the research infrastructure required to support the e-Science revolution
  - Service Oriented Middleware Services and Components plus high bandwidth academic research networks
  - Powerful new tools to analyze research data and for knowledge management and discovery
  - Open Access federation of research repositories containing full text and data

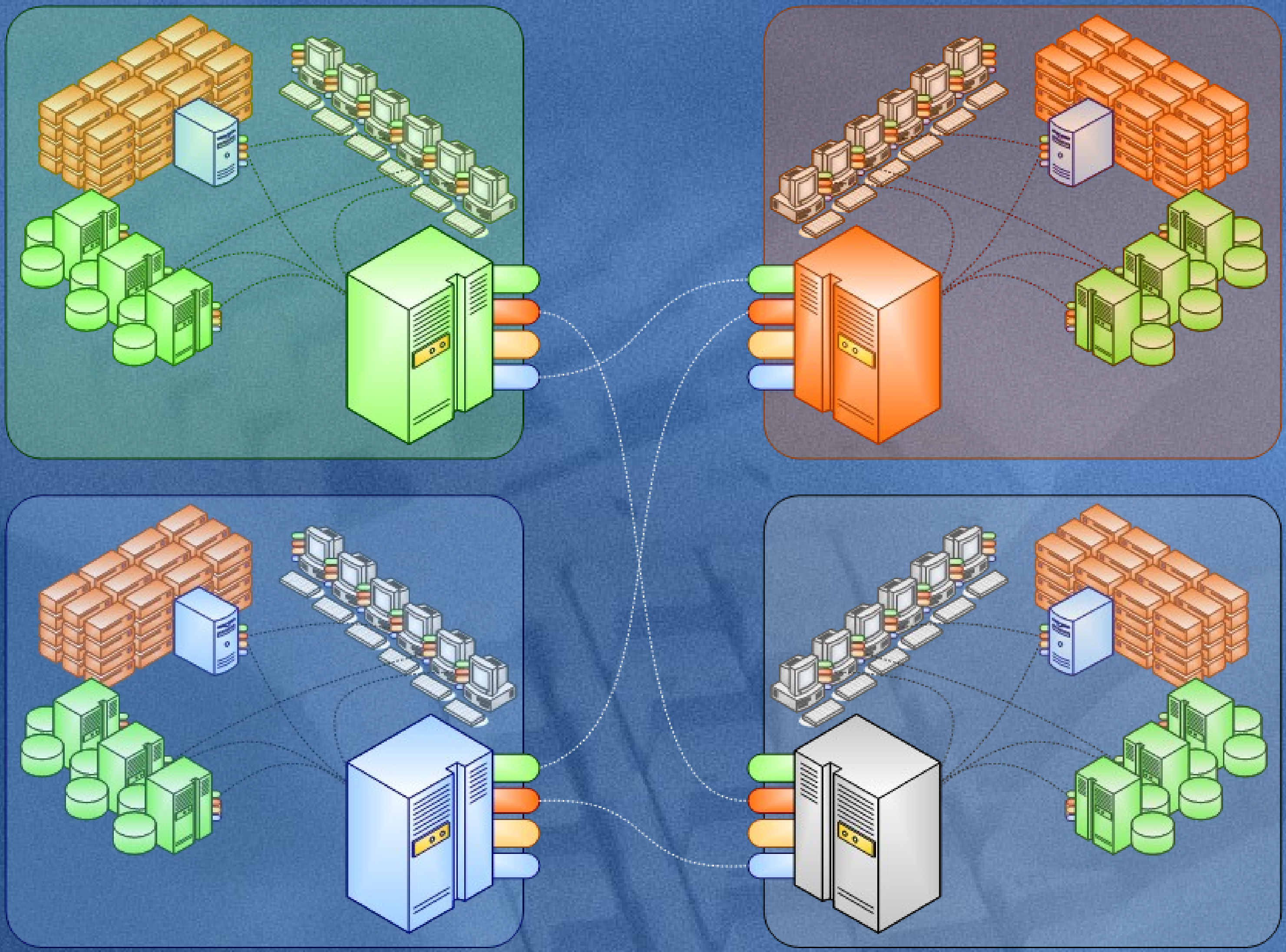
# Grids in Industry (1)

- IBM Enterprise Grid Solutions
  - "Virtualization" – the driving force behind Grid computing
  - Infrastructure optimization
    - Consolidate workload management
    - Provide capacity for high-demand applications
  - Resilient highly-available infrastructure
    - Balance workloads
    - Enable recovery from failure
- Emphasize reduction in TCO and better ROI for businesses
  - Different motivations in academia?

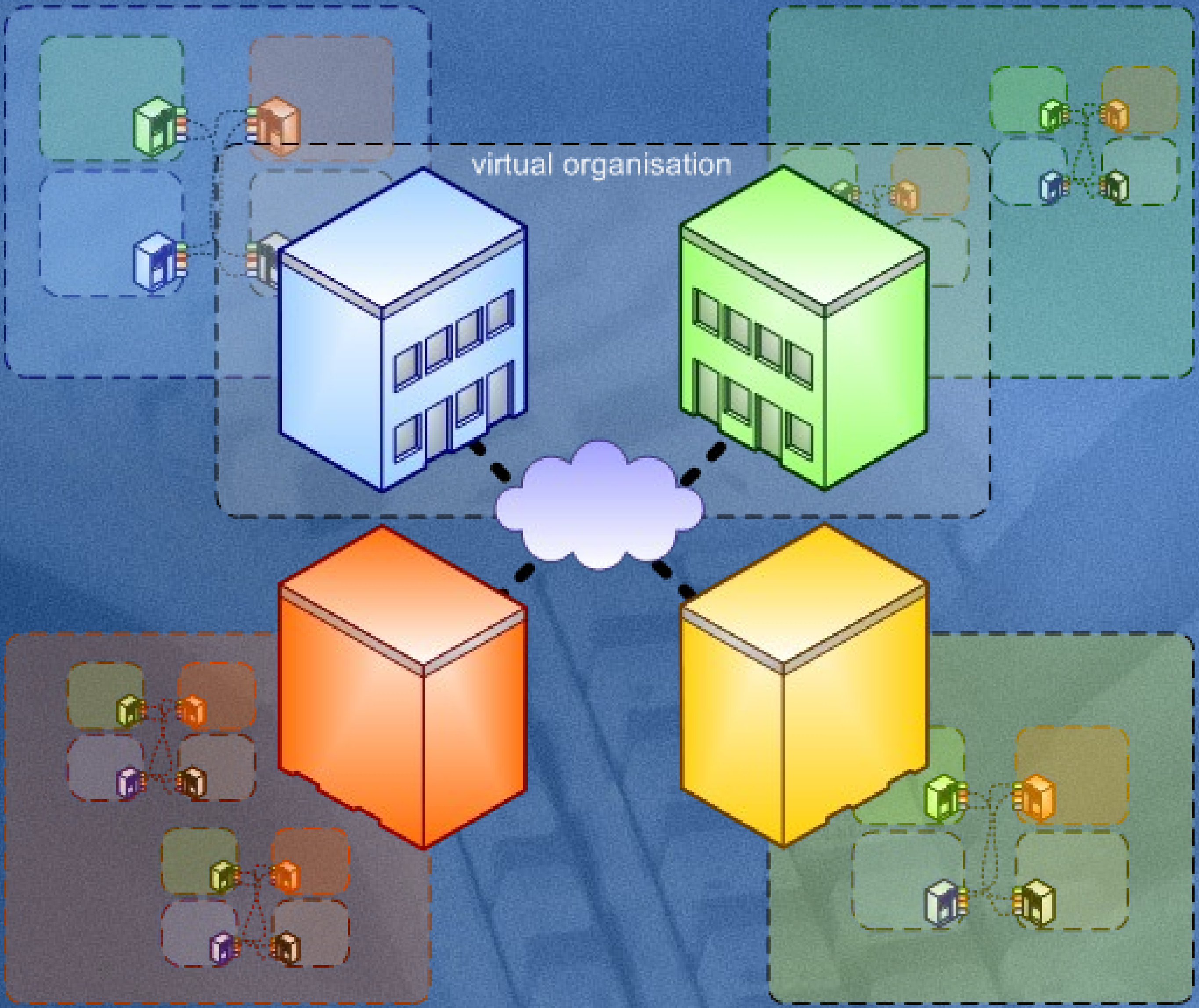
# Cluster-based HPC



# Intra-Organization HPC



# Virtual Organizations



GRID today

The Leading Source for Global News and Information from the evolving Grid ecosystem, including Grid, SOA, Virtualization, Storage, Networking and Service-Oriented IT

September 25, 2006

Home Page

*Special Features:*

## Getting Beyond the Compute Grid: The Challenge of 'Grid 2.0'

### By William Fellows, Principal Analyst, The 451 Group

Early adopters of Grid computing have indicated that many are looking for ways to move beyond the use of grids specifically for computational tasks. They want to eliminate silos and organize IT around shared resources, simplify access to data, and provide a single, consistent view of the business for the entire enterprise. The challenge they face is essentially getting "beyond the compute grid," a topic we've looked at in a number of reports.

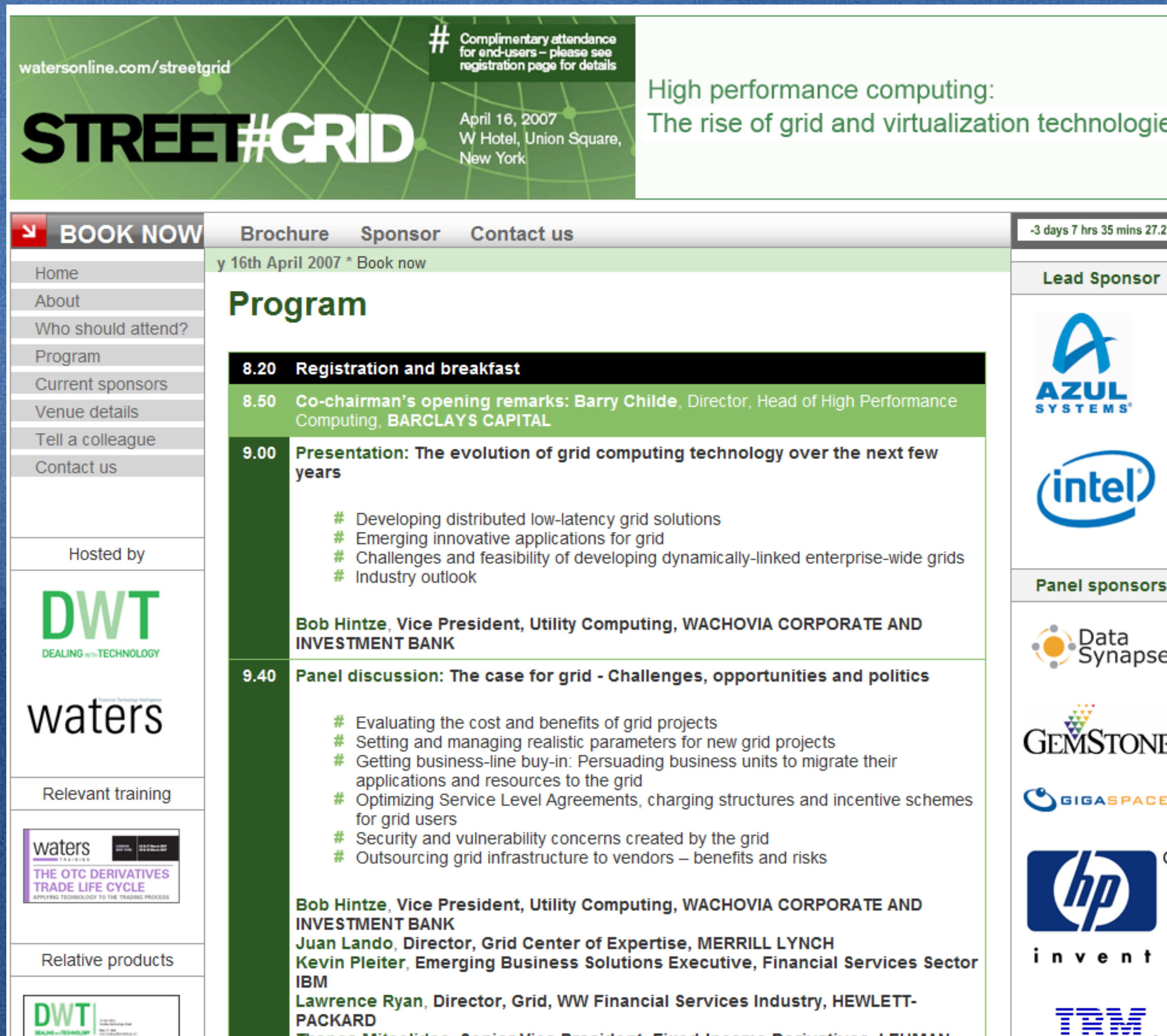
Grid computing supports the key functional drivers required to achieve these goals, including virtualization, service-oriented architecture (SOA) and utility delivery models. To do this, grids need to be absorbed into the enterprise infrastructure itself and not used only as a stand-alone computational engine. In such a strategic role, grids must support automated data, storage and service activities just as capably as handling computational tasks. These challenges are being addressed by what is being called "Grid 2.0."

If "Grid 1.0" is principally concerned with the virtualization, aggregation and sharing of compute resources, Grid 2.0 is focused on the virtualization, aggregation and sharing of all compute, storage, network and data resources. It is both service-oriented -- uses Web services and provides access to IT as a service -- and automated.

# The 451 Group Analysis

- See 5 levels of Grids:
  - Level 1 – Proofs of concept, trials
  - Level 2 – Single application running
  - Level 3 – Siloed Grids, multiple applications
  - Level 4 – Inter-departmental Grids
  - Level 5 – Enterprise-wide Grids
- Early adopters of computing Grids:
  - Financial services
  - Pharmaceutical companies
  - Oil & Gas
  - Manufacturing
  - TelCos

# Financial Grids



watersonline.com/streetgrid

STREET#GRID

# Complimentary attendance for end-users – please see registration page for details

April 16, 2007
W Hotel, Union Square,
New York

High performance computing:
The rise of grid and virtualization technologies

BOOK NOW | Brochure | Sponsor | Contact us

-3 days 7 hrs 35 mins 27.27

y 16th April 2007 * Book now

- Home
- About
- Who should attend?
- Program
- Current sponsors
- Venue details
- Tell a colleague
- Contact us

## Program

**8.20** **Registration and breakfast**

**8.50** **Co-chairman's opening remarks: Barry Childe**, Director, Head of High Performance Computing, **BARCLAYS CAPITAL**

**9.00** **Presentation: The evolution of grid computing technology over the next few years**

# Developing distributed low-latency grid solutions
# Emerging innovative applications for grid
# Challenges and feasibility of developing dynamically-linked enterprise-wide grids
# Industry outlook

**Bob Hintze, Vice President, Utility Computing, WACHOVIA CORPORATE AND INVESTMENT BANK**

**9.40** **Panel discussion: The case for grid - Challenges, opportunities and politics**

# Evaluating the cost and benefits of grid projects
# Setting and managing realistic parameters for new grid projects
# Getting business-line buy-in: Persuading business units to migrate their applications and resources to the grid
# Optimizing Service Level Agreements, charging structures and incentive schemes for grid users
# Security and vulnerability concerns created by the grid
# Outsourcing grid infrastructure to vendors – benefits and risks

**Bob Hintze, Vice President, Utility Computing, WACHOVIA CORPORATE AND INVESTMENT BANK**
**Juan Lando, Director, Grid Center of Expertise, MERRILL LYNCH**
**Kevin Pleiter, Emerging Business Solutions Executive, Financial Services Sector IBM**
**Lawrence Ryan, Director, Grid, WW Financial Services Industry, HEWLETT-PACKARD**
**Thanos Mitsolides, Senior Vice President, Fixed-Income Derivatives, LEHMAN**

Hosted by

DWT DEALING with TECHNOLOGY

waters

Relevant training

waters THE OTC DERIVATIVES TRADE LIFE CYCLE

Relative products

DWT

Lead Sponsor

AZUL SYSTEMS

intel

Panel sponsors

Data Synapse

GEMSTONE

GIGASPACES

hp invent

IBM

# Grids in industry (2)

- Google, Amazon, Yahoo, eBay and Microsoft as major 'Cloud Platform' providers
  - All have infrastructures of hundreds of thousands of servers
  - Many large data centers, distributed across multiple continents
  - Have developed proprietary technologies for job scheduling, data sharing and management
  - Care about power consumption, fault tolerance, scalability, operational costs, performance, etc.
  - They are living the "Grid dream" on a daily basis

# Google as an example …

- Estimated 450,000* servers distributed around the world
- Google File System - highly distributed, resilient to failures, parallel, etc.
- Schedulers and load balancers for the distribution of work
  - Use their ‘Map-Reduce’ middleware as parallel computational model
- What is financial/competitive advantage for them to adopt standards for these in-house developed solutions?

* source: Wikipedia

# Amazon Web Services: Simple Storage Service S3

- S3 is storage for the Internet
  - Designed to make web-scale computing easier for developers
- Provides a simple Web Services interface to store and retrieve any amount of data from anywhere on the Web
  - ‘CRUD’ philosophy – Create, Read, Update and Delete operations
- Uses simple standards-based REST and SOAP Web Service interfaces
  - Built to be flexible so that protocol or functional layers can easily be added

# Amazon S3 Functionality

- Intentionally built with a minimal feature set
  - Write, read, and delete objects containing from 1 byte to 5 gigabytes of data each
- Can store unlimited number of objects
  - Each object is stored and retrieved via a unique, developer-assigned key
- Authentication mechanisms provided
  - Objects can be made private or public, and rights can be granted to specific users
- Default download protocol is HTTP
  - BitTorrent protocol interface is provided to lower costs for high-scale distribution

# Amazon Web Services: Elastic Compute Cloud EC2

- Compute on demand service that works seamlessly with their S3 storage service
- Create Amazon Machine Image (AMI) containing application, libraries and data
- Use EC2 Web Service to configure security and network access
- Compute cost ‘subsidized’ by Christmas!

# EC2 Web Service

- Use EC2 to start, terminate and monitor as many instances of your AMI as you want
- Each instance has:
  - 1.7 GHz x86 Processor
  - 1.75 GB RAM
  - 160 GB local disk
  - 250 MB/s network bandwidth
- Used by Catlett and Beckman as capacity computing alternative to TeraGrid 'SPRUCE' capability computing for emergency urgent response

# Industry and Standards?

- Why don't they care about Grid standards?
  - See competitive advantage in their proprietary infrastructure, proprietary services
- No need for interoperation between companies
  - Will not see map-reduce tasks from Google being offloaded to Amazon's infrastructure any time soon
  - Security and privacy are still major concerns
  - Probably still cheaper for them to maintain their own infrastructure than to offload to someone else's
- There are some possibilities for standardization
  - WS-Management standards could reduce the operational cost of large collections of commodity infrastructure components - motherboards, disks, network hubs, etc.

# Customer demand can drive interoperability solutions

## Example of Security

- Customers want secure access/confidentiality no matter what service they use
- No competitive advantage in doing simple security in a proprietary way
  - Customer demand is forcing Microsoft’s Passport, Google’s ID, Liberty Alliance, OpenID to interoperate
- More complicated ‘VO’ scenarios are yet to be proven widely useful
  - Cross organization trust, credential delegation, id federation
  - Work has already started in this space e.g. WS-Trust, WS-Policy
  - VOs required for academic/research collaborations
  - Perhaps B2B scenarios like Supply Chains could provide commercial driver for their wide adoption

# Open Access Research Repositories

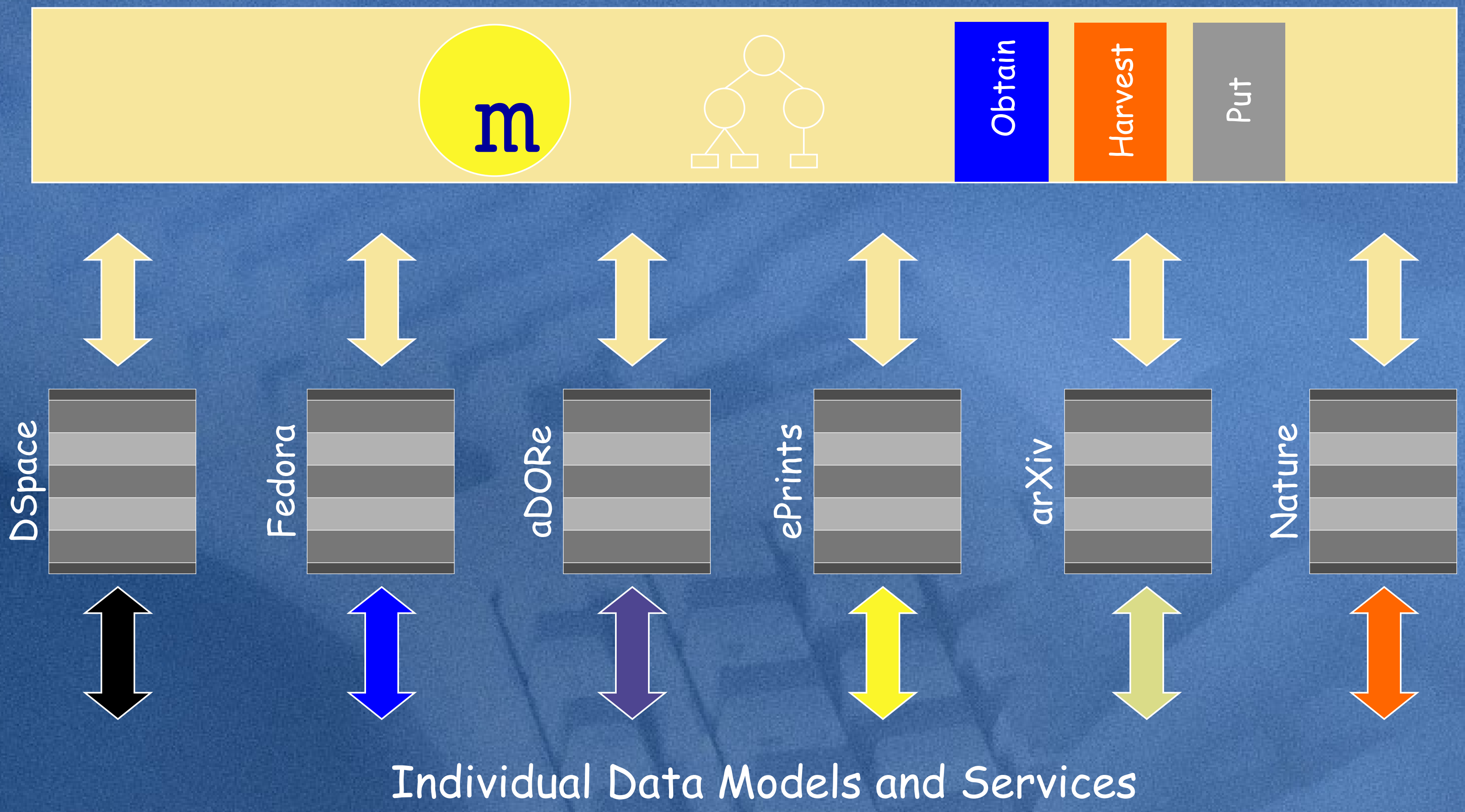
- Repositories will contain not only full text versions of research papers but also ‘grey’ literature such as technical reports and theses
- In addition, repositories in the future will also contain data, images and software
- There will be many types of repository software and more powerful interoperability protocols such as OAI-ORE
- Libraries and researchers can add value by creating composite services as ‘eResearch Mashups’
- Currently over 1400 OAI-compliant repositories and more than 20 different types of software

# Institutional Repositories?

http://www.webometrics.info/top3000.asp

| UNIVERSITY | | GOOGLE SCHOLAR RANK |
|---|---|---|
| HARVARD UNIVERSITY | | 1 |
| UNIVERSITY OF CHICAGO | | 2 |
| JOHNS HOPKINS UNIVERSITY | | 3 |
| MASSACHUSETTS INSTITUTE OF TECH | | 4 |
| STANFORD UNIVERSITY | | 5 |
| TOKYO UNIVERSITY | | 6 |
| UNIVERSITY OF PENNSYLVANIA | | 7 |
| UNIVERSITY OF SAO PAULO | | 9 |
| UNIVERSITY OF CALIFORNIA BERKELEY | | 10 |
| UNIVERSITY OF WISCONSIN MADISON | | 11 |
| CARNEGIE MELLON UNIVERSITY | | 12 |
| UNIVERSITY OF CALGARY | | 13 |
| VIRGINIA POLYTECHNIC INSTITUTE | | 14 |
| UNIVERSITY OF SOUTHAMPTON | | 15 |
| UNIVERSITY COMPLUTENSE MADRID | | 17 |
| STATE UNIVERSITY OF CAMPINAS | | 18 |
| UNIVERSITY OF TEXAS AUSTIN | | 19 |
| UNIVERSITY OF ILLINOIS (U C) | | 21 |
| UNIVERSITY OF NEBRASKA LINCOLN | | 22 |
| UNIVERSITY OF MARYLAND | | 23 |
| CALIFORNIA INSTITUTE OF TECHNOLOGY | | 23 |
| GEORGIA INSTITUTE OF TECHNOLOGY | | 25 |
| UNIVERSITY OF MICHIGAN | | 26 |
| UNIVERSITY OF ARIZONA | | 27 |
| UNIVERSITY OF EDINBURGH | | 28 |
| CORNELL UNIVERSITY | | 29 |
| TEXAS A&M UNIVERSITY | | 30 |

# ORE - Augmenting interoperability



# Sustainability?

- Interoperable network of National, Institutional and Subject Repositories
  - National Repositories are supported by Governments
  - Institutional Repositories are the major role for university libraries in the future
  - Subject  Repositories will be supported by Research Funding Agencies
- Deep interoperability such as OAI-ORE will be essential given hetereogeneity of repository software

# Towards a Sustainable Research Infrastructure?

- IT Infrastructures are still considered a competitive advantage
  - Optimization and better utilization of the infrastructure brings more money
- Companies are trying to take advantage of their proprietary infrastructures to build new services and compete
  - Why would they want to standardize now?
- Could some commercial 'Cloud Services' be a part of the sustainable e-Infrastructure?
- Research Repository component of the infrastructure has 'natural' sustainability model?

Microsoft®

Your potential. Our passion.™