

# WORKSHOP:

## How to use HPC & Cloud Clusters to enhance your research outcomes

Paul Coddington, Deputy Director eRSA  
9 March 2016

## Workshop Schedule

9:30 – 10:40	Introduction to eRSA and Introduction to High Performance Computing
10:40 – 11:00	Break and refreshments
11:10 – 12:30	Using Tizard
12:30 – 1:30	Lunch break
1:30 – 3:00	Hands-on session <ul style="list-style-type: none"><li>• Unix tutorial</li><li>• Running jobs on Tizard</li></ul>
3:00 – 4:00	Help with running your own applications on Tizard

# Introduction to eRSA

- What is eResearch?
- eResearch services
- National eResearch initiatives
- Introduction to eRSA
- eRSA services

## What is eRSA?

- An incorporated joint venture of the three South Australian universities.
- The point of focus in S.A. for eResearch infrastructure and support.
- The South Australian partner in national eResearch initiatives.
- A mechanism for attracting funding for shared eResearch infrastructure to SA, e.g. through ARC LIEF grants and federal eResearch infrastructure programs



University of  
South Australia



THE UNIVERSITY  
*of* ADELAIDE



Flinders  
UNIVERSITY

The mission of eRSA is to **enable discovery, innovation and collaboration by providing eResearch facilities, services, training and expertise for South Australian researchers.**

## eRSA

- 20 staff, 15 technical staff
- Expertise in high-performance computing, data storage, data management, cloud, software development, sysadmin, user support
- Director reports to Board of 3 uni DVCRs and independent chair
- Offices at Uni of Adelaide Thebarton campus
- Facilities in data centres at Uni of Adelaide North Terrace campus and UniSA City West

## What is eResearch?

- The application of advanced Information and Communication Technologies (ICT) to research
- Usually in the context of accessing external resources (compute, data, web services, etc)
  - i.e. not being restricted by what can be done on a standalone personal computer
- Often called e-Science (e.g. in the UK)
  - but also used for the humanities, so the Australian government uses the term eResearch
- Has grown out of a number of advances in distributed computing, digital data repositories, web applications, grid/cloud computing, high-speed networks, etc

## Why eResearch?

- We are in the digital information age
  - computers are changing almost all aspects of society, including how research is done
- New electronics technologies and digitization of information have led to a “data deluge”
  - many instruments can now generate TBytes/day
- Research is increasingly collaborative, across disciplines and geographic locations
  - increasing need for remote communication and data sharing, enabled by high-speed networks, the Internet and the web
- Computational science and simulation
  - third paradigm augmenting theory and experiment



## eResearch Services

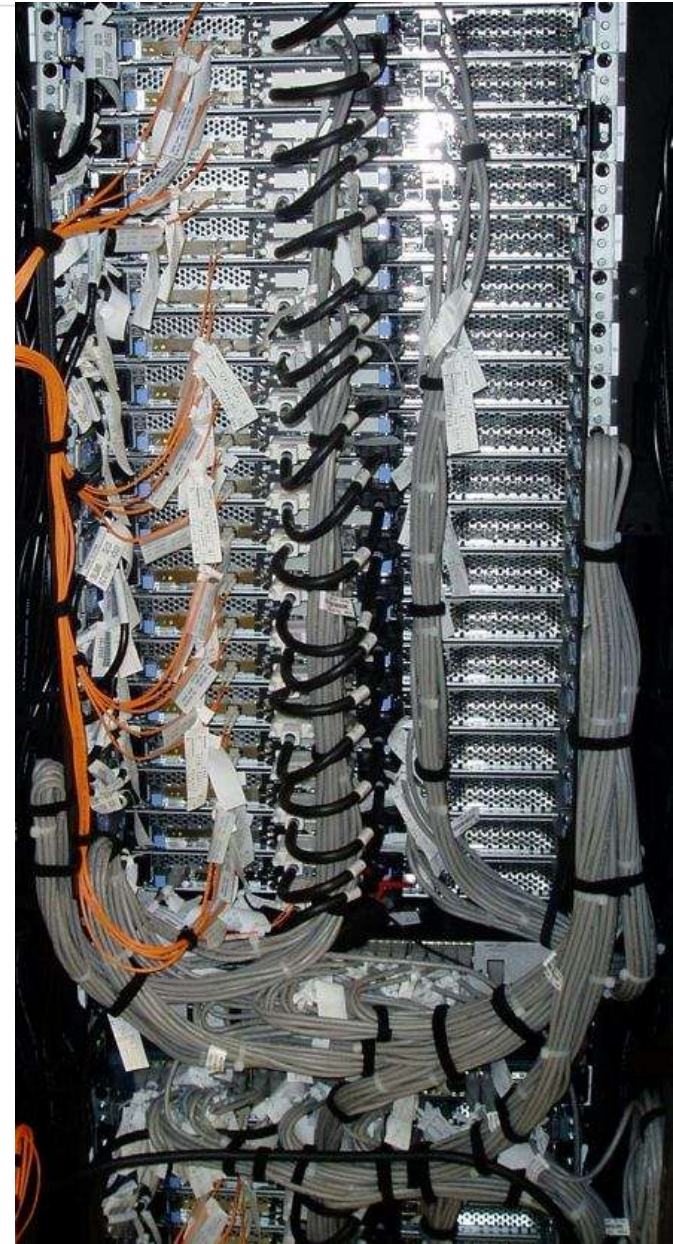
- Research collaboration tools
- Videocollaboration
- Data storage
- Data management and sharing
- High-performance computing
- Cloud computing
- Web applications
- Visualization

# Data Management and Sharing

- Management and sharing of research data is currently often rather ad-hoc
  - Data stored on researchers' PCs, CDs, DVDs
  - Sharing via email, DVD, hard disk, etc
- This works up to a point, but is not scalable
- Can't guarantee data will be accessible or understandable long-term, e.g. if researchers or students leave
- Want sustainable data repositories
  - data online, easily discoverable and accessible, backed-up, with associated metadata
- International and national moves to require that data from publicly-funded research be made publicly available

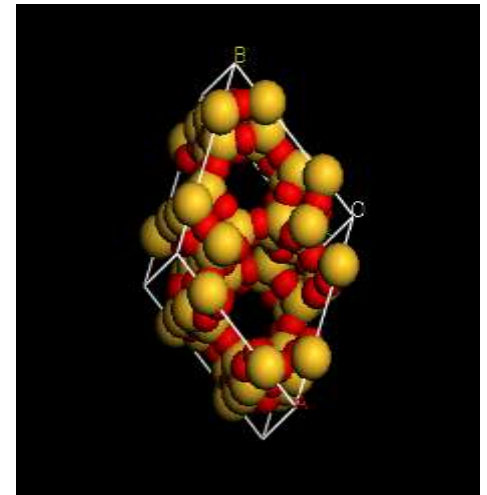
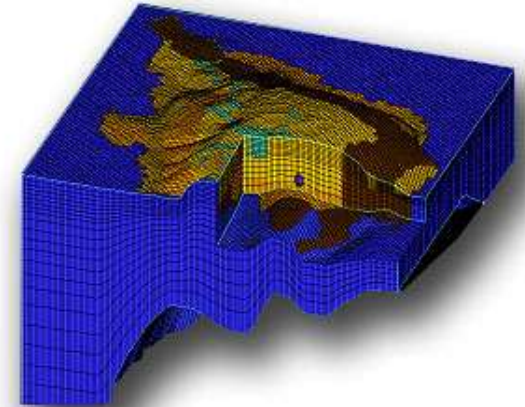
# High-Performance Computing

- Supercomputers provide hundreds or thousands of times more compute power and memory than a PC
- HPC enables much faster data processing, analysis, simulation and modelling
- Also enables researchers to tackle larger, more complex problems than could otherwise be done
- Has been used for many years in physics, chemistry, engineering
- Recently much broader usage, e.g. molecular biology, analysis of large data sets in many fields



# Visualization

- It is often difficult for humans to extract information from large data sets without having some visual representation of the data
- Visualization tools can help a researcher (or a group of researchers) understand their data, and explain their results to others in a convincing and engaging way
- Stereoscopic visualizations of 3D objects can help understand them, e.g. molecular structure and function of proteins, building design



## National eResearch Initiatives

- Australian National Data Service (ANDS)
  - Processes and tools for managing data, especially the creation of searchable metadata
- Research Data Storage Infrastructure (RDSI) and Research Data Services (RDS)
  - Storage and services for research data
- National eResearch Collaboration Tools and Resources (NeCTAR)
  - Research cloud infrastructure and tools
- National Computational Infrastructure (NCI)
  - National supercomputing facilities

## eRSA and National Initiatives

- eRSA is involved in these national initiatives, usually as the SA partner
- SA member of AeRO (Australian eResearch Organisations)
- Involved in several ANDS projects
- SA node of RDSI
  - Multi-Pbyte storage facility
- SA node of NeCTAR Australian Research Cloud
  - 3000 core node of 30,000 core national cloud

## eRSA Services

- Large-scale data storage facilities
- Data sharing and data repositories
- High-performance computing facilities
- Hosting of dedicated compute facilities
- Cloud computing and virtual machines
- eResearch expertise and consulting
- Helpdesk, user support and training
- Software development
- Development and/or hosting of web applications
- Visualization services

How much does it cost?

Nothing!\*

\*

Some conditions apply ...



# Funding

- Main funding for eRSA comes from:
  - The three S.A. universities
  - S.A. Government
  - National eResearch funding schemes (RDSI, RDS, NeCTAR, ANDS, etc)
  - ARC Linkage Infrastructure, Equipment and Facilities (LIEF) grants for facilities
- eRSA also does contract work, e.g. development of software and services, consultancies, etc for university researchers and external organizations

## Fees

- This funding enables eRSA to provide most of its basic services at no cost to researchers from the 3 SA universities or their collaborators
- Services that are not free are usually at low cost for uni researchers. This includes:
  - Dedicated compute nodes or facilities
  - Software or database development
  - Significant ongoing dedicated support effort

# Data Storage and Management

- eRSA services allow researchers to:
  - store your data using reliable, professionally-maintained storage systems
  - access the data easily from the desktop or the web
  - back-up large amounts of data
  - integrate your data with related national or international data repositories
  - publish your data so it can be easily discovered and accessed
  - enable data sharing across distributed locations
  - simplify the data archiving process

## Data Storage and Management

- Recently significant increased eRSA storage and backup capacity through RDSI project
  - 3 Petabytes disk storage
  - 1.5 Petabytes Tape silo
  - Extensible as needed
  - Accessible from HPC systems
  - Internet access for sharing
- Standard mechanisms for data access from PC
  - sftp, scp, rsync, http
  - Other mechanisms for simpler access in train



## Cloud computing

- Cloud provides on-demand *virtual machines*
- Faster and easier to provision and manage than your own physical server
- Nectar Australian Research Cloud provides cloud infrastructure for research
  - Similar to Amazon, Microsoft, Google, etc clouds
  - 3000 cores in SA, almost 30,000 nationally
- Virtual servers to run web sites, applications, databases, compute jobs, etc.
- Large, on-demand, elastic compute resource

## HPC Facilities at eRSA

- Great success in procuring world-class HPC for SA
  - 1999 - Perseus - largest cluster in Australia
  - 2000 - Orion - #1 in Australia
  - 2003 - Hydra - #2 in Australia
  - 2004 - Aquila - large shared memory
  - 2007 - Corvus - #2 in Australia
  - 2012 - Tizard
- Supported several hundred users in the past 10 years
- Over 100 active users every quarter



## HPC Users and Applications

- Computational physics
- Computational chemistry, e.g. molecular structure
- Life sciences, e.g. genomics, proteomics, phylogenetics
- Computational Fluid Dynamics
- Geology and Geophysics, e.g. seismic analysis
- Materials science
- Petroleum engineering
- Electrical and mechanical engineering
- Optimization problems
- Computer science, e.g. parallel algorithm development
- Economics
- And many others...



## Software development

- eRSA has a software development team
- Specialise in developing custom eResearch services or service integration, web application development, web databases
- Provide an existing, experienced team for research software development



## Getting Started

- Talk to us about your eResearch requirements and how we can help you.
- Check the information on our website [www.ersa.edu.au](http://www.ersa.edu.au)
- Contact our service desk [servicedesk@ersa.edu.au](mailto:servicedesk@ersa.edu.au) and the appropriate person will contact you to discuss your requirements
- Talk to the Deputy Director, Paul Coddington  
[paul.Coddington@ersa.edu.au](mailto:paul.Coddington@ersa.edu.au)
- Fill in a account request form on our website



**e R S A**

Advancing Research Innovation

## Upcoming workshops

**How to use HPC and Cloud Clusters to enhance your research outcomes**

*9 March, University of Adelaide*

**R-Studio in the Cloud**

15 April, University of Adelaide

*Talk to us after the workshop to register*