Linux and Grid Computing in Higher Ed

Rudolf Simmer

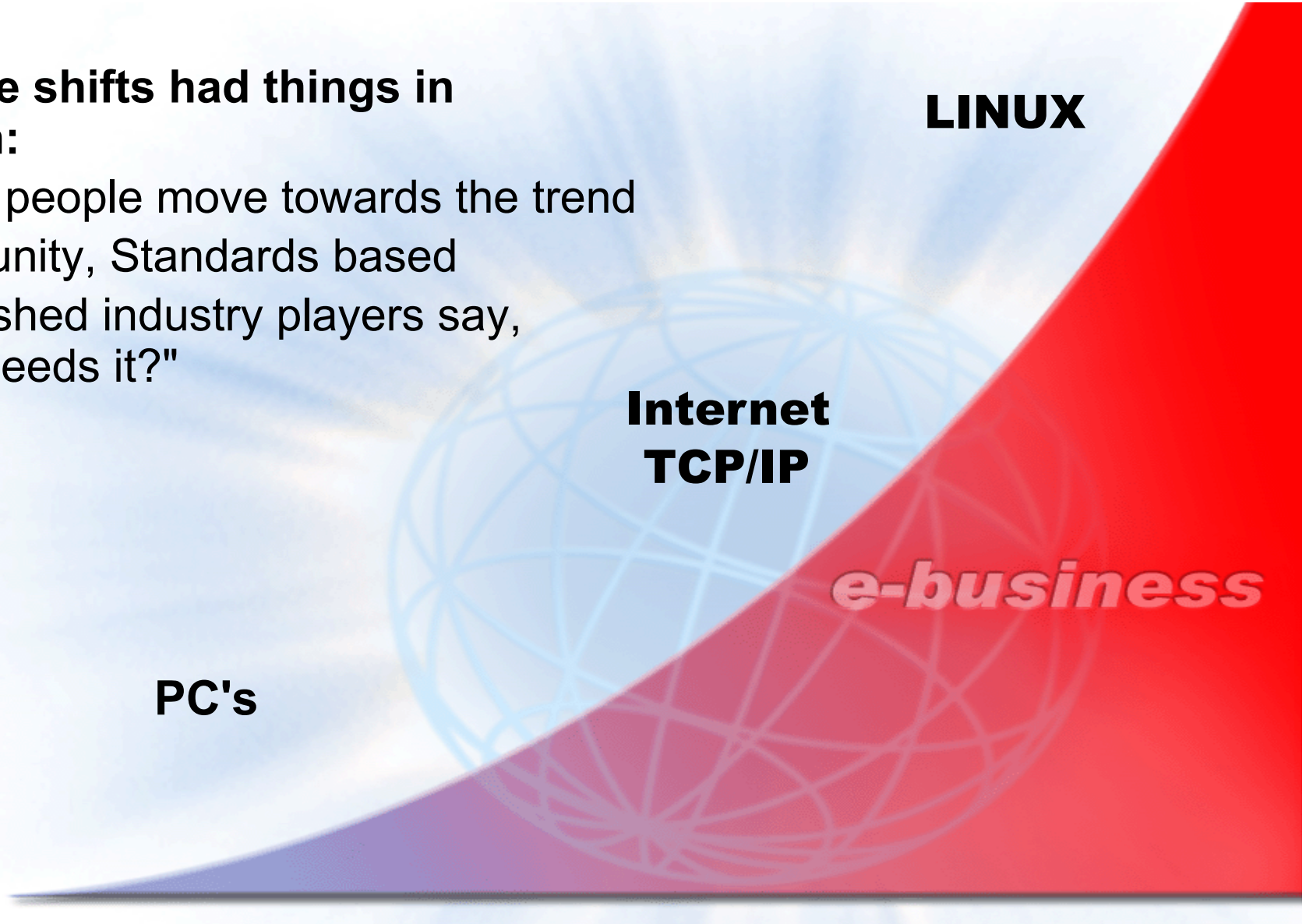*IBM Central Europe, Middle East, Africa*

Agenda

- **Linux Trends & Myths**

- **Linux in Higher Ed**

- **GRID Computing**

Finding a Trend
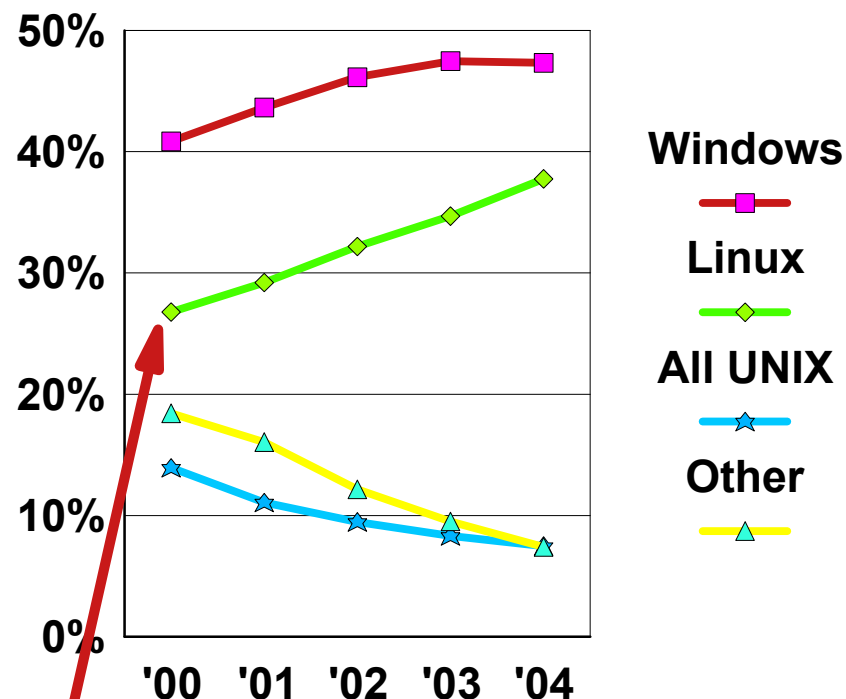
The three shifts had things in common:

- $$ and people move towards the trend
- Community, Standards based
- Established industry players say, "Who needs it?"

LINUX

Internet
TCP/IP

e-business

PC's

IBM

Linux Momentum

- Linux will definitely have the <u>fastest growth</u> in 2002, at almost 50% over 2001
  Gartner Group, December 2001

- Linux will have a "<u>breakout year</u>" in 2002. Now it seems clear that Linux has become a viable <u>alternative for enterprise use</u>.
  IDC, January 2002

- Exploring Linux - especially on the server side - <u>should be part of an organization's IT investment</u> portfolio. ...viewed as an insurance policy to <u>hedge against future Microsoft incursions into your wallet</u>.
  META Group, August 2001
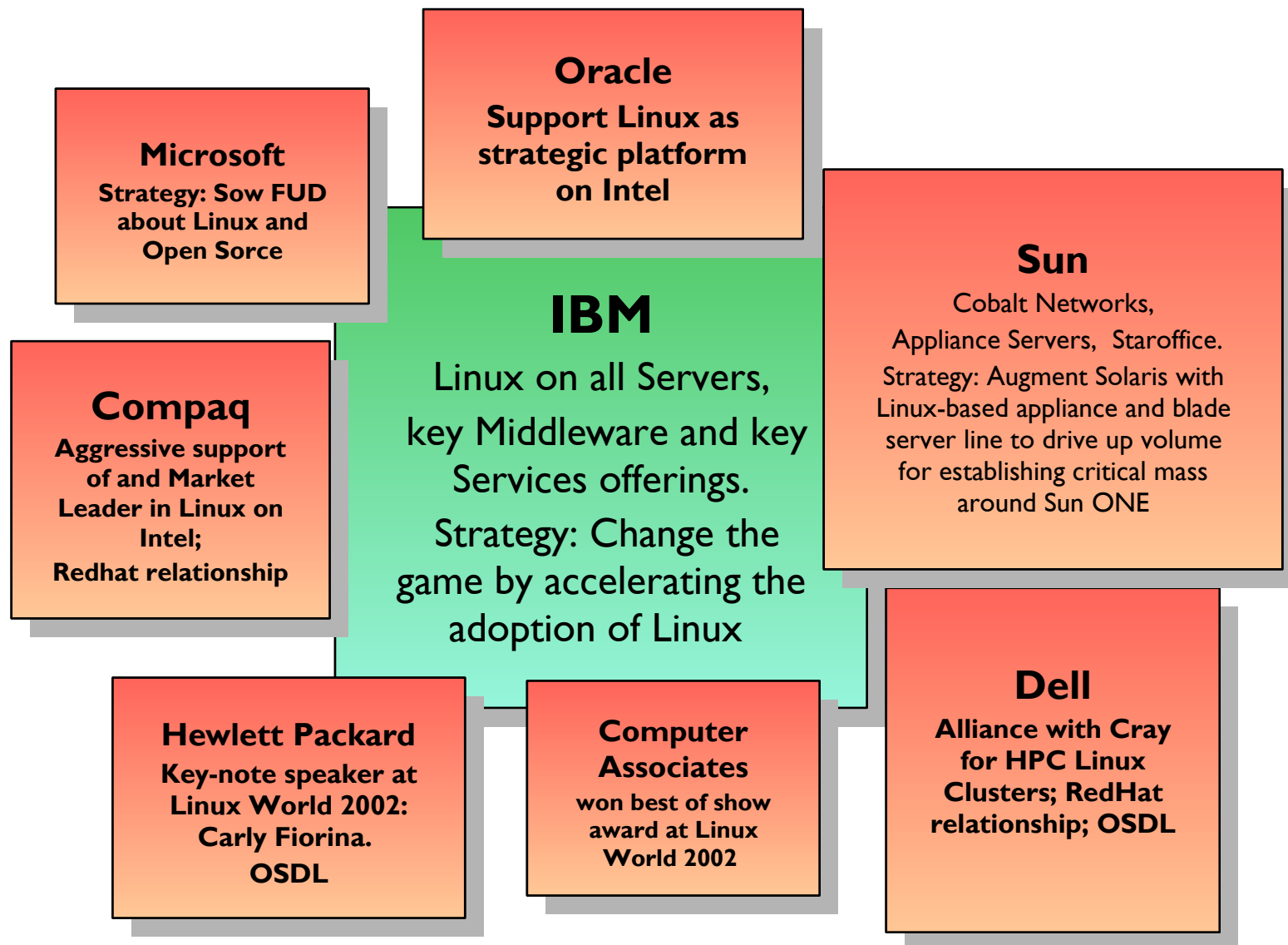
WW New Server OS Share (units)

Windows

Linux

All UNIX

Other

IDC, July 2001

Linux 23.6% Growth

Who is (and who is not) playing

Microsoft
Strategy: Sow FUD about Linux and Open Sorce

Oracle
Support Linux as strategic platform on Intel

IBM
Linux on all Servers, key Middleware and key Services offerings.
Strategy: Change the game by accelerating the adoption of Linux

Sun
Cobalt Networks, Appliance Servers, Staroffice.
Strategy: Augment Solaris with Linux-based appliance and blade server line to drive up volume for establishing critical mass around Sun ONE

Compaq
Aggressive support of and Market Leader in Linux on Intel;
Redhat relationship

Hewlett Packard
Key-note speaker at Linux World 2002: Carly Fiorina.
OSDL

Computer Associates
won best of show award at Linux World 2002

Dell
Alliance with Cray for HPC Linux Clusters; RedHat relationship; OSDL

IBM

Linux Distributors Combine their Power

UnitedLinux = Open Industry Consortium

- **Concept:**
  - Binary-compatible Linux distribution, branded "UnitedLinux"
  - Initial Members: Caldera, SuSE, Turbolinux, Conectiva
    - Open to others

- **Business Model:**
  - Distributors add software and services
  - Distributors maintain brand names with "UnitedLinux Inside"
  - UnitedLinux promotes the brand, issue memberships, certifications, manages requirements
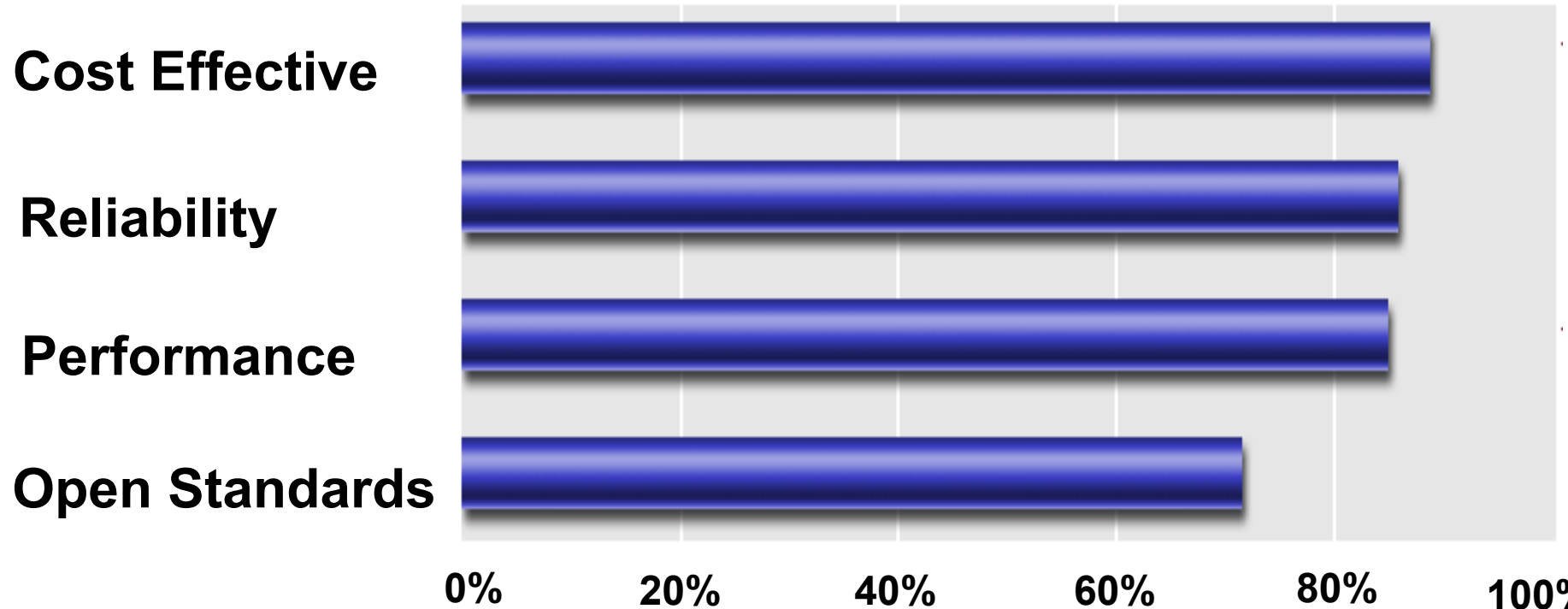 - SuSE acts as UnitedLinux systems integrator

- **Implementation:**
  - Supported by all IBM @server platforms, key middleware
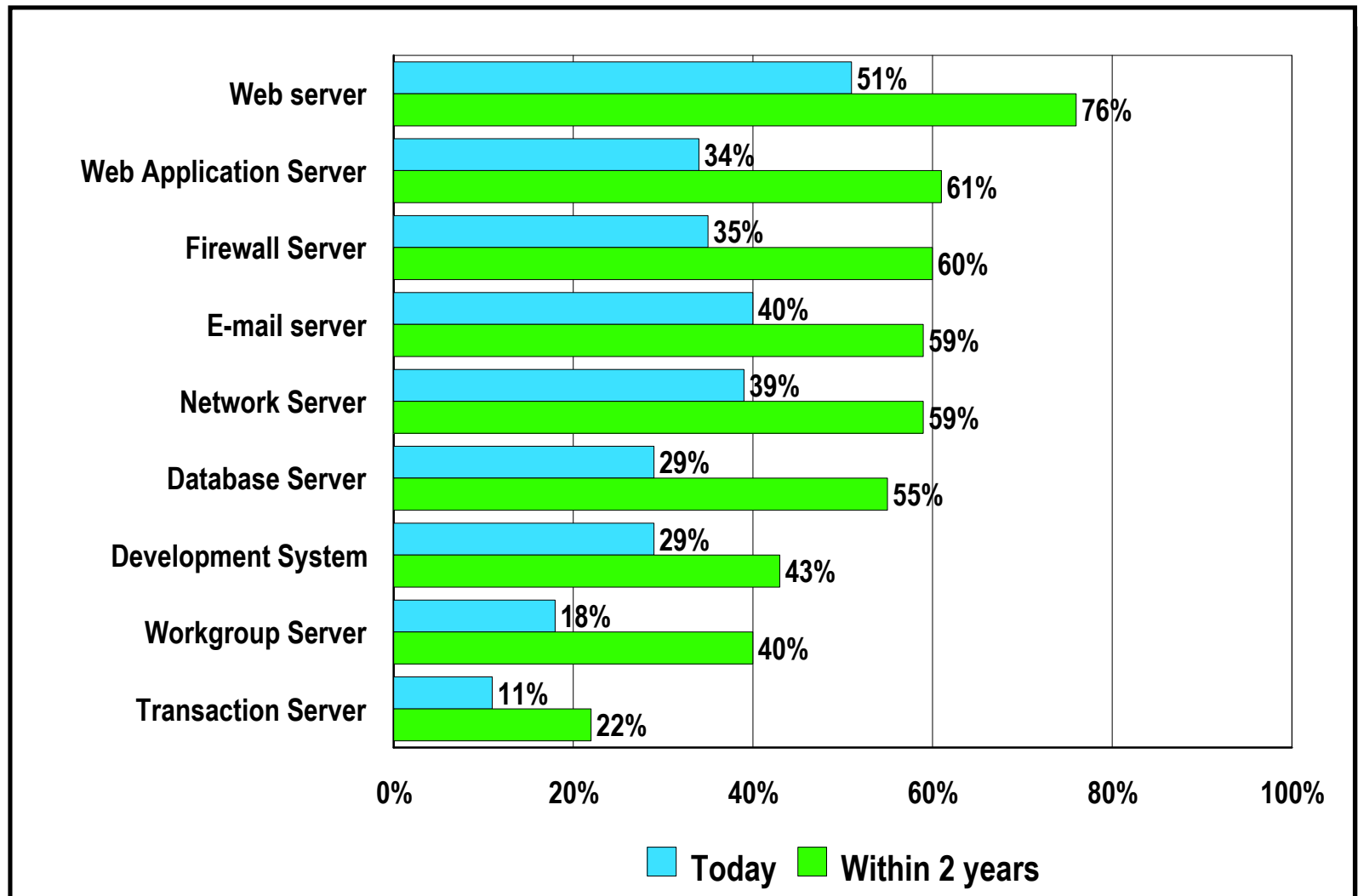 - Worldwide language support, standards based (e.g. LSB 1.1, LI18NUX)

Linux Value

Cost Effective

Reliability

Performance

Open Standards

0%　20%　40%　60%　80%　100%

"It's going to be almost 30 times cheaper to run and maintain" (than Sun systems)

Josh Levine
CTO, e*Trade

Linux Application Deployment

Linux Application Deployment bar chart showing "Today" (blue) and "Within 2 years" (green) deployment percentages:

Application	Today	Within 2 years
Web server	51%	76%
Web Application Server	34%	61%
Firewall Server	35%	60%
E-mail server	40%	59%
Network Server	39%	59%
Database Server	29%	55%
Development System	29%	43%
Workgroup Server	18%	40%
Transaction Server	11%	22%

IBM Market Research
February 2001

Linux: Myth or Reality

- ■ *Myth #1: Linux is "Not Ready for Primetime"*

  - ► *Fact: Linux is rock solid in many mission critical applications*
    - – *Internet Banking: Banco Mercantil*
    - – *TelCo: Telia*
    - – *Retail: Lawson*

- ■ *Myth #2: Linux does not scale*

  - ► *Fact: Linux sets records in horizontal scalability, rapidly improving SMP*
    - – *Seismic: 1024-way xSeries Cluster at Shell*

- ■ *Myth #3: Linux lacks Business Applications*

  - ► *Fact: ISV's adopting Linux.*
    - – *2400+ Linux apps in IBM Global Solutions Directory: www.software.ibm.com/soltions/isv*
    - – *#1 site on developerworks portal, Linux zone=400k hits per month: www.ibm.com/developerworks/linux*

- ■ *Myth #4: Linux is not secure*

  - ► *Fact: Linux is architected for Security; Open Source Development enhances it.*

IBM

Linux Factoids- Industry Traction

- In 2001, IBM invested $1 billion in hardware, software and services related to Linux. IBM recouped most of that investment by the end of 2001.

- IBM is now engaged with more than 2,500 Linux customers worldwide

- 120,000 competitive servers displaced by Linux in 2001, with solutions ranging from web serving to some of the largest supercomputers in the world doing genomic analysis

- IBM is committed to using Linux inside IBM with more than 1000 servers running Linux, including the IBM website, www.ibm.com.

- 11% of the MIPS we shipped on the mainframe in 2001 were based on Linux.

- 5,000+ IBMers worldwide working in porting centers, research, services and development labs dedicated with Linux

- $300 million investment in Global Services infrastructure to educate, design, implement, operate and maintain Linux systems in the enterprise

- Over 2,800 ISV applications enabled for Linux

IBM

IBM's Commitment to Linux - Hardware & Software

xSeries
- 90-day no-charge start-up support
- from 4 major distributions
- Pre-install available
- 125+ ServerProven applications

zSeries
- Runs native, LPAR, on VM or new VIF

Thinkpad
- Certification on all models
- Pre-install available

pSeries
- Selected models, LPAR

Desktop
- Select models certified

iSeries (SOD)
- LPAR

Intellistation
- All 2D models certified

Linux Cluster
- e1300

Network Station
- Red Hat toolkit on web for 2200 & 2800

IBM MQSeries 5.2

NetVista
- Internet appliance device

IBM DB2 UDB 7.1
- EEE (01/01)
- EE (Intel and /390 based servers)

IBM Java Virtual Machine 1.1.8
- All 4 distributions

Lotus Domino 5.03
- All 4 distributions

IBM VisualAge for Java Pro
- with JDK 1.3

IBM WebSphere Commerce Suite (Start V 4.1)

IBM WebSphere Application Server 3.5
- AE on Intel and /390 based servers; SE on Intel based servers
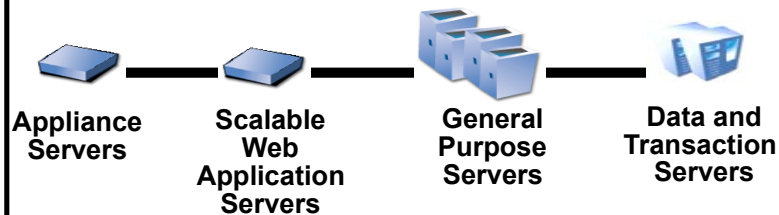- HomePage Builder V4

IBM

Industry's Broadest Linux Server Line

Linux for IBM @server xSeries
The Point of Entry - Where Industry Standards Meet Enterprise Capabilities

Appliance Servers — Scalable Web Application Servers — General Purpose Servers — Data and Transaction Servers

xSeries 100
- *Web servers*
- *NAS servers*

xSeries 300, 330, 340 and 342
- *Rack optimised*
- *Clusters*

xSeries 250
xSeries 350
xSeries 370
xSeries 380
- *X-architecture*
- *Clusters*

xSeries 200s
- *Price/performance*
- *High Availability*
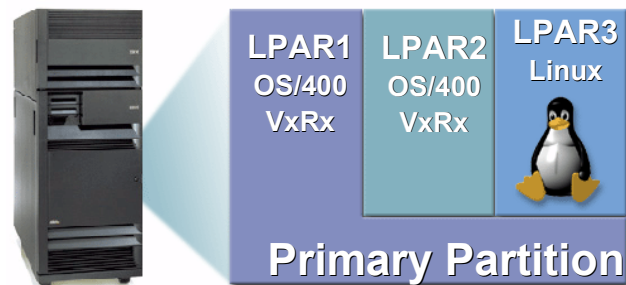
Linux for IBM @server pSeries & RS/6000
The Point of Integration - Where Linux Meets UNIX

- **Native Linux for RS/6000 (32-bit)**
- **Native Linux for pSeries (64-bit)**
- **Exploit Power3 / Power4 Floating Point, 64-Bit Performance, I/O Bandwidth and RAS**
- **AIX Toolbox for Linux Applications in AIX 5L**

AIX
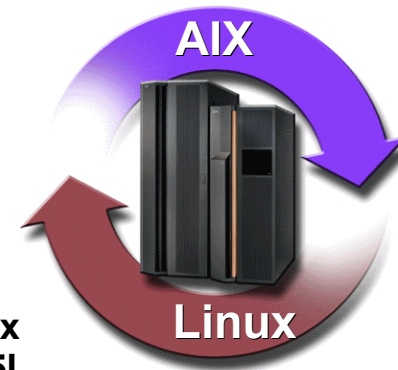
Linux

Linux for IBM @server iSeries
The Point of Coexistence - Where Linux Complements Integrated e-business Solutions

LPAR1 OS/400 VxRx | LPAR2 OS/400 VxRx | LPAR3 Linux
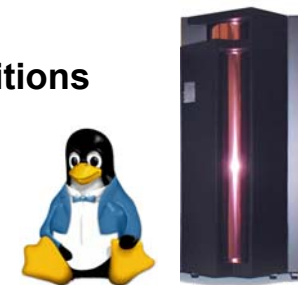
Primary Partition

- **Linux in a partition**
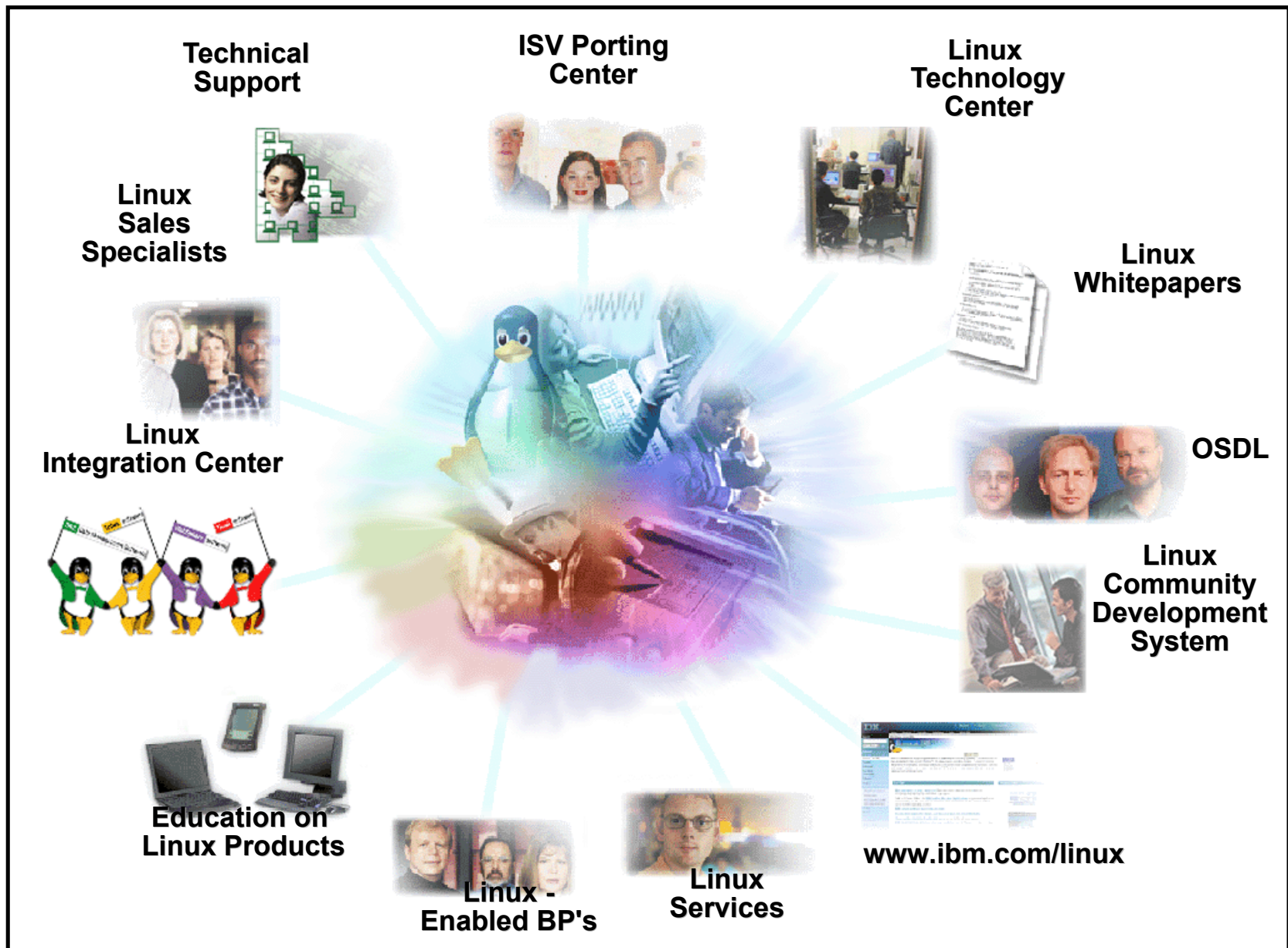- **Integrates new e-business applications**

Linux for IBM @server zSeries and S/390
The point of consolidation - Linux Ascends to the Mainframe

- **Pure Linux OS**
- **Exploits zSeries hardware**
- **Scalable, protected partitions**
- **Shared infrastructure**
- **Reduced total cost of ownership**

IBM

IBM's Commitment to Linux

Technical Support

ISV Porting Center

Linux Technology Center

Linux Sales Specialists

Linux Whitepapers

Linux Integration Center

OSDL

Linux Community Development System

Education on Linux Products

Linux - Enabled BP's

Linux Services

www.ibm.com/linux

Linux Value Proposition for Higher Education

Customer Pains

- **Distinquishing Institution**
- **Reduced budgets**
- **Attract and retain key personnel**
- **Maintenance/management of heterogenous systems**
- **Soaring costs and real estate of growing system capacity**
- **Universal student access**

Value Proposition

- **Lower total cost of ownership**
- **More available compute power**
- **Scalability and reliability**
- **Flexibility and openness**
- **Leading edge technologies**

IBM Solution Elements

- **eServer**
  - ▶ **xSeries and IntelliStation**
  - ▶ **zSeries**
  - ▶ **pSeries**
  - ▶ **iSeries**
- **IGS**
  - ▶ **Migration services from SUN**
  - ▶ **Linux Systems Management**
  - ▶ **Linux Cluster Management**

Target Customers

- **Researchers**
- **Administrative depts**
- **Academic depts**
- **Medical, Bio, GRID Research**

Target ISVs

- **Blackboard,WebCT**
- **Oracle, Peoplesoft,SAP**
- **ExLibris, Ameritec,Endeavor**
- **Datatel, SCT, Jenzabar**

Proof Points

- **See following pages of University references & users**
  - ▶ **Clusters**
  - ▶ **zSeries**

IBM

What are universities doing w/Linux?

Customer References: ibm.com/linux

Teaching

Daito Bunka University, Japan

Established a large classroom system of Linux/Window2000 double boot system- 1000 IBM Thinkpads, 14 Netfinity 5500s, 2 xSeries 240 servers. File, mail, and Web servers for the universities 14,000 students

Colorado State University
Knowledge to Go Places

In a partnership w/ IGS, Colorado state uses the S/390 VM/Linux environment to provide virtual Linux instances for students to learn and also enables other universities to use the system for Linux research projects

MARIST COLLEGE

One of the first schools to integrate technology into the classroom and wire dorms for internet access, Marist uses the S/390 VM/Linux environment give each computer science student their own virtual Linux system. Marist is also working on a videostreaming capability to the digital library on S/390 and has a Linux lab set up.

Indian Institute of Technology

IIT,Kharagpur uses RS/6000 workstations with AIX and Linux for a computing facility used by students and faculty for writing and compiling programs across computer science, mathematics, geology, and mechanical departments.

KYOTO SANGYO UNIVERSITY

Kyoto Sangyo University in Japan uses Netfinity servers and xSeries 330 servers to create an 'intelligent space' in the classroom for multiple applications. A dual boot of Linux and Windows NT provides a maintenance friendly, scalable and reliable platform for computer use in the classrooms.

Administrative

Beijing University

One of the largest universities in China uses the S/390 and Linux for a popular BBS bulletin board application to support students

Tamkang University Taiwan

TKU consolidated mainframe administrative and e-mail systems use Websphere Application Server on its Multiprise 3000 Linux to serve the universities' 29,000 staff and students.

IBM

What are universities doing w/Linux?

Popular/typical Linux Applications

Research

MAUI HIGH PERFORMANCE COMPUTING CENTER

The approximately 700 users at the Maui Center have a 256 dual-processor node xSeries cluster to perform military research such as modeling and simulation projects while other reseachers use help predict and mitigate natural disasters such as hurricanes, wildfires, and tsunamis

University of Oxford

University of Oxford uses xSeries and datastore for the High Energy Physics GRID infrastructure.

PENN

The Universities National Scalable Cluster Lab is building a computer GRID for breast cancer diagnosis and screening. The purpose of the GRID is to ensure that every patient's vital data is provided to authorized physicians quickly, efficiently, and securely

Mississippi State UNIVERSITY

Engineering Research Center(ERC) expands its Linux Cluster by adding 128 nodes xSeries 330 servers using XCAT for systems management and internal Computational Fluid Dynamics codes

NCSA

The National Center for Supercomputing at the University of Illinois-Urbana-Champaign provides a supercomputing infrastructure for scientific and engineering community in the US. Using 512 x300 thin servers and a second cluster with 160 IntelliStations, NCSA is one of the first high-performance computing centers to move aggressively toward parallel computing

NC STATE UNIVERSITY

Using a S/390 Linux environment, NCState performs many research projects including proactive management of security, e-Education, open AFS and others.

Institute of High Performance Computing iHPC

The IHPC in Singapore is an R&D institute that is using Intel based servers and IU xSeries clusters to perform technical computing in fluid dynamics simulation

UNIVERSITY OF Nebraska

The University of Nebraska uses their zSeries Linux server for many research projects including Videostreaming over Internet2 and wireless QOS. Nebraska has also enabled other universities to experiment with a S390 Linux environment on their machine.

IBM

Linux Shared Research/ Joint Study Projects

Additional research in universities using Linux clusters

Boston University's Center for Computational Science and the Office of Information Technology for eBusiness supports the school's projects in deep computing and visualization.

SJSU envisioned ways of using new technology - specifically Linux - to enhance its students' learning environment while expanding the services provided to end users. Choosing Linux for zSeries for the platform's ability to support hundreds of virtual servers on demand for student and faculty projects, the university installed a zSeries Multiprise 300M model H30 server running SuSE Linux with TotalStorage Enterprise Storage Server in a SAN. Once the solution is in production, faculty members will use the platform for administrative and teaching projects, and students will use it as a learning tool. Other universities in the California State University system are expected to follow SJSU's leadership in Linux.

University of California at San Diego is using LInux clusters in support of the Cal-(IT)2 - 1st Testbed in the area of Environmental Sensing. Cal-(IT)2 is a partnership with UCSD and UC at Irvine with the mission to enable anywhere/anytime access to the internet.

The Academic Computing and Dept of Electrical Engineering is providing support for high performance computing projects in Bioinfomatics and for specific projects using IBM's General Parallel File System in a heterogeneous cluster environment.

Rensselear Polytechnic Institute is developing a high performance web site for bioinfomatics. The goal is to extend world leadership in bioinformatics service in RNA secondary structure prediction to the broad spectrum of services covered by the research activities in the Joint Center for Bioinfomatics at RPI.

The Chemistry Department is using a 64-bit Linux processor IBM PowerPC cluster to research the enhancement of General Atomic and Molecular Electronic Structure System (GAMESS) cluster application

In an e-Education partnership with IBM, UNC will focus on research in software technology and applications
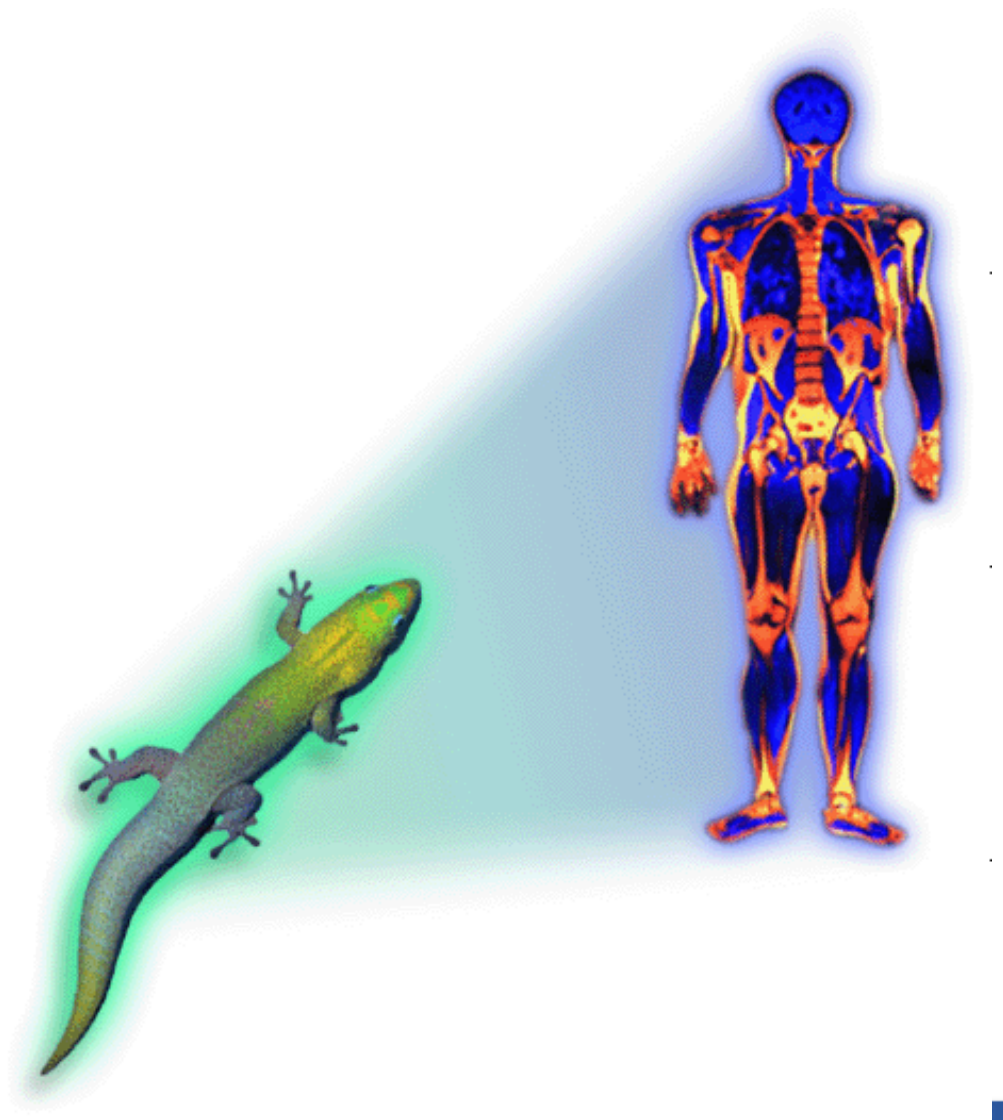
IBM

The Next Big Thing

Autonomic Computing

- **Self-optimizing**

- **Self-configuring**

- **Self-healing**

- **Self-protecting**

Project eLiza™
IBM's blueprint for self-managing systems

IBM @server platforms incorporate self-managing technology

- End-to-end availability and security
- Lower costs for maintaining and deploying systems
  - Self-configuring
  - Self-healing
  - Self-protecting
  - Self-optimizing

Self-managing systems reduce downtime, operating costs and administrative requirements
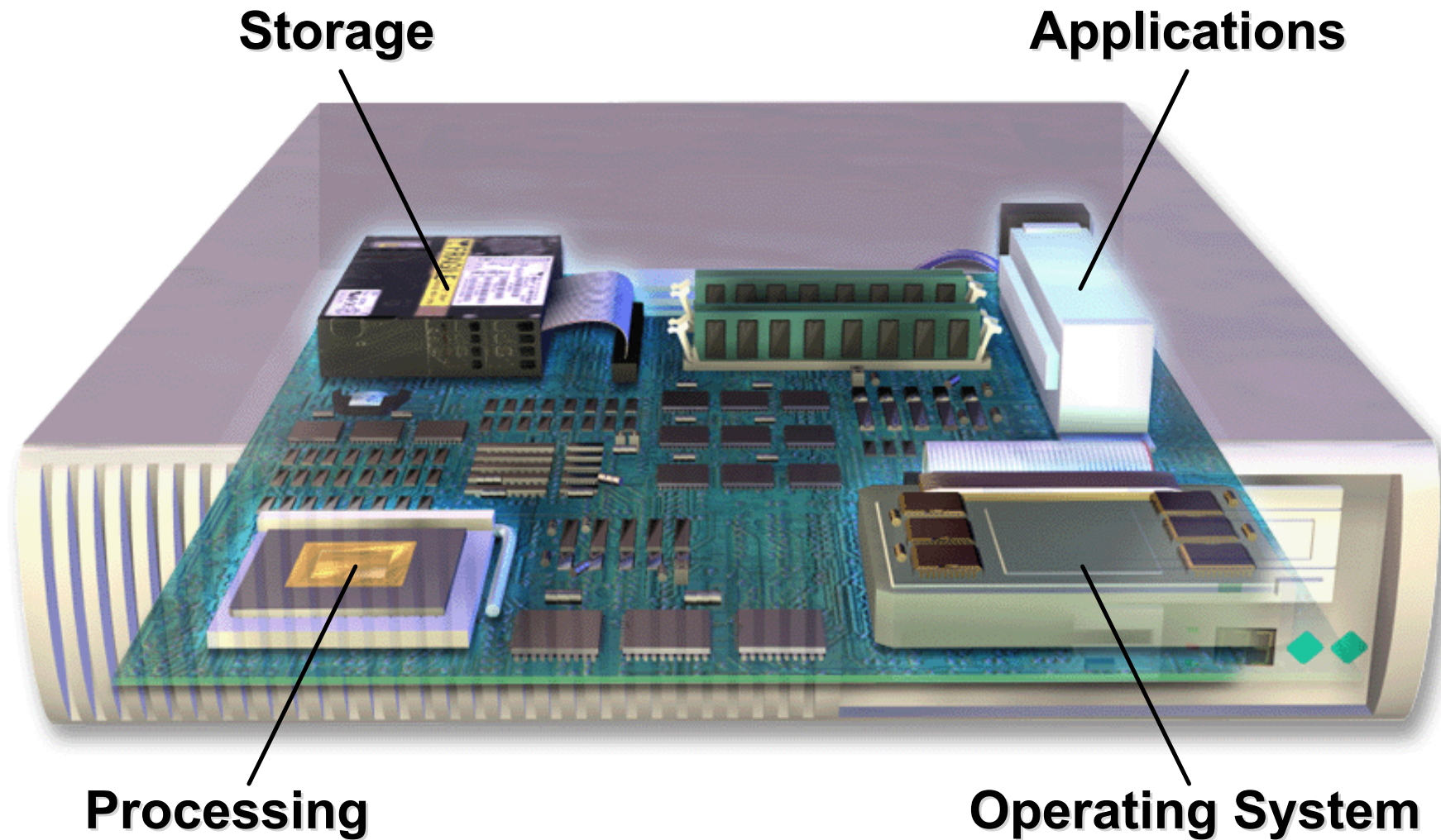
Cost of skills:

- Using existing technology, for every US $1 spent on server HW, US $10 will be spent on people to manage the HW by 2004

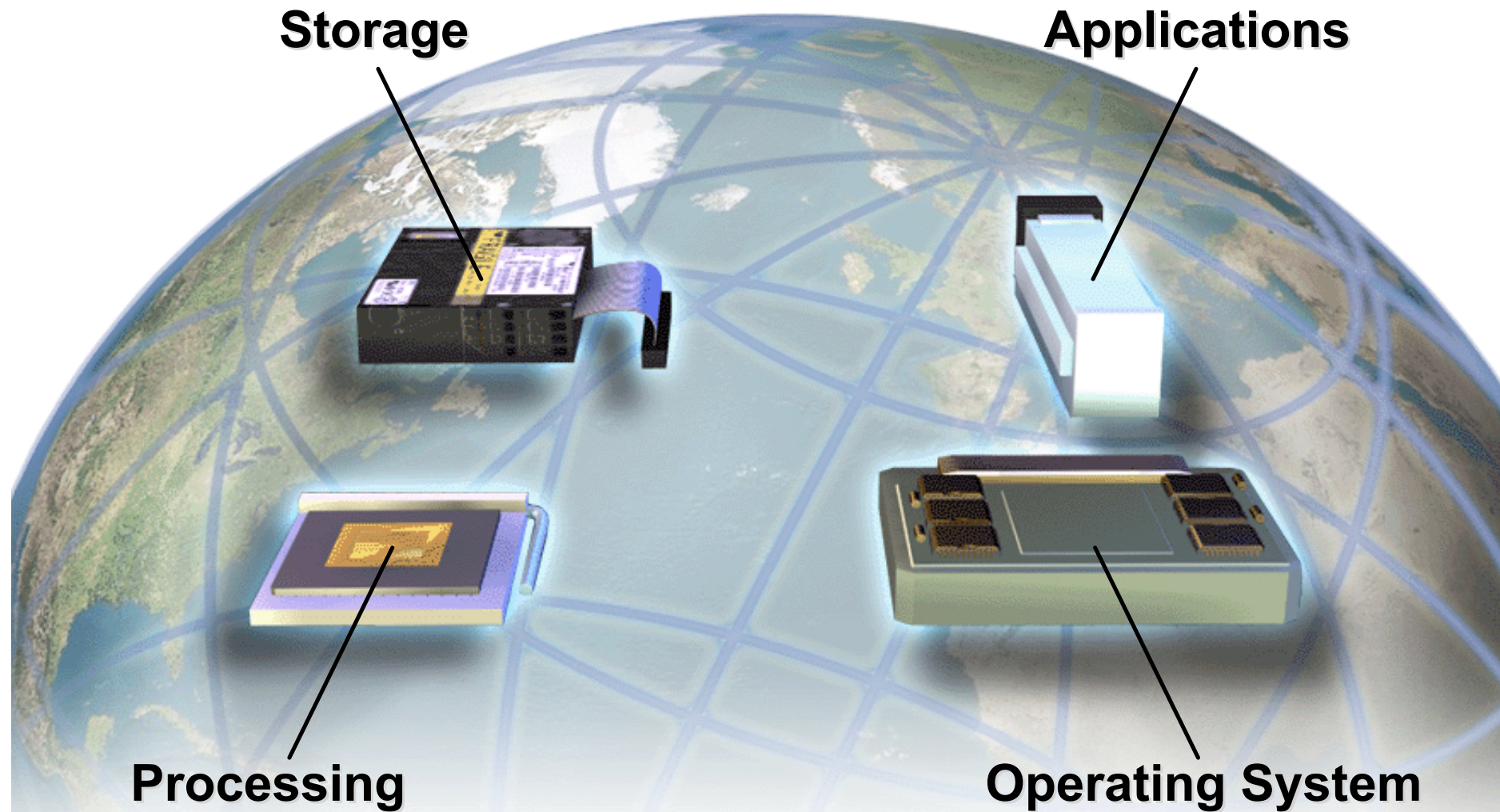**Cost of downtime:**

- US $10,000 per minute for e-commerce

The Standish Group, 2001

IBM

The Next Big Thing
The Internet as a Computing Platform

The Next Big Thing
The Internet as a Computing Platform

Storage

Applications

Processing

Operating System

One virtual computing platform,
'limitless' global resources

The Internet as a Computing Platform

The Grid

- **Shared applications**

- **Shared data**

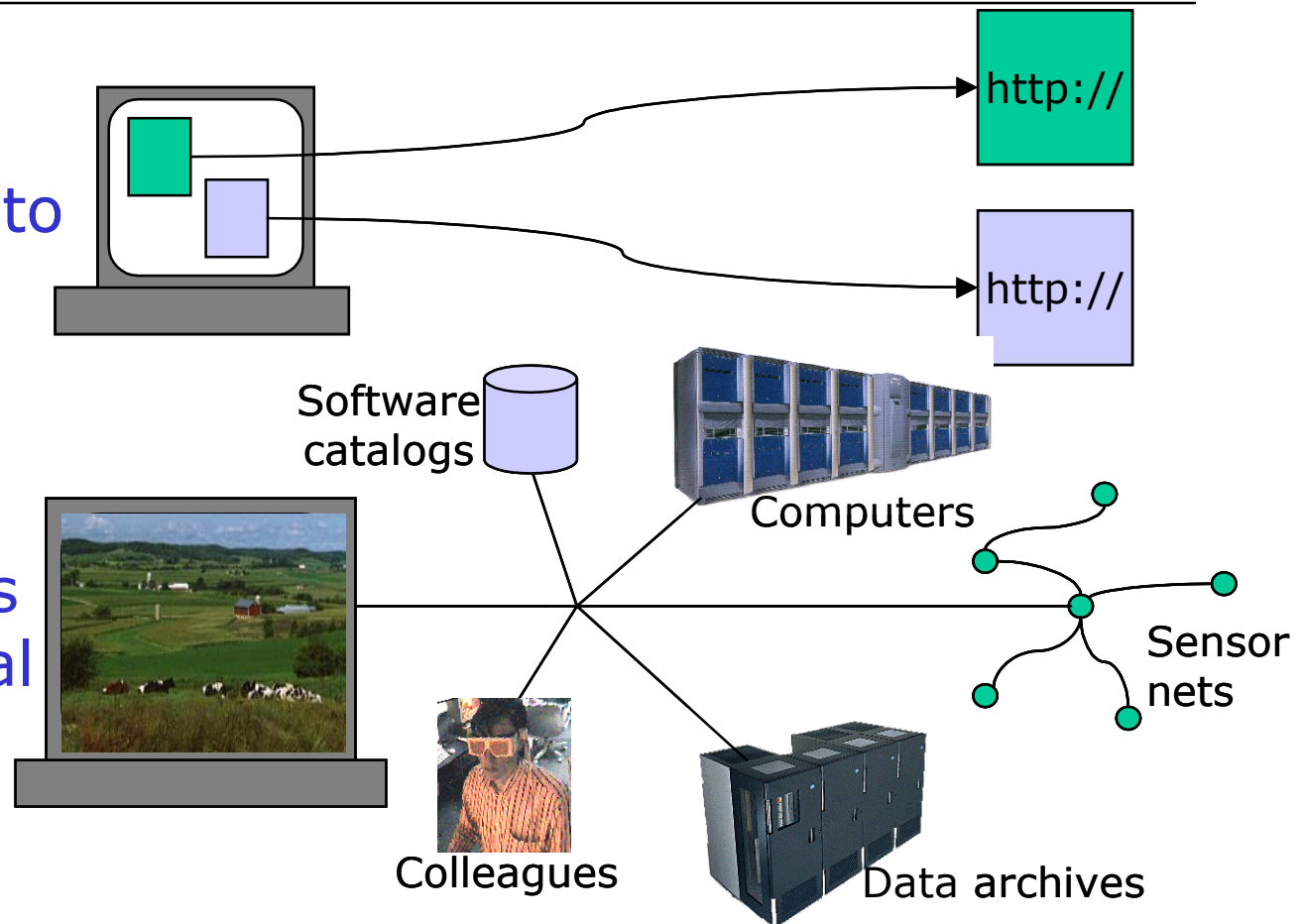- **Shared processing**

- **Shared storage ...**

...Over the Internet

The Grid

What is it?

Web: Uniform naming/access to documents

Grid: Uniform, high-perf access to computational resources

http://

http://

Software catalogs

Computers

Sensor nets

Colleagues

Data archives

On-demand creation of powerful virtual computing systems

A new open-source-based model of computing, Grids are clusters of servers joined together over the Internet, using standard protocols and other open technologies, including Linux. Just as the World Wide Web allows people to share content via standard Internet protocols, Grid computing allows widely-dispersed organizations to share applications, data and resources using emerging Grid protocols.
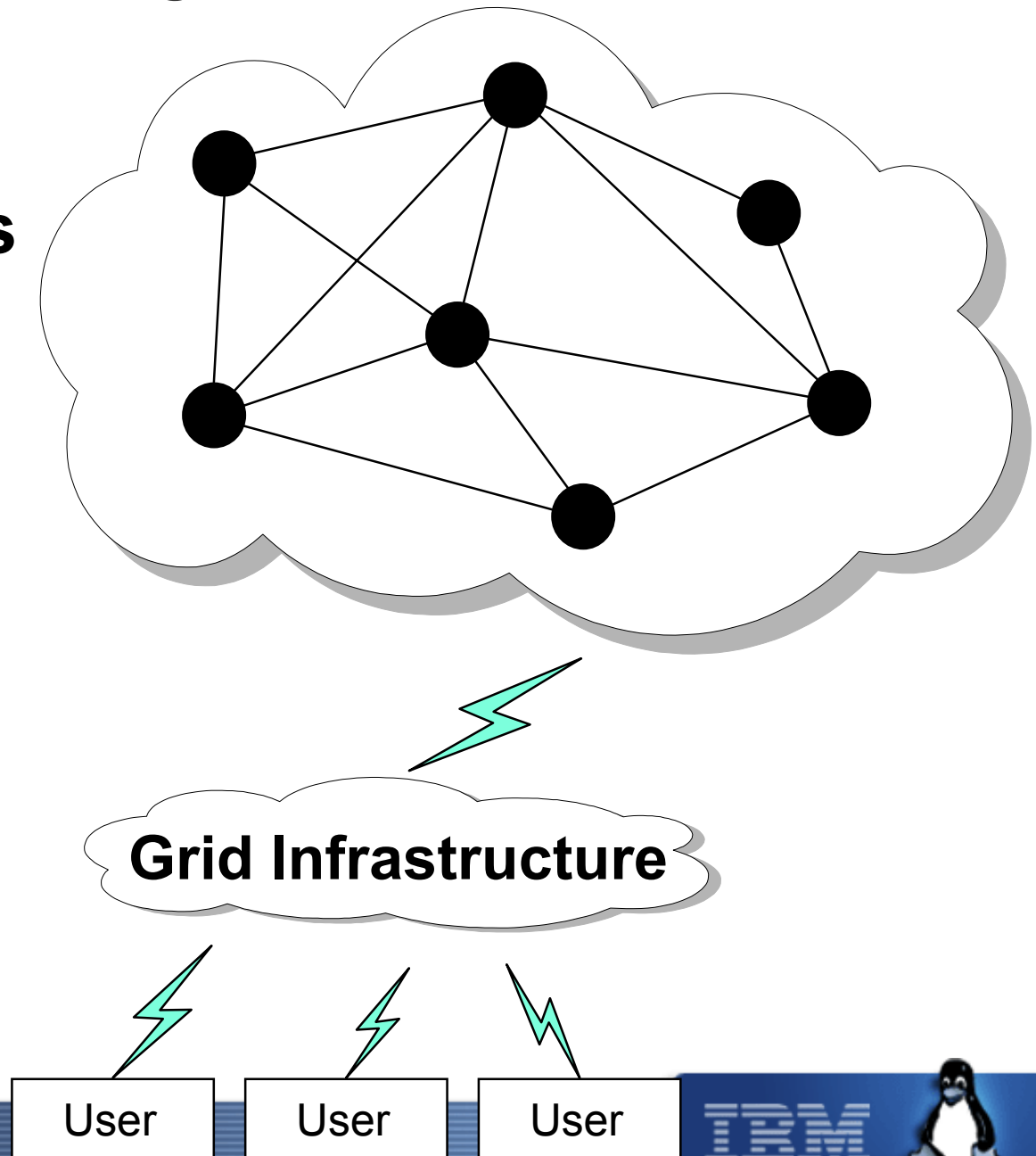
IBM

Grid Computing

Large Scale Resource Sharing

Enables dynamic, virtual organizations

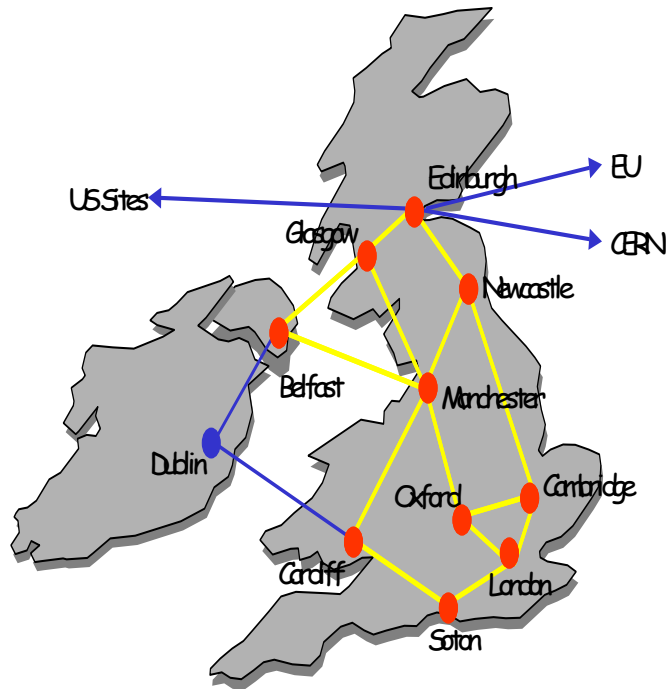**Access to Data**

Access to Computation and Storage Resources

Access to Services

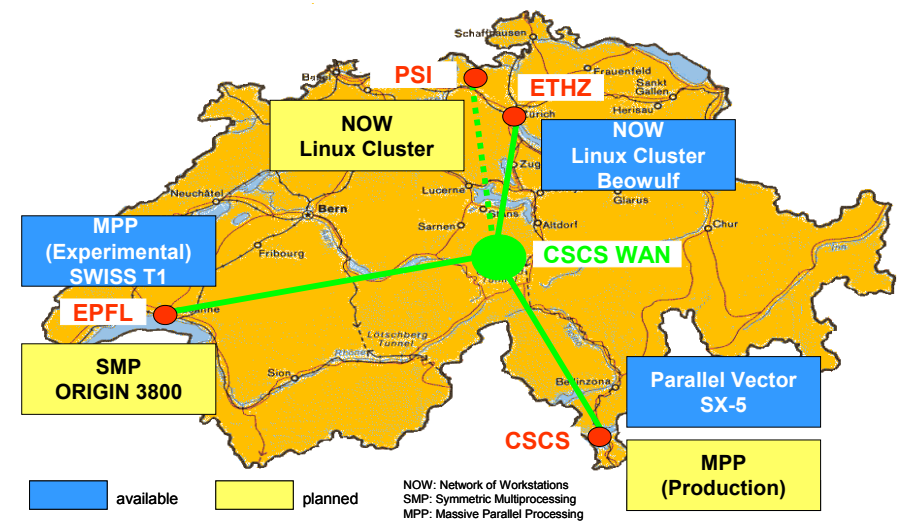**Collaboration between people and organizations**

Grid Infrastructure

User User User

IBM

Research initiatives

Grid in universities and government funded programs

US Sites
EU
Edinburgh
CERN
Glasgow
Newcastle
Belfast
Manchester
Dublin
Oxford
Cambridge
Cardiff
London
Soton

DAS-2 (Dutch National Grid)

PSI
ETHZ
NOW Linux Cluster
NOW Linux Cluster Beowulf
MPP (Experimental) SWISS T1
CSCS WAN
EPFL
SMP ORIGIN 3800
Parallel Vector SX-5
CSCS
MPP (Production)

available   planned

NOW: Network of Workstations
SMP: Symmetric Multiprocessing
MPP: Massive Parallel Processing

Distributed Terascale Facility

INFORMATION POWER GRID
Engineering and Research Site

Data GRID

IBM

GRID Computing

Benefits

- ▶ **Resource aggregation** -- allows users to treat geographically dispersed systems as one virtual computer with efficient resource management.

- ▶ **Database-sharing** -- allows access to any remote database within a Grid. This is particularly useful for life sciences firms that need to share human genome data with other firms. Engineering and financial firms could also benefit significantly.

- ▶ **Collaboration** -- allows widely dispersed organizations to work together on a project -- sharing everything from engineering blueprints to software applications.

- ▶ The ability to form virtual organizations over the Internet that collaborate on common problems, by enabling them to share applications and data over the Internet, much as they would if they were sharing a single (virtual) computer.

- ▶ The ability to **tackle very large problems demanding huge computing resources**, by enabling the aggregation of computing power, storage and other resources over the Internet, such as problems in the scientific and engineering worlds where a single application might require teraflops of power and petabytes of storage, or problems in the e-business world where an application might be accessed by many millions of people around the world.

- ▶ The ability to **lower the total cost of computing**, by enabling the sharing, efficient optimization and overall management of those computing resources over the Internet.

- ■ **What is IBM's involvement in Grid Initiatives?**
  - ▶**IBM Computing GRIDs**
    - ● **Distributed Teragrid Facility**
    - ● **UK national GRID**
    - ● **Netherlands GRID project**
    - ● **MDS Proteomics, MCNC**
  - ▶**IBM Data GRIDs**
    - ● **North Carolina BioGrid (Genomics)**
    - ● **University of Pennsylvania (Mammography)**
    - ● **San Diego Super Computer Center (Brain Mapping)**
  - ▶**IBM Web/Application Services GRIDs**
    - ● **Galileo**
    - ● **Storebrand ASA**
    - ● **Hewitt Associates**
    - ● **Bekins**

- ■ **For more information: *ibm.com/linux/grid***

Start with Grid Computing today

- **How to start a Grid Project today ?**
  - ► **Improve utilisation of your servers**
    - **Analyse utilisation**
    - **Load Balancing**
  - ► **Implement a Computing Portal**
    - **"Check-In" to heterogenous Compute Infrastructure**
    - **Improves Management of Server Infrastructure**

- **IBM Grid Innovation Center**
  - ► **first and only Grid Technology Center WW**
  - ► **opened in Montpellier in April 2002**
  - ► **Solutions of Industry Partners available: Globus Project, Avaki, Platform Computing, Nice, Science & Computing, Unicore, etc...**

The Internet as a Computing Platform

e-Business on Demand

- Access new capability more quickly

- Better performance

- Reduce up-front investment

- Gain expertise not available internally

E-BUSINESS ON DEMAND™
THE NEXT UTILITY™

IBM