The APAC National Grid Program in Australia

“providing advanced computing, information and grid infrastructure for eResearch”

Glenn Moloney
University of Melbourne
for the
Australian Partnership for Advanced Computing
APAC National Grid

- Darwin
- GrangeNet Backbone
- Centie/GrangeNet Link
- AARNet Links
- Internet2
- Canarie
- Geant
- APAN
- • 10 Gbps
- • IPv6
- • Multicast

Darwin

GrangeNet Backbone
Centie/GrangeNet Link
AARNet Links

Perth
IVEC
CSIRO

Brisbane
QPSF

Canberra
ANU

Sydney
ac3

APAC
National Facility

Adelaide
SAPAC

Melbourne
VPAC

CSIRO

CSIRO

Hobart
TPAC

CSIRO

Glenn Moloney
The APAC Grid Program in Australia
ISGC, Taipei, 2006
2
Australian Partnership for Advanced Computing

“providing advanced computing, information and grid infrastructure for eResearch”

The APAC Partners:

• **AC3**: Australian Centre for Advanced Computing and Communications in NSW
• **CSIRO**: Commonwealth Science and Industry Research Organisation
• **QPSF**: Queensland Parallel Supercomputing Foundation
• **IVEC**: Interactive Virtual Environments Centre in WA
• **SAPAC**: South Australian Partnership for Advanced Computing
• **ANUSF**: The Australian National University
• **TPAC**: The University of Tasmania
• **VPAC**: Victorian Partnership for Advanced Computing
National Role of APAC

Advanced Computing Infrastructure
- peak computing facilities

Information Infrastructure
- support for community-based data collections
- management of large-scale data collections (archiving)

Grid Infrastructure
- access to national computing and information infrastructure
  - access to federated computing and information systems
- advanced collaborative services for research groups
  - collaborative visualisation, computational steering, tele-presence, virtual organisation support
- support Australian participation in international research programs
  - eg, astronomy, high-energy physics, earth systems, geosciences
The APAC Grid Program

Australian government provided AU$29m for stage 2 of APAC:

- Providing the advanced computing and grid infrastructure for eResearch

- AU$12.5m for upgrade of National Facility
  - Commissioned mid 2005
  - Canberra

- National grid infrastructure projects:
  - Computing infrastructure
  - Information infrastructure
  - User Interface and Visualisation

- Application support projects:
  - Astronomy (Virtual Observatory)
  - Computational chemistry
  - Theoretical and experimental high energy physics
    - International LatticeGrid, ATLAS, Belle
  - Geosciences
  - Bioinformatics
APAC National Facility

Usage
• mainly biology, chemistry, physics
• currently 247 projects and 722 users (27 universities)

Computing Systems
• SGI Altix 3700 Bx2 system: 1680 processors
• Dell Linux cluster: 150 processors

Mass Data Storage System (MDSS)
• Storagetek (robotic silo) HSM tape library
  – Petabyte capable storage

Visualisation Systems
• Virtual reality systems, Access Grid rooms

Staff
• User support, Systems support
• Computational tools and techniques
• Large-scale data collection management

http://nf.apac.edu.au
Global Connectivity

10Gbps ring
APAC National Grid

- Basic Services
  - single ‘sign-on’ to the facilities
  - portals to the computing and data systems
  - access to software on the most appropriate system
  - resource discovery and monitoring

one virtual system of computational facilities
APAC Grid Deployment

APAC National Grid.v1 – Single Sign-on, data sharing
Base: VDT (GT2.4.3, Monalisa, Ganglia), GridSphere, SRB, OpenDAP, Nimrod, LCG
Use APAC CA
VO model: follow Grid3
Manually configured solutions

APAC National Grid.v2
– Add portals and workflow support
Base: VDT-> GT4, Gridsphere, SRB OpenDAP, Nimrod, LCG
VO Model: not yet determined

Use National CAs
Auto configuration

APAC National Grid.v3
Interoperability:
Align with OSG, EGEE
Use aarnet3 backbone

2005 2006
APAC Grid Gatekeeper Machines

Each partner site has a 'gateway' machine which 'hosts' Grid front-ends to the available resources

Xen Virtual Machine Monitor
University of Cambridge Computer Laboratory

Hardware:
Dual Xeon 2.8GHz, 4Gb RAM, 300Gb mirrored SCSI disk, 5 GigE network cards
(1 mgmt, 2 data VM, 2 other VM's)

Grid front-ends:
• Globus 2 (VDT-1.2.4), Globus 4 (VDT-1.4 ??),
• Storage Resource Broker 3.3.1, LCG, Nimrod/G
National Grid Infrastructure

Portal Tools:
- GridSphere

Workflow Tools:
- Kepler?

Security:
- APAC CA
- MyProxy
- VOMRS

Systems:
- Gateways
- Partners’ Facilities

Network:
- GrangeNet
- APAC VPN (AARNet)

a virtual system of computing, data storage and visualisation facilities
Resource Discovery:
- APAC Software Registry
- MDS
- INCA?

Job Submission:
- Command Line
- Portals

Job Monitoring:
- Scope
- MonaLisa?

Job Management:
- Globus, Nimrod, PBS

Computing Systems:
- Peak
- Mid-range
- Special

APAC National Grid Computing Grid Infrastructure

QPSF

IVEC

MonaLisa?

Scope

Globus, Nimrod, PBS

APAC National Facility

QPSF

(JCU)

ac3

ANU

CSIRO

SAPAC

VPAC

TPAC

Glenn Moloney

The APAC Grid Program in Australia

ISGC, Taipei, 2006
Data Access:
- OGSA-DAI
- Web services
- OPeNDAP

Data Management:
- Globus
- SRB
- SRM

Data Transfer:
- RFT
- GridFTP
- Global File System

Mass Data Storage Systems:
- Tape – based (silos)
- Disc-based
Visualisation Services:
*Prism and VisServer
*Visualisation Software

Collaboration Tools:
*AG Whiteboard

Facilities:
*Access Grids
*Virtual Reality Systems

APAC National Grid
Collaboration Support Infrastructure

- **QPSF**: (JCU)
- **APAC National Facility**: IVEC
- **SAPAC**: ANU
- **CSIRO**: ac3
- **TPAC**: IVEC

Glenn Moloney
The APAC Grid Program in Australia
ISGC, Taipei, 2006
Delivering National Grid Services

Data Centres

Research Teams

grid-based portals
distributed computation
federated data access
remote control
collaboratories

Other Grids:
Institutional
National
International

Sensor Networks

Instruments
Astronomy and Astrophysics

- **MACHO Project Data**
  - Largest online astro data set in Australia (~10TB)
  - Hosted by APAC as part of IVO collection
  - Mapping metadata to VOTable 1.0 standard

- **Australian Virtual Observatory**
  - Provide uniform access to key data collections
    - 2dFGRS, HIPASS, ATCA-OA, SUMSS, MACHO, TNO...
  - Grids for theoretical astrophysics simulations
    - Portals for job configuration, submission and monitoring
    - MLAPM, GCD+, Zeus-MP, LensView, (x)oopic, Swift,

- **International Virtual Observatory**
  - SIAP service for ATCA Phoenix Deep Field Survey
    - SIAP is an International Virtual Observatory protocol
Accelerate progress on genome annotation, for genomes of national economic significance
Support lead discovery through molecular docking

- Data update and synchronisation services, including the BioMirror
- Grid-wide compute services for Ensembl, Blast, RepeatMasker and Glimmer
- Grid-wide compute services for molecular docking including support for analysis workflows
Computational Chemistry

**Unified Grid-based portal to chemistry software**

- Portal to computational chemistry software on APAC Grid
- Uniform access to software on a computer system
- Gaussian, Amber, Gamess-US, Gromacs, Mopac and Molpro
Earth Systems Science

Access to Data Products

- Inter-governmental Panel Climate Change scenarios of future climate (3TB)
- Ocean Colour Products of Australasian and Antarctic region (10TB)
- 1/8 degree ocean simulations (4TB)
- Weather research products (4TB)
- Earth Systems Simulations
- Terrestrial Land Surface Data

Grid Services

- Globus based version of OPeNDAP (UCAR/NCAR/URI)
- Server side analysis tools for data sets: GRADS, NOMADS
- Client side visualisation from on-line servers
- THREDDS (catalogues of OPeNDAP repositories)
Geosciences

Develop systems that support the real-time steering of complex geoscience analysis

This requires:

• Workflow support for mantle convection modelling with components running on distributed grid resources
• Portlets for compute services including ‘snark’ and ‘Finley’
• Hypothesis exploration through real-time ensemble management
High-Energy Particle Physics

**Belle Physics Collaboration**
- K.E.K. B-factory detector
  - Tsukuba, Japan
- Matter/Anti-matter investigations
- 45 Institutions, 400 users worldwide
  - ~1 PB data currently
- Australian grid for KEK-B data
  - Data grid centred on APAC National Facility

**Atlas Experiment**
- Large Hadron Collider (LHC) at CERN
  - Operational in 2007
- Deploying LCG/EGEE infrastructure on APAC Grid
APAC High Energy Physics project

Our scientific goals:
• Ensure expertise and infrastructure for Australian physicists to analyse data sets from the ATLAS experiment
• Deploy data grid technologies within the Belle collaboration

Aims of the project:
• Establish an Australian Data Grid infrastructure for applications in experimental high energy physics.
• Deploy LCG grid facility in Australia for ATLAS data analysis
• Deployment of a grid-based international network of regional data centres for the processing of data from the Belle experiment.
Progress to date: ATLAS

Dec 2004: Complete ATLAS Data Challenge 2
January 2005: Deployment of LCG-2 at University of Melbourne Physics
April 2005: First deployment of LCG-2 at VPAC and University of Melbourne
August 2005: Deployment of LCG XEN virtual machine on APAC grid gateway machine at VPAC
December 2005: Preparation of initial Australian Tier 2 facility at University of Melbourne
May 2006: Tier 2 Site Functionality Tests commence

Marco will talk on Australian Tier 2 Status and Plans this afternoon.
Progress to date: **Belle**

June 2004: Establishment of SRB federation between KEK and ANUSF

June 2004: Commenced distributed Monte Carlo production for Belle (utilising ANUSF/KEK SRB federation) non-grid system for computer resources

Dec 2004: Leading role in deployment of Belle SRB federation: KEK, Krakow, Korea, Taiwan, Beijing

May 2005: Completed first phase of Belle MC simulation (220 million events, 4.5 Terabytes of data, 195,000 CPU hours).

June 2005: Deployment of prototype Belle Analysis Data grid (APAC, VPAC, Unimelb). SRB, LCG and globus resources

Belle Monte Carlo on APAC Grid

- GQSched Resource Broker decomposes requests for analysis on sets of files:
  - Single job description – Many job submissions
- Data available from Belle SRB federation
- Data staging (3 models on the APAC Grid)
  - Handled within job during execution
  - Handled by separate process on another job manager
  - Handled by a separate process on another PBS queue

Second round of off-site Belle Monte Carlo production about to commence
The APAC Grid Program

The APAC grid program has been active in deploying a grid infrastructure in Australia

• Focussed on needs of Application Projects
• Interoperability – must work closely with international grids
• Tyranny of distance is being tamed: high bandwidth international connections

But – we need to do more:
• improved international collaboration
• more efficient deployment
• Operations: we are just beginning