NRC·CNRC

CADC and CANFAR Overview

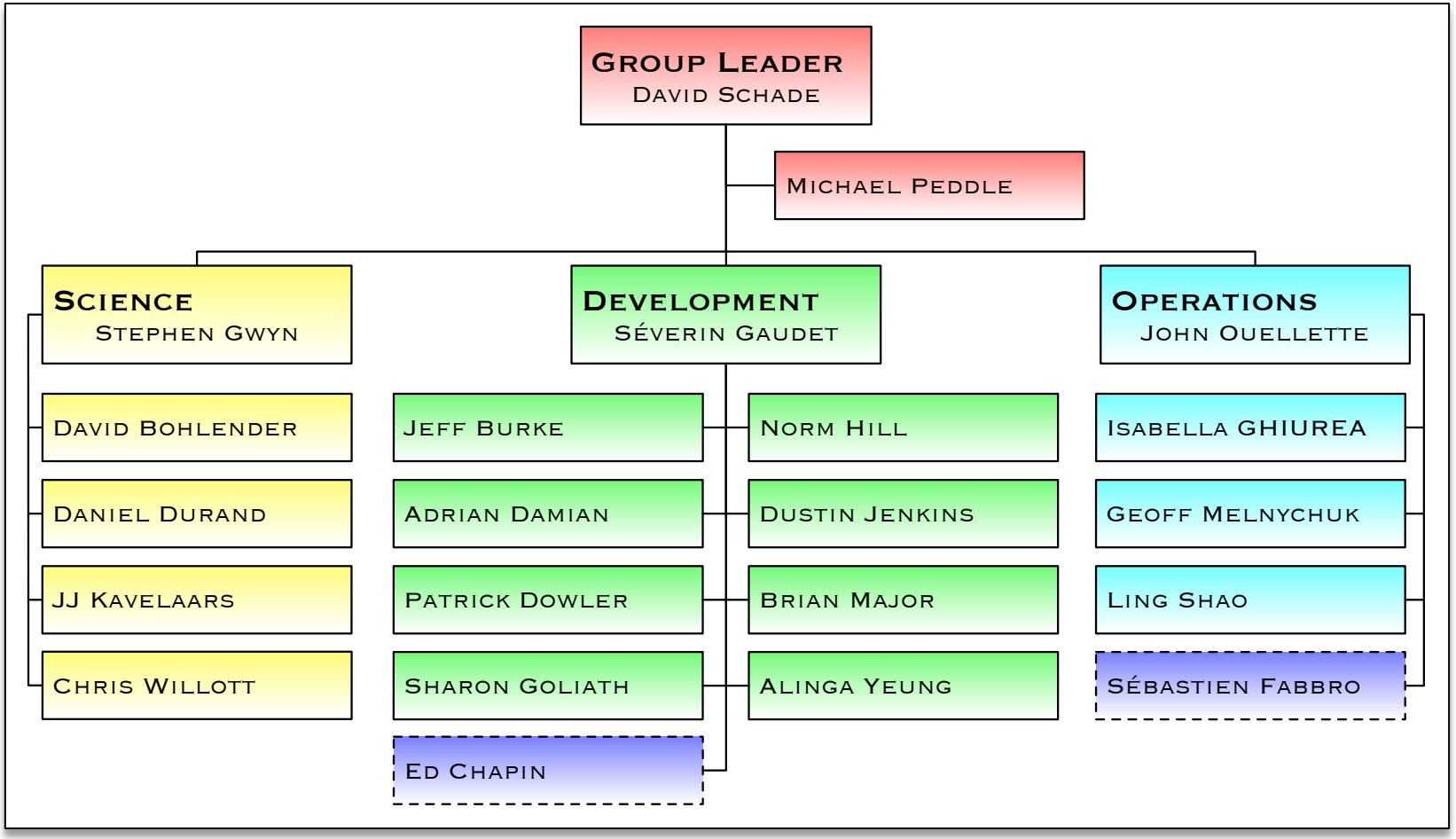
Séverin Gaudet
Canadian Astronomy Data Centre

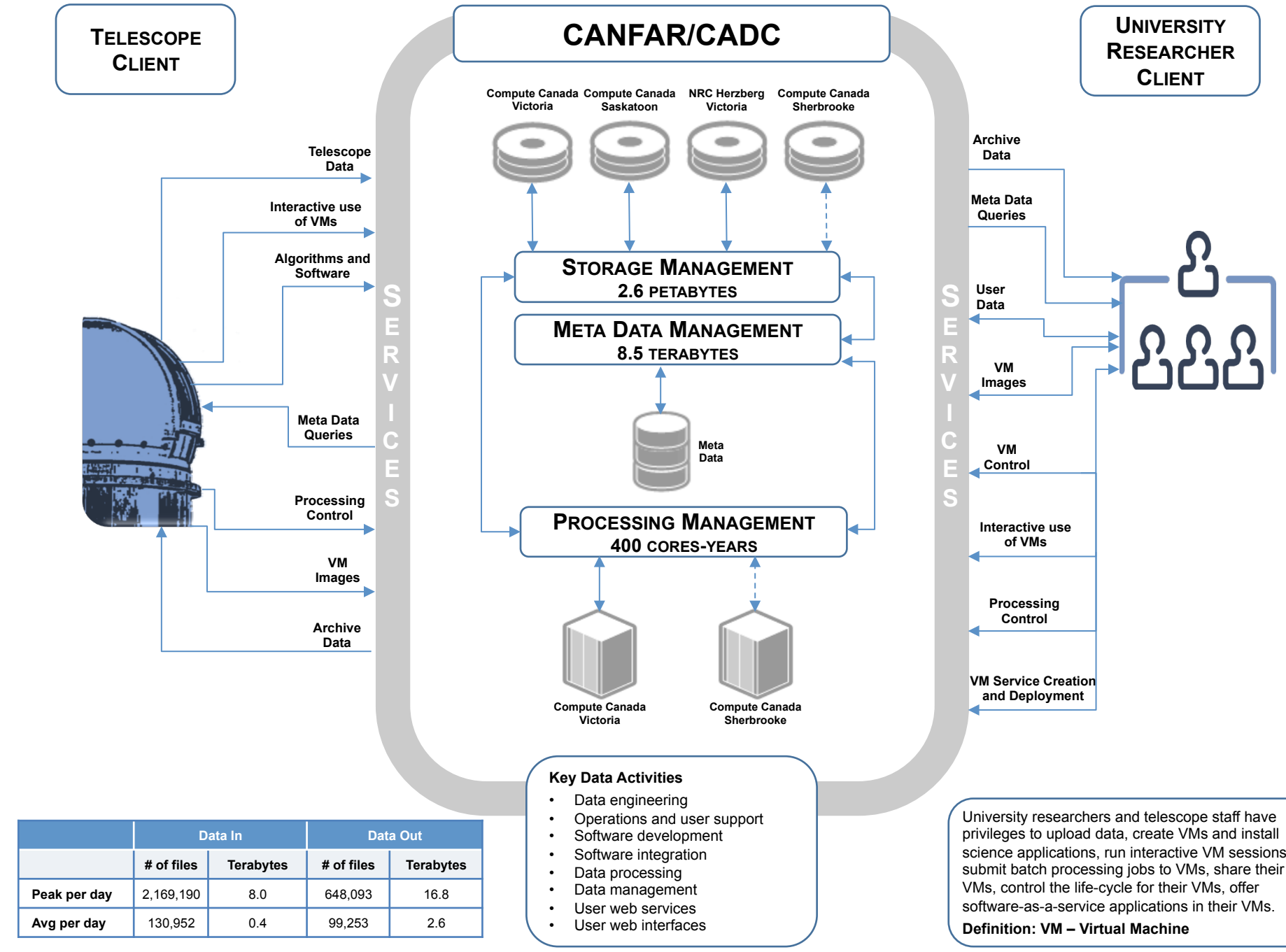
Canadian Astronomy Data Centre

- Telescope collections:
  - Multiple missions, facilities and wavelengths
  - Pointed and survey observations
  - 12 telescopes
  - 6 advanced data collections
- Services
  - Community projects
- 20+2 staff
- Many international collaborations

- Development and operations hub for CANFAR

CADC/CANFAR Staffing





	Data In		Data Out	
	# of files	Terabytes	# of files	Terabytes
Peak per day	2,169,190	8.0	648,093	16.8
Avg per day	130,952	0.4	99,253	2.6

International Virtual Observatory Alliance (VO)

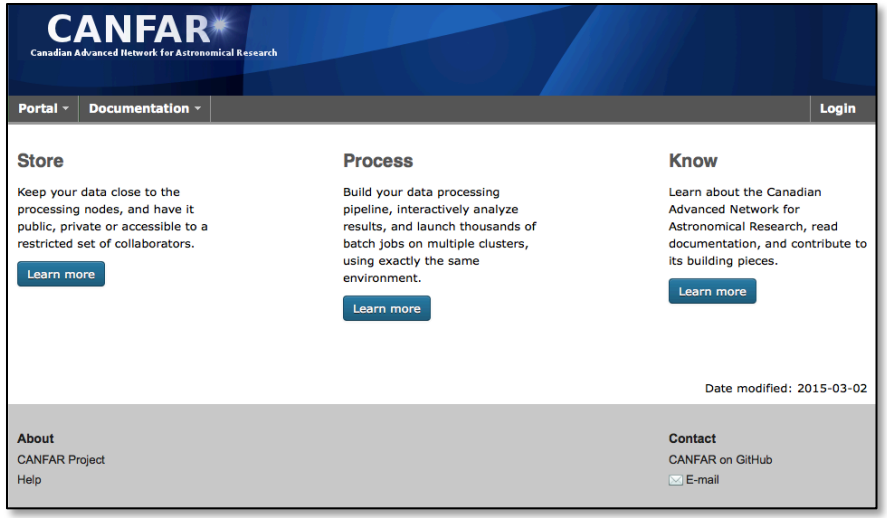
- *"Allow **astronomers** to interrogate multiple data centers in a seamless and transparent way"*
- *"Give **data centers** a standard framework for publishing and delivering services using their data.*

CANFAR/CADC 2014

- Size:
  - 233M files (932M files)
  - 597 TiB (2.3 PiB)
- Users
  - Authenticated access: 762
  - Anonymous access: 7,544
  - Registered: 7,018
- Data handled in the last year
  - TiB: 1,106
  - Files: 91M

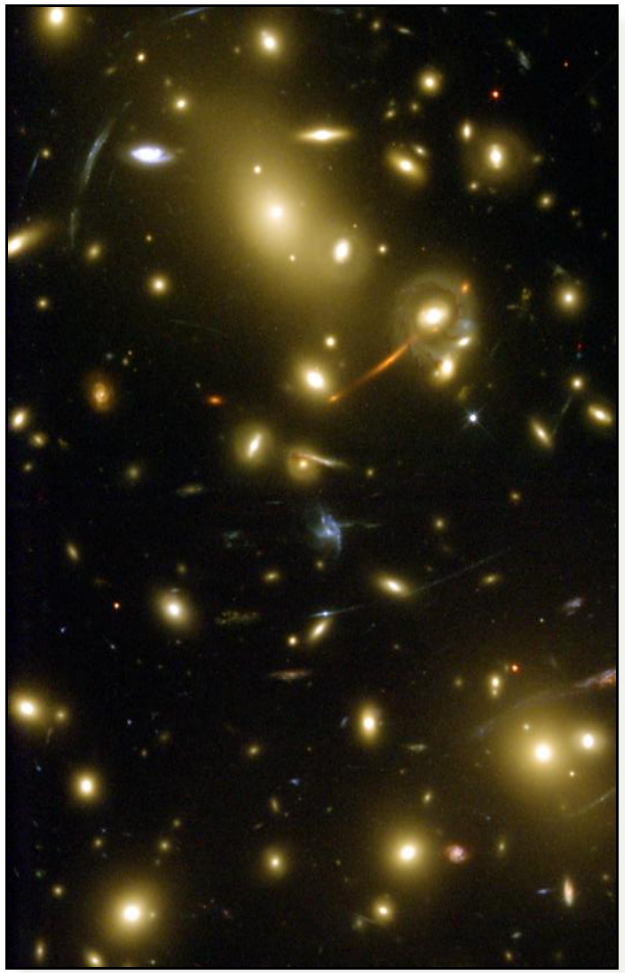
Canadian Advanced Network for Astronomical Research

- A cloud ecosystem for data intensive astronomy
- User services
  - Store and share data
  - Create and configure VMs
  - Run interactive VMs
  - Run persistent VMs
  - Batch processing with VMs
- Using research cloud resources
  - Compute Canada
- Integrated authentication and authorization



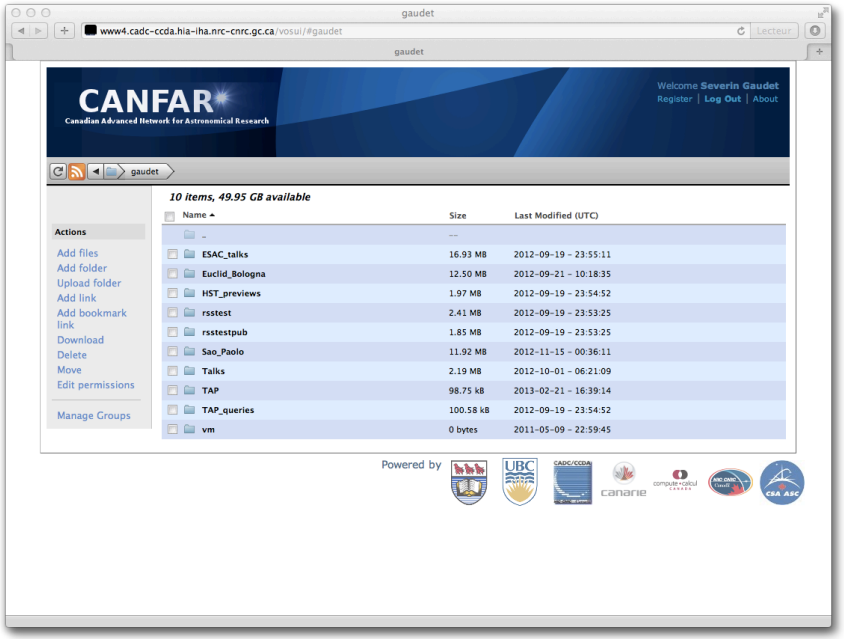
Context:
Science Team Support

- Virtual organisations
  - Forming around a given multi-year survey project
  - Handling large datasets
  - Faced with acquiring and building project infrastructure
- Require infrastructure
  - Larger datasets
  - Data management, data distribution, data processing
  - Challenging a team's ability to produce and maintain infrastructure
- One time use of project-specific infrastructure
- Strong central institution not always present



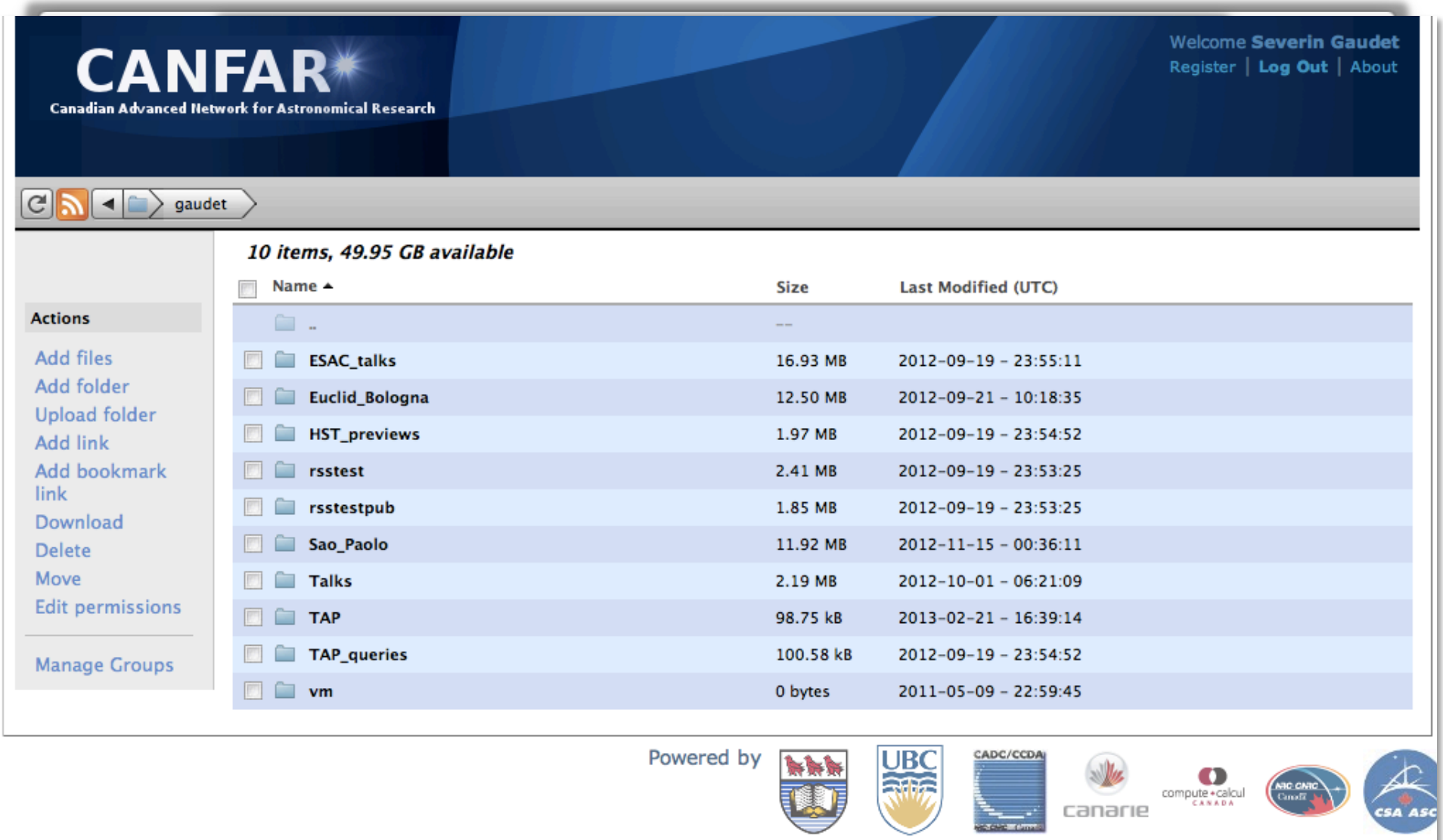
Virtual File System: VOSpace

- User storage
- Cloud processing file persistence
- Asynchronous query input and output

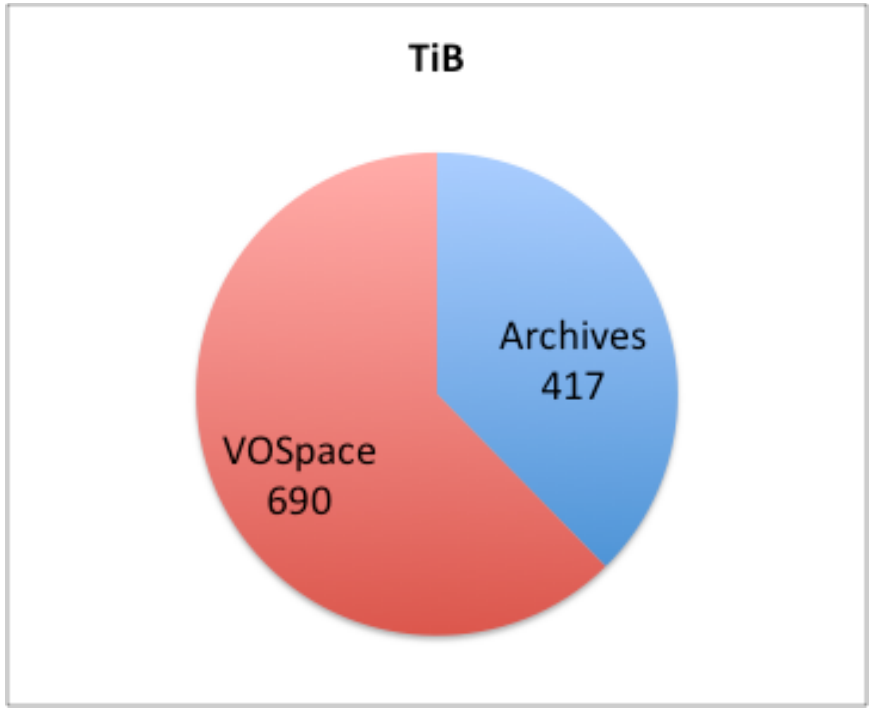


- Browser UI
- Python clients:
  - Command line, e.g., vcp
  - Mountable file system: mountvofs
- https://github.com/canfar/vos

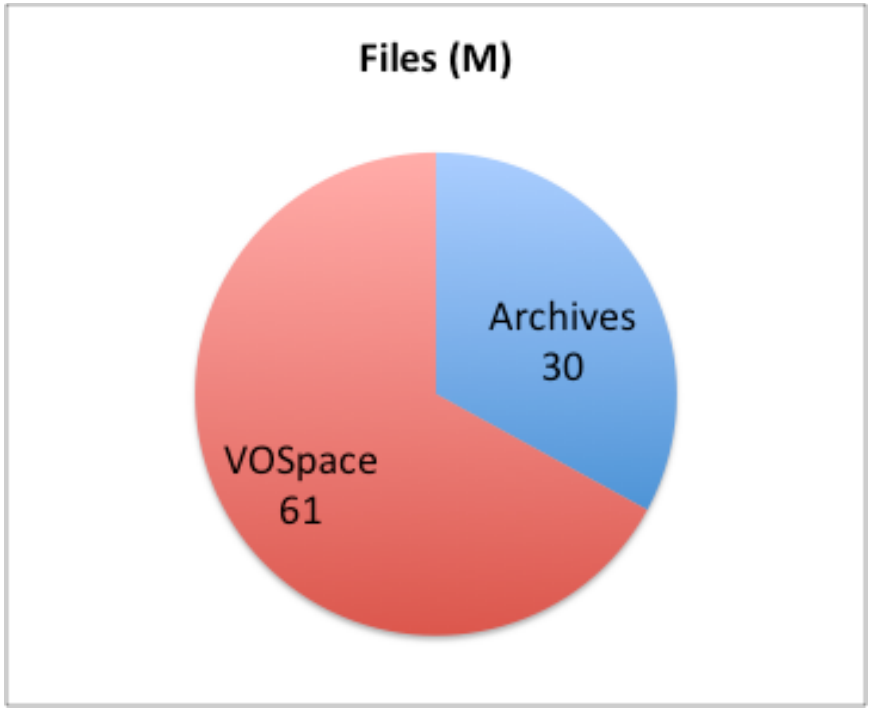
- Full access control
- Notifications via RSS feeds


CANFAR
Canadian Advanced Network for Astronomical Research
Welcome Severin Gaudet
Register | Log Out | About
gaudet
10 items, 49.95 GB available
Name
Size
Last Modified (UTC)
..
--
ESAC_talks
16.93 MB
2012-09-19 - 23:55:11
Euclid_Bologna
12.50 MB
2012-09-21 - 10:18:35
HST_previews
1.97 MB
2012-09-19 - 23:54:52
rsstest
2.41 MB
2012-09-19 - 23:53:25
rsstestpub
1.85 MB
2012-09-19 - 23:53:25
Sao_Paolo
11.92 MB
2012-11-15 - 00:36:11
Talks
2.19 MB
2012-10-01 - 06:21:09
TAP
98.75 kB
2013-02-21 - 16:39:14
TAP_queries
100.58 kB
2012-09-19 - 23:54:52
vm
0 bytes
2011-05-09 - 22:59:45
Actions
Add files
Add folder
Upload folder
Add link
Add bookmark link
Download
Delete
Move
Edit permissions
Manage Groups
Powered by
UBC
CADC/CCDA
canarie
compute • calcul
CANADA
CSA ASC

VOSpace Usage in 2014

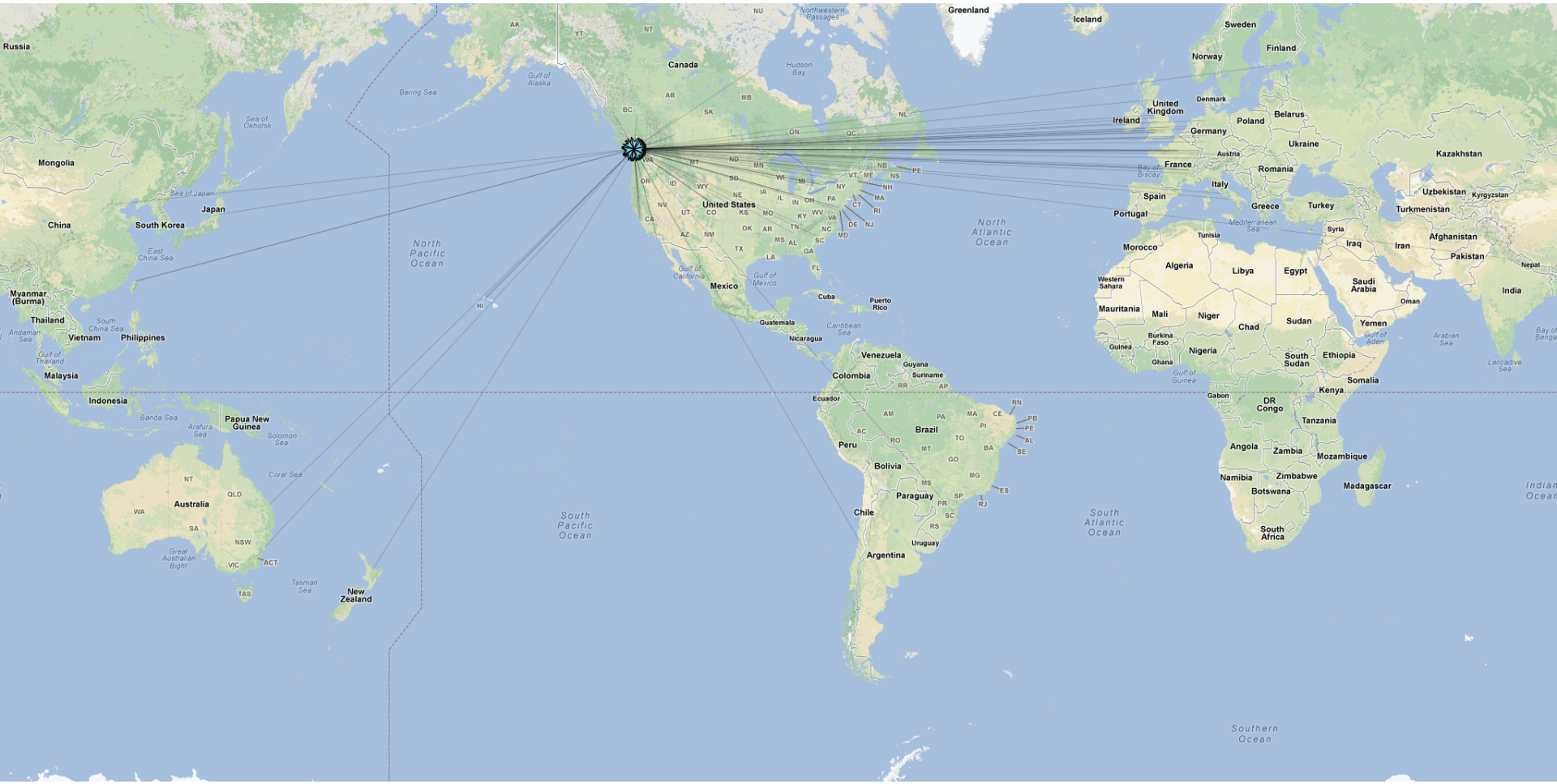


62% VOSpace
Average per week: 13.3
Peak week: 39.0

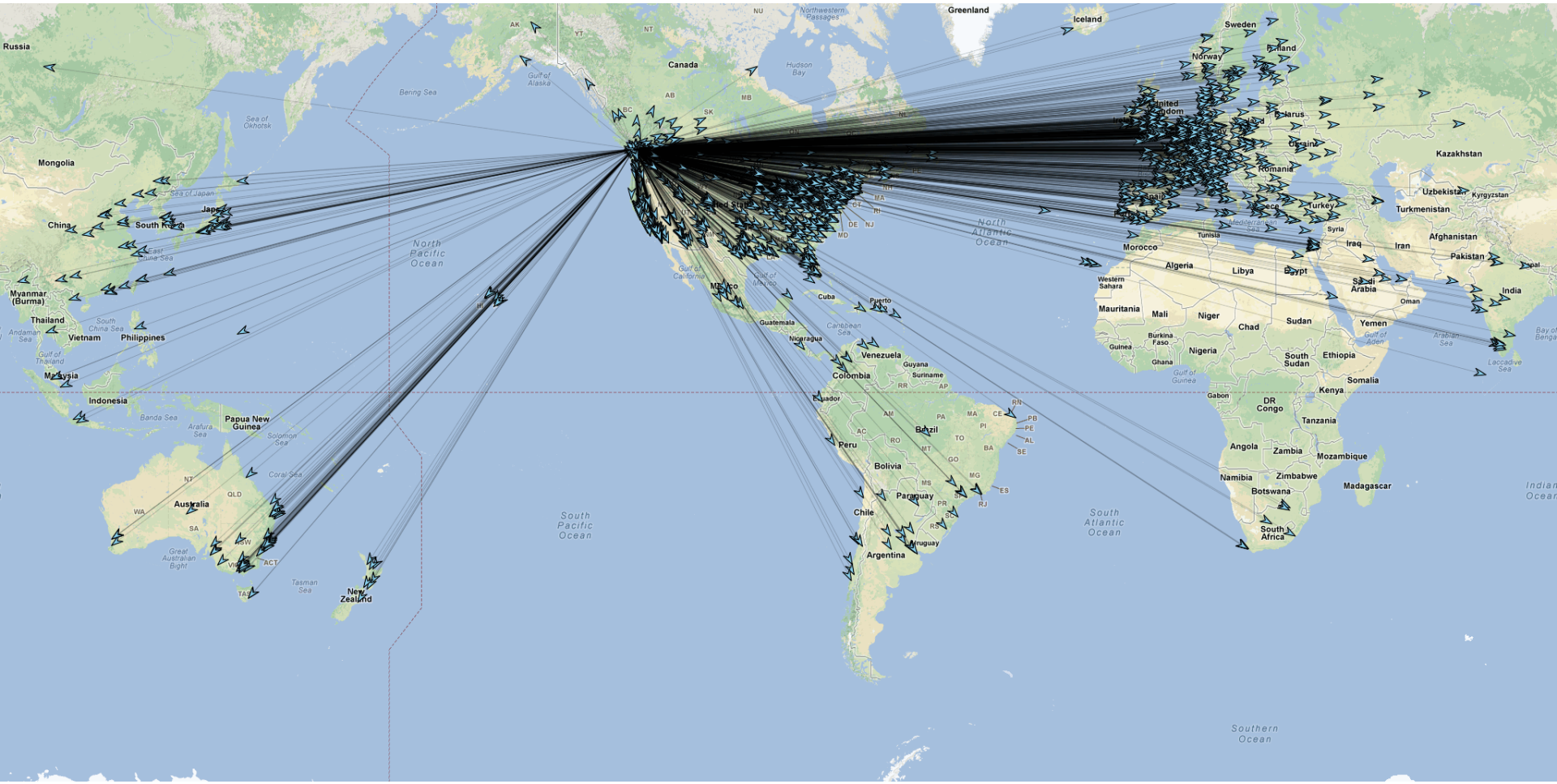


67% VOSpace
Average per week: 1.2M
Peak week: 11.7M

Geography of VOSpace PUTs

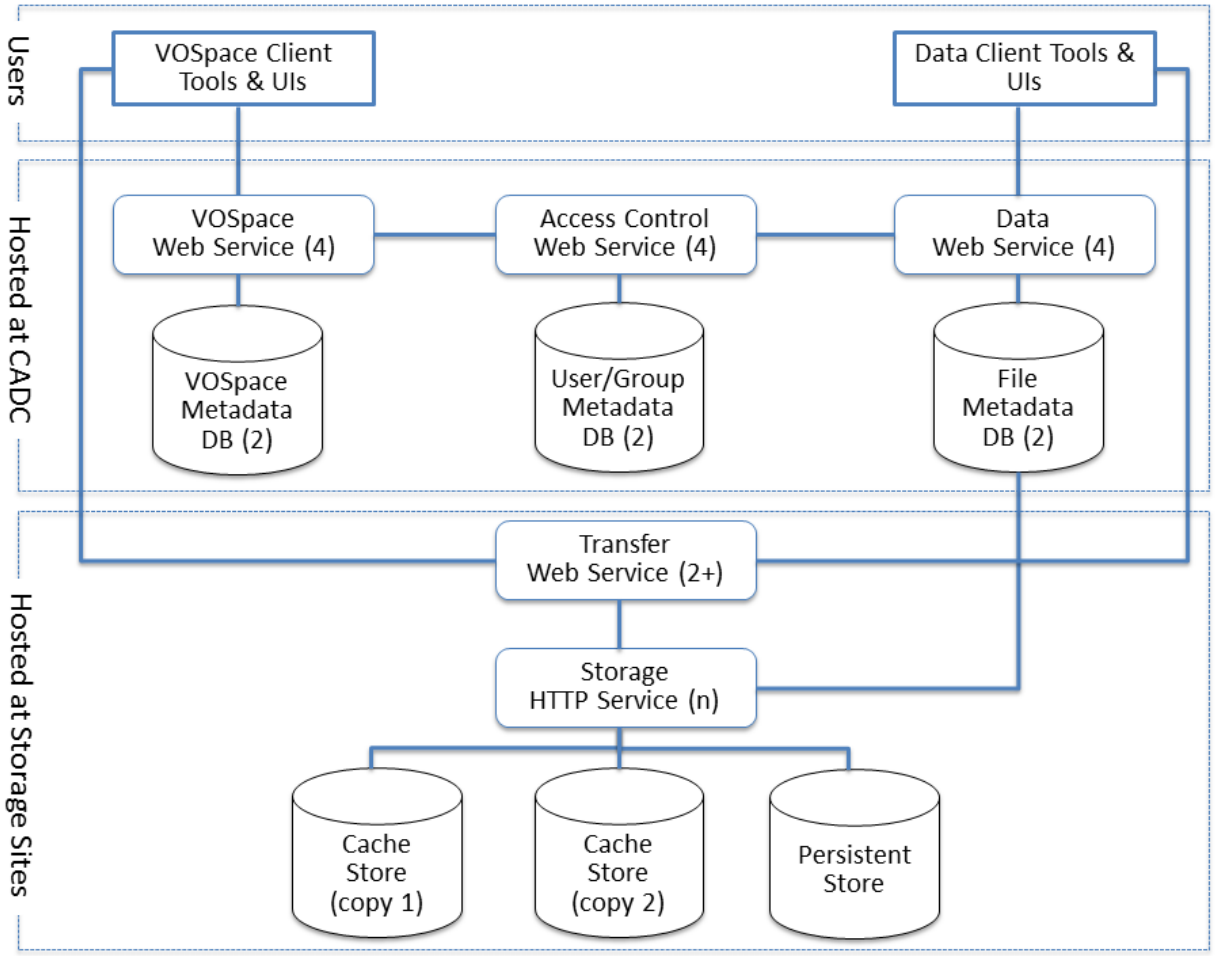


Geography of VOSpace GETs



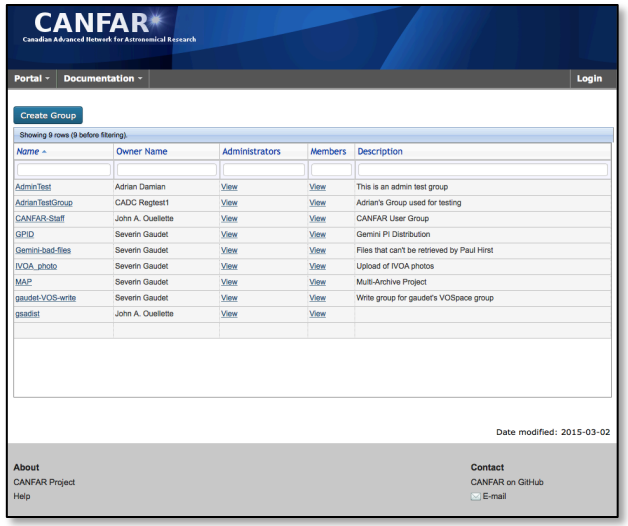
Virtual Object Store

- Storage web services using several distributed storage resources
- Optimization and QoS strategies not user nor provider dependent
- Same system for both archive and user files



Access Control

- Telescope, project team or user managed
- Processing, storage, querying
- X.509 certificates
  - **Not** user facing
  - Self-signed
  - Platform service accepted by resource providers
- Based on VO:
  - Single Sign-On
  - Credential Delegation Service



CANFAR
Canadian Advanced Network for Astronomical Research

Portal | Documentation | Login

Create Group

Showing 9 rows (9 before filtering).

Name	Owner Name	Administrators	Members	Description
AdminTest	Adrian Damian	View	View	This is an admin test group
AdrianTestGroup	CADC Regtest1	View	View	Adrian's Group used for testing
CANFAR-Staff	John A. Ouellette	View	View	CANFAR User Group
GPID	Severin Gaudet	View	View	Gemini PI Distribution
Gemini-bad-files	Severin Gaudet	View	View	Files that can't be retrieved by Paul Hirst
IVOA_photo	Severin Gaudet	View	View	Upload of IVOA photos
MAP	Severin Gaudet	View	View	Multi-Archive Project
gaudet-VOS-write	Severin Gaudet	View	View	Write group for gaudet's VOSpace group
gsadist	John A. Ouellette	View	View	

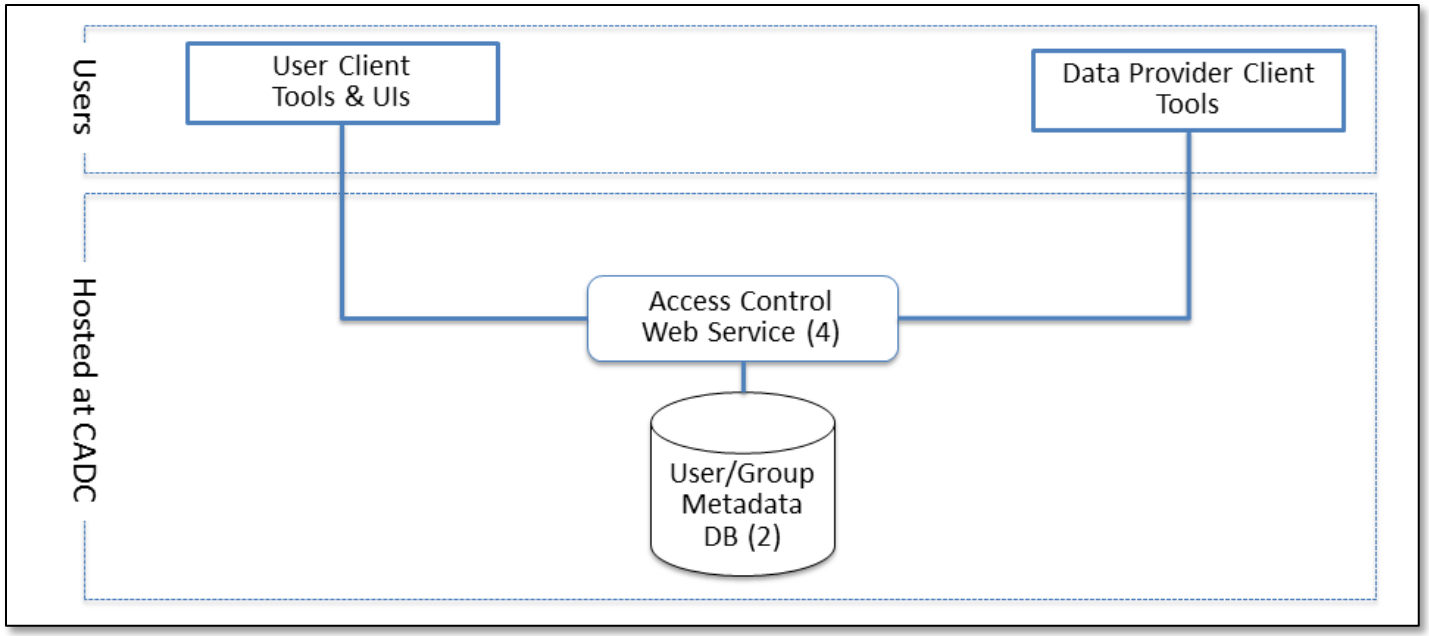
Date modified: 2015-03-02

About
CANFAR Project
Help

Contact
CANFAR on GitHub
E-mail

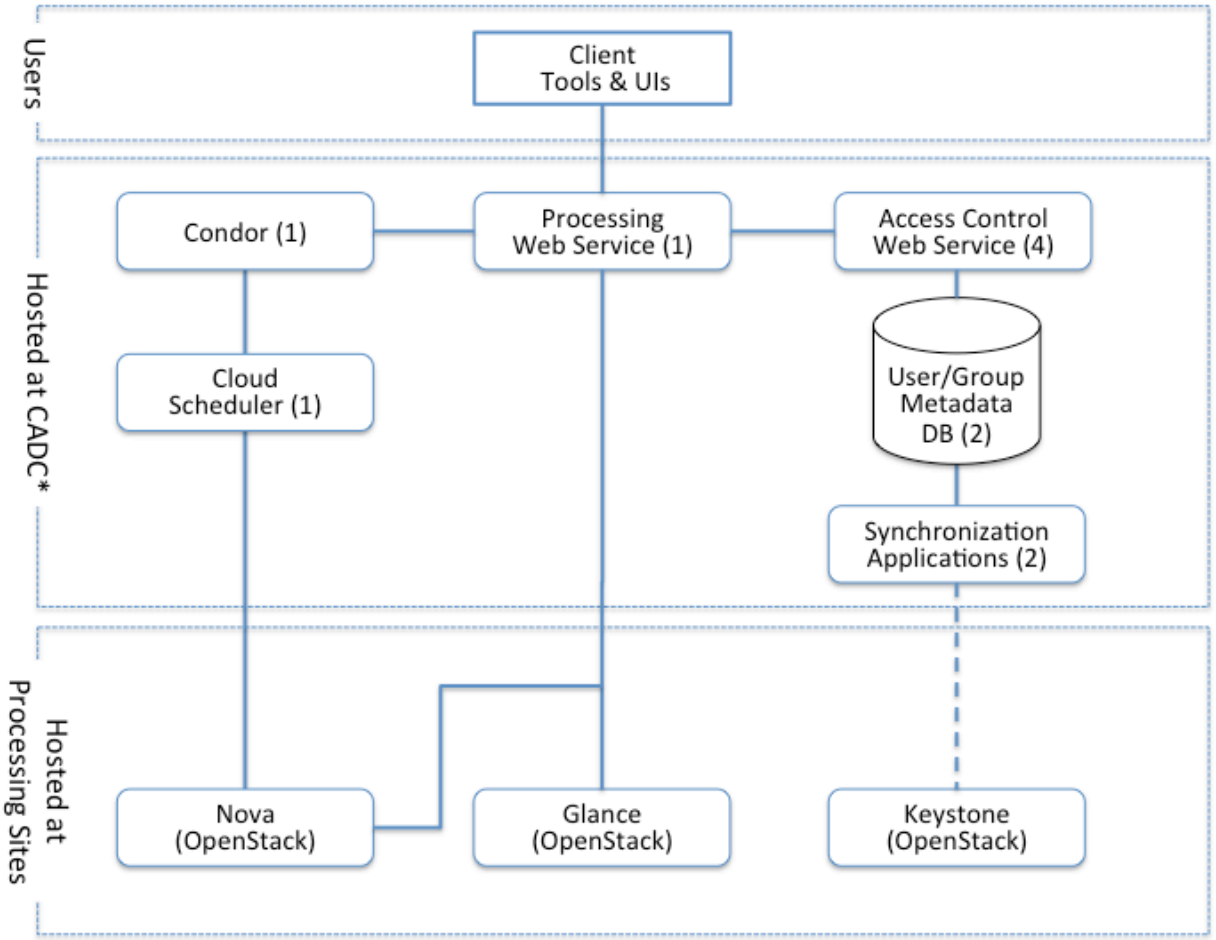
Access Control

- It's the glue!
- It's the integration challenge!
- It has to be done



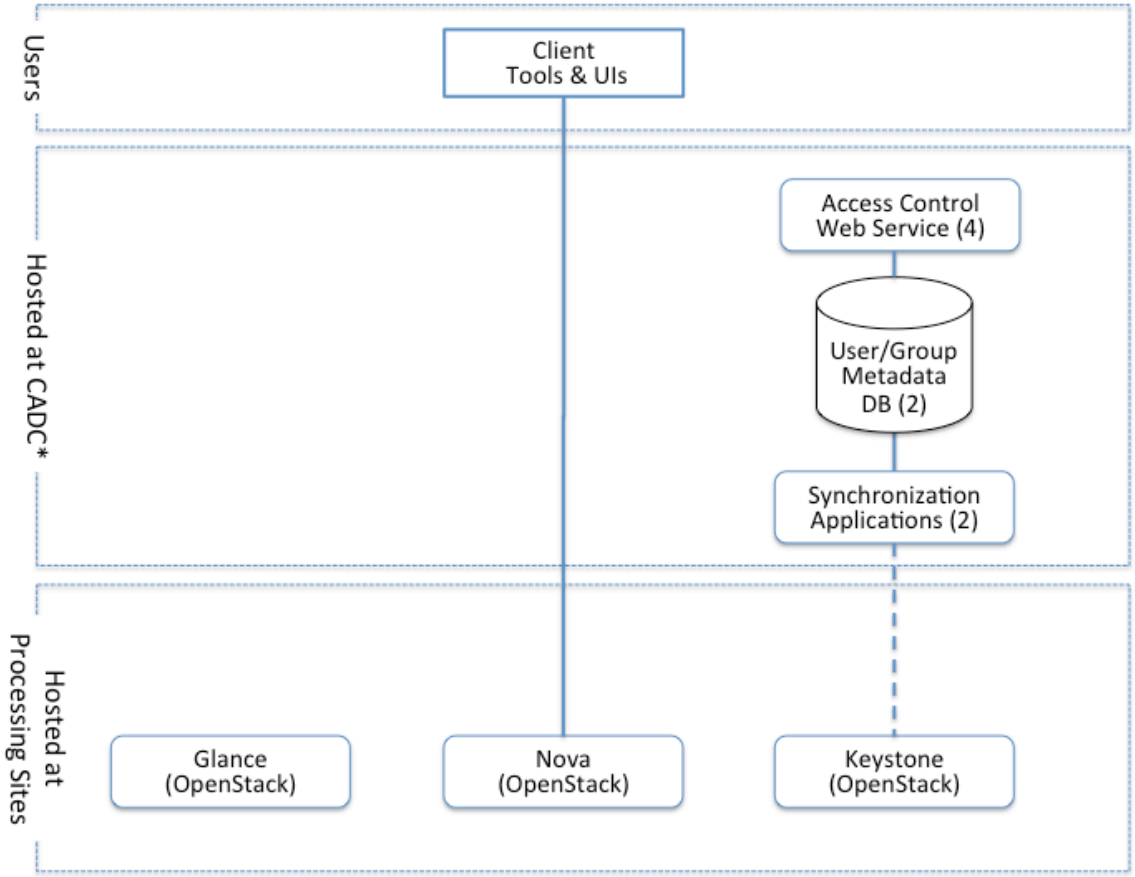
Batch Processing Service

- For embarrassingly parallel processing
- Formation of virtual clusters
- User provided VMs with Condor client

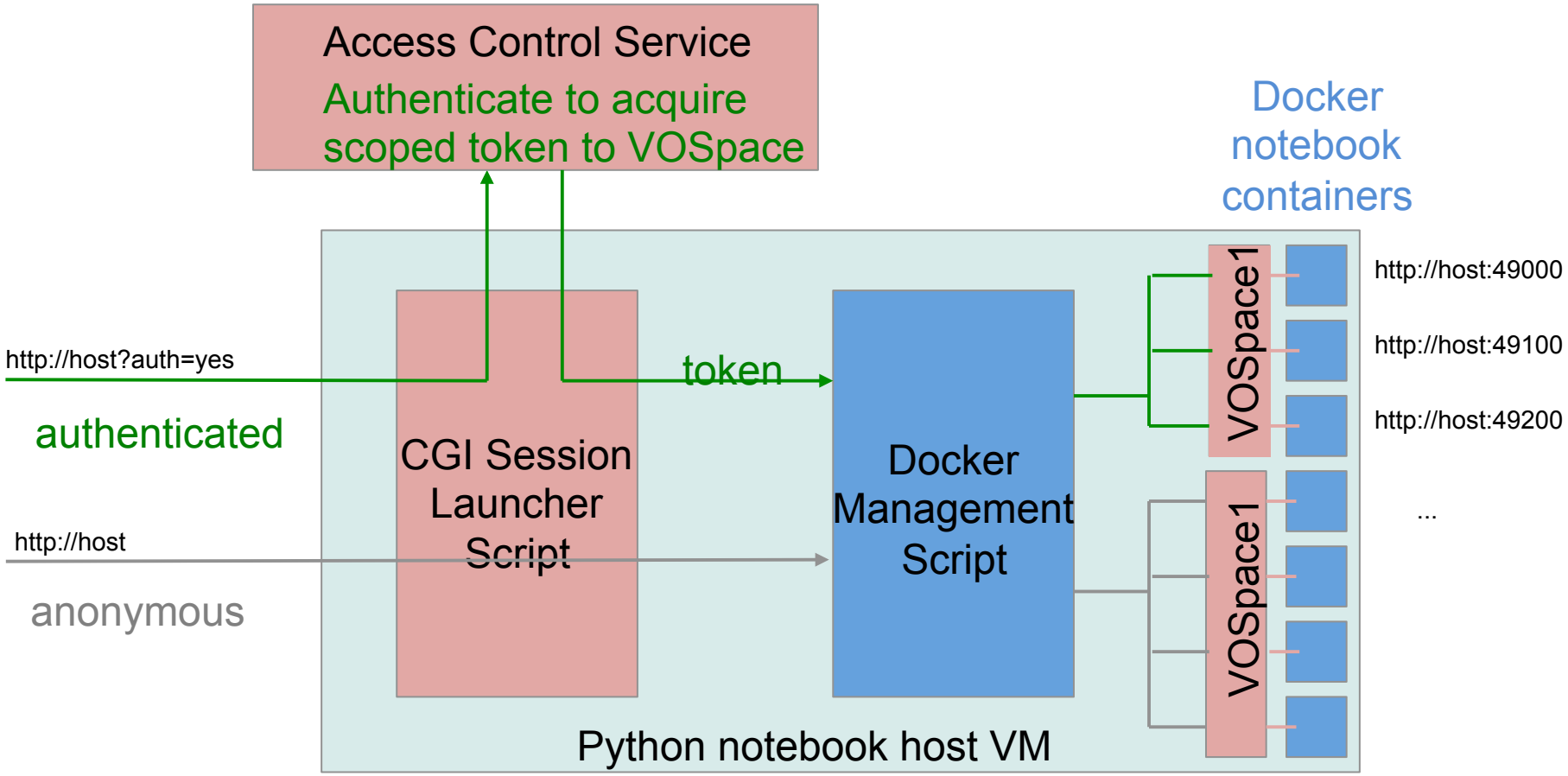


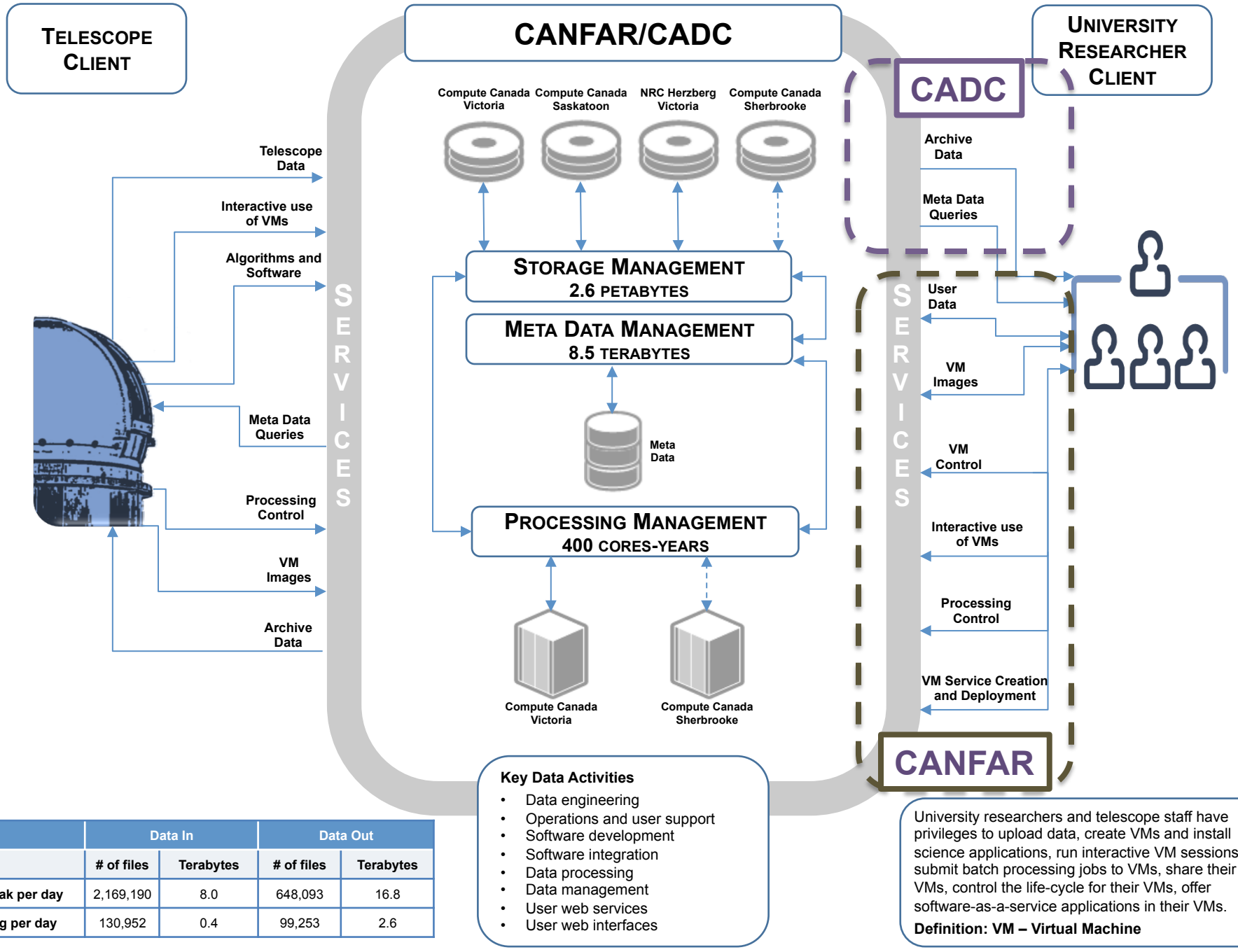
Interactive and persistent VMs

- Interactive VMs
  - Configuration
  - Testing
  - Interactive
- Persistent VMs
  - SaaS
  - Container deployment
  - Discovery agents

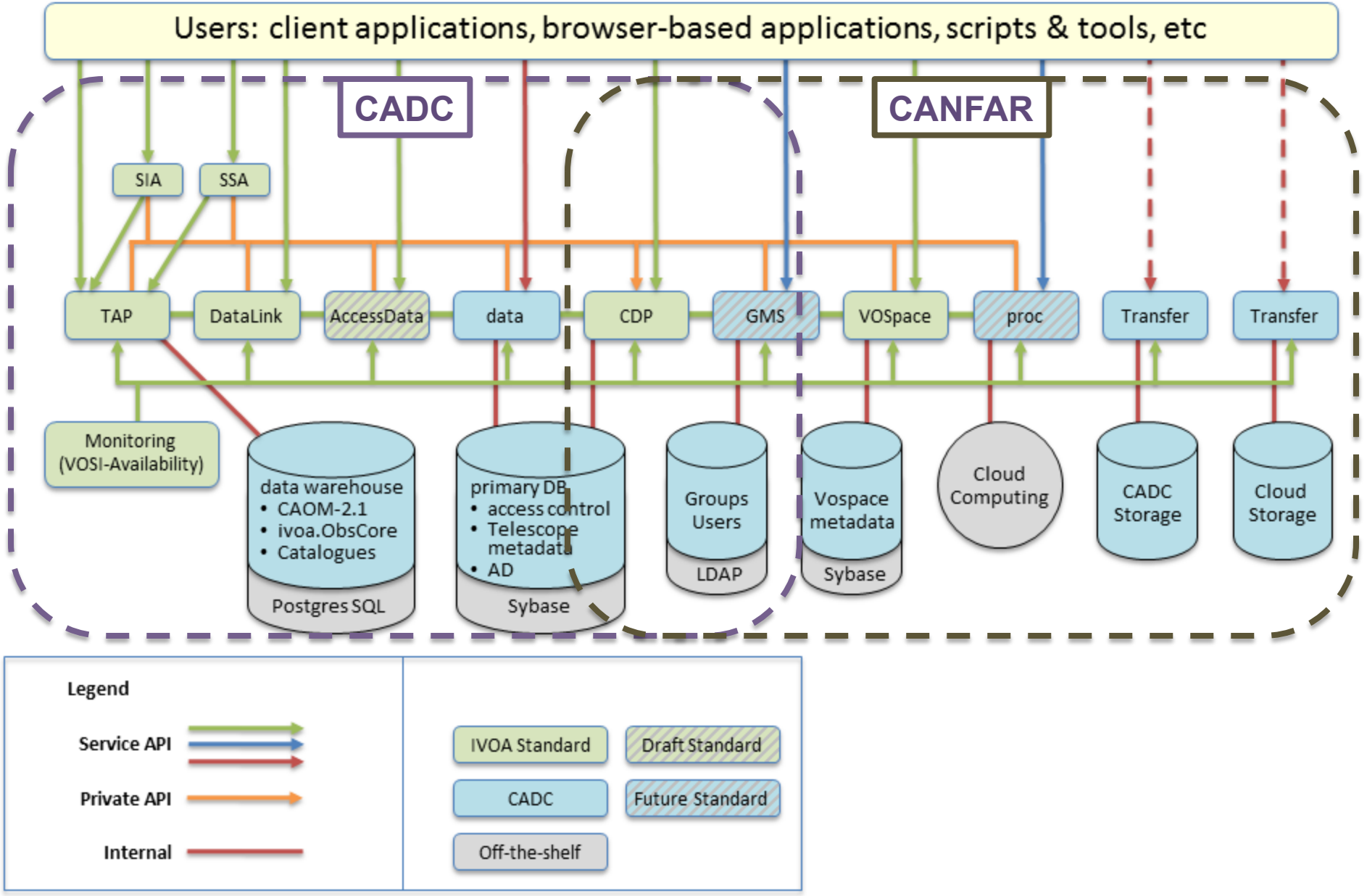


Integrated Session Management





	Data In		Data Out	
	# of files	Terabytes	# of files	Terabytes
Peak per day	2,169,190	8.0	648,093	16.8
Avg per day	130,952	0.4	99,253	2.6


Users: client applications, browser-based applications, scripts & tools, etc
CADC
CANFAR
SIA
SSA
TAP
DataLink
AccessData
data
CDP
GMS
VOSpace
proc
Transfer
Transfer
Monitoring
(VOSI-Availability)
data warehouse
• CAOM-2.1
• ivoa.ObsCore
• Catalogues
Postgres SQL
primary DB
• access control
• Telescope metadata
• AD
Sybase
Groups
Users
LDAP
Vospace
metadata
Sybase
Cloud
Computing
CADC
Storage
Cloud
Storage
Legend
Service API
Private API
Internal
IVOA Standard
Draft Standard
CADC
Future Standard
Off-the-shelf