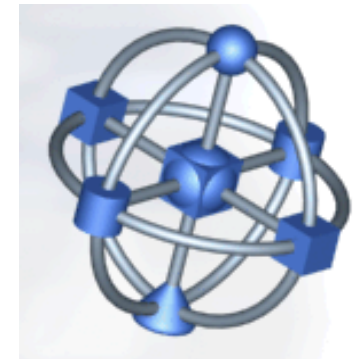# Industrial Grid Applications

Jean-Pierre Prost (jpprost@fr.ibm.com)
IBM EMEA Design Center for ebusiness on demand
Montpellier, France

Agenda

- **IBM Grid Offerings**
- **Case Studies**
  - eDiaMoND (life sciences)
  - IBM LabGrid (engineering)
  - Grid@Shell (petroleum)
  - Financial Services Company (finance)
- **Other References**

IBM Grid Focus Areas

	Research & Development Grids	Engineering & Design Grids	Business Analytics Grids	Enterprise Optimization Grids	Government Development Grids
Description	Accelerate and enhance the R&D process by enabling the sharing data and computing power seamlessly for research intensive applications	Share data and computing power, for computing intensive engineering and scientific applications, to accelerate product design	Enable faster and more comprehensive business planning and analysis through the sharing of data and computing power	Optimize computing and data assets to improve utilization, efficiency and business continuity	Create large-scale IT infrastructures to drive economic development and/or enable new collaborative government services

Productivity	Primary Buyers: LOB decision makers - Secondary Buyers: IT as an enabler
Efficiency	Primary Buyers: IT decision makers - Secondary Buyers: LOB influencers

IBM Grid Offerings

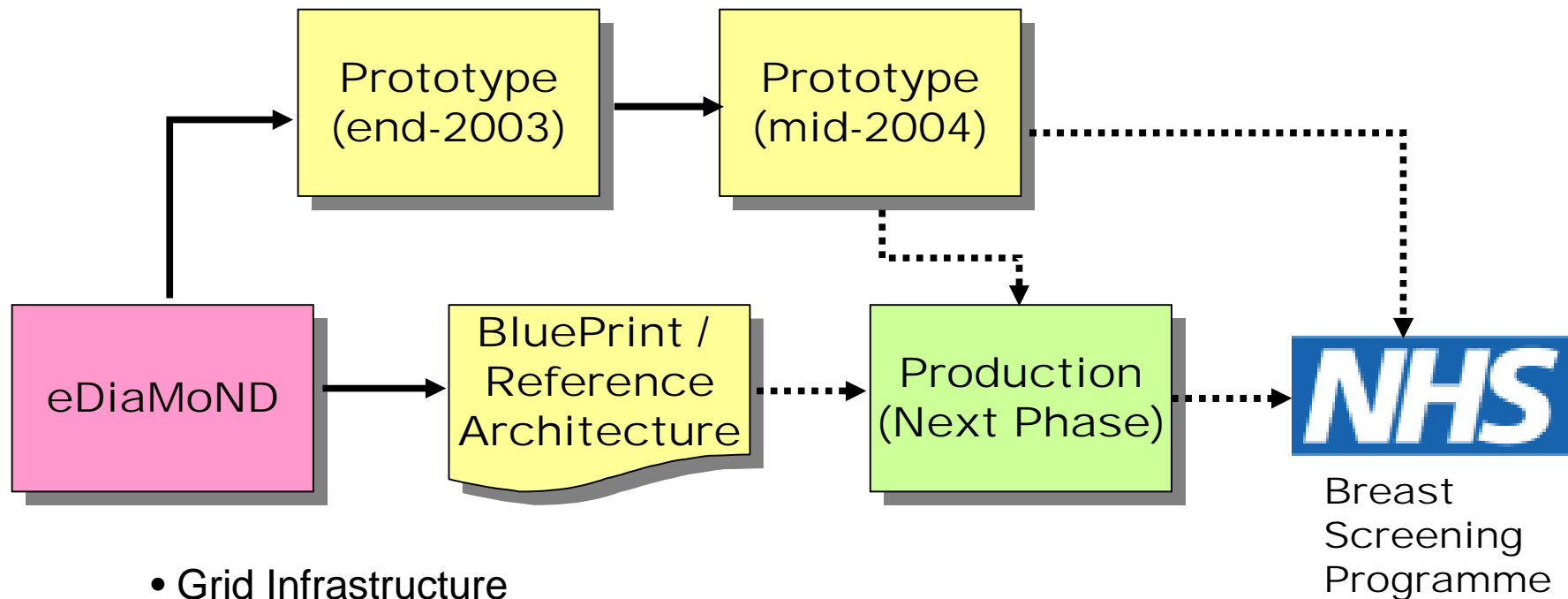| Research & Development	Engineering & Design	Business Analytics	Enterprise Optimization	Government Development
•**Life Sciences: IBM Grid Offering for Information Accessibility** •Higher Education: IBM Grid Offering for University Research Collaboration •Agricultural Chemical: IBM Grid Offering for Information Access •Grid Innovation Workshop/Modules	▪Aerospace: IBM Grid Offering for Engineering Design ▪Aerospace: IBM Grid Offering for Design Collaboration ▪Automotive: IBM Grid Offering for Design Collaboration ▪Automotive: IBM Grid Offering for Engineering Design ▪**Electronics: IBM Grid Offering for Engineering Design** ▪Electronics: IBM Grid Offering for Design Collaboration •Grid Innovation Workshop/Modules	▪**Financial Services: IBM Grid Offering for Analytics Acceleration** •Life Sciences: IBM Grid Offering for Analytics Acceleration •**Petroleum: IBM Grid Offering for Geophysical Analysis: Upstream Petroleum** •Agricultural Chemical: IBM Grid Offering for Analytics Acceleration •Grid Innovation Workshop/Modules	▪Financial: IBM Grid Offering for IT Optimization ▪Petroleum: IBM Grid Offering for IT Optimization •Grid Innovation Workshop/Modules	▪Government: IBM Grid Offering for Information Access •Grid Innovation Workshop/ Modules

eDiaMoND (UK)

eDiaMoND – Partners

eDiaMoND – Project Deliverables

```
          ┌──────────────┐      ┌──────────────┐
          │  Prototype   │ ───→ │  Prototype   │
          │  (end-2003)  │      │  (mid-2004)  │
          └──────────────┘      └──────────────┘
┌──────────┐    ┌──────────────┐    ┌──────────────┐    ┌──────┐
│ eDiaMoND │ →  │  BluePrint / │ ⋯→ │  Production  │ ⋯→ │ NHS  │
│          │    │  Reference   │    │ (Next Phase) │    └──────┘
│          │    │ Architecture │    │              │   Breast
└──────────┘    └──────────────┘    └──────────────┘   Screening
                                                        Programme
```
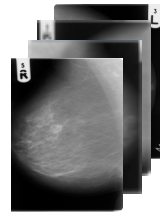
- Grid Infrastructure

- Grid-connected Workstation

- Database for Storage & Retrieval of Images & Metadata

- Computation for CADe, CADi and Statistical Analyses

- Required Hardware, Software & Network for given Service Levels

UK Breast Screening – Today
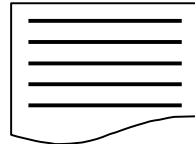
Paper

Film

Began in 1988

Women 50-64
Screened
Every 3 Years
1 View/Breast

~100 Breast
Screening
Programmes
   -Scotland,
   -Wales,
   -Northern Ireland
   -England

1,300,000 - Screened in 2001-02
65,000 - Recalled for Assessment
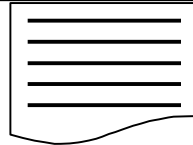8,545 – Cancers detected
300 - Lives per year Saved

230 - Radiologists (Double Reading)   Statistics from NHS Cancer Screening web site
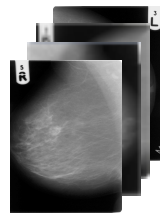
UK Breast Screening – Challenges

Digital

Digital

2,000,000 - Screened every Year
120,000 - Recalled for Assessment
10,000 - Cancers
1,250 - Lives Saved

230 - Radiologists (Double Reading)
50% - Workload Increase

Began in 1988

Women 50-70
Screened
Every 3 Years
2 Views/Breast
+ Demographic
Increase

Scotland,
Wales,
Northern Ireland
England
(8 Regions)

~100 Breast
Screening
Programmes

Scotland

Northern
Ireland

Northern
& Yorkshire

North
West

Trent

West
Midlands

Eastern

Wales

London

South West

South East

UK Breast Screening – Workflow
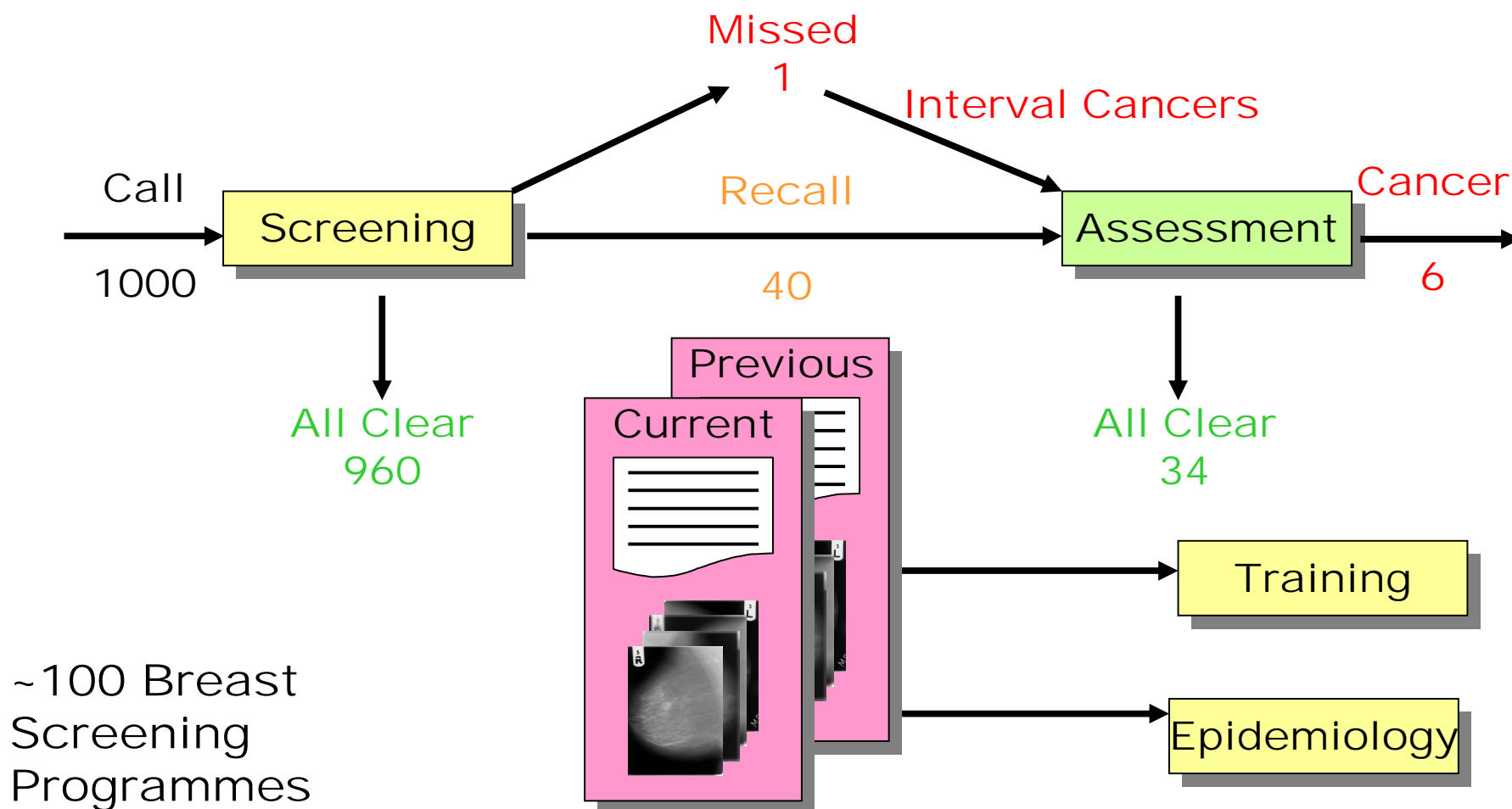
eDiaMoND – Scope

Workstation

Grid

32 MB / Image

256 TB / Year

Screening

Training

Epidemiology

4 Breast
Screening
Programmes

Previous

Current

Data

Compute

Standard
Mammo
Format

CADe
CADi

Data
Mining

CADe=computer aided detection

CADi= computer aided diagnosis
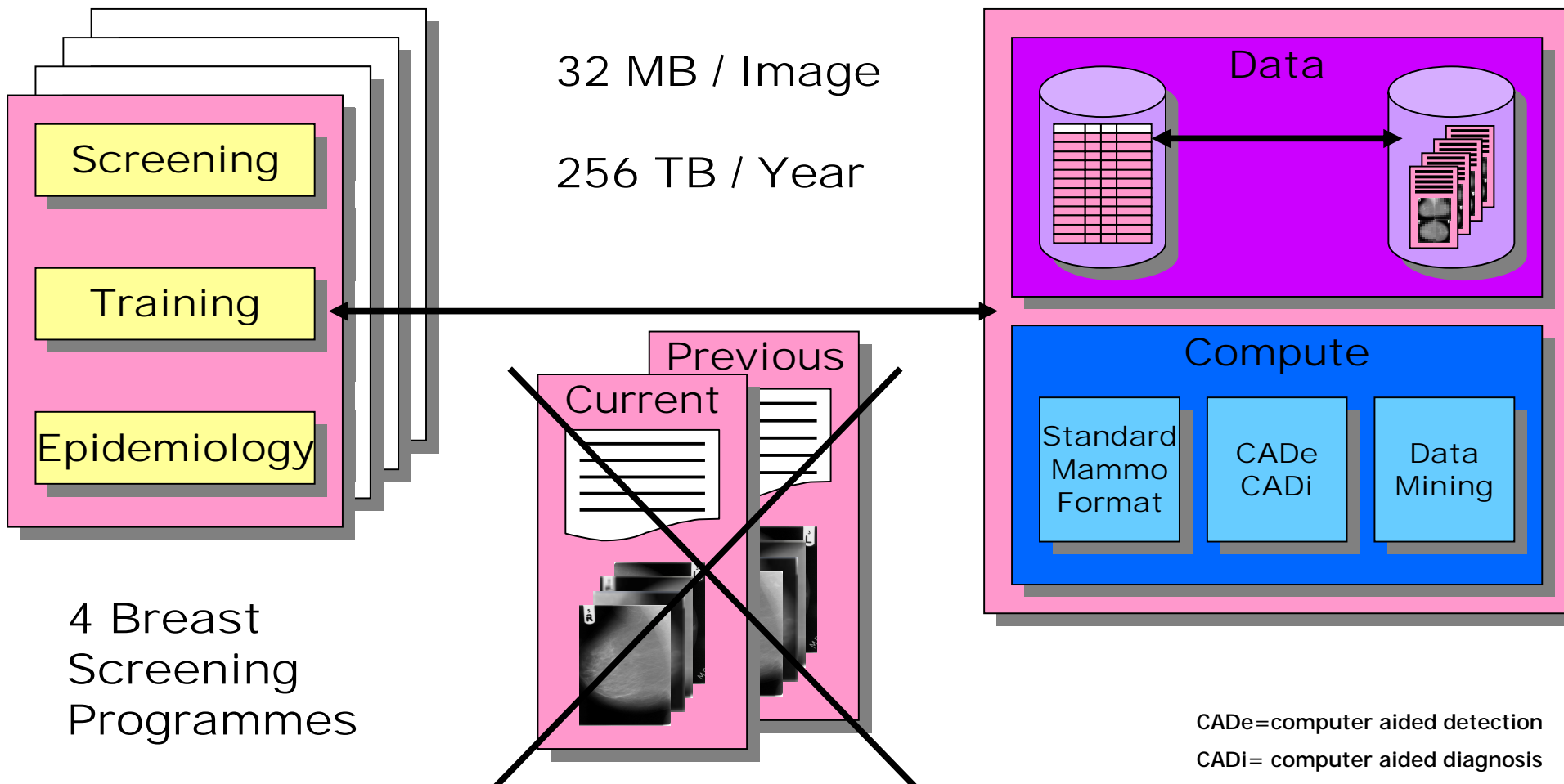
eDiaMoND – Data

Data

Patient	Age	…	Image
107258	55	…	1.dcm
236008	62	…	2.dcm
700266	59	…	3.dcm
895301	58	…	4.dcm
………	…	…	……..
………	…	…	……..
………	…	…	……..
………	…	…	……..
………	…	…	……..
………	…	…	……..
………	…	…	……..
………	…	…	……..

Images

DICOM

DICOM

DICOM

DICOM

Grid

Data
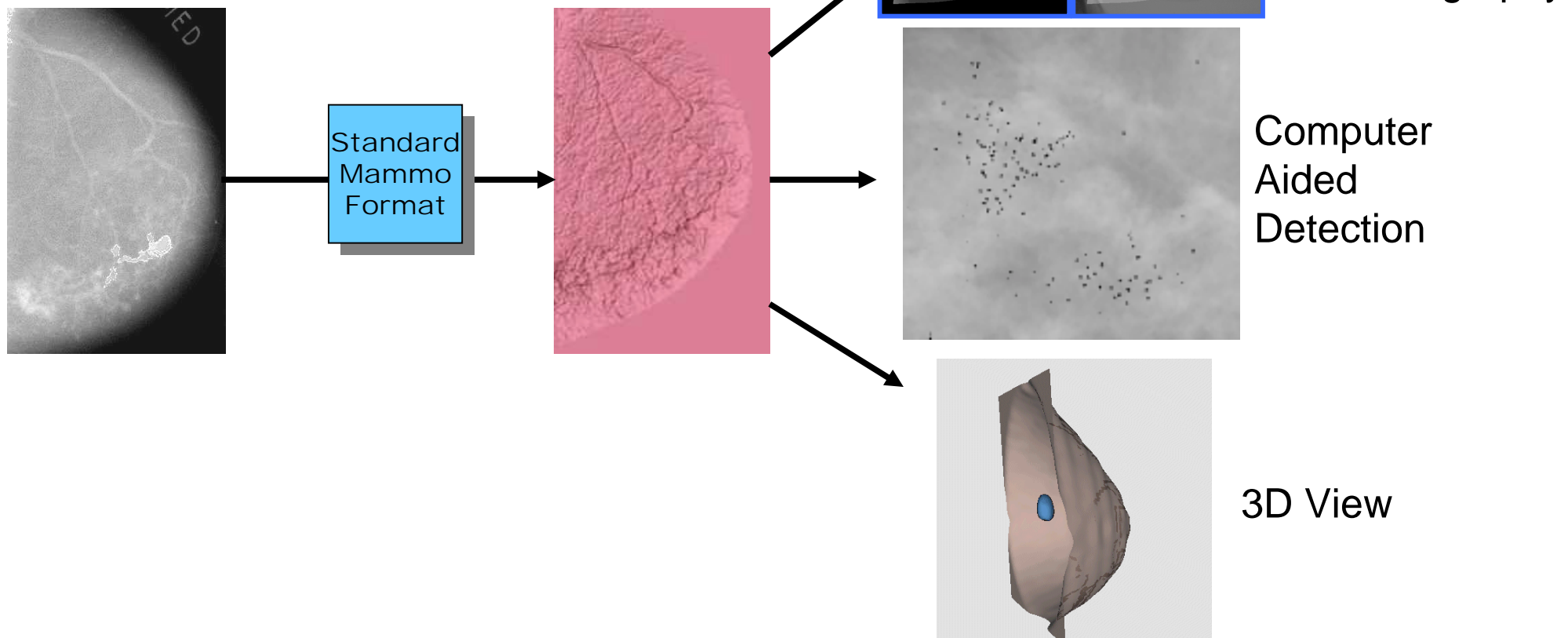
Compute

| Standard Mammo Format | CADe CADi | Data Mining |
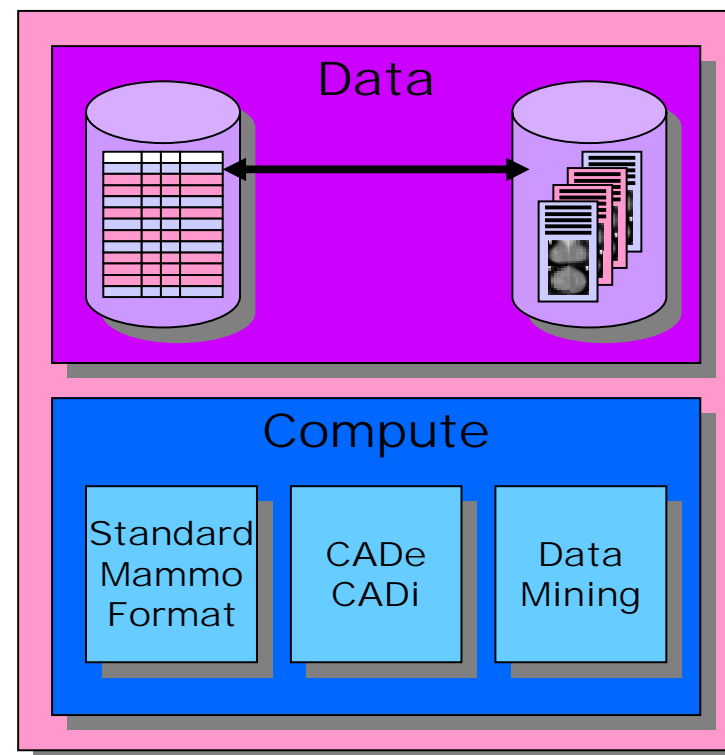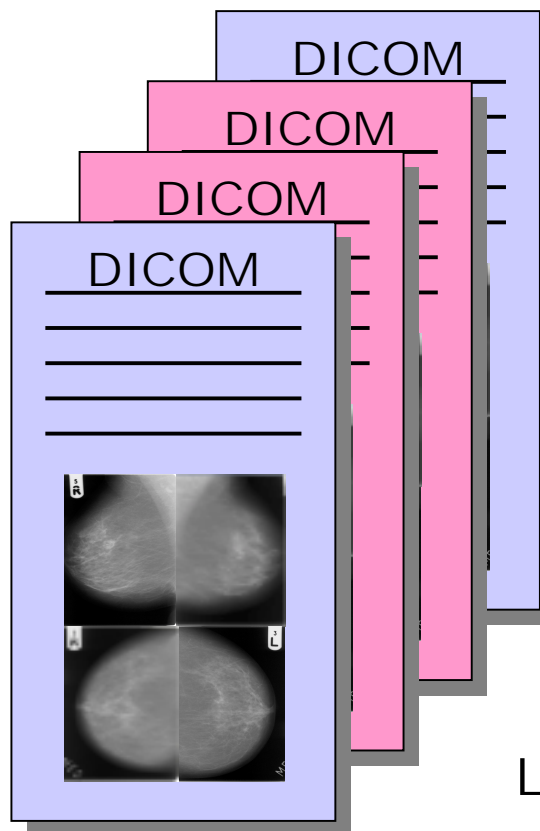
Logical View is One Resource

eDiaMoND – Compute

Mammograms have different appearances, depending on image settings and acquisition systems

Standard Mammo Format

Temporal mammography

Computer Aided Detection

3D View

eDiaMoND – Non-Functional requirements

Screening

Training

Epidemiology

Ethics

Legal

Security

Performance

Scalability

Manageability

Auditability

......

Anonymisation

Lossless Compression

Encryption

256MB & 5 secs response

~100 Centres

Systems Administration

Non-Repudiation

eDiaMoND – Prototype

Development OUCL

OUCL LAN	Oxford / Churchill	Edinburgh	Aberdeen (demo purposes)

OUCL LAN

Oxford LAN

Edinburgh LAN

Aberdeen LAN

OUCL LAN

Oxford LAN

Edinburgh LAN

Aberdeen LAN

JANET Network

IBM LAN

Mirada LAN

UCL LAN

KCL LAN

IBM LAN

Mirada LAN

UCL LAN

KCL LAN

Development (IBM)

elop (Mirada)

UCL / St Georges

KCL / Guys

Grid Boundary
Dev teams

Server

Workstation

T221

eDiaMoND – Blueprint

Clinical Knowledge

eDiaMoND

NDMA Experience

MOMS Experience

Blueprint

Data
- X-Ray
- MRI
- PET
- Ultrasound
- ….…...

Hardware
- Server
- Workstation
- Online Storage
- Offline Storage
- High Resol'n Screen

Standards
- DICOM
- JPEG2000
- XML
- OGSA
- OGSA-DAI

Network
- 1 Gigabit
- 10 Gigabit
- VPN
- ……...

Software
- DB2
- Content Manager
- MQ
- TSM
- Rational Rose

Database Schema
- >3NF Data
- DICOM Files
- Federated

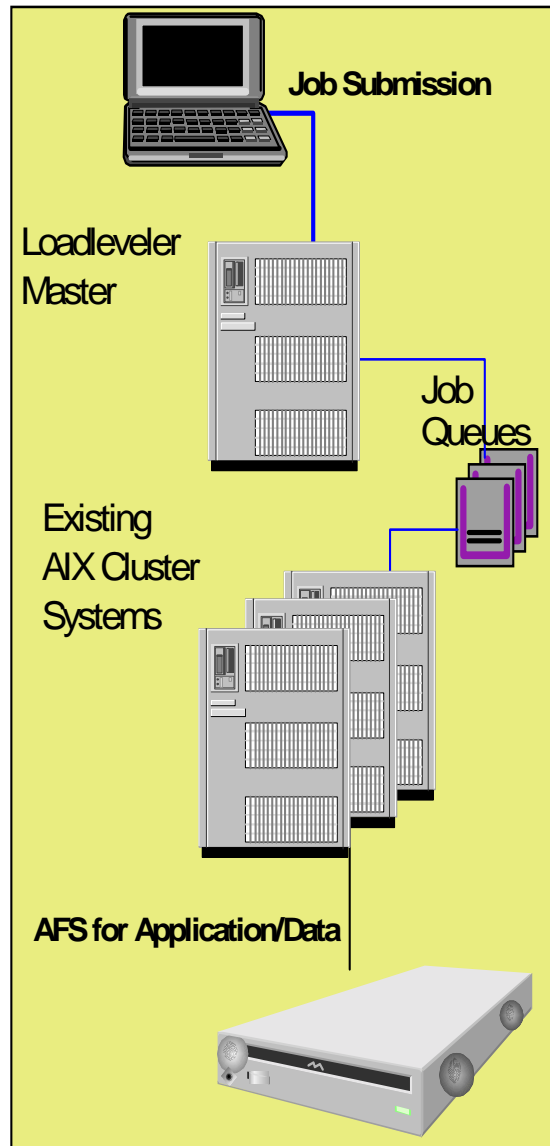# IBM LabGrid (Germany)

IBM LabGrid: Application Description

Grid Application:
Functional Verification of
zSeries Processor Units

- **Random unit simulation, simulation environment/model on AFS (shadowed once a day to participating AFS cells)**

- **Input data: None (Random start seed)**

- **Output data, Testcase passed:**
  Statistics, summary, coverage, ~150 kB

- **Output data, Testcase failed:**
  Statistics, summary, coverage, debug info, ~1.5 MB

- **Testcase runtime determined by clock cycles simulated (130k cycles, <10min runtime, depending on machine)**

IBM LabGrid: Before Grid Usage

Job Submission

Loadleveler
Master

Job
Queues

Existing
AIX Cluster
Systems

AFS for Application/Data

Execution environment is ONE cluster of AIX systems controlled by LoadLeveler
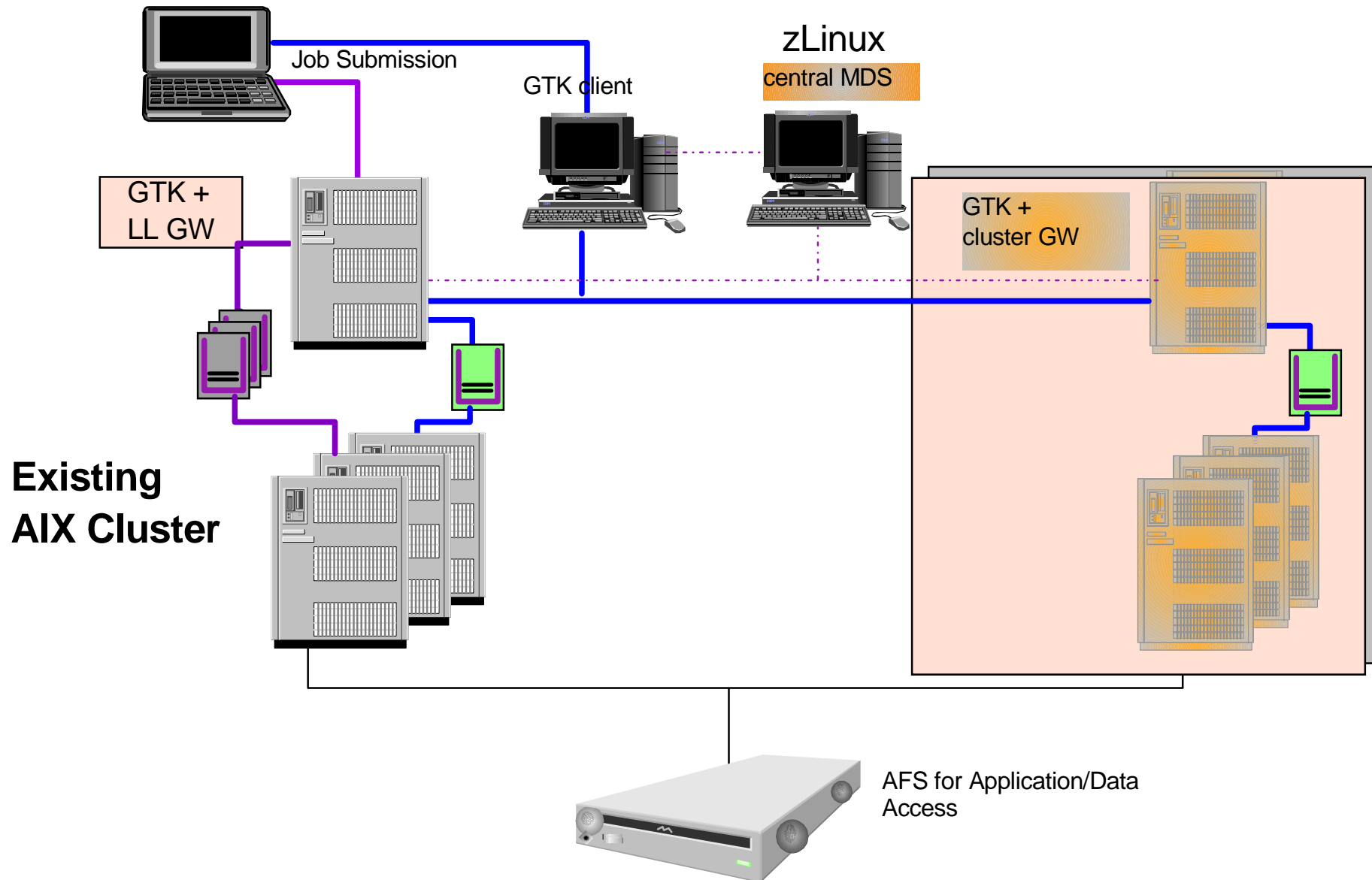
Access to applications and simulation data through AFS

Many different simulation applications

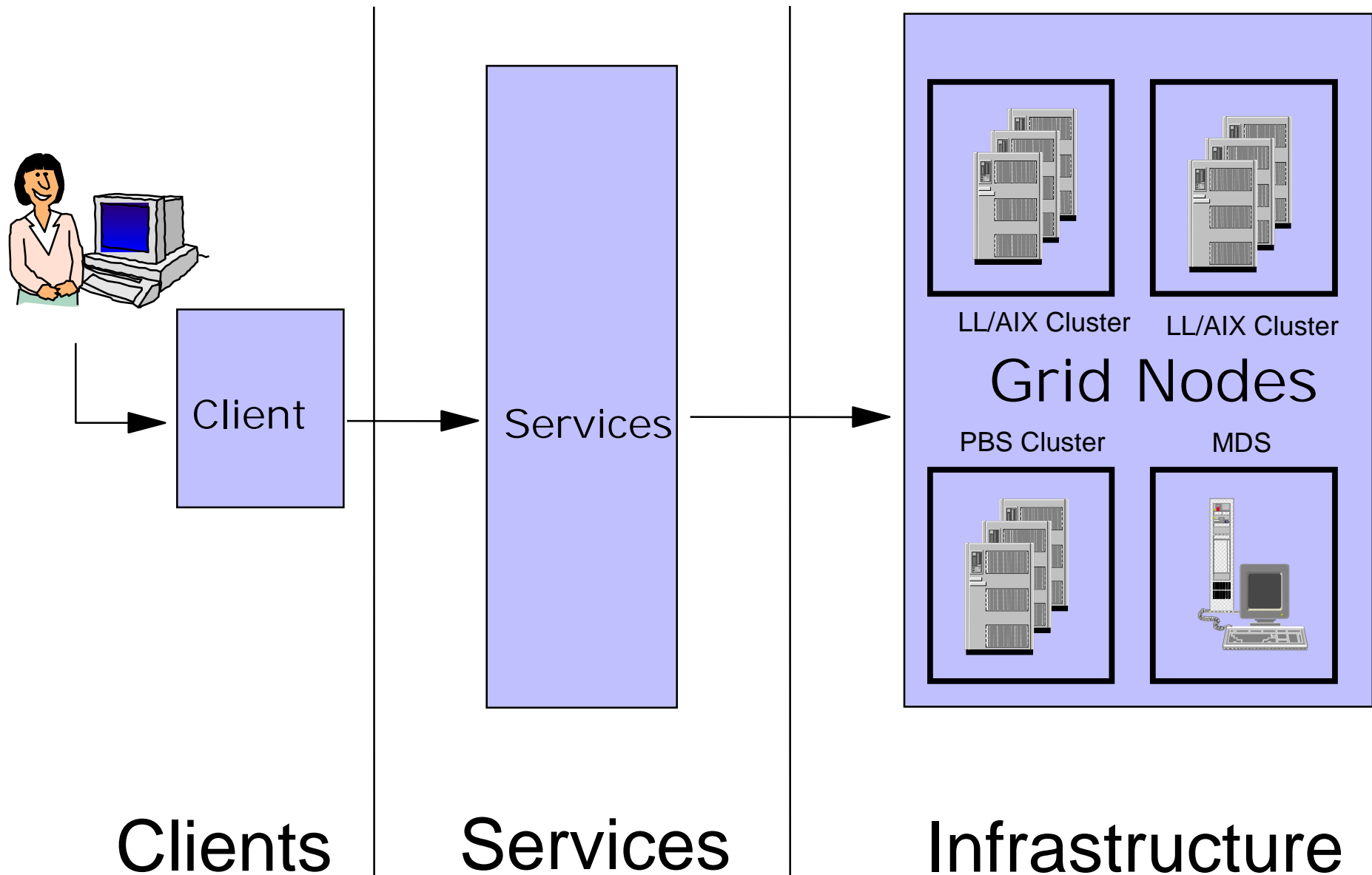**Jobs are submitted by users and automated job generators**
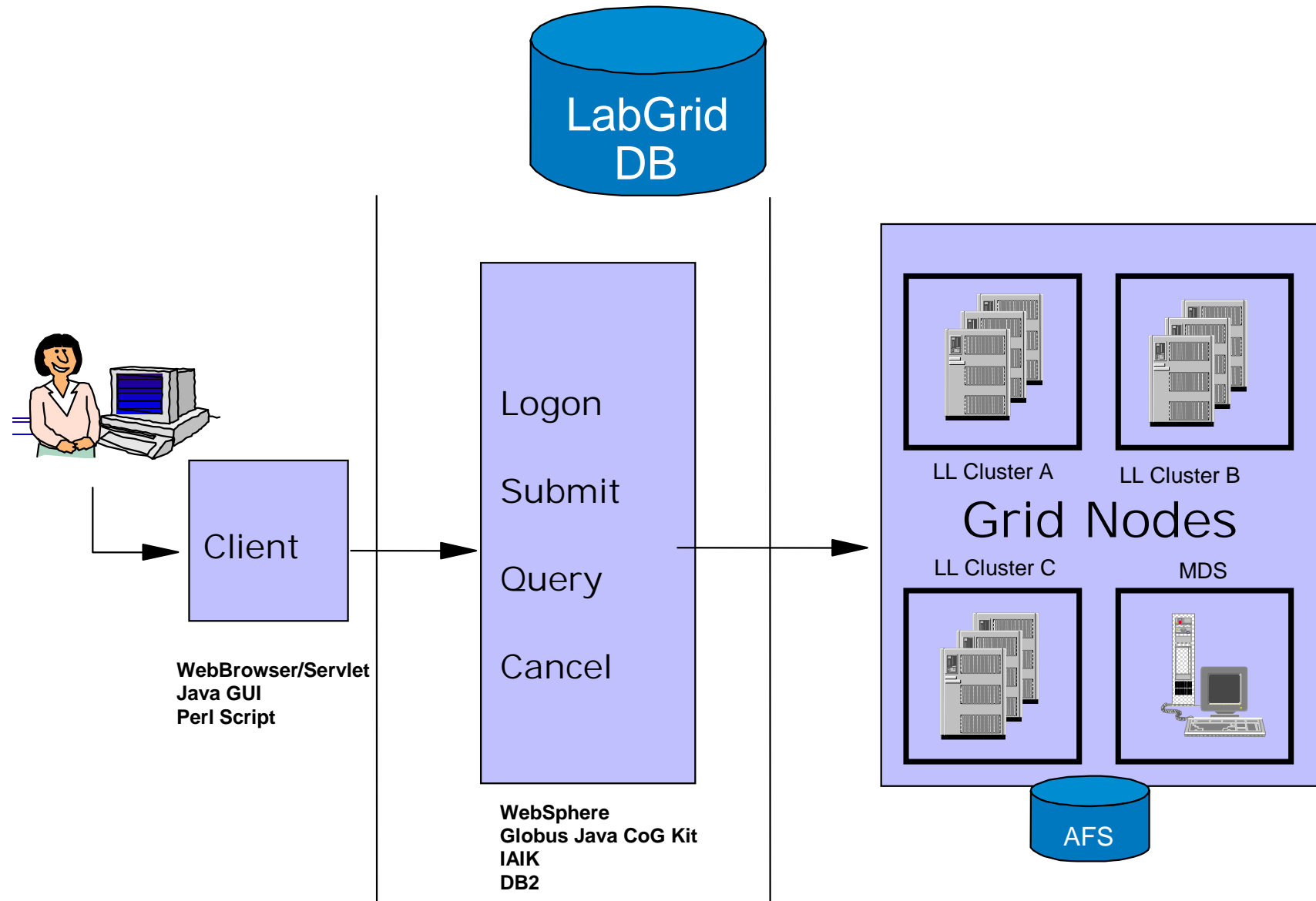
IBM LabGrid: Grid Exploitation Today

Job Submission

GTK client

zLinux

central MDS

GTK +
LL GW

GTK +
cluster GW

**Existing
AIX Cluster**

AFS for Application/Data
Access

IBM LabGrid: Infrastructure Overview

Client

Services

LL/AIX Cluster   LL/AIX Cluster

Grid Nodes

PBS Cluster   MDS

Clients | Services | Infrastructure

IBM LabGrid: Technical View

LabGrid
DB

Client

WebBrowser/Servlet
Java GUI
Perl Script

Logon

Submit

Query

Cancel

WebSphere
Globus Java CoG Kit
IAIK
DB2

LL Cluster A

LL Cluster B

Grid Nodes

LL Cluster C
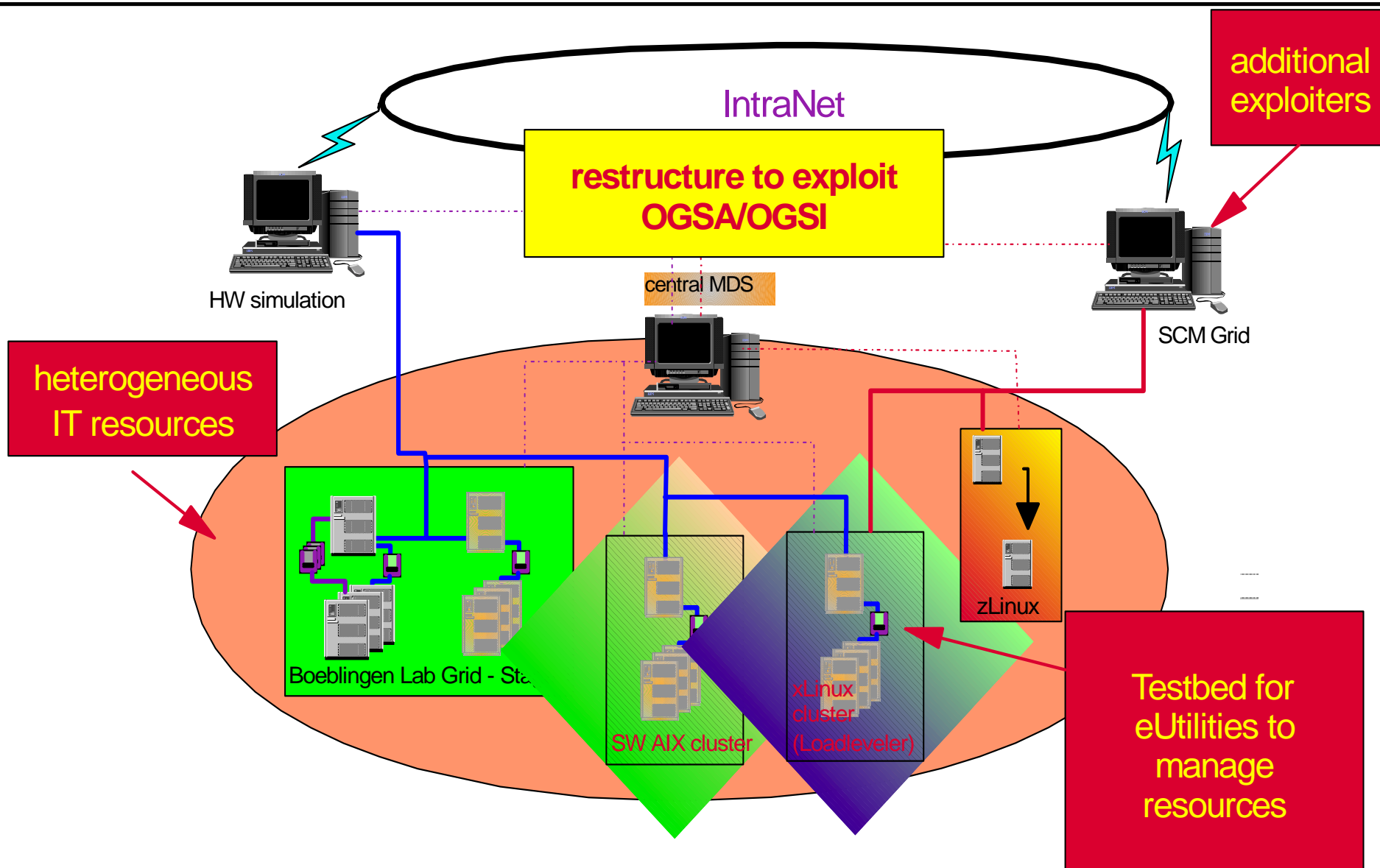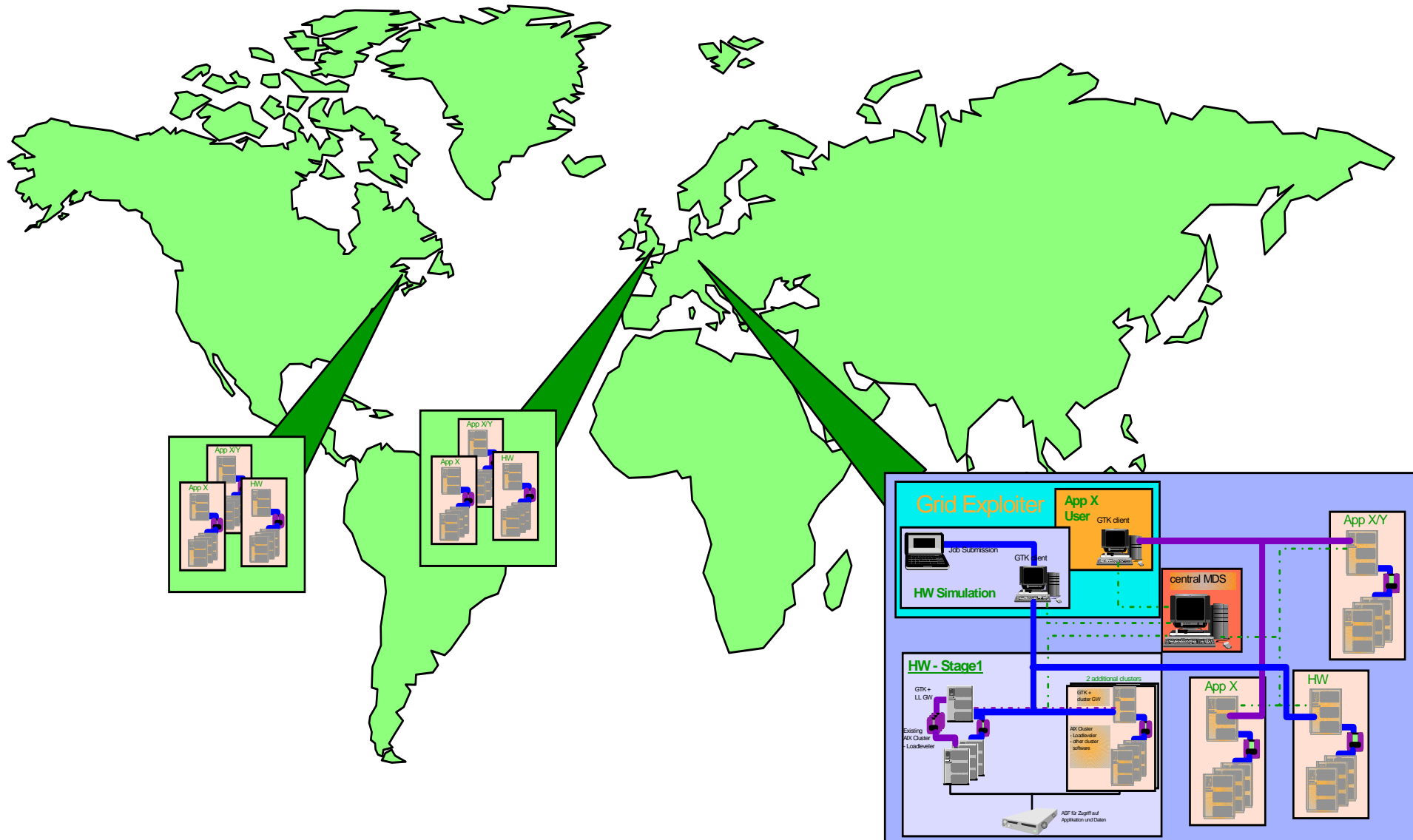
MDS

AFS

IBM LabGrid: Plans for 2003

IBM LabGrid: Additional Sites

IBM LabGrid: Business Value

- **Increased simulation capacity results in better quality due to higher coverage in simulation**
- **Cross-area cooperation enhances flexibility in IT environment**
- **Cost savings due to exploitation of system capacity on low-utilized servers**
- **Recovery in case of cluster problems**
- **Experiences in a production-like environment**
- **Develop skills in the area of grid computing**
- **Re-usable parts for scheduling and monitoring**
- **Evaluation environment for new eUtility components**
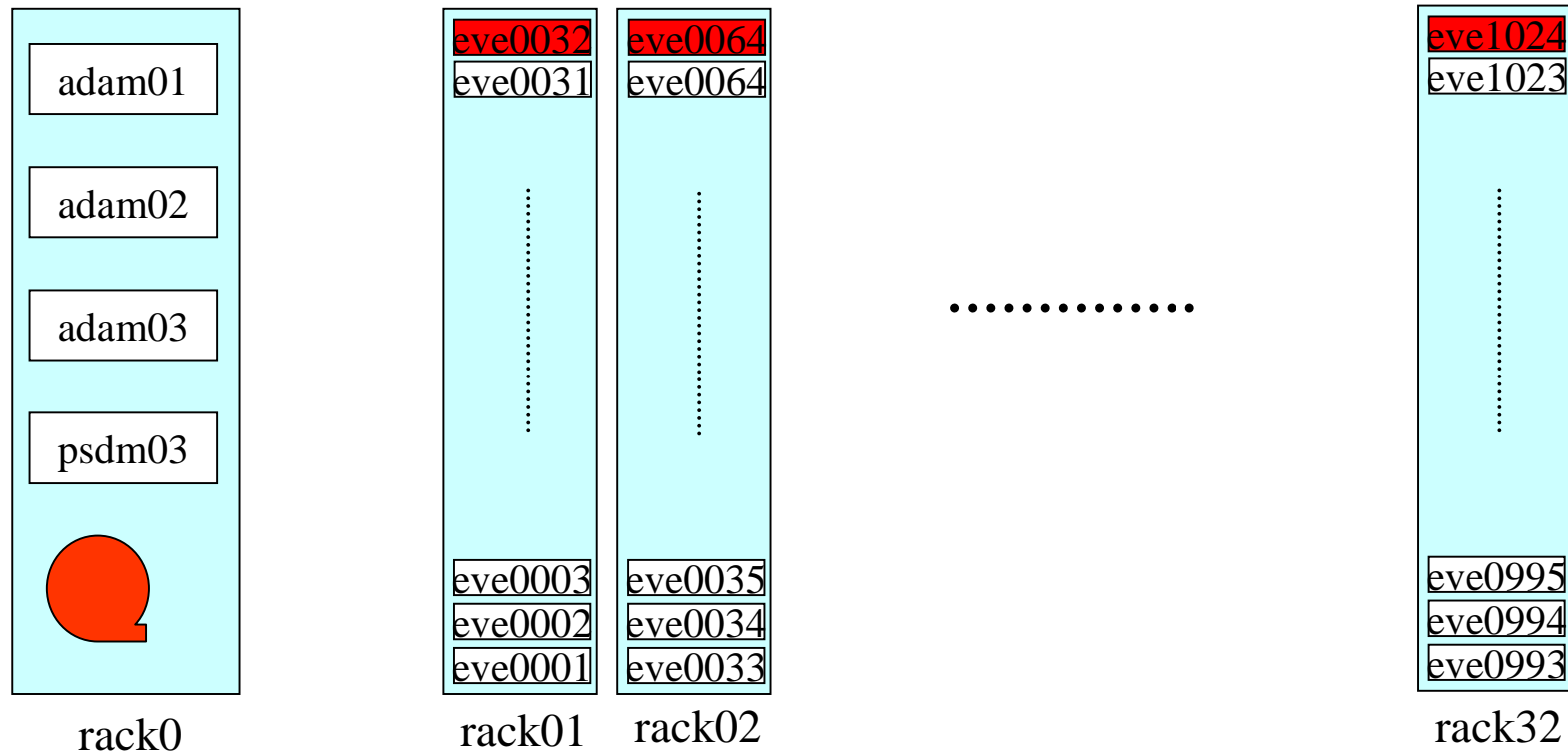
Grid@Shell (The Netherlands)

Grid@Shell: Pre-Stack Depth Migration

- **Objective of PSDM:**
  create a structural image of the <u>sub</u>surface using seismic <u>surface</u> measurements

- **Required operations for the Shell implementation of PSDM:**
  - calculation of traveltime tables (raytracing)
  - summation of contributions from seismic input data to output points in image space

Grid@Shell: Genesis Linux cluster

rack0	rack01	rack02		rack32
adam01	eve0032	eve0064		eve1024
adam02	eve0031	eve0064	....	eve1023
adam03				
psdm03	eve0003	eve0035		eve0995
	eve0002	eve0034		eve0994
	eve0001	eve0033		eve0993

32 racks of 32 nodes (1024 nodes), with:

1,0 GHz Intel Pentium III CPU

512 MB memory
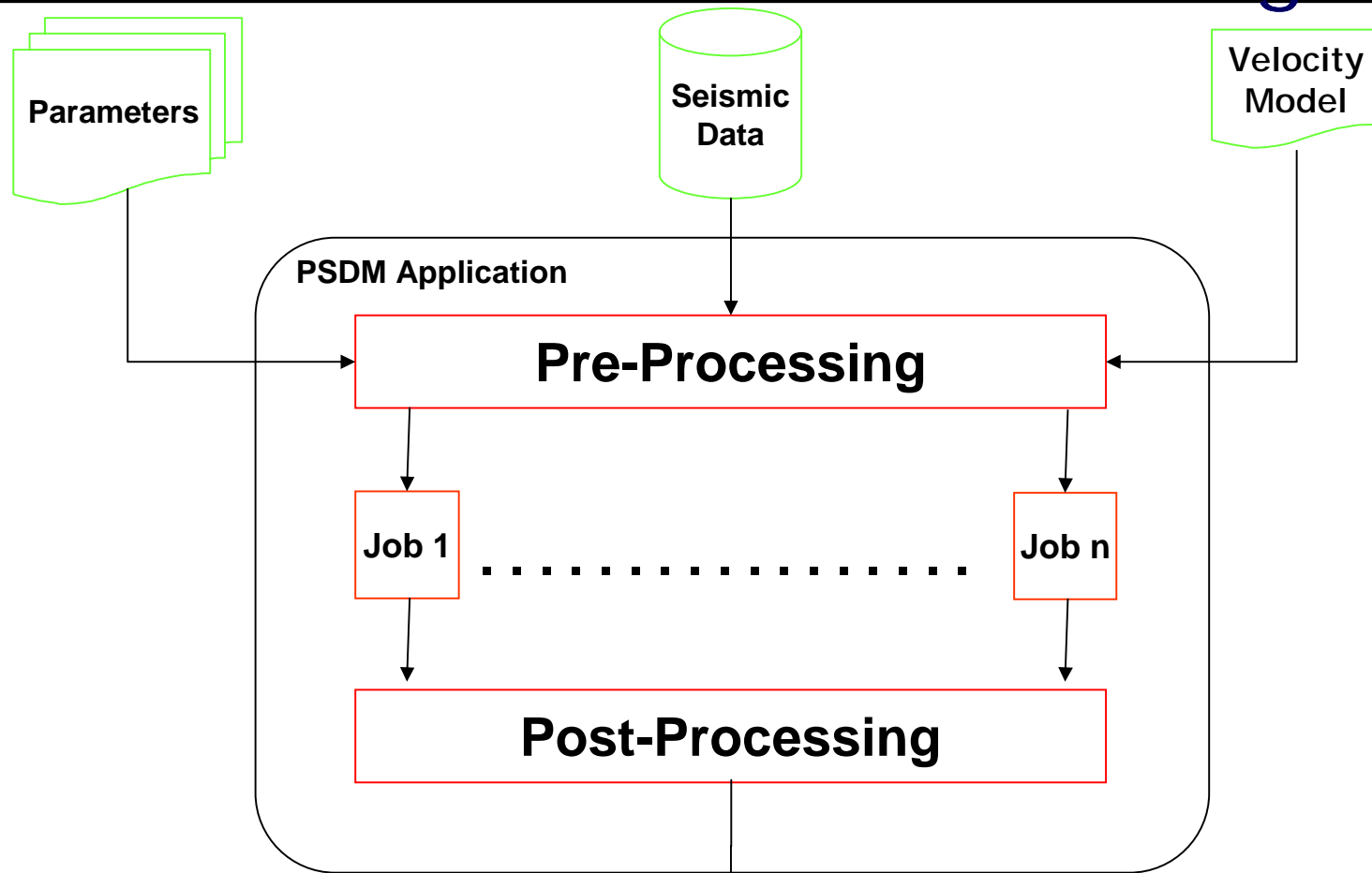
55 GB local storage space

Red Hat 7.1 installed

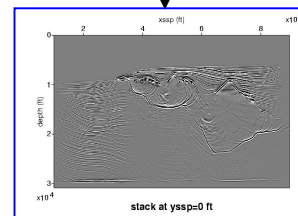# Grid@Shell: Implementation of PSDM

- **PSDM jobs are split up according to output image cells**

- **Output image cell size determined by:**
  - memory limitations
  - available resources
  - run-time requirements

Grid@Shell: PSDM Parallel Processing

Parameters

Seismic Data

Velocity Model

PSDM Application

Pre-Processing

Job 1 . . . . **Job n**

Post-Processing

xssp (ft)

stack at yssp=0 ft

Red : execution

Green : Input

Blue : output
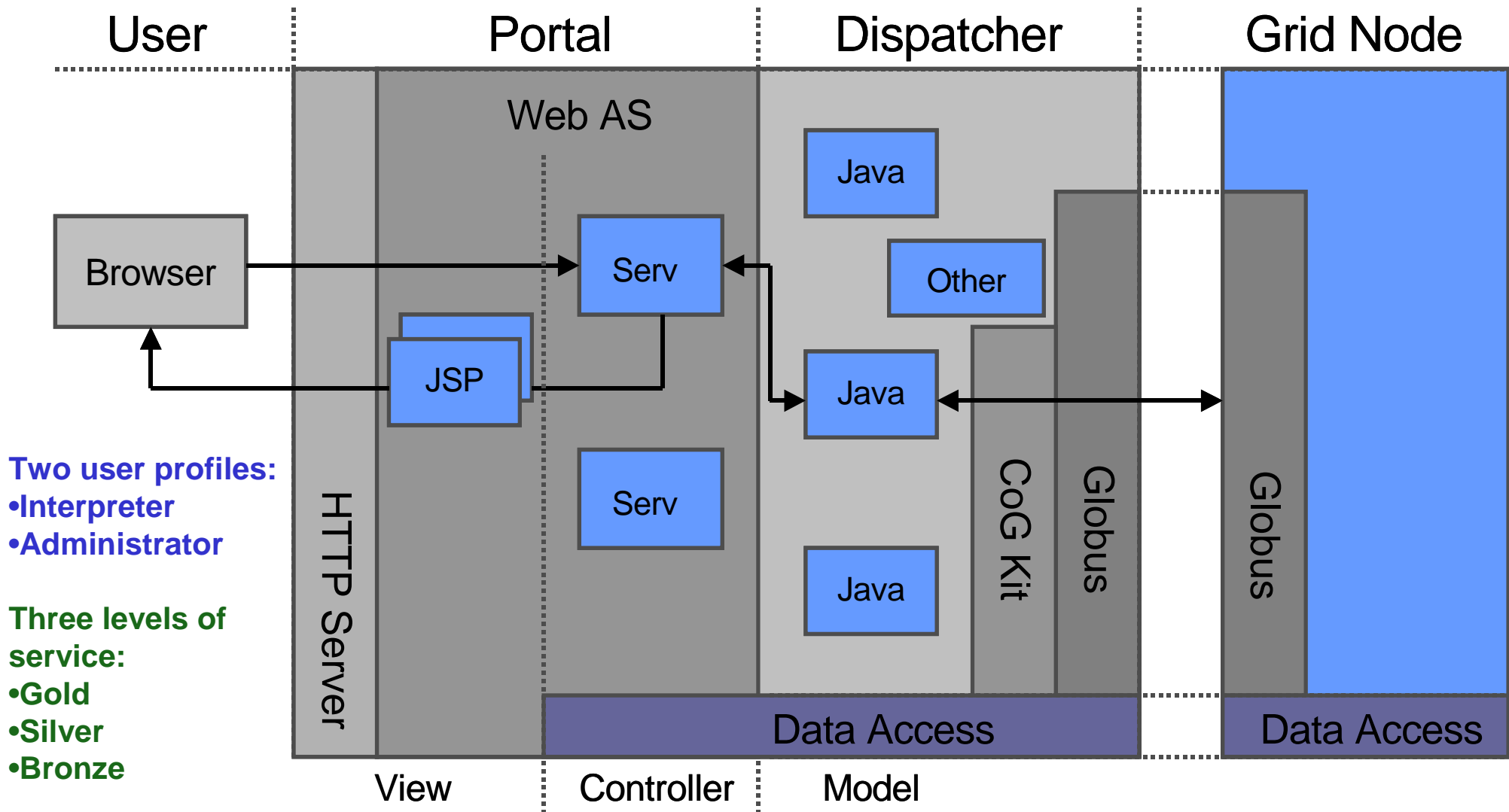
Grid@Shell: PSDM Grid Enablement Objectives

- **Provide PSDM as a service with various levels (QoS)**
- **Provide storage and computing on demand**
- **Allow for dynamic provisioning (adding and removing resources) and discovery**
- **Provide accounting and cost estimation prior to run**
- **Ease of use (browser-based interface)**
- **Job management portal**
- **Grid management portal**
- **Extensible to other Grid Enabled Applications (GEAs)**

Grid@Shell: Solution Architecture Overview

User   **Portal**   **Dispatcher**   **Grid Node**

Web AS

Java

Browser → Serv

JSP

Other

Java

Two user profiles:
- **Interpreter**
- **Administrator**

Serv

Java

HTTP Server

CoG Kit

Globus

Globus

Three levels of service:
- **Gold**
- **Silver**
- **Bronze**

Data Access

Data Access

View   Controller   Model

Grid@Shell: Portal Screen Copy

Grid@Shell: Portal Site Map

Welcome

| Grid Users | Register | Unregister | Login | Change Password |

| Data Storage | Book data storage capacity | Upload data | Download data | Inspect data |

| GEA Run | Book a GEA Run | Monitor a GEA Run | View results of a GEA Run |

| Monitor & Manage | View the Grid | Change a GEA Run | Get a report | Add new IT Resource | Remove IT Resource |

GEA Run Views
System Views
Application Views

Activity Report
Failure statistics

Stop Run
Allocate new nodes
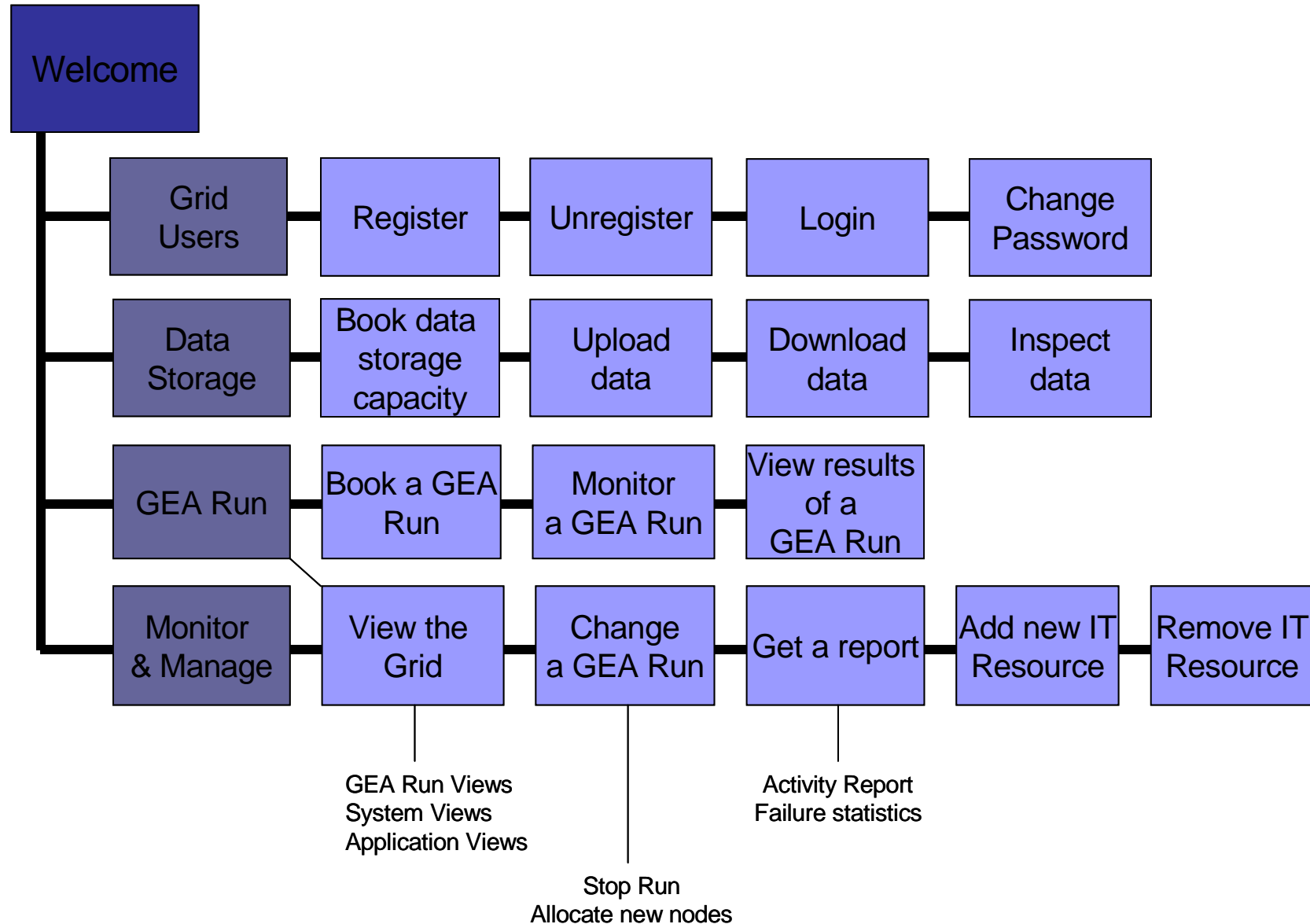
Royal Dutch/Shell

Challenge

- **Improve accuracy and speed of summarization and scientific modeling applications**

Solution

- ▪ **IBM** @server
- ▪ **Linux**
- ▪ **Globus Toolkit**

"Grid computing is important to Shell because it offers the potential to create a truly unlimited resource, with a uniform interface to a variety of services. This is a significant opportunity for Shell to engage its independent companies in closer cooperation." J.N. Buur, Principal Research Physicist, Shell International Exploration and Production B.V.

Technology Benefits:
- ▪ More robust, scalable IT infrastructure that adjusts as volumes fluctuate
- ▪ Open standards permit easy integration of existing software

Business Benefits:
- ▪ Cut processing time of seismic data, while improving the quality of the output
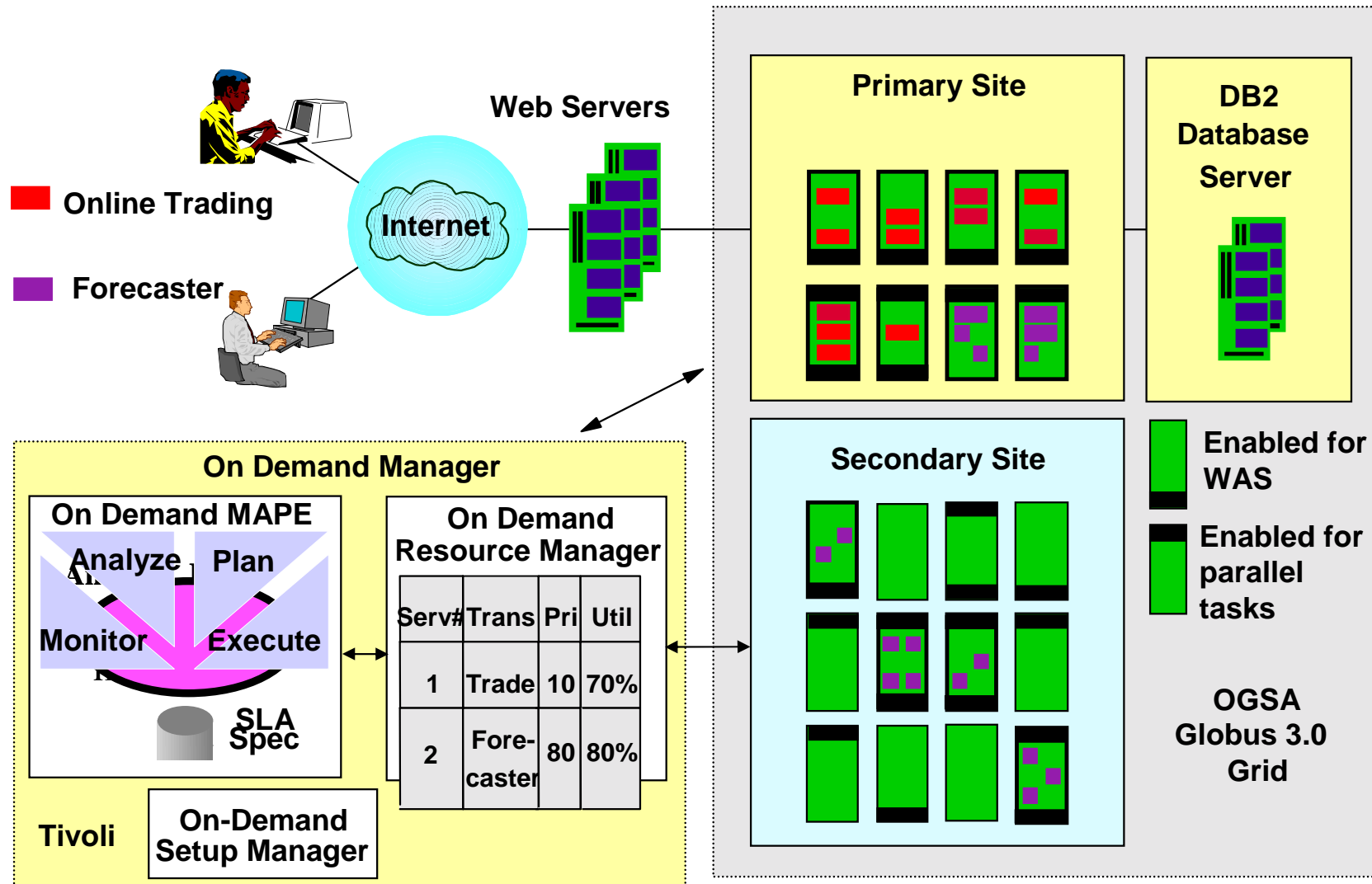- ▪ Focus employees on key scientific, not IT problems

Financial Services Company

Proposed Architecture

Online Trading

Forecaster

Internet

Web Servers

Primary Site

DB2 Database Server

Secondary Site

On Demand Manager

On Demand MAPE

Analyze Plan

Monitor Execute

SLA Spec

Tivoli

On-Demand Setup Manager

On Demand Resource Manager

Serv#	Trans	Pri	Util
1	Trade	10	70%
2	Fore-caster	80	80%

Enabled for WAS

Enabled for parallel tasks

OGSA Globus 3.0 Grid

Other References

Aventis

Challenge

- **Distributed, diverse data sources across continents**
- **Limited ability to consolidate, construct and analyze data sets**

Solution

- **Linux**
- **IBM** *e*server
- **IBM Discovery Link**

Technology Benefits:

- Using IBM DiscoveryLink to bring together data sources in one coherent view

Business Benefits:

- Significant increase in researcher productivity due to improve collaboration

- Better data quality and currency

Charles Schwab

Challenge

- **Reduce the processing time on an existing wealth management application to improve customer service.**

Solution

- **IBM@server**
- **Linux**
- **Globus Toolkit**
- **IBM Infrastructure Technology Services**
- **IBM Research**

"We believe that Grid computing … has the potential to greatly improve our quality of service and be a truly disruptive technology."
Oren Leiman, Managing Director, Charles Schwab
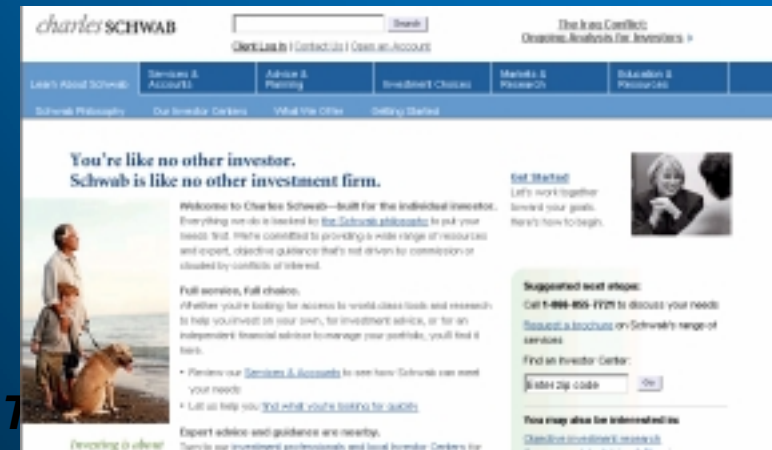
Technology Benefits:
- Reduced processing time from four minutes to fifteen seconds…
- Leverages existing infrastructure…
- Grid enabling many more applications

Business Benefits:
- Increase customer satisfaction by responding to inquiries in real time…
- Enabling Schwab to move from a low cost transactional broker to an advice based wealth manager

RBC Insurance

Challenge

- **Reduce the time it takes for an insurance policy valuation application to run**

Solution

- **IBM @server**
- **IBM Infrastructure Technology Services**
- **Platform Computing (ISV)**

RBC Royal Bank

Technology Benefits:
- Reduced processing time from eighteen hours to thirty-four minutes
- Automated job-scheduling
- Expanding implementation

Business Benefits:
- Can run more complex scenarios to reduce risk exposure
- Actuaries can spend less time scheduling application

IBM

Challenge

- **Microprocessor Design**
- **Benchmarking & Testing**
- **Server Design**

Solution

- **IBM @server**
- **Globus Toolkit**
- **IBM Global Services**

Microprocessor Design Grid

- Chip simulation driving 80% resource utilization
- Lower error rates in microprocessor designs
- Reduced development cycle, improved ROI and design engineer productivity

Benchmarking/Testing Grid

- Allows for larger scaling tests at lower costs by pooling all the servers across multiple sites

Z Series Design Grid

- Production environment is adjusted to average workload, lowering fixed cost
- Increased computing power for HW simulations
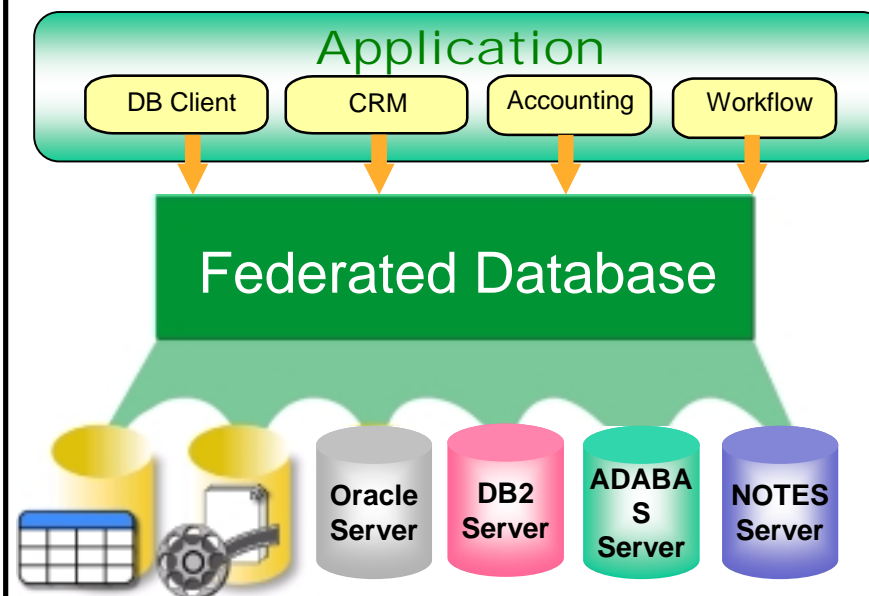- 40% increase in productivity of hardware engineers

Kansai Electric Power Co,

Challenge

- **Japan's second largest electric utility company has various information in a heterogeneous, distributed database environment**

- **Integrate information beyond departments and affiliated companies to enable information sharing**

Solution

- **Create virtual database federated from heterogeneous database environment**
- **IBM DB2 Data Federation Technology**
- **Wrapper to access other RDBs including legacy database**

Application

DB Client | CRM | Accounting | Workflow

Federated Database

Oracle Server | DB2 Server | ADABAS Server | NOTES Server

Technology Benefits:
- Virtualize various data sources across the enterprise
- Enable information sharing using existing systems including legacy data base
- Enable to develop new businesses more rapidly at a minimum cost

Butterfly.net

Challenge

- **Scalable, resilient infrastructure for running massive multiplayer games**

Solution

- **IBM** *@*server
- **Globus Toolkit**
- **IBM WebSphere Application Server**
- **DB2 Universal Database**
- **IBM e-business Hosting Services**

Technology Benefits:
- Improved end-user experience supporting over one million simultaneous sessions

Business Benefits:
- Developers avoid huge upfront costs
- Announced with Sony Computer Entertainment
- The Butterfly Grid for PlayStation2 unveiled at Games Developers' Conference in San Jose March 2nd

TIGER

Challenge

- **The Taiwanese government is building a grid between their leading academic and research institutions for research and collaboration in the areas of nanotechnology and life sciences**

Solution

- **IBM and NCHC building National Grid Test Bed**
- **IBM is assisting in the planning and implementation of the grid infrastructure.**
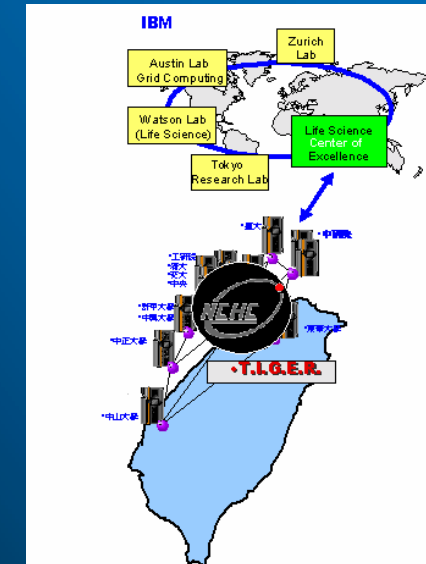
Technology Benefits

- Integrate in-country academic and research computing resources
- Test implementations and investigations into billing and provisioning systems will take place

Business Benefits

- Stimulate research in Life Sciences and Nanotechnology

IBM Commitment & Focus

Commitment

- Open standards

- R&D and investments in grid and related technologies

- Industry-leading partners

- Multiplatform experience and expertise

- Worldwide grid strategy, design, implementation and integration services

Focus

- Industry-specific offerings

- Product development roadmaps

- Building an ecosystem

- Building grids for commercial and public organizations

- Integrated solutions: HW, SW, Services and Partners

Web Resources

- **IBM Grid Computing: www.ibm.com/grid**

- **IBM Alphaworks: www.ibm.com/alphaworks**

  – IBM Grid Toolbox:
    - Globus Toolkit 2.2.4 for IBM pSeries (AIX 5.1 and 5.2), iSeries and zSeries (SLES8), xSeries (RedHat Linux 7.3)
    - Better documentation, customization scripts, and LoadLeveler enablement

  – Grid Application Framework for Java (GAF4J)
    - Lightweight framework that enables the development of Java applications that can distribute execution threads as tasks over grid resources

  – Emerging Technologies Toolkit
    - Architectural overview of web services, grid and autonomic technologies (SOAP engine – Apache AXIS, XML parser, UDDI client API, GT3 alpha 2, ReGS) - requires Java SDK 1.3.1 or 1.4 on Windows or Linux

  – Business Workload Manager Demo and Prototype
    - Performance monitoring and allocation of computing resources for distributed transactions based on application instrumentation

  – Optimal Grid
    - Creation and management of connected parallel applications on the grid

Web Resources (continued)

- **IBM Developerworks:** **www-106.ibm.com/developerworks/grid**
  - Globus Toolkit 3.0 and OGSI Architecture
    - Overview of the Globus GT3 software architecture and simple Grid Service illustration with WSDL
  - A developer's overview of OGSI
    - Examples to explain core concepts of OGSI
  - Introduction to GT3 Alpha 3
    - Tutorial for programmers and system administrators for GT3 installation and deployment onto WAS V.5.0
  - Grid Toolbox tutorial
  - Grid computing explained
    - Irving Wladawsky-Berger explains Grid Computing

Web Resources (continued)

- White Paper about the IBM Server Allocation for Websphere Application Server:
  ftp://ftp.software.ibm.com/software/info/openenvironment/Server_Allocation_hvws_wp.pdf

- Presentation of the IBM on demand operating environment virtualization component (relates to grid computing):
  http://www-3.ibm.com/software/info/openenvironment/virtualization.html