The University of Manchester CIR Campus Ecosystem — User View

What is the Ecosystem?

The Computationally-Intensive Research Ecosystem is a response to feedback asking for an *integrated* system of infrastructure designed to address all aspects of research groups' computational work and requirements. It comprises:

- traditional batch computational facilities; a facility for interactive computation, e.g., for development work;
- high-capacity, resilient storage;
- computational facilities and storage linked via a dedicated, secure, fast network for data-transfer;
- a virtual machine service for research groups;
- a cluster for sharing emerging tech. hardware.

Local Versus Centrally-Run Infrastructure

Before 2010, many small "beowulf" HPC clusters existed on campus. Some were well-run by academics and postgrads; others were not. All took time to administer which was better spent on research; most had many "spare" CPU cycles. Since then, most such beowulfs have been decommissioned and contributions have been made to the CSF instead — "HPC" infrastructure has been centralised. Academics now have access to a professionally-run, campus service, with all the benefits that brings.

Following the success of this strategy, academics are now encouraged to make use of, and contribute to — buy into — other centralised research infrastructure run by IT Services, as introduced here.

The Research Infrastructure Team

The RI Team administer IT infrastructure for computationally-intensive research (CIR) — many of the facilities mentioned here: The CSF, Redqueen, The iCSF/Incline and Zrek; also the Research Virtual Desktop Service, SSH gateway and SSHFS service. In addition, we are the business owner of the Research Data Storage Service, Research Data Network and the Research Virtual Machine Service. If you are interested in finding out more about any of these services, please email

> its-ri-team@manchester.ac.uk

Alternatively, please visit our Web site

> http://ri.itservices.manchester.ac.uk/
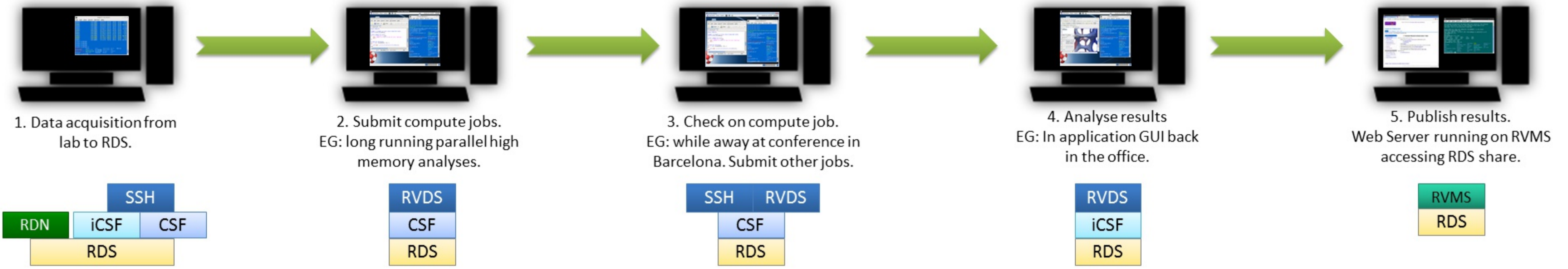
Tailored Workshops

We also offer workshops tailored to individual research groups aimed at increasing researchers' productivity by use of the facilities introduced here.

The EPS Faculty IS Team

The Condor pool is maintained and supported by IT Services in the Faculty of EPS.

The FLS Faculty IS Team

Kadmon is maintained and supported by the FLS Faculty IS Team.

---

1. Data acquisition from lab to RDS.

SSH	
RDN	iCSF	CSF
RDS		

2. Submit compute jobs. EG: long running parallel high memory analyses.

RVDS
CSF
RDS

3. Check on compute job. EG: while away at conference in Barcelona. Submit other jobs.

SSH	RVDS
CSF	
RDS	

4. Analyse results EG: In application GUI back in the office.

RVDS
iCSF
RDS

5. Publish results. Web Server running on RVMS accessing RDS share.

RVMS
RDS

Example Use Case

1. Rusty's data exits his group's DNA sequencer straight onto storage provided by the RDS, over fast, dedicated networking infrastructure (the RDN). The data is now visible on the CSF and iCSF.

2. By using the Research Virtual Desktop Service (RVDS), he defines a series of computational jobs to process the data and submits them to the batch system.

3. Later, from home, Rusty re-connects to his RVDS session to monitor his jobs to ensure all is well — or make any necessary tweaks. Over the next few days, from a conference in Barcelona, Rusty checks

progress again using the RVDS, from his laptop, and also the SSH Gateway, from his phone; he clears some jobs which have failed and submits additional, corrected work.

4. Back at the office, batch jobs finished, and using the same RVDS session, Rusty starts GUI-based, interactive post-processing on Incline (aka the iCSF) — no need to move data as all the same RDS-based filesystems are available on all ecosystem compute resources.

5. Finally, the results are ready and made available to the public via a Web server running on the Research Virtual Machine Service (RVMS) — accessing the same RDS share.

Working from Office, Home and Barcelona

Components of the CIR ecosystem are accessible directly only from on the University campus. Off-campus access is supported by a variety of globally-accessible services:

SSH Gateway

University staff and postgrads can login to the SSH gateway from anywhere in the world and then hop to any component of the campus ecosystem.

SSHFS (experimental)

RDS storage is not accessible off-campus. An experimental SSHFS service offers off-campus access to RDS shares used on ecosystem computational facilities. (We are not able to offer off-campus access to other RDS shares currently.)

Research Virtual Desktop Service (RVDS, experimental)

This service allows users to: access the CSF (Danzek), the iCSF (Incline), Redqueen and Zrek from off-campus; do interactive/GUI-based work over relatively slow connections; and re-connect to the same desktop session from office, home and elsewhere.

Campus CIR Ecosystem

Research Virtual Machine Service

A virtual machine service is planned for research groups on which PIs and their team may have OS administrator/root privileges. This will be located on the same resilient, professionally-managed infrastructure as other IT Services VMs. It is expected that the primary use of this service will be for public-facing Web servers which access RDS data.

(For licensing reasons, only Linux is available.)

Campus Gateways		
X2GO / NX	Cmd-line	"mounts"
RVDS	SSH	SSHFS

RVMS → VMs

Campus only

Campus Network

Computation — Interactive and Batch

There are five computational facilities in the UoM Campus CIR Ecosystem:

The Computational Shared Facility (CSF, aka Danzek)

The Computational Shared Facility is the University flagship "HPC" cluster. It is used for a wide variety of work: parallel computation using many CPUs or GPGPUs; high-throughput work (running lots of small jobs); work requiring large amounts of memory (RAM) or access to high-capacity (disk) storage with fast I/O.

The CSF is primarily used for "batch" computing.

The Interactive CSF (iCSF, aka Incline)

The iCSF, aka Incline, is designed specifically for computationally-intensive interactive and GUI-based work — hence the name: the interactive-CSF. It is expected that Incline will be used closely with the Research Virtual Desktop Service (see *Gateways to the Ecosystem*).

The EPS Condor Pool

Condor is a system that makes use of hundreds of desktop PCs around the University to provide a high-throughput computing environment. It is freely available to all researchers at the University and complements more traditional cluster systems such as the CSF. Users of the CSF can choose to keep their data on the CSF/Isilon and submit jobs to Condor via a CSF-Condor gateway server.

Condor	CSF	Redqueen	iCSF	Zrek
Submitter	Job Queue	Job Queue	Interactive	Job Queue
Campus PC clusters	Compute nodes	Compute nodes	Compute nodes	FPGA, GPU, Xeon Phi

Redqueen

Redqueen is a smaller "HPC" cluster which is used by the RI Team for testing and development work. It is also available to researchers. We are sometimes able to accommodate requests on Redqueen which cannot be met on the CSF, for example, running of exceptionally-long jobs (over a week).
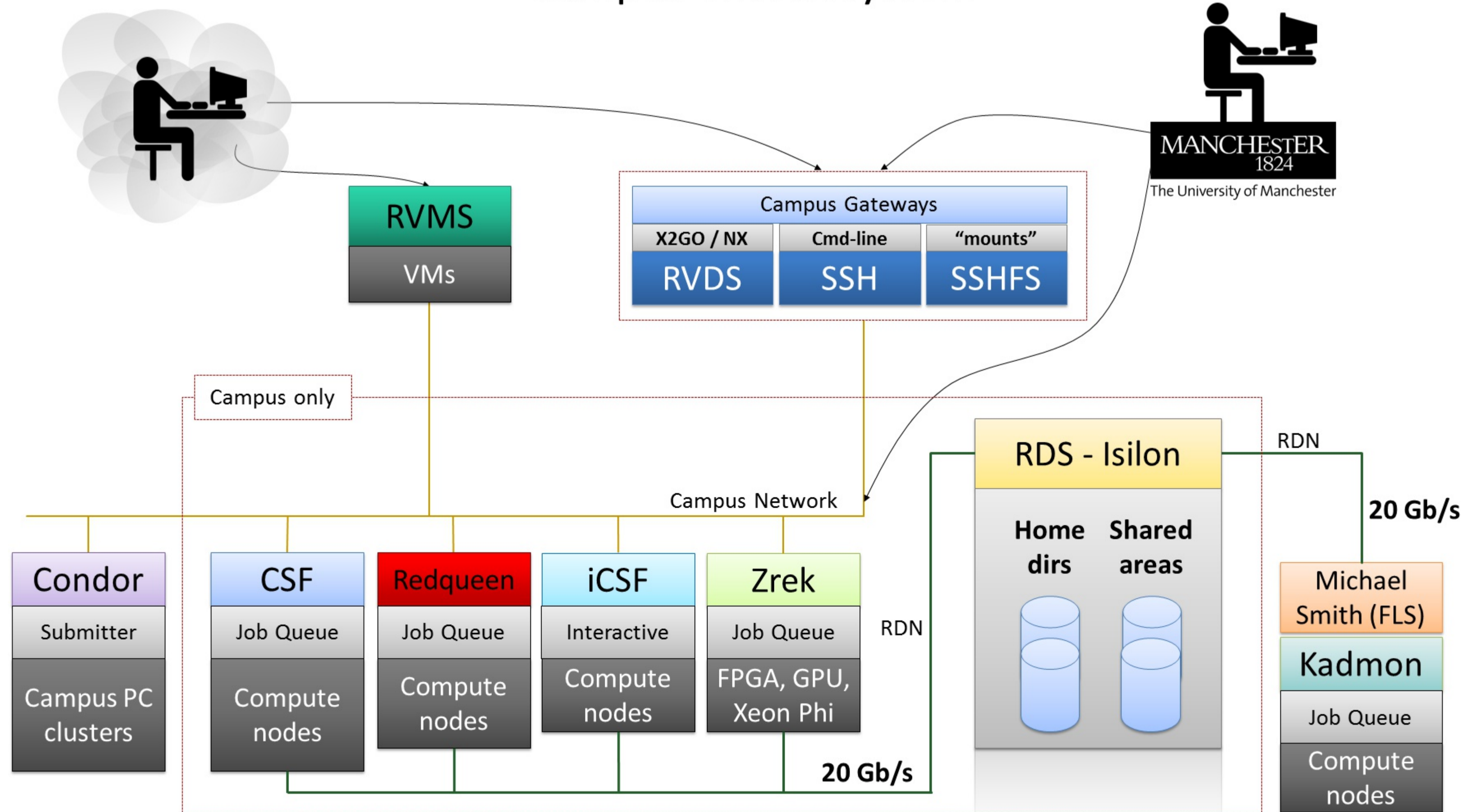
Zrek — Emerging Technology Cluster

We currently administer several, disparate servers hosting new and emerging technologies including GPUs, FPGAs and Xeon-Phi co-processors. These will soon be brought together to form a loose-knit cluster.

Kadmon

Kadmon is an FLS-specific, contribution-based cluster.

RDS - Isilon

RDN

Home dirs	Shared areas

20 Gb/s

Michael Smith (FLS)

Kadmon → Job Queue → Compute nodes

RDN

20 Gb/s

Storage and Data Transfer

All computational facilities within the ecosystem access common filesystems, i.e., users see the same files when they login to each:

- each user has the same home-directory on all facilities;
- large shared-areas for data storage are available to research groups as part of the RDS (see below);
- a fast, dedicated network (the RDN, see below) links the RDS and the computational facilities so that transfer of large datasets is possible.

The Research Data Storage Service (RDS, aka Isilon)

IT Services provides centrally-hosted and administered data storage for research staff and students. Some storage is available to each academic-led research project at no charge. Further storage will be charged for. This storage is commonly referred to as Isilon.

The storage provided by this service is accessible from desktop and laptop machines on campus and may also be accessed from on-campus research computing systems (including the CSF and the iCSF). For off-campus access, use the VPN or the SSHFS service.

Files stored on this service can be considered secure. For example, files corrupted or accidentally deleted can be recovered for up to 28 days.

To find out more about the RDS, please visit

> http://www.rds.itservices.manchester.ac.uk/

The Research Data Network (RDN)

Many users of CIR have large quantities of data which must be moved from experimental instrument to RDS and/or from RDS to computational cluster. This requirement is satisfied by the RDN which connects all nodes on all facilities within the ecosystem to the RDS using fast dedicated hardware on a secure network.

MANCHESTER 1824
The University of Manchester