Grid Portals: bridging the gap between the users & the Grid

Rajesh (Raj) Chhabra

QPSF Grid Manager- HPC, QUT, Brisbane, Australia

Project Leader-User Interface & Visualisation Infrastructure Projects- APAC Grid Program

r.chhabra@qut.edu.au

Osaka University (organized under PRIUS program)

22nd May 2006

AGENDA

- **Section 1**
  - **Overview of Grid computing & Grid applications**
    - **Group Activity**
- **Section 2**
  - **Latest technologies for building Grid Portals & GridSphere toolkit**
    - **Live Demo**
- **Section 3**
  - **Grid Portals Architecture with GridSphere (technical session)**
    - **Group Activity (combined in the lecture)**
- **Section 4**
  - **Project to be carried out at QUT under PRIUS attachment**
    - **Closing Remarks**

Section1

Overview of Grid computing

&

Grid applications

Grid Computing

- **Grid can be understood with reference to World Wide Web**

- **In simple terms-**
  Web →
  Information: anywhere, anytime
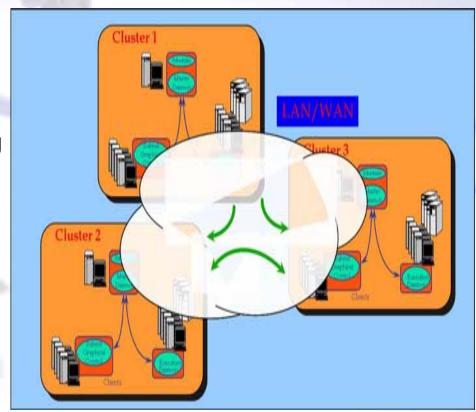
  Grid →
  Computing/Software/Data/Devices/ People: anywhere, anytime

- **The word 'Grid' itself is an analogy to the electricity grid.**

Many definitions of Grid

- **Cluster of clusters/ supercomputers**
  - **Various clusters/ supercomputers from different/same organizations are connected together using some form of middleware**

  - **Computational grid**

  - **This is a common understanding of Grid in the academic world**

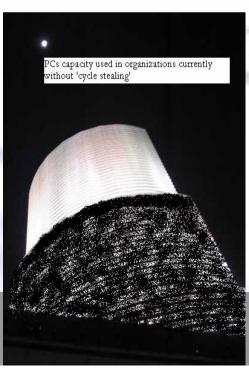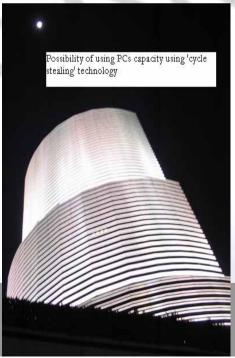 - **An important aspect but not the complete Scenario**

Many definitions of Grid..

- ### Cycle Stealing
    - **To harness (utilize) the power of many unused CPUs of desktop computers loosely connected in a network (eg. comp labs & offices in Osaka University)**
    - **This is a common understanding of Grid in the business world**
    - **Easy to explain and understand the concept but again it's one part of the whole picture**
    - **Relatively much easier to implement (less issues on management policies, firewalls etc)**
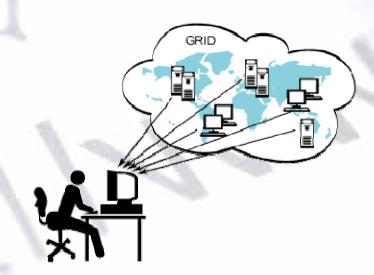   - **Popular project- SETI@home**

PCs capacity used in organizations currently without 'cycle stealing'

Possibility of using PCs capacity using 'cycle stealing' technology

Many definitions of Grid..

- **Distributed Computing**
- **Peer to Peer (P2P)**
- **Internet2**
- **Web Services**

- **All of these various technologies contribute to the vision of the Grid**

- **So what actually is Grid? (Vision- Where are we heading?)**
  - **Virtualization of resources**
  - **On demand provisioning of resources & services**
  - **Seamless sharing of resources & services**

Some common terms

- **Computational Grid**
  - Large set of distributed heterogeneous computational resources connected together to form a virtual pool of computing resources

- **Data Grid**
  - Large set of distributed heterogeneous Data Collections connected together to form a virtual pool of data collections

- **Sensor/instrument Grid**
  - Large set of distributed sensors (instruments) connected together to provide aggregated information
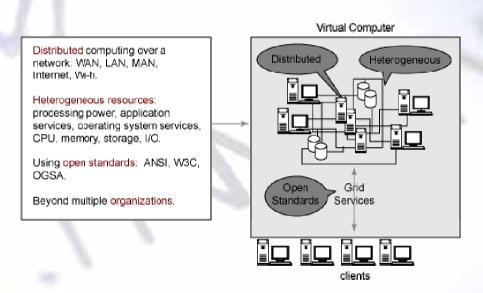
- **Access Grid**
  - Large scale of videoconferencing system capable of connecting multiple sites with high quality video and audio

- **Similarity among the terms?**
  - Large set/scale, Distributed, Heterogeneous, Connected, Virtual
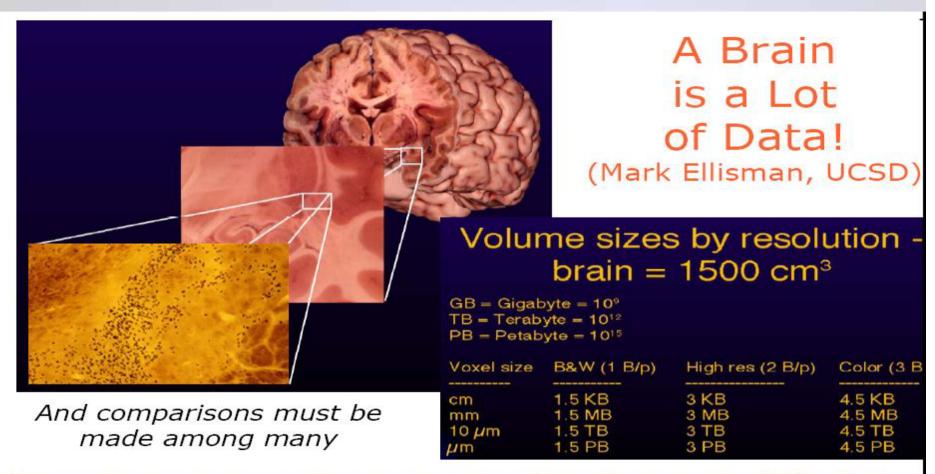
Path required to accomplish the vision for the Grid

- **Path – How to get there?**
  - **Along with long list of technical requirement that we need there are two critical non technical requirments**
    - **Standardization (to allow interoperability)**
      - **GGF, OASIS etc play an important part**
      - **OGSA –Open Grid Service Architecture**
      - **WSRF- Web Services Resource Framework etc**
    - **Management policies**
      - **Opening a firewall between two organizations- its 99% policy and 1% technology!**

Distributed computing over a network: WAN, LAN, MAN, Internet, Wi-fi.

Heterogeneous resources: processing power, application services, operating system services, CPU, memory, storage, I/O.

Using open standards: ANSI, W3C, OGSA.

Beyond multiple organizations.

Virtual Computer

Distributed   Heterogeneous

Open Standards   Grid Services

clients

More Grids

- **BioGrid**
    - **http://www.biogrid.jp/ (grid program)**
    - **http://biogrid.icm.edu.pl/ (grid program)**

- **World Community Grid**
    - **http://www.worldcommunitygrid.org/ (program to utilize cycle-stealing concept)**

- **CancerGrid**
    - **http://www.cancergrid.org/ (grid program)**

- **SARS Grid**
    - **http://www.anl.gov/Media_Center/logos21-2/grid.htm (project initiated during the SARS outbreak in Taiwan)**
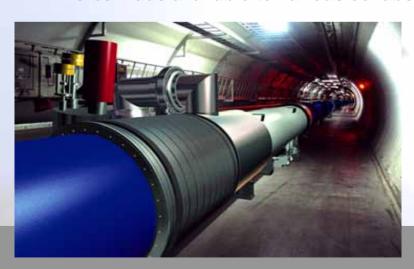
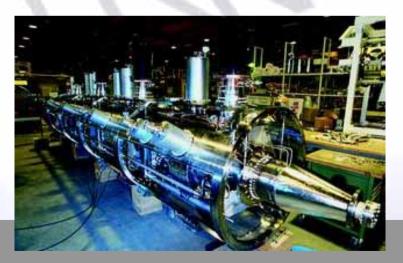- **EuroGrid, AP Grid (http://www.apgrid.org/) etc**

Why Grid? (Brain Scans)

A Brain
is a Lot
of Data!
(Mark Ellisman, UCSD)

And comparisons must be
made among many

Volume sizes by resolution -
brain = 1500 cm³

GB = Gigabyte = 10⁹
TB = Terabyte = 10¹²
PB = Petabyte = 10¹⁵

Voxel size	B&W (1 B/p)	High res (2 B/p)	Color (3 B
cm	1.5 KB	3 KB	4.5 KB
mm	1.5 MB	3 MB	4.5 MB
10 μm	1.5 TB	3 TB	4.5 TB
μm	1.5 PB	3 PB	4.5 PB

We need to get to one micron to know location of every cell. We're just nov
starting to get to 10 microns – Grids will help get us there and further

Why Grid? ..(LHC)

- **The Large Hadron Collider (LHC)**
  - **A particle accelerator and collider located at CERN, near Geneva, Switzerland**
- **Collaboration**
  - **Over two thousand physicists from 34 countries, universities and laboratories**
- **Operational in 2007**
  - **It will collide hadrons (protons) in the 27 km circumference tunnel!**
- **Data**
  - **15PB/year through 5 projects**
  - **To be made available to various collaborators across the world**

Why Grid? (local issues for a Uni)
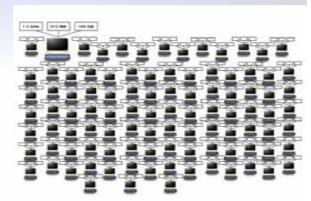
- **Computing Infrastructure**
  - Various computing platforms, different architectures required for different problems
  - More the merrier (enough is not enough)
  - Implementing the grid infrastructure allows to access large pool of resources by joining into various Grid programs
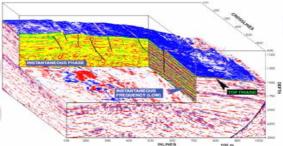
- **Data**
  - Data requirements are ever increasing in all the universities
  - Again implementing grid infrastructure allows better management of data and makes it easier to share

- **Collaboration**
  - Increasingly the projects are spanned across the institutions and countries
  - Bigger challenges require collaborations from various organizations
  - Large amount of data has to be shared across multiple sites often across the countries.
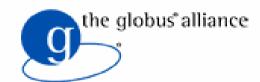 - Grid computing provides the mechanisms required for sharing the resources effectively

Grid computing examples

- **Some typical Grid computing examples**

    - **Running multiple computing jobs simultaneously at various sites say Australia, US, Taiwan & UK.**

    - **Running a computing job in Australia with data from multiple repositories of SDSC (US) & NCHC (Taiwan)**

    - **Using an electron microscope of Osaka University from SDSC with the results being stored simultaneously at Osaka & San Deigo**

Globus Toolkit

- **An open source software toolkit**
  - **Started around 10 years ago in Argonne National Lab (USA)**
  - **Now being developed by the Globus Alliance and many others all over the world**
  - **De facto standard for building Grid systems and applications.**

- **Provides**
  - **A layer between the various computing infrastructure**
  - **Security (GSI)**
  - **Job submission mechanism (GRAM)**
  - **Data transfer (GridFTP)**
  - **Information Service (MDS)**
  - **and more..**

Running a typical grid job

- **Running multiple computing jobs** (current mechanism -it will change in the very near future)
  - **Certificate based Authentication (GSI-Single sign on)**
    - **Proxy generation**

  - **Resources discovery (MDS- GIIS, GRIS)**
    - **Verification of resources availability/Selection of the resources where the job is going to be submitted based on criteria such as quota, software availability, architecture etc**

  - **Data transfer (making data available before run time) (Gridftp)**
    - **Upload file/files to the resources**

  - **Job submission (GRAM- Globus Gatekeeper, Job Manager)**
    - **Job submission at multiple resources as selected earlier**

  - **Job monitoring (GRAM- Globus Gatekeeper, Job Manager)**
    - **To monitor the status of the job submitted at the various resources**

  - **View results**
    - **To view the results – visualize an image or a data file etc**

GROUP ACTIVITY

Section 2

Latest Technologies for building Grid portals

&

GridSphere

Web Portal

- **What is a web portal?**
  - **Gateway to provide a variety of services to users**
    - **Yahoo Portal- email, personal groups, personalized news channels, stocks, weather etc**
  - **Single sign-on**
  - **Customizable user environment**
- **Various technologies (Perl/CGI, PHP)**
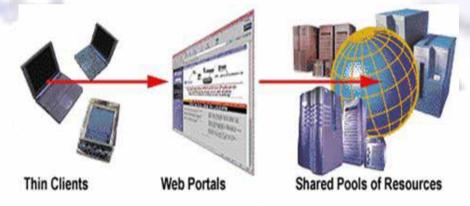- **Number of organizations, research groups implementing portals**
  - **Re-inventing the wheel**

Why Portals?

- **Ease of use**
  - Users comfortable with Unix/Linux command line in the room?
  - Users comfortable with web browsers?
- **Single Sign-on**
  - One login, multiple resources behind
- **Unified Interface**
  - Multiple resources, Single interface
- **And many other benefits that web offers like**
  - Anywhere, anytime etc

What is Grid Portal?

- **A web portal providing Grid services**

- **Grid portal's common components**
  - **Login, job submission (globus), file upload, file download, job monitoring etc**

- **What if these common components can be**
  - **Based on standard technology**
  - **completely modular**
  - **reused easily?**
- **JSR168 is one such standard!**

Thin Clients

Web Portals

Shared Pools of Resources

Courtesy http://www.sun.com

Portlet standard JSR168

- **Java Specification Request (JSR) 168 (Portlet) specification – Java community standard method for creating the components (portlet) of a portal thus enabling interoperability between portlets and portal**

- **Put forward by IBM & SUN with expert group forming members from Apache & BEA as well**

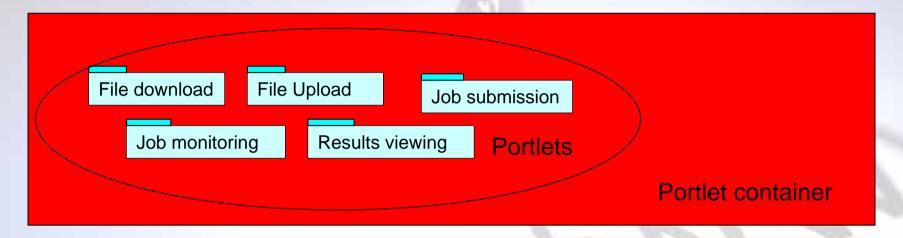- **The specification's final release came towards the end of 2003 and has been supported by many leading industry players**
http://www.jcp.org/en/jsr/detail?id=168

Supporting this JSR:

Accenture
Apache Software Foundation
BEA
Boeing
Borland
Bowstreet
Cap Gemini Ernst & Young
Citrix
Computer Associates
CoreMedia
DaimlerChrysler
Documentum
Enformia Ltd
Hewlett-Packard
Interwoven
Macromedia
McDonald Bradley
Novell
Oracle
Plumtree
SAP
Sybase
Tarantella, Inc
Vignette

Portlet Containers

Portlets

Portlet container

- **Various portlet containers, few popular ones…**
  - **GridSphere**
  - **Jetspeed-2**
  - **LifeRay**
  - **Websphere (IBM)**
  - **Oracle 10g portal etc..**
- **Other popular project**
  - **OGCE – NSF (National Science Foundation) NMI (National middleware Initiative) funded Open Grid Computing Environment Portal (OGCE) Project http://www.collab-ogce.org/ogce2/ supports JSR168 standard and package many gridportlets**
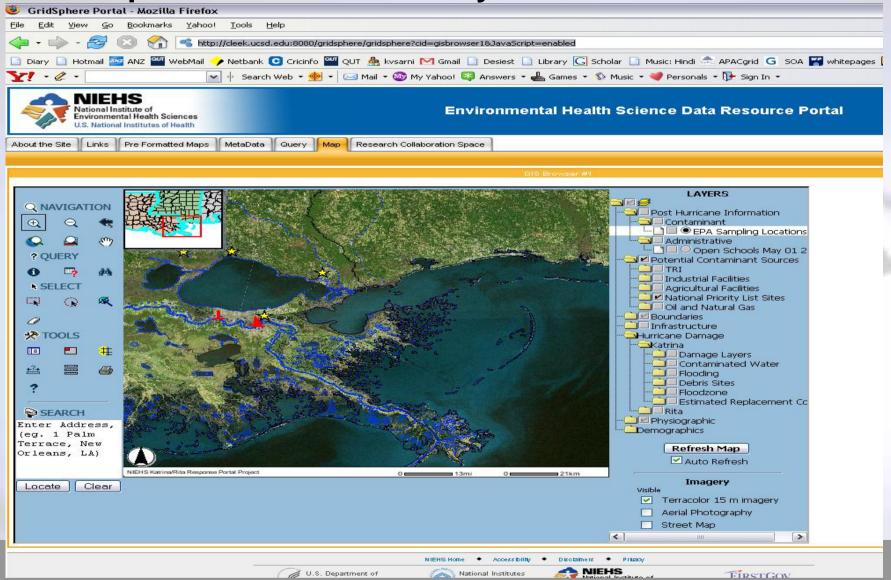
GridSphere

- **Part of GridLab projects funded by EU**

- **100% JSR168 compliant**

- **Open source & free to use**

- **Gridportlets- the most important offering of GridSphere project for Grid users**
  - Login, logout
  - File upload & download **(GridFTP based)**
  - **Globus 2.4 based job submission**
  - Job monitoring
  - Credential management **(GSI based & MyProxy complaint)**
  - Profile personalization and Layout customization
  - Administration portlets for creation of users, groups, portlet management and portal layout customization
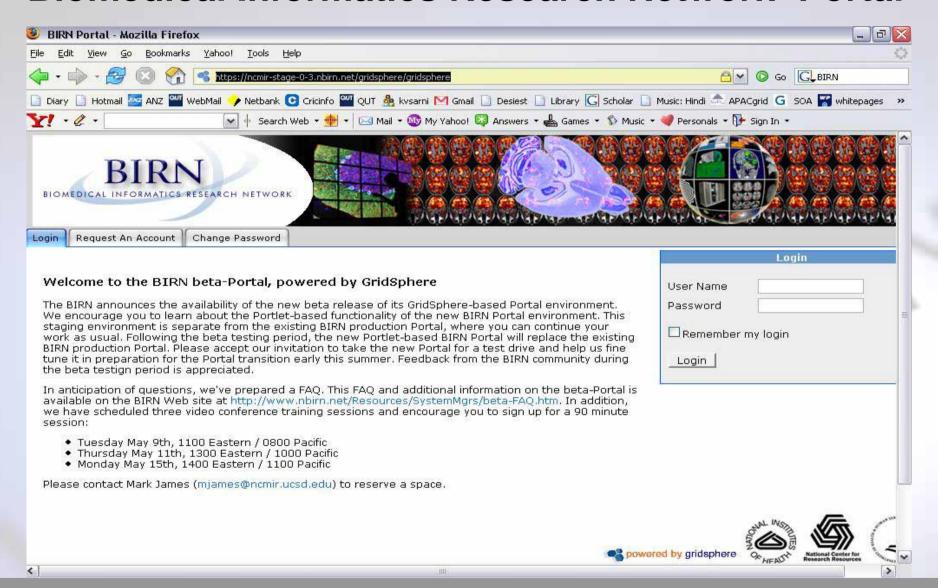 - **Persistence (using Hibernate)**

GridSphere based portals

- **7 Grid portals based on GridSphere in Australia and more in the pipeline**

- **Many around the world**

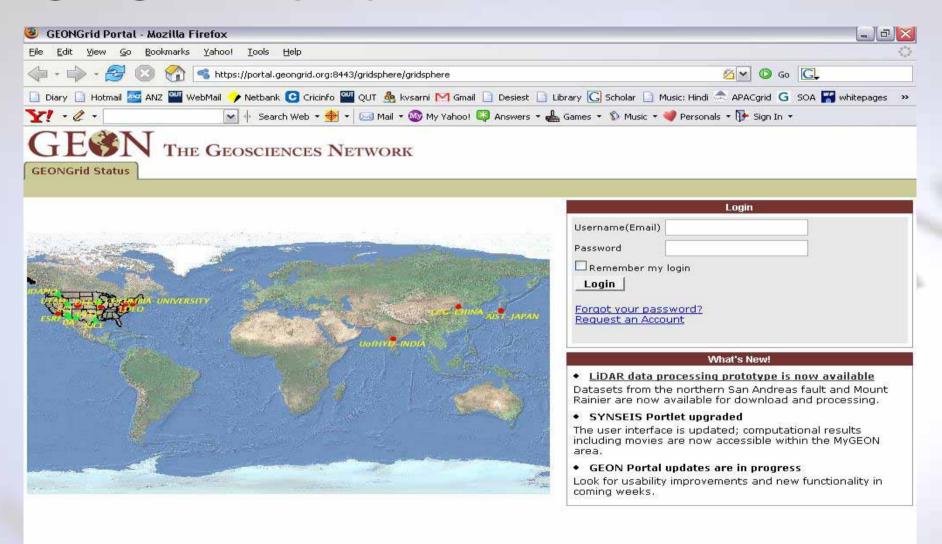- **PRAGMA members largely involved in building GridSphere based portal**

National Institute of Environmental Health Science Portal- in response to Katrina/Rita cyclones
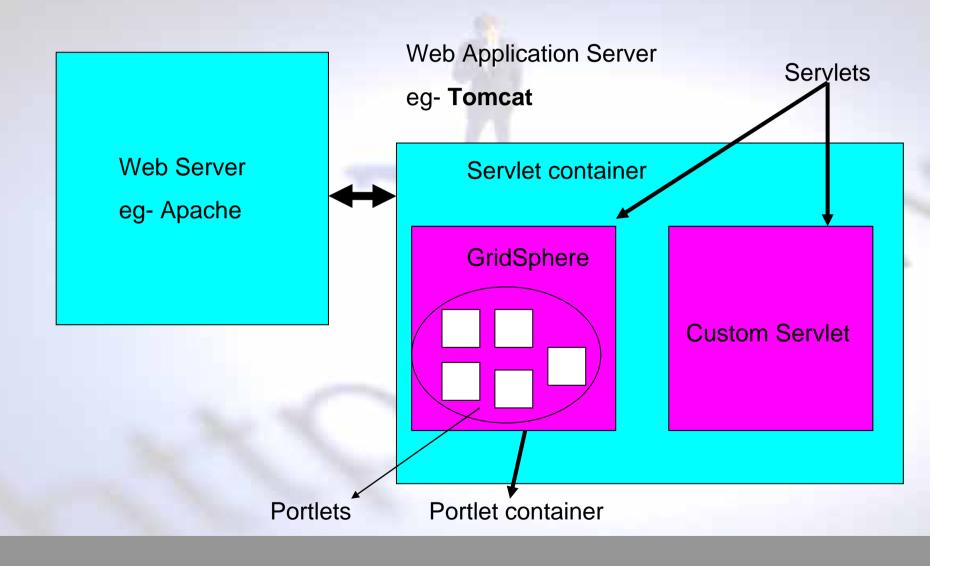
Biomedical Informatics Research Network- Portal

GEON- Portal

Section 3

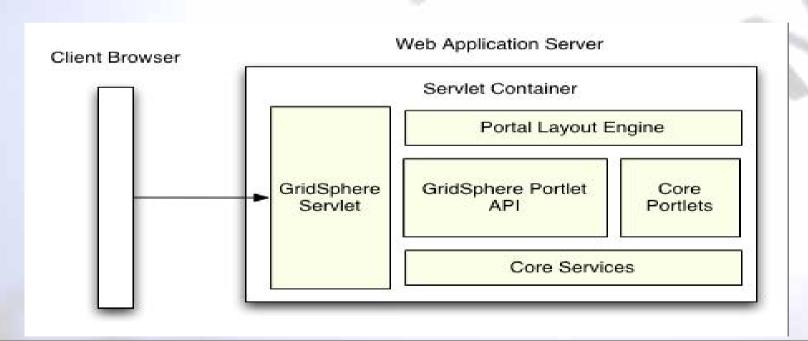# Grid Portals Architecture with GridSphere

Our GridSphere based Grid Portal Architecture

Web Application Server

eg- **Tomcat**

Servlets

Web Server

eg- Apache

Servlet container
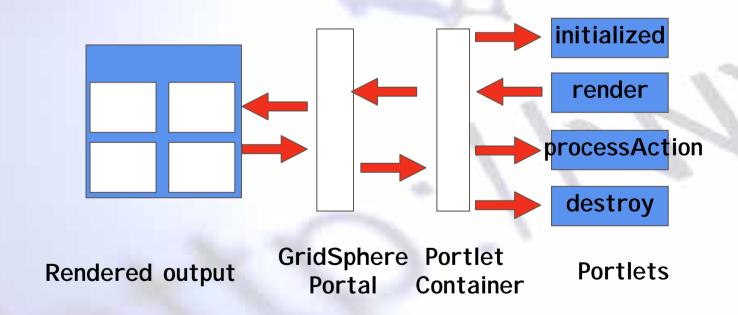
GridSphere

Custom Servlet

Portlets

Portlet container

GridSphere Architecture..

- **The GridSphere portlet container is implemented as a web application and requires a hosting environment such as the Jakarta Tomcat container.**

- **Many additional libraries are used and deployed to the servlet container during installation.**
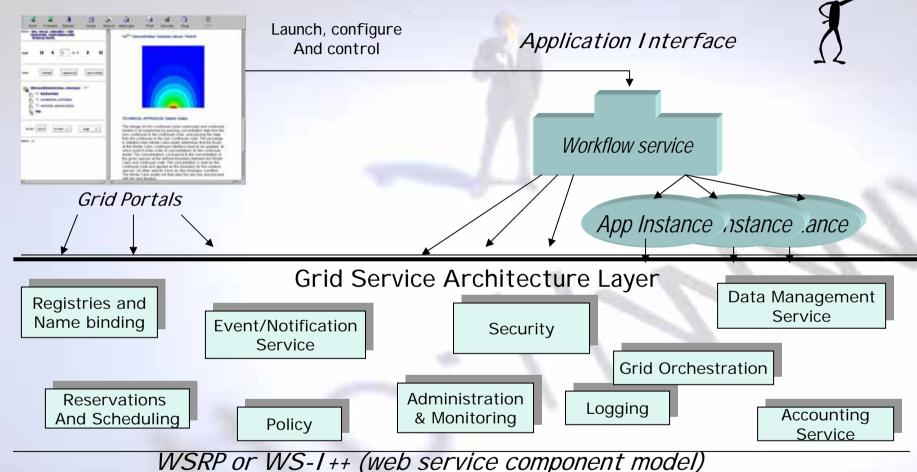

Courtesy of GridSphere team

GridSphere Architecture..

Rendered output   GridSphere Portal   Portlet Container   Portlets

initialized

render

processAction

destroy

Courtesy of GridSphere team

Grid Portals & Workflow

Launch, configure
And control

Application Interface

Workflow service

Grid Portals

App Instance *nstance* *ance*

Grid Service Architecture Layer

Registries and
Name binding

Event/Notification
Service

Security

Data Management
Service

Grid Orchestration

Reservations
And Scheduling

Policy

Administration
& Monitoring

Logging

Accounting
Service

WSRP or WS-I++ (web service component model)

Resource layer

1000s of PCs ->massive supercomputers

Online instruments

Section 4

Project to be carried out at QUT under PRIUS program

PRAGMA: Pacific Rim Applications Grid Middleware Assembly

- **Participants from:**
  - **Asia Pacific and growing..**

PRAGMA..

- **Common pool of computing resources, expertise, collaborations, student exchange etc**

- **PRIME – NSF funded student exchange program**

- **PRIUS – Japanese equivalent program in Osaka University of student exchange funded by Ministry of Education, Culture, Sports, Science and Technology (MEXT)**

Grid Portals in PRAGMA

- **Major applications group in PRAGMA- BioInformatics**

- **Grid Portals from S.Korea, Taiwan, Australia, Malaysia, Switzerland/USA, Singapore etc**

- **Different applications portal being developed at various groups with some overlap.**

Similarities among these Portals

- **Bio-Science domain**

- **Based on JSR168 standard**

- **Most of them using GridSphere toolkit**

- **Commonly used applications in the Bio-Science community eg. NAMD, Blast**

- **These portal users can use the common shared pool of PRAGMA resources**

What is involved?

- **Combine the portlets from various grid portals to make an integrated Bio-Science Portal**

  - **Evaluating the available portlets within the PRAGMA group**

  - **Establishing common needs of the groups -What portlets will be required commonly?**

  - **Testing modularity - Ease of stripping the portlet from their main portal**

  - **Customizing portlet interface**

  - **Integration of the portlets into 1 portal**

  - **Connecting the PRAGMA resources at the backend**

  - **Hosting and testing the Portal**

  - **Presentation in the PRAGMA 11(15th-16th Oct 2006) in Osaka (if we can start the project by Aug 2006)**

Support provided

- **Complete supervision**
  - **Technical**
  - **Administration**
  - **Contacts**
  - **Books**

- **Computing & Office facilities**

- **Help in finding accommodation and other logistics**

QUT, Brisbane

Benefits in this project

- **Experience on cutting edge technology**
  - Grid computing, Portals, Portlets

- **Exposure to foreign working environment**

- **Contacts within PRAGMA community**
  - To work with major biotechnology groups in PRAGMA

- **Good employment prospects after the degree**
  - Reference of working abroad

- **Chance to practice English everyday**
  - Added advantage of English fluency in the ever-increasing globalized world

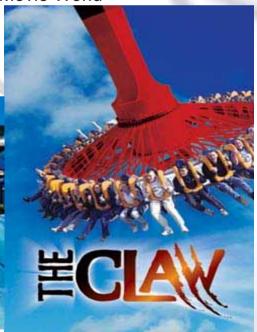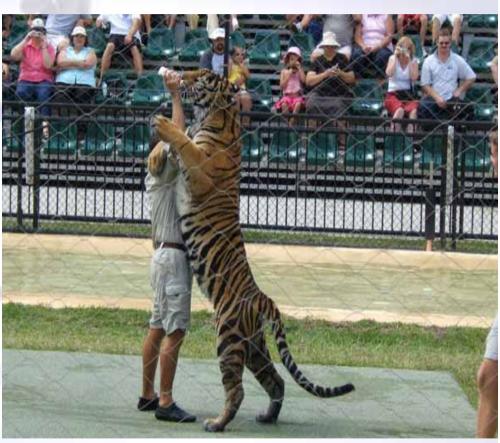# Other benefits –Australian Sea Food
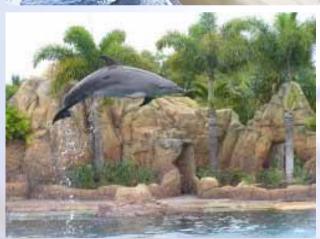
GoldCoast, Theme Parks & Beaches

GoldCoast

Movie World

Sea World

Wet n Wild (theme park)

Dream World

Australia Zoo (Steve Irwin)

Closing Remarks

- **Thanks to-**
  - **PRIUS & PRAGMA**
  - **Associate Prof Susumu Date**
  - **Prof Shinji Shimojo**
  - **Ms Kae Nakagawa**
  - **Ms Nahoka Kataoka**
  - **Prima Tours Inc – Hitoshi Onishi**
  - **QUT, QPSF, APAC**

- **Visa Process**
  - **Student's resume**
  - **Supporting letter from QUT & Osaka University**
  - **Medical, Police clearance etc as required by Aus Consulate**
    http://www.australia.or.jp/
  - **Twin 21 MID Tower, 16th Floor, 2-1-61, Shiromi, Chuo-ku Osaka 540-6126, Japan**
  - **Telephone 81 6-6941-9271**

Courtesy of images

- **Clusters of clusters – Uni of Melbourne**
- **Cycle stealing http://www.columbia.edu/~man75/images/cycle.jpg**
- **Miami building at night - http://www.toddicus.com/?x=2005_01_01_news_archive.html**
- **Grid virtualization - http://www-128.ibm.com/developerworks/grid/library/gr-info1/infogridp1_fig1.gif**
- **LHC http://www.nevis.columbia.edu/~conrad/visuals/lhc.jpg**
- **Grid http://images.genesi.lu/slides/grid/Grid0.png**
- **Data http://www-sst.unil.ch/research/seismic/laurent/cubtronq_gd.gif**
- **Collaboration http://blogs.salon.com/0002007/images/collaboration.jpg**
- **Globus http://www.globus.org/**
- **Rock portal- http://www.cmnh.org/pressroom/images/portal.jpg**
- **QUT-**
    - **http://upload.wikimedia.org/wikipedia/en/b/b0/Qut_theater.jpg**
    - **http://upload.wikimedia.org/wikipedia/en/b/ba/Qut.jpg**
    - **http://upload.wikimedia.org/wikipedia/en/d/dc/QUT-Gardens-Point-entrance.jpg**

Courtesy of pictures..

- **Food-**
  - http://www.coolum.com.au/imgs/dining_scene_body2.jpg
  - http://www.proserpineecotours.com.au/images/home_mudcrab.jpg
- **GoldCoast-**
  - http://www.dfat.gov.au/aii/publications/_lib/img/06/Gold-Coast-Aerial-shot.jpg http://www.travel-australia.biz/queensland/graphics/surfers.jpg
  - http://national.atdw.com.au/multimedia/tq/500972_1.jpg
  - http://2005.schoolies.com/images/the_claw.jpg
  - http://www.australianexplorer.com/competition/entries/280205211306p2090200.jpg
  - http://www.mortenmeyer.com/GoldCoast/Batman.jpg
- **Australia Zoo**
  - http://www.rfleming.net/diarypics/Crikeysteve.jpg
  - http://www.jeremyparsons.com/personal/picture_pages/2005_Brisbane_and_Noosa/Tiger_Feeding@Australia_Zoo.jpg

Other References

- **GridSphere project- www.gridsphere.org**
- **http://cleek.ucsd.edu:8080/gridsphere/gridsphere?cid=gisbrowser1&JavaScript=enabled**
- **https://ncmir-stage-0-3.nbirn.net/gridsphere/gridsphere**
- **https://portal.geongrid.org:8443/gridsphere/gridsphere http://www.extreme.indiana.edu/~gannon/GridPortals.ppt**
- **www.collab-ogce.org/GGF15Workshop/documents/GridPortals-intro.ppt**