



Science and
Technology
Facilities Council

JASMIN



Natural
Environment
Research Council

JASMIN Overview

Dr Matt Pritchard

JASMIN Operations Manager

What is JASMIN for?

JASMIN

Supports data analysis for NERC environmental science community

- Large scale, data-intensive science

Designed for performance

- Tailored to needs of academic community

Compute co-located with the data

- CEDA Archive data (curated)
- Self-managed Group Workspaces (not curated)

Flexible compute capabilities

- Interactive and batch compute
- JASMIN Cloud provides autonomy, scalability



How does JASMIN achieve this?

JASMIN

Operated by STFC on behalf of NERC

- Architecture: CEDA & SCD
- Physical infrastructure: SCD
- User services: CEDA

Large-scale, high-performance storage
and compute

- Centered more around **storage & data analysis** than traditional supercomputer
- More flexible environment than supercomputer
- Wide range of workflows

CEDA

Centre for Environmental Data Analysis (RAL Space)

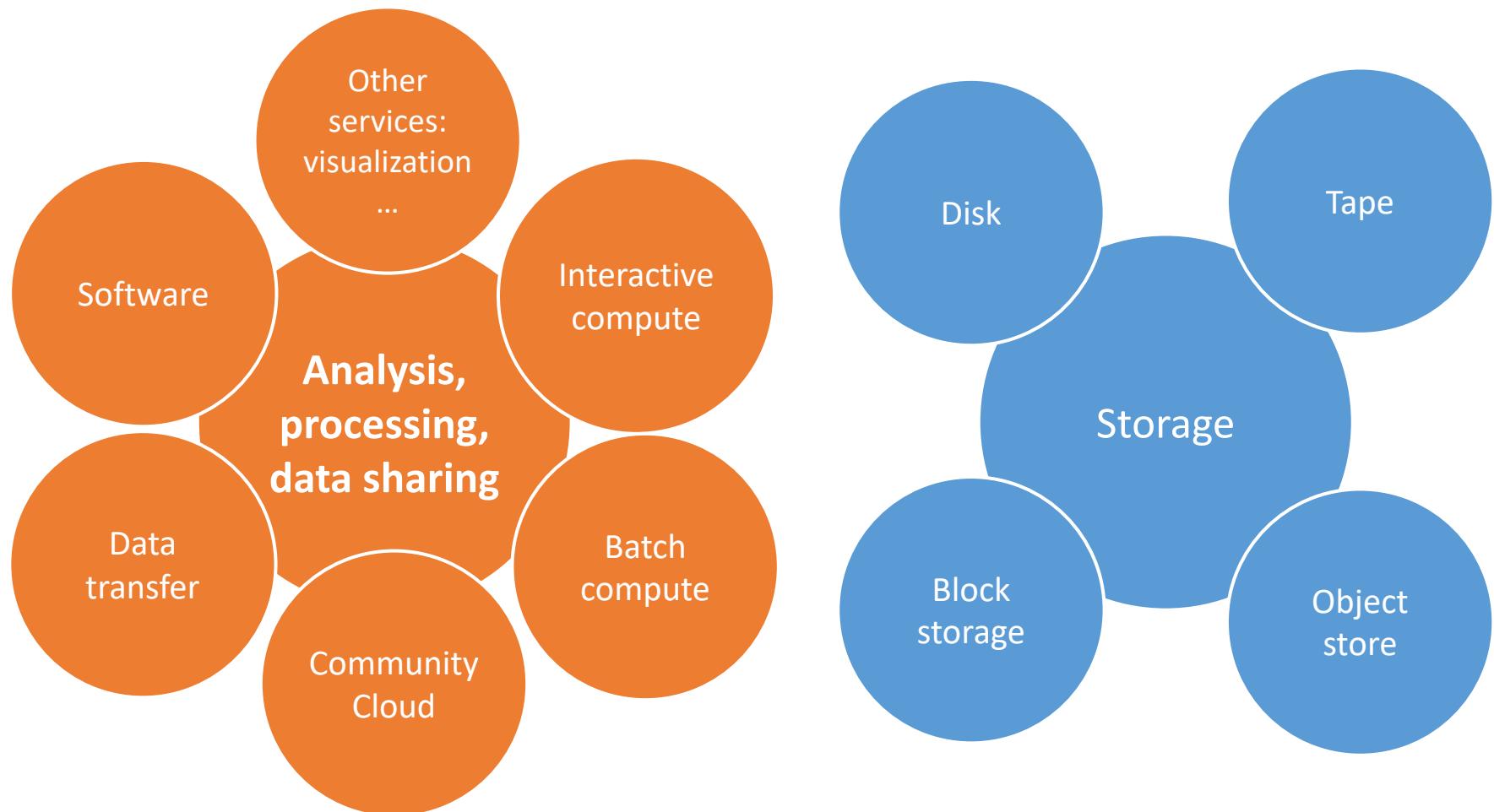
SCD

STFC Scientific Computing Department



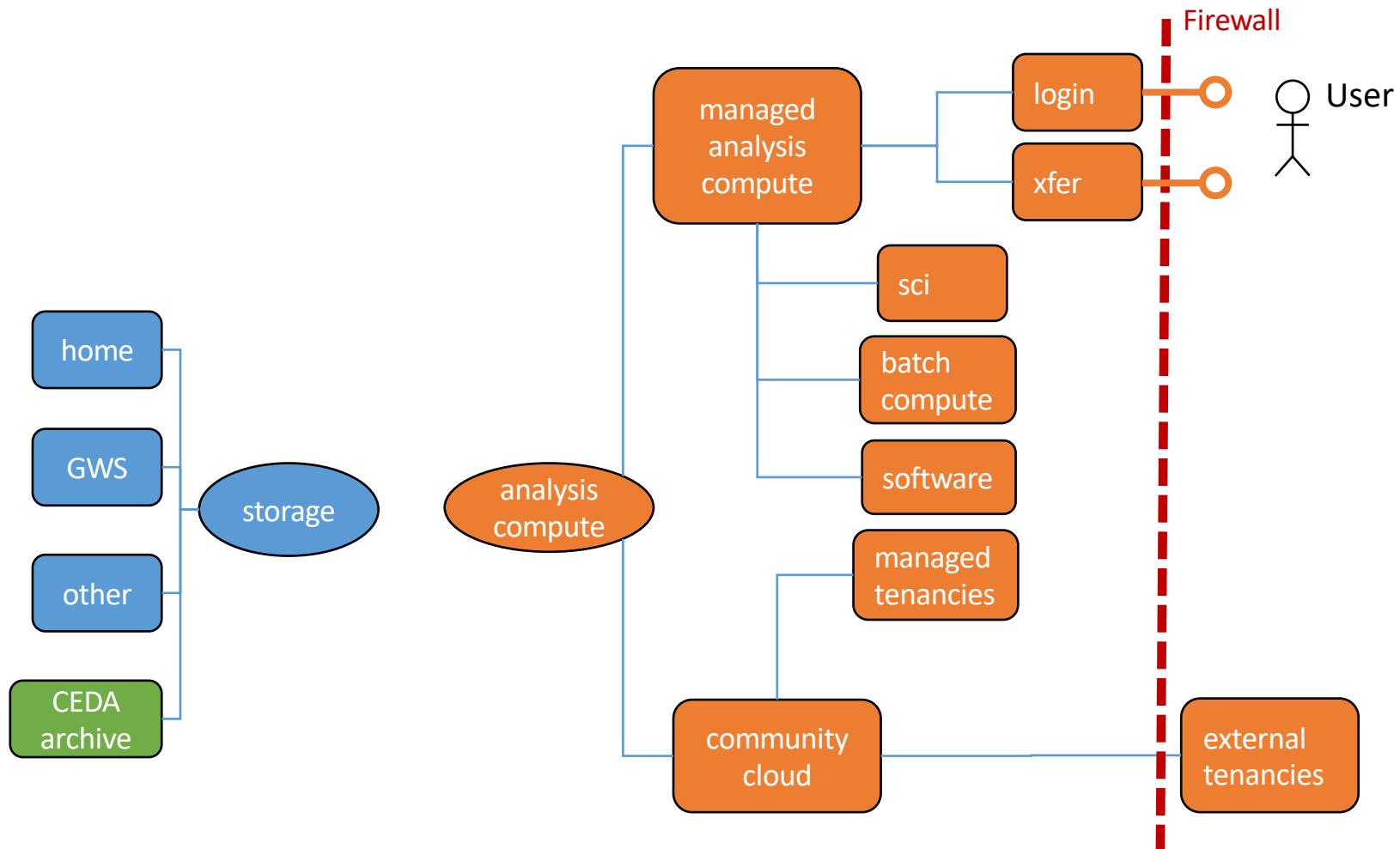
JASMIN components - 1

JASMIN

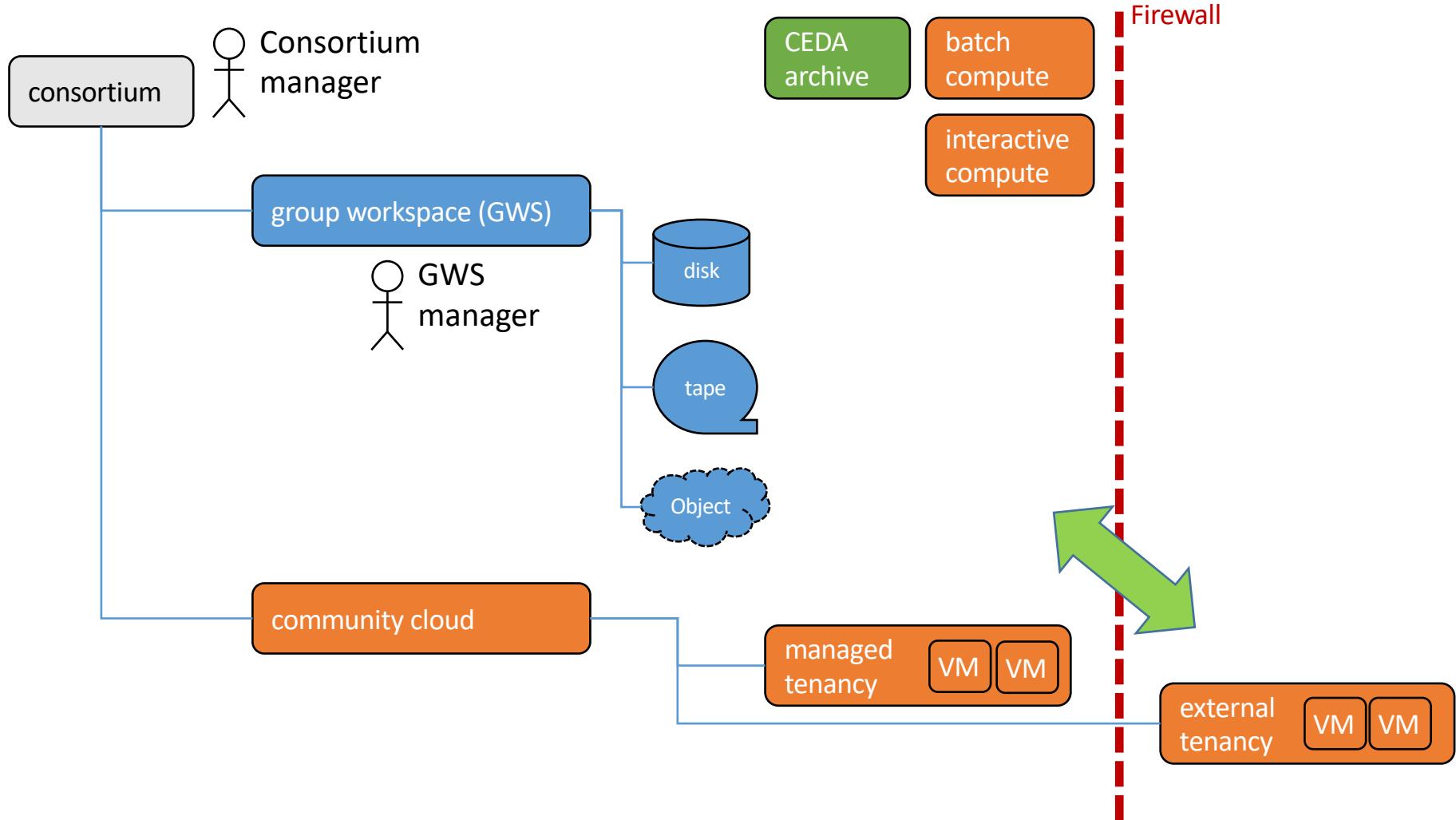


JASMIN components - 2

JASMIN



Resources for Projects

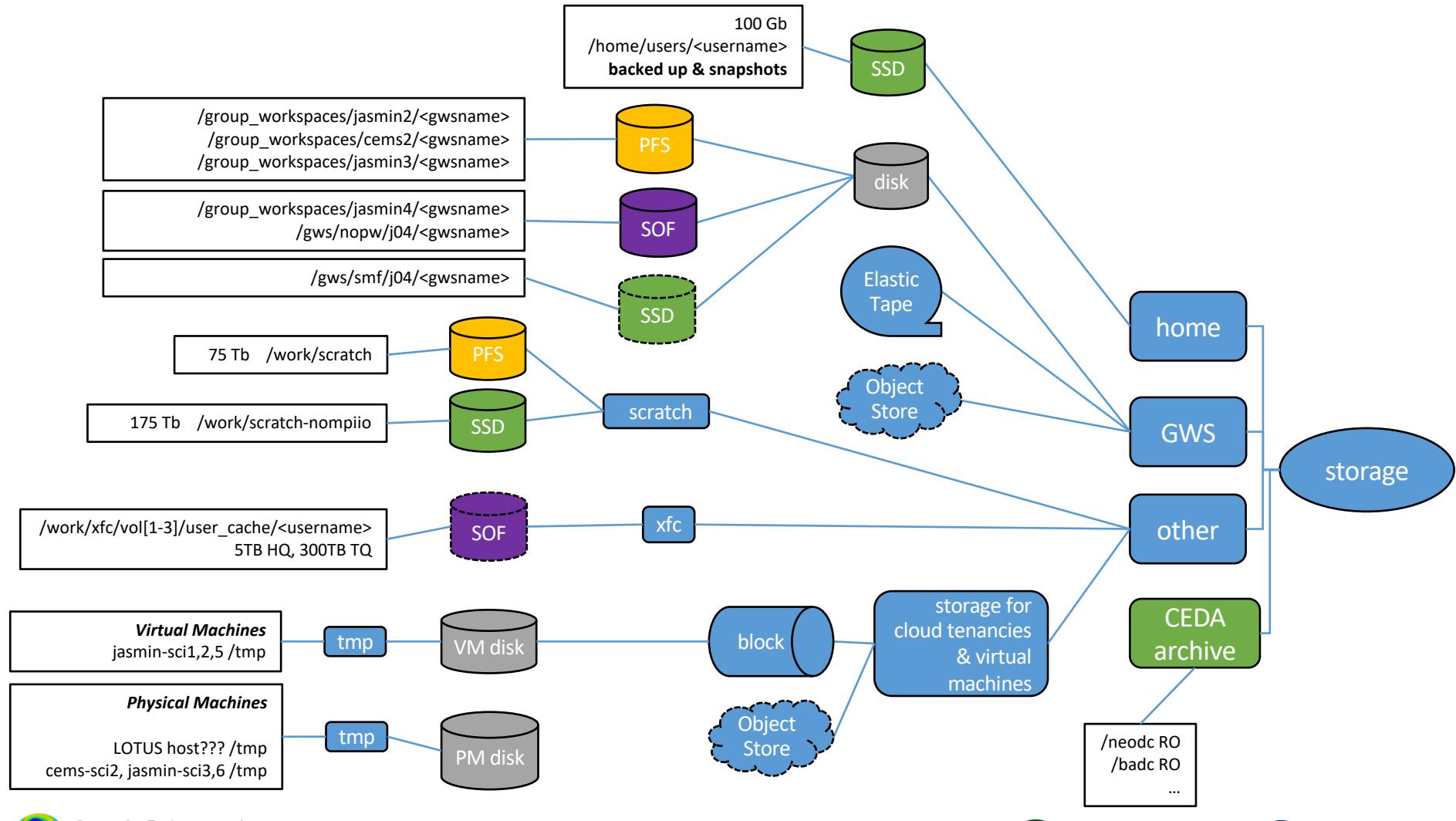


Storage types

	Features	PFS	SOF	SSD
disk	Parallel R	✓	✓	✓
	Parallel W	✓	✗	✗
	Small files	✗	✗*	✓
	Mount type	permanent	automount	automount
tape	Near-line tape	Interfaces for: CEDA Archive data (staging & backup) JASMIN Group Workspace data (Elastic Tape)		
Object	Object store	S3 interface, like Amazon S3 Several planned uses		
block	Block storage	Used for provisioning virtual machine disks JASMIN virtualization estate JASMIN community cloud Other specialist storage : databases		

* Coming soon

Storage deployments



Collaborative workspace storage provided to a “Project”

- GWS Manager is responsible for
 - Data (not CEDA/JASMIN team)
 - Granting access to their GWS (not CEDA/JASMIN team)
- **Not backed up**
 - Use high-performance, reliable storage systems
 - But tools provided for managing secondary copy (Elastic Tape, JDMA)
- For science-related data only
 - No personal / sensitive / confidential data (true for all of JASMIN)
- Not permanent
 - For duration of project
 - Speak to CEDA Archive team for long-term curation of data
- Consortium Manager
 - Responsible for allocating Projects with GWS resources (CEDA implements!)

<https://help.jasmin.ac.uk/category/195-short-term-project-storage>

What's mounted where?

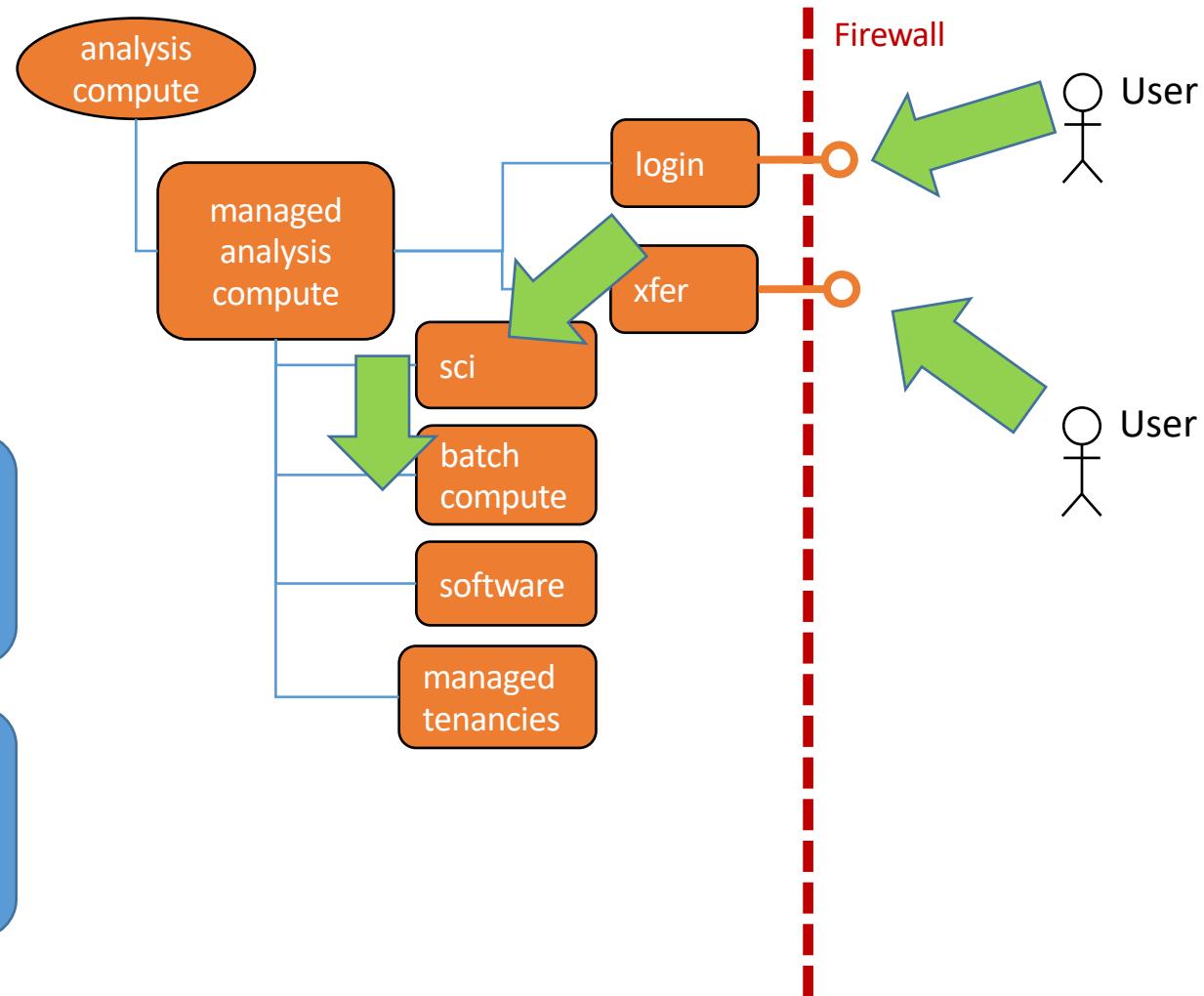
Storage	login	sci	xfer	LOTUS
/home/users/<username>	✓	✓	✓	✓
/group_workspaces/jasmin2/<gwsname>	✗	✓	✓	✓
/group_workspaces/jasmin4/<gwsname>	✗	✓	✓	✓
/gws/nopw/j04/<gwsname>	✗	✓	✓	✓
/gws/smf/j04/<gwsname>	✗	✓	✓	✓
/work/xfc/volX	✗	✓	✓	✓
/work/scratch	✗	✓	✗	✓
/work/scratch-nompiio	✗	✓	✗	✓
/badc, /neodc (CEDA Archive)	✗	✓ Read Only	✓ Read Only	✓ Read Only

Interactive compute

- Login servers
- Xfer servers
- Sci servers

Use login servers to access most services from outside

Transfer servers can be accessed directly from outside



Current specifications

CPU	RAM	Hosts	Type
8	32 GB	jasmin-sci[1,2,4,5].ceda.ac.uk cems-sci1.cems.rl.ac.uk	Virtual
24	384 GB	jasmin-sci6	Physical
48	2117 GB	jasmin-sci3, cems-sci2	Physical

Intended usage of Scientific analysis servers

- Interactive processing
- Testing & prototyping (prior to use LOTUS)
- For job submission to LOTUS (LSF client)
- Common software stack installed

Interactive compute – organized by community

- **New**

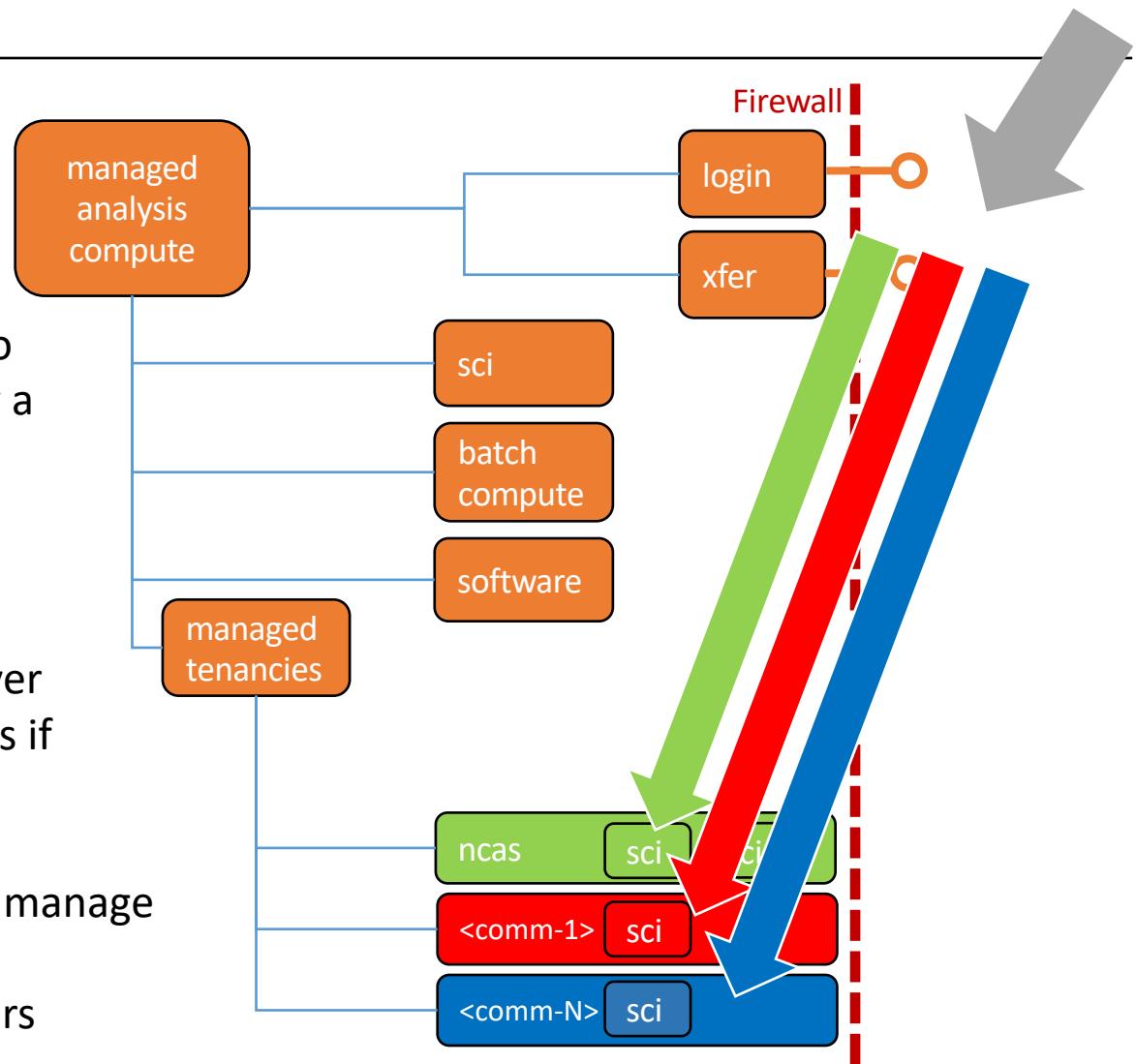
JASMIN users organized into communities able to supply a manager

System manager from each community can:

- Start / stop / restart server
- Deploy additional servers if resources allow

Benefit to you & us:

- Each community able to manage its own machines & behaviour of its own users



Batch compute cluster: LOTUS

JASMIN

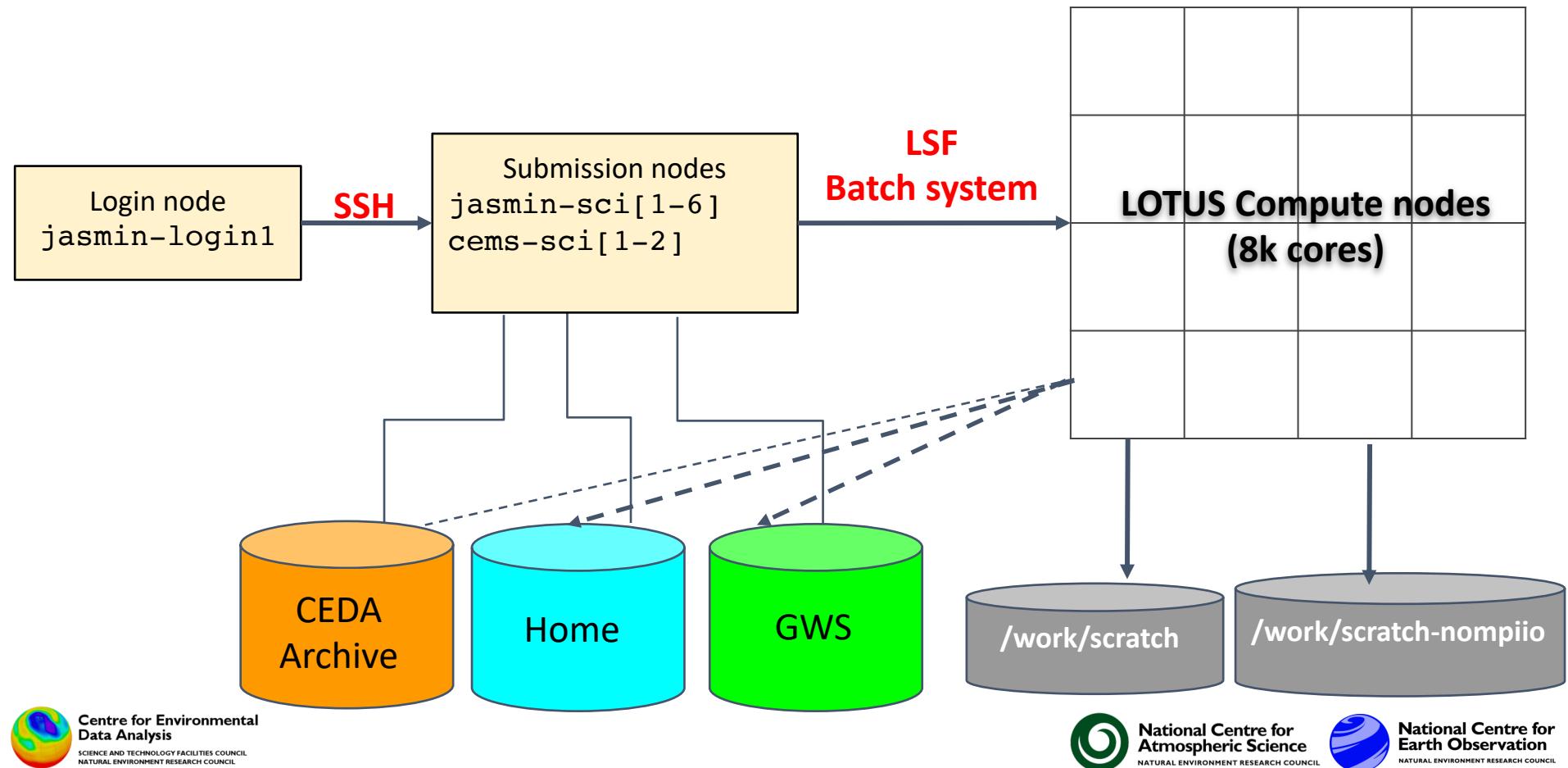
- Cluster of over 400 hosts, ~8000 cores
- Intel processor models:
 - Ivybridge, Haswell, Broadwell, Skylake
- Hosts with a range of memory sizes
 - 128GB to 2048GB
- Scheduling system applies “fair-share” scheduling (IBM LSF)
- Shares access to
 - Same software as interactive compute
 - Same storage resources
- Hardware interchangeable with cloud infrastructure hosts



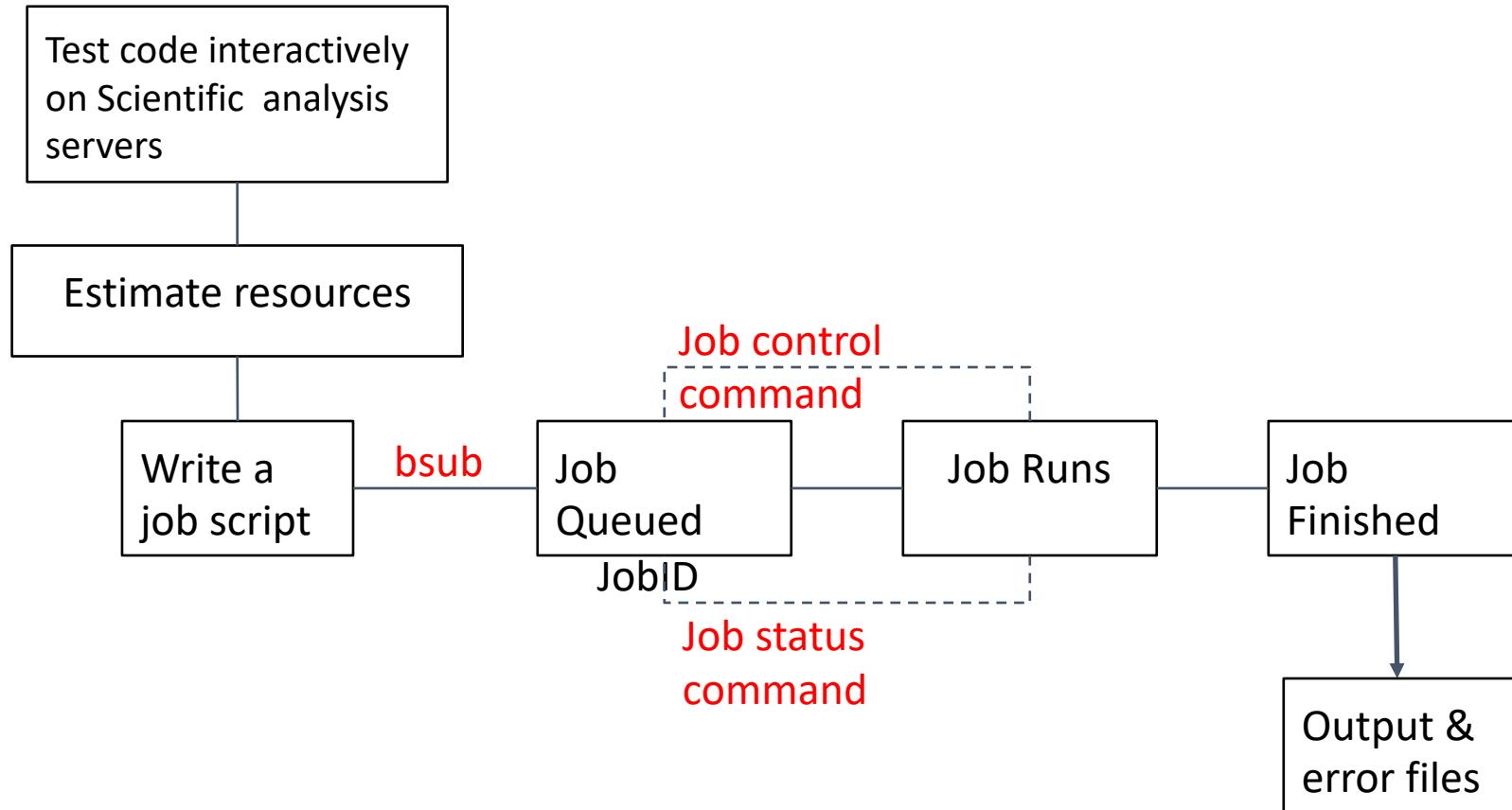
Schematic view of LOTUS on JASMIN

JASMIN

Components: submission nodes, compute nodes, storage & LSF



Scientific analysis servers



Access to CEDA Archive data

JASMIN

“Moving the compute to the data...”

7 Petabytes of CEDA archive data is available online in JASMIN

- File systems mounted read-only throughout
- Can be processed in-place (no need to copy)
- Some large datasets are kept partly near-line
 - Near-Line Archive (NLA) service to stage data
- Access permissions reflect those of your **CEDA** account (must be linked)

CEDA Archive is a “tenant” of JASMIN

- CEDA Archive services (catalogue, download) hosted on JASMIN

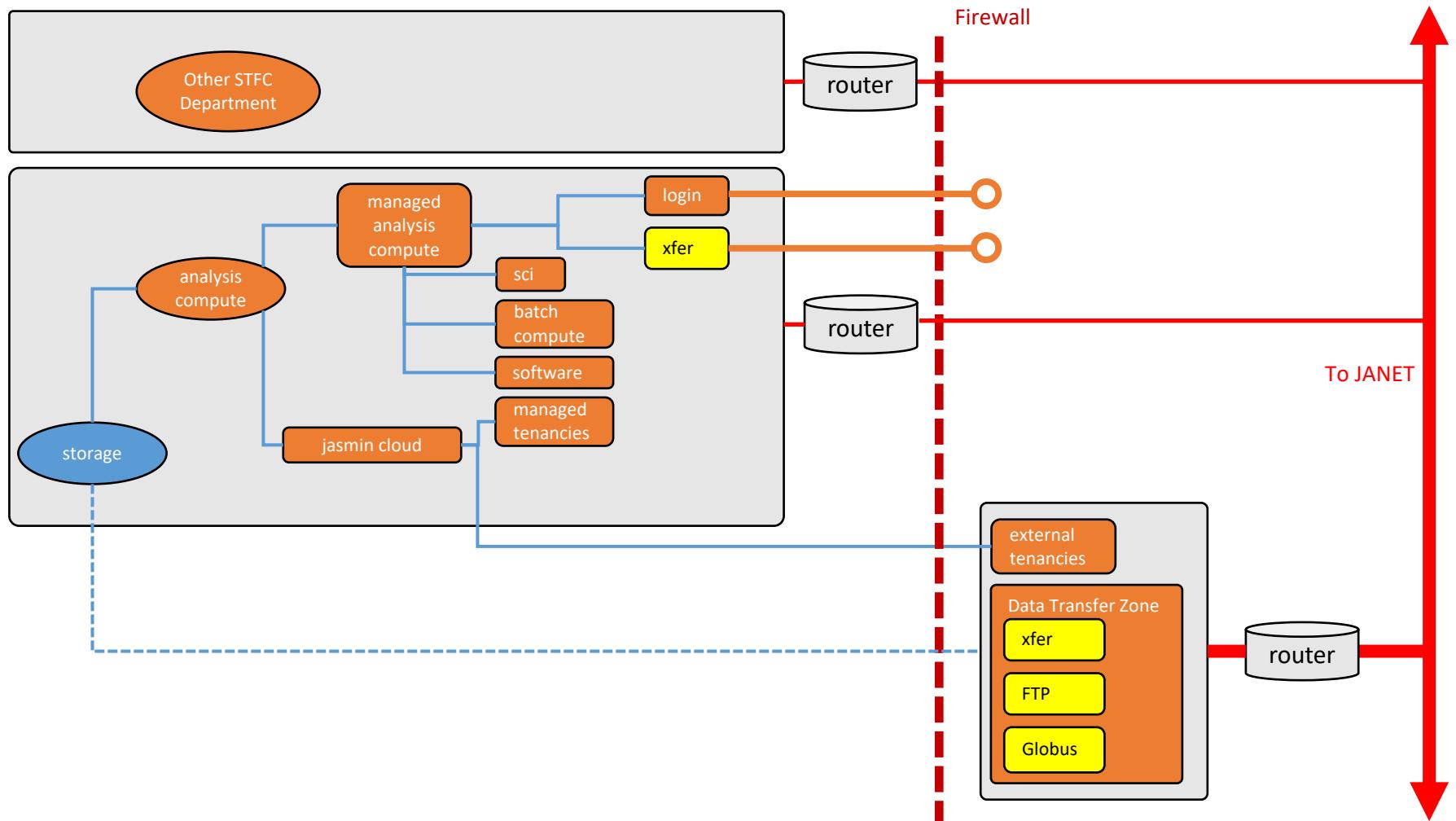
JASMIN ***is not*** the CEDA Archive

- Leaving your data in your GWS ***is not*** long term data curation



Data transfer services

JASMIN



- Range of tools provided for transfer of data into/out of JASMIN
 - All xfer servers
 - scp, rsync, sftp : push or pull data
 - ftp client : pull only
 - lftp
 - High Performance services
 - Xfer servers in “Data Transfer Zone”
 - jasmin-xfer[23] : same tools, faster servers + gridftp client tools
 - Globus endpoint & Gridftp server
 - Globus = online service for efficient bulk data transfers over gridftp
 - Other data-intensive services
 - ESGF, CEDA FTP download service, Archive pull services
 - <https://help.jasmin.ac.uk/category/217-data-transfer>

- Analysis software common to batch & interactive compute
 - Stack developed with user community
 - JASPY
- Additional tools for compiling and building software
- Restricted software
- Data movement software

<https://help.jasmin.ac.uk/article/273-software-on-jasmin>

- Your JASMIN account
 - JASMIN account = profile to attach privileges to
 - <https://accounts.jasmin.ac.uk>
 - “jasmin-login” = what you need for system access
 - (not everyone needs this!)
 - Apply for other roles (GWSs, hpxfer, ..) on accounts portal
 - GWS roles approved by GWS manager (not CEDA/JASMN team)
 - Reminder emails
 - Once a year, need to re-confirm account
 - Each role: 2 month / 2 week / 2 day reminder before expiry
 - You can renew in advance
 - Different to your CEDA account
 - Needs to be linked

- Information
 - Events like this!
 - <https://help.jasmin.ac.uk>
 - Getting started
 - <https://help.jasmin.ac.uk/article/189-get-started-with-jasmin>
 - YouTube channel
 - <https://www.youtube.com/channel/UC11nPZVvjDLjYIS7NvbnlmQ>
 - [@cedanews](#) on twitter
 - emails
- Helpdesk
 - support@ceda.ac.uk
 - CEDA team does outward user support
 - SCD team looks after the infrastructure
 - [Template for support queries](#)

Expectations

- JASMIN is a research infrastructure
 - At cutting edge of data-intensive environmental informatics
 - Not a 24x7x365 enterprise
 - Small team
 - More flexible than a supercomputer
 - Relies on you being sensible
 - You need to know what you're doing
 - Test stuff before scaling up!
 - Build failure-tolerance into your workflows
 - Please acknowledge JASMIN
 - <https://help.jasmin.ac.uk/article/4693-acknowledging-jasmin>
 - Tell your funding body what you've done with JASMIN
 - ...And tell us! (impact stories)
 - Helps ensure continued funding
 - Makes the case for improving the service further ☺