Research Data Facility (RDF)

Introduction and Layout

Andy Turner, EPCC
a.turner@epcc.ed.ac.uk

Reusing this material

www.epcc.ed.ac.uk
www.archer.ac.uk

Outline

- ARCHER/RDF
  - Layout
- Data Analytic Cluster (DAC)
  - Hardware
  - Software
  - Visualisation
  - Running Jobs
- Data Transfer Nodes

ARCHER and RDF

ARCHER

- UK National Supercomputer
- Large parallel compute resource
  - Cray XC30 system
  - 118,080 Intel Xeon cores
  - High performance interconnect
- Designed for large parallel calculations
- Two file systems
  - /home – Store source code, key project data, etc.
  - /work – Input and output from calculations, not long-term storage

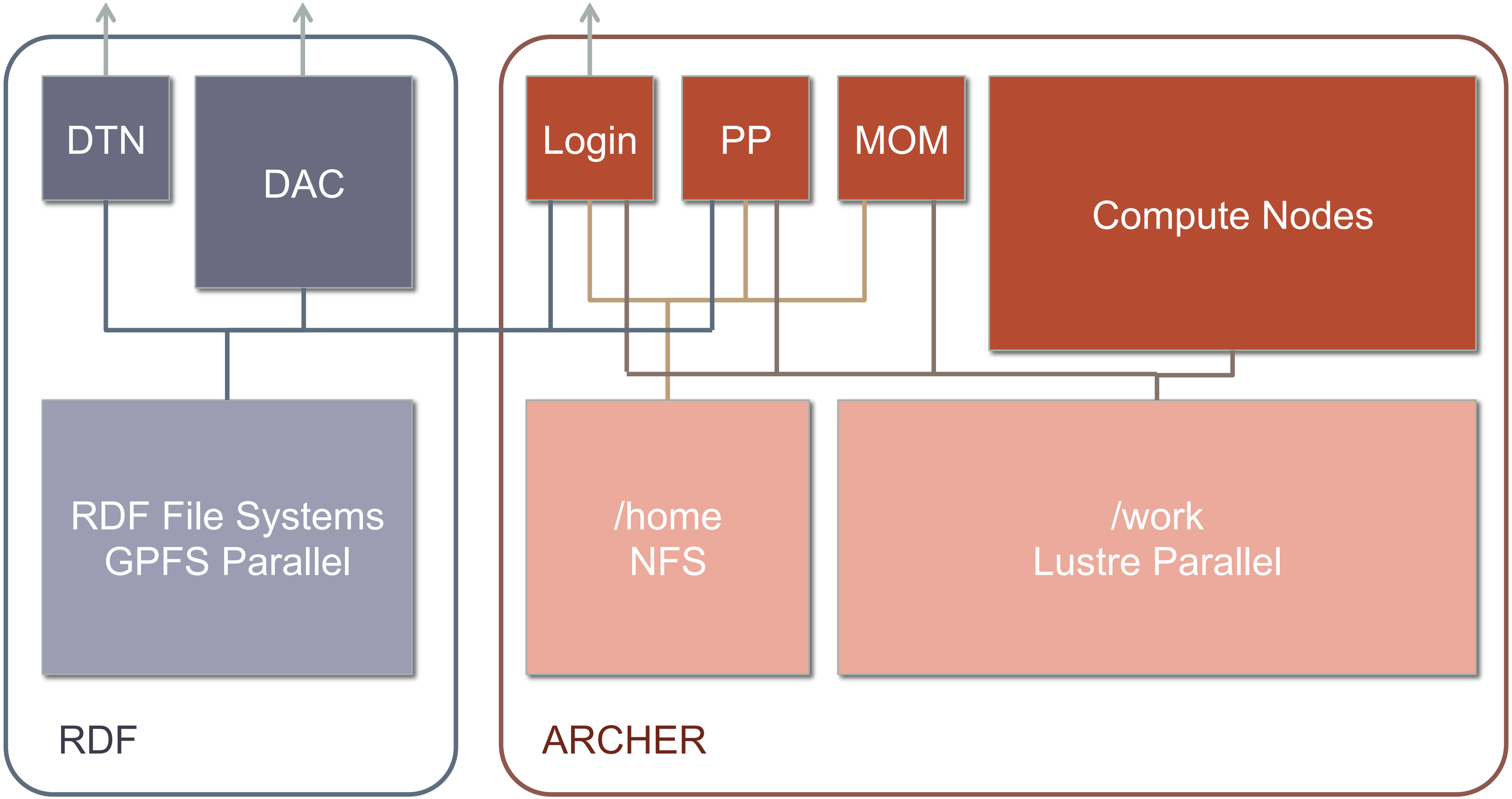
RDF

- Large scale data storage (~20 PiB)
  - For data under active use, i.e. not an archive
  - Multiple file systems available depending on project
- Modest data analysis compute resource
  - Standard Linux cluster
  - High-bandwidth connection to disks
- Data transfer resources

Terminology

- ARCHER
  - Login – Login nodes
  - PP – Serial Pre-/Post-processing nodes
  - MOM – PBS job launcher nodes
  - /home – Standard NFS file system
  - /work – Lustre parallel file system
    - ARCHER installation is a Sonexion Lustre file system
- RDF
  - DAC – Data Analytic Cluster
  - DTN – Data Transfer Node
  - GPFS – General Parallel File System
    - RDF parallel file system technology from IBM
    - Multiple file systems available on RDF GPFs

Overview


DTN
DAC
Login
PP
MOM
Compute Nodes
RDF File Systems
GPFS Parallel
/home
NFS
/work
Lustre Parallel
RDF
ARCHER

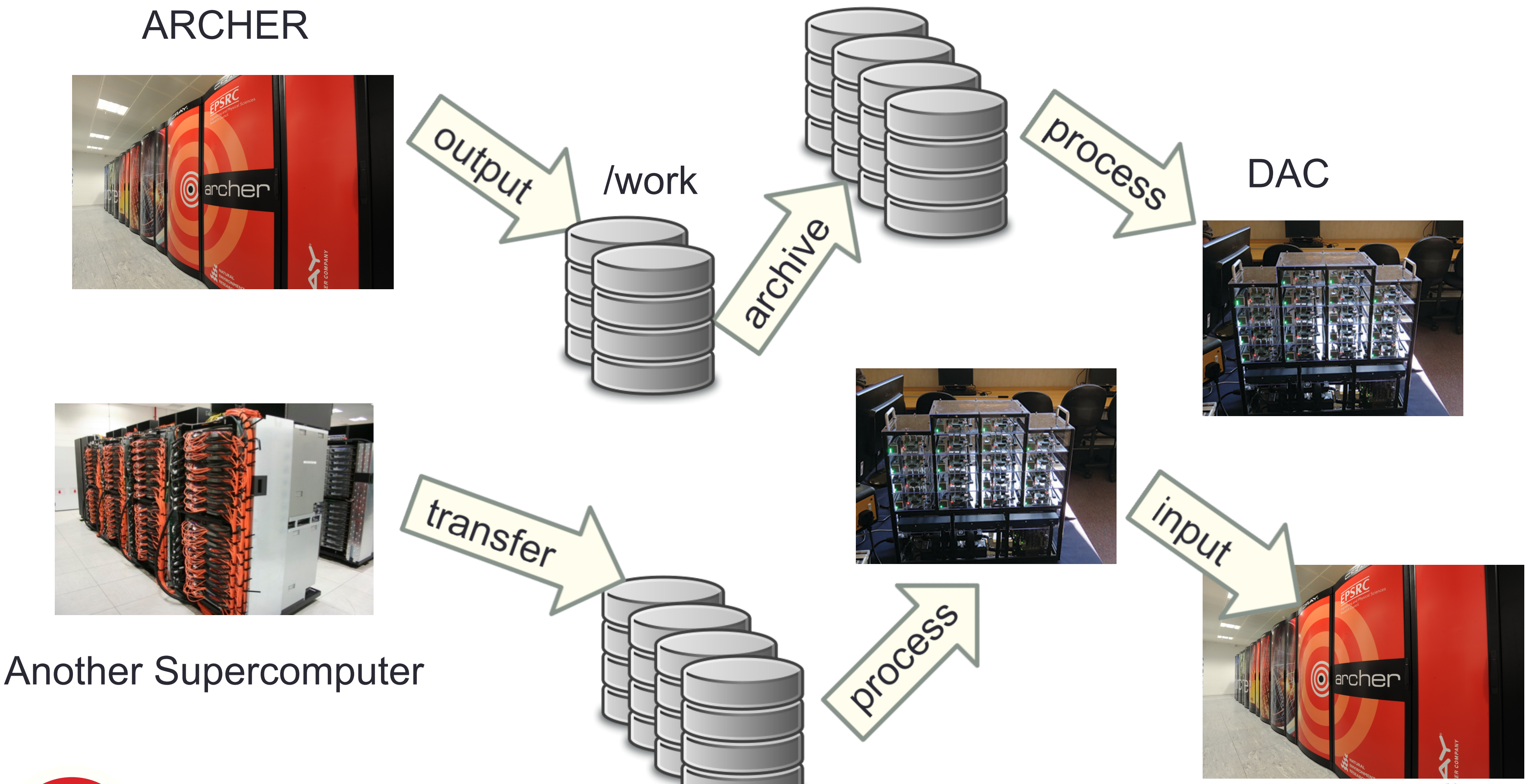
Data Analytic Cluster (DAC)

login.rdf.ac.uk

Hardware

- 1 login node
  - two Intel Ivy Bridge 10-core processors, 128 GB memory
- 12 standard compute nodes
  - two Intel Ivy Bridge 10-core processors, 128 GB memory
- 2 high-memory compute nodes
  - with four Intel Westmere 8-core processors, 2 TB memory
- HyperThreads are enabled on all nodes
  - standard compute nodes each have 40 CPUs available
  - high-memory compute nodes each have 64 CPUs available.
- All DAC nodes have high-bandwidth, direct Infiniband connections to the UK-RDF disks.

DAC use cases



RDF

ARCHER

/work

DAC

Another Supercomputer

Why use the DAC?

- Fastest connection to RDF disks
  - much faster than ARCHER

- Fast connection to external networks
  - via DTN nodes
  - e.g. PRACE network, NERC Jasmine system

- Easier and more flexible than ARCHER compute nodes
  - more powerful than ARCHER post-processing nodes
  - currently free to use!

Software - Compilers and MPI

- GCC
  - gcc – C
  - gfortran – Fortram
  - g++ - C++
- OpenMP
  - compile and link with –fopenmp flag
- MPI – OpenMPI library
  - Module: “openmpi-x86_64” or “openmpi/1.10.2-gcc-5.1.0”
  - compile: mpicc, mpif90, mpic++
  - run: mpiexec –n <nproc> --oversubscribe mympiprogram

Software - Python

- Python 2.* available via the Anaconda distribution
  - module load anaconda

- Python 3 also available
  - module load anaconda/2.2.0-python3

- Parallel python
  - MPI provided by anaconda: from mpi4py import MPI
  - load normal MPI module
  - mpixec –n 4 python myjob.py

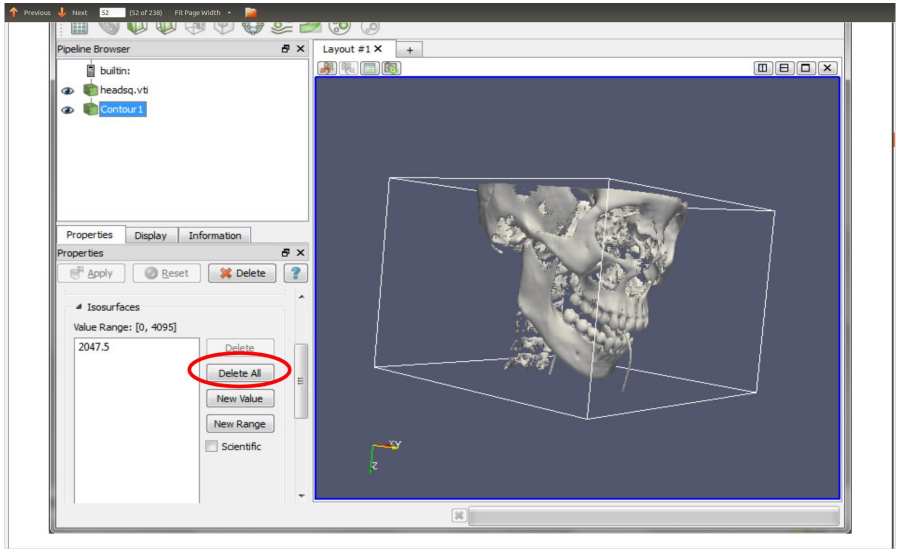
Other software

- Statistics
  - "R" is available by default (no module)
- Data Formats; HDF5 and NetCDF
  - serial versions available by default
  - parallel hdf5 available via standard wrappers, e.g. h5pcc and h5pfc
  - parallel netcdf requires a module + flags – see documentation
- Linear algebra
  - BLAS and LAPACK available by default
  - for parallel, link with: -lmpiblacs -lscalapack

Visualisation – Paraview/VisIt

- Paraview and VisIt available

- Can also be used for parallel visualisation



- Paraview works in client/server mode
  - run paraview GUI as a client
  - run parallel paraview server “pvserver”
  - connect the two via a socket

Paraview - Parallel Visualisation

- See http://www.archer.ac.uk/documentation/rdf-guide/cluster.php#paraview

```
 -bash-4.1$ hostname rdf-comp-ns10
 -bash-4.1$ qsub -IXV -lwalltime=3:00:00,ncpus=16
 -bash-4.1$ module load paraview-parallel
 -bash-4.1$ mpirun -np 16 pvserver --mpi --use-
offscreen-rendering --reverse-connection --server-
port=11112 --client-host=rdf-comp-ns10
```

- Assumes a paraview GUI listening on port 11112
  - run GUI on the login node
  - see: File -> Connect

Paraview - Remote visualisation

- Exporting graphical display slow over network

- Assuming you have paraview on your laptop ...
  - run GUI locally
  - connect to parallel pserver running on DAC

- Requires *port forwarding*
  - see http://www.archer.ac.uk/documentation/rdf-guide/cluster.php#portfwd
  - some compatibility restrictions on paraview versions ...

module load anaconda/2.2.0-python3

Running Jobs – Batch system

- Torque batch system
  - Similar to PBS – qsub, qstat, qdel, …
- Request walltime and cores
  - #PBS -l ncpus=1
  - #PBS -l walltime=1:0:0
- Specify project (use is uncharged):
  - #PBS -A t01
- Jobs cannot use multiple nodes
  - Max. of 40 cores on standard nodes
  - Max. of 64 cores on high memory nodes

module load anaconda/2.2.0-python3

Running Jobs - Interactive access

- Often useful to have a shell on the compute nodes
  - testing
  - debugging
  - visualisation
  - ...
- Submit an interactive job, e.g.
  - qsub -IXV -lwalltime=3:00:00,ncpus=16
  - wait for prompt ...
- Notes
  - you start off back in your home directory
  - remember to reload your modules!

Data Transfer Nodes (DTNs)

dtn01.rdf.ac.uk
dtn02.rdf.ac.uk

Moving Data – Supported Protocols

- Basic serial transfers:
  - scp/sftp
- Parallel transfers:
  - bbcp
- Certificate-based methods:
  - GridFTP
  - Globus Online
- In all cases, software must also be installed at remote end
- Parallel methods can give performance but are more difficult to set up

Summary

Summary

- RDF provides complimentary functionality to ARCHER
  - Large disk resource
  - Data analytic capability
  - Data transfer
- Data Analytic Cluster
  - Fast I/O performance
  - Standard tools and codes
- Data Transfer Nodes
  - High-bandwidth network connections
  - Variety of data transfer software