(Some)

Grid Activities in Australia

Geoffrey Taylor
The University of Melbourne

gntaylor@unimelb.edu.au
Contents

Network Infrastructure (Aarnet)

National Research Infrastructure (ARCS)

High Energy Physics (ATLAS, Belle) Grid

Example of State-based Initiative (VeRSI)

HPC centres versus networked resources
AARNet Pty Ltd (APL) is the company that operates Australia's Academic and Research Network (AARNet). It is a not-for-profit company limited by shares. The shareholders are 38 Australian Universities and the CSIRO.

The network is built on a highly resilient backbone stretching from Brisbane to Perth.

AARNet has achieved high speed access across Australia based on STM-64c (10Gbps) circuits from Nextgen Networks with high speed links to multiple "Tier One" Internet Providers in the USA and SX TransPORT, the joint initiative of AARNet and Southern Cross Cable Networks.
Australian Research Collaboration Service (ARCS)

“The ARCS Missions is to enable and enhance research through the provision of long-term eResearch support, services and tools.”

Central goals include improved and transparent access to:

- Online real-time collaboration & collaboration workspaces;
- Web applications & services;
- ARCS Data Fabric: Data movement, federation & digital repositories;
- ARCS Compute Cloud: Seamless resource sharing;
- Authorization expertise, exemplars, tools and services;
- Virtual environments;
- Scientific instruments - remote access and operation.

Thanks to Tony Williams, Exec. Director, ARCS
Target User Communities

Including but not limited to:

- Biomolecular Platforms (NCRIS 5.1);
- Characterisation (NCRIS 5.3);
- Astronomy (NCRIS 5.10);
- Integrated Marine Observing Systems (IMOS) (NCRIS 5.12);
- Structure and Evolution of the Australian Continent (NCRIS 5.13);
- Climate Modeling;
- Computational Chemistry;
- Earth System Science;
- High-Energy Physics;

- Gravity Wave (LIGO)
- Biological Systems (NCRIS 5.2);
- Australian Biosecurity Intelligence Network (ABIN) (NCRIS 5.8);
- Terrestrial Ecosystems Research Network (TERN) (NCRIS 5.11);
- Population Health Research Network (PHRN) (NCRIS);
- Humanities;
- Social Sciences;
- Other research and discipline-based communities;
- Other eResearch service providers.
Videocollaboration: ARCS provides support for EVO (in collaboration with AARNET) and Access Grid technologies - including High Definition;

Instant Messaging (IM): Jabber is flexible IM system like MSN, ICQ etc (ARCS runs an ARCS Jabber server)

Helpdesk: End-user support is provided via telephone 1800 TOARCS (answered all business hours) & email request tracker system help@arcs.org.au.

Web collaboration tools: Research web-based workspaces and tools - Sakai, Drupal, Plone, Wiki, Twiki, Google Calendar, Google Apps,...
Collaboration Services - Some Use Cases

**Sakai** - 129 users... 5 sites
- DataMinx - NeAT Project for Characterisation
- IPPN - International Pediatric Pulmonary Network
- NCRIS Roadmap Review 2008
- TERN - Terrestrial Ecosystem Research Network
- ARCS

**EVO** – 236 users... 3 active communities
- Auscope
- National Taxonomy Research and Information Network (TRIN)
- CSIRO Computational and Simulation Sciences
- All Neat Project Committees being contacted and set up
ARCS Data Services

Provide support and expertise in all questions regarding data transfer;

Have local staff in all states;

Can setup large scale data transfers, one off or continuously, national and international;

Can help to organize large scale storage at several locations;

Provide all-purpose data service: ARCS Data Fabric.
ARCS Data Fabric

- Is a system to store, manage and share data;
- Is available to every researcher working in Australia;
- Has storage and access nodes in most states;
- Provides 25GB of free storage to every researcher (can be extended typically at hardware cost only to arbitrary size at any MARCS or elsewhere);
- Can be accessed with GUI and CLI tools;
- Can interface with Grid compute job.
ARCS Systems Services

- 12 ARCS staff hosted at local sites.

- Host basic ARCS infrastructure for services
  - Mail, including lists
  - Jabber
  - Wiki
  - Development repositories (Trac/Subversion)
  - Request Tracker

- Develop and document best practices for:
  - Use of OS Virtualisation (Xen/VMWare ...)
  - Deploying services – generally on CentOS Linux
  - Monitoring
ARCS Systems Services (ii)

Major role is to manage the ARCS National Grid and to develop the ARCS Compute Cloud

- **A number of HPC, data and other facilities available to Australian researchers.**
- **Standard interface provided by “The ARCS Grid”**
- **Develop and support easy to use tools for job submission, and data staging, e.g., “Grisu”**
- **Publish information about the resources available – Monitoring and Discovery Service (MDS)**

Deploy other services as required by Data, Collaboration or Authorisation teams

- **Sakai, SRB (ARCS Data Fabric), IdP, SLCS certificates etc**
ARCS Authorisation Services

- Provides ARCS authentication and authorisation infrastructure, services, tools, exemplars, expertise and best practice advice;

- Assists with interfacing to the Australian Access Federation (AAF) and any other relevant Identity Providers (IdPs);

- Develop suggested template usage policies for eResearch Service Providers (SPs) including ARCS, ANDS and discipline-based services;

- Pro-actively work with other SPs teams to allow all of their services to make effective use of Auth Services;

- Provide broad support for any research-oriented organisations that wants to access these services.
The Melbourne Australia-ATLAS site

✧ The site was created as part of our commitment to the ATLAS experiment (ATLAS-Australia Tier-2), and to service the needs of our local

✧ HEP users at Melbourne and Sydney (“Tier-3” or ”User Interface” facility)

✧ Australian Tier-2 served by ASGC Tier-1 site

✧ Australian Belle (KEK) resource. (Superbelle coming!)

✧ EGEE biomedical community has queue access on a fair-share basis.
Hardware

**TIER2 Hardware**
- 80 Processors – 150 kSI2K
- 120TB Storage in disk via DPM

<table>
<thead>
<tr>
<th></th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU (kSI2K)</td>
<td>40</td>
<td>150</td>
<td>200</td>
<td>450</td>
<td>860</td>
</tr>
<tr>
<td>Disk (TB)</td>
<td>20</td>
<td>70</td>
<td>110</td>
<td>280</td>
<td>370</td>
</tr>
</tbody>
</table>

**TIER3/TIER2.5 Hardware**
- 16 Processors in a PBS queue
  - Used for short run jobs and testing potential TIER2 jobs
- A 5TB NAS serving the home area to the User Interface (UI)
- NAS backed up to tape via Tivoli Storage Manager
- 15TB of TIER2 storage reserved for local users
Virtualization

- We are using Xen to virtualize core components
- Virtual servers for:
  - PXE/TFTP server, Kickstart, Cfengine
  - DNS & DHCP
  - Repository mirrors
    - SL, Glite. Jpackage, Local extras
  - Syslog-ng (All nodes and servers log to a central syslog-ng server)
  - Ganglia
  - Backup (All SE/MON databases dumped twice daily)
- All virtual servers backup to tape via Tivoli Storage Manager
- We are setting up our new CREAM-CE as a virtual server
- Xen 1500 MTU restrictions may prevent us virtualizing some services...
Network: Perth - Singapore path

The Perth-Singapore path
- The 620 megabit is split over four 155 megabit links
- Shortcomings in the routing hardware prevent data streams from passing down separate paths
- The path is currently degraded, one of the four links is missing

Currently 155 megabit limited!
Data transfers: Testing the U.S. path

**Speed**
- Ensuring we have sufficient bandwidth to transfer the ALTAS data is critical

**Testing the U.S. Path**
- We have currently switched to the non-preferred U.S. path for data transfers
- This allows us to test the path in the case that the Singapore path fails at any point, as one of it’s links has now

**Incoming data rates as seen by disk servers during February 2009**

~20MBytes/Sec is sustainable
Network:
The light at the end of the tunnel

The hardware on one of the Singapore links has been upgraded to STM-4, making a single link 655 megabit.
VeRSI is providing a coordinated and accelerated approach to the uptake of eResearch by Victorian researchers.

- Establishing relationships with the research community.
- Providing shared resources and infrastructure necessary for eResearch.
- Developing tangible examples of how eResearch can benefit research.
VeRSI

Synchrotron Science

Security & Access

Federated data storage

Capability Projects

Awareness & Outreach

http://versi.edu.au
VeRSI – eResearch Tools

Virtual Organisations for Collaboration

Elements of the VeRSI VOs

Identity and Authorisation / Shibboleth

Collaborative Tools and Applications

http://versi.edu.au
VeRSI – Enabling Technology

Data Storage

- dedicated data storage system for VeRSI projects

User Access to, and support for, VeRSI Services

- Access/user accounts: http://portal.versi.edu.au
- User support: http://support.versi.edu.au

Security & Access

- Via Shibboleth identity

Awareness

- VeRSI collaborators & researchers
- research institute
- research supporters
- the public.

http://versi.edu.au
VeRSI – Capability Projects

These capability projects assist through the development of storage repositories, portals, applications and instrumental interfaces for eleven research groups:

http://versi.edu.au
VeRSI – eResearch Demonstrator

Australian Synchrotron Virtual BeamLine

Allows researchers using the protein crystallography beam-line to:

- view the experiment from a location remote to the synchrotron.
- manipulate samples via motor controls, acquire data, transfer data between storage and compute resources.
- collaborate using voice, video and shared applications.

The VBL acts as a model for the design and development of VBLs on other experimental beamlines.
Some Comments

- Grid Solutions not yet widely embraced in Australia.
- High-end users still typically source their own compute and data storage needs.

HPC/Peak Computing centres develop:

- **New AUD$100M VLSCI Peak Compute Centre in Melbourne**
- **NCI(ANU) to replace Large SGI system with SUN system**
  - *(Bureau of Meteorology/CSIRO in same purchase – AUD$30M)*

In HEP we need large systems and Grid infrastructure.

- HEP community participating in (still limited) glite rollout.

Biomedical and Geoscience projects will benefit – data grid a benefit for large distributed data sets.

Grid projects part of larger push in eResearch.
Thank you