

CADC and CANFAR: Extending the role of the data centre

Séverin Gaudet
Canadian Astronomy Data Centre



February 2012



Canadian Astronomy Data Centre

- Heterogeneous collection:
 - Multiple missions, facilities and wavelengths
 - Pointed and survey observations
 - 12 telescopes
 - 6 advanced data collections
- Common data model
 - Single query interface
 - Virtual Observatory interfaces
- Services
 - Community projects
 - CANFAR processing and storage
- 20+2 staff
- Many international collaborations

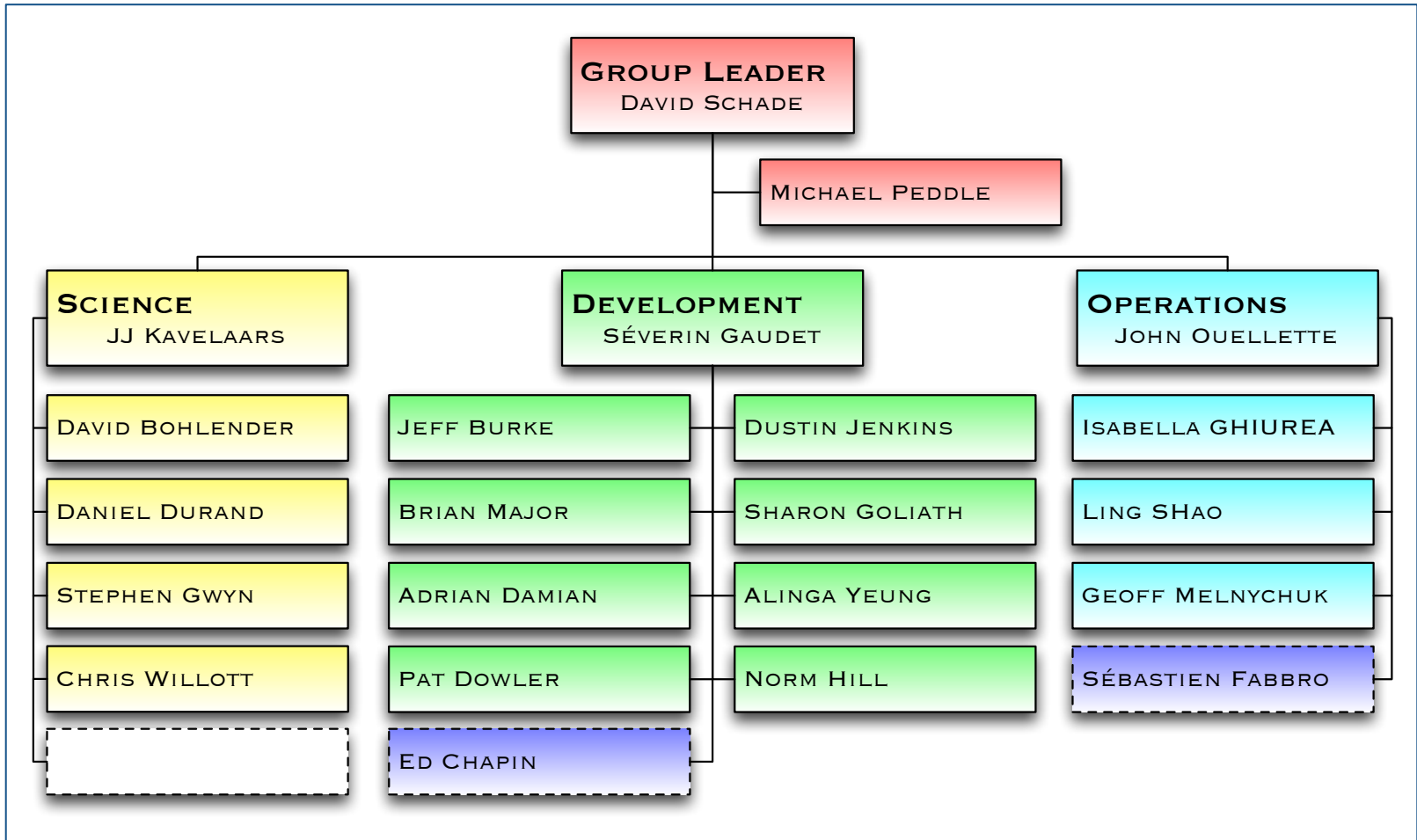
The screenshot shows the Canadian Astronomy Data Centre (CADC) website. At the top, there is a navigation bar with links for 'Telescope Data Products', 'Advanced Data Products', 'Services', 'Advanced Search', and 'Login'. A search bar is prominently displayed with the text 'Search for data by target' and a 'Search' button. Below the search bar, the page is organized into three main columns: 'Telescope Data Products', 'Advanced Data Products', and 'Services'. Each column contains a grid of icons representing various astronomical facilities and projects. The 'Telescope Data Products' column includes Gemini, CFHT, JCMT, HST, BLAST, MOST, DAO, MACHO, OMM, ELISE, and UKIRT. The 'Advanced Data Products' column includes MegaPipe, HLA, IRIS, CGPS, CHILS, and WIRWolf. The 'Services' column includes Meetings, Community, SSOIS, and CANFAR. At the bottom of the page, there is a footer with links for 'Terms and conditions', 'Transparency', 'News', and 'Contact us'. The date 'Date modified: 2014-04-28' is also visible.

CADC Usage Numbers

- Size:
 - 66M files
 - 434 TB
- Users
 - Authenticated access: 511
 - Anonymous access: 3,234
 - Registered: 6,207
- Data handled in the last year
 - TB: 353
 - Files: 18,221,690



Staffing





The Canadian Astronomy Data Centre

If you have used CADCC facilities for your research, please include the following acknowledgment:

This research used the facilities of the Canadian Astronomy Data Centre operated

by the National Research Council of Canada with the support of the Canadian Space Agency.

[About](#) | [Register](#) | [Contact](#)

Search by Target:

All Archives ▾

[Advanced Search](#)

[English](#) | [Français](#)

NRC is currently renovating this web site to improve its functionality in both official languages. Due to the complexity of the site, these improvements are taking place in stages. We regret the inconvenience to our users and will update this notice to report our progress.

In the meanwhile, if you are unable to access the information you require, please contact us at cadc@nrc.gc.ca for assistance in the official language of your choice.

Thank you for your patience.



Advanced Query Service



Astronomy Meetings



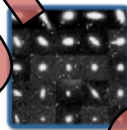
CFHT Legacy Survey



CFHT MegaCat Stacks



Canadian Virtual Observatory Services



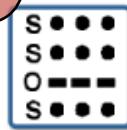
Community Projects



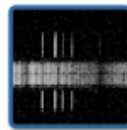
NRC/CFHT WISE/SDSS stacks



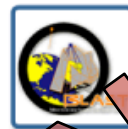
Programmatic Archive Access



Solar System Object Search



DAO Spectroscopic Plate Archive



BLAST



CFHT



CGPS



CFHT



FUSE



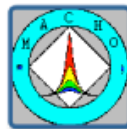
GSA



HST



JCMT



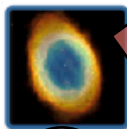
MACHO



MOST



AstroCat



Digitized Sky Survey



Guide Stars



U.S. Naval Observatory A2 Catalog



VizieR



CANFAR



No Interoperability!

[Telescope Data Products](#)[Advanced Data Products](#)[Services](#)[Advanced Search](#)[Login](#)

CADC Home

[Advanced Search](#)

Telescope Data Products



[Gemini](#)



[CFHT](#)



[JCMT](#)



[HST](#)



[BLAST](#)



[MOST](#)



[DAO](#)



[MACHO](#)



[OMM](#)



[FUSE](#)



[UKIRT](#)

Advanced Data Products



[MegaPipe](#)



[HLA](#)



[IRIS](#)



[CGPS](#)



[CFHTLS](#)



[WIRwolf](#)

Services



[Meetings](#)



[Community](#)



[SSOIS](#)



[CANFAR](#)

Date modified: 2014-04-28

[Terms and conditions](#) | [Transparency](#)



About us

[Our mandate](#)

[Acknowledgements](#)

News

Contact us

[Email](#)

[Address](#)

Advanced Search

- Enabled by the Common Archive Observation Model
 - CAOM
- Single query interface to “all” CADC collections
- With proprietary metadata and data access
- Many years in the making
- Phase 1 complete
- [Demo?](#)

Canadian Astronomy Data Centre

Canada

Telescope Data Products | Advanced Data Products | Services | Advanced Search | Login

CADC Home > Advanced Search

Advanced Search

Search Results Error ADQL Help

Search Reset

Observation Constraints

- ▶ Observation ID
- ▶ P.I. Name
- ▶ Proposal ID
- ▶ Proposal Title
- ▶ Proposal Keywords

Science and Calibration data

Spatial Constraints

- ▶ Target
- ▶ Pixel Scale
- Do Spatial Cutout

Temporal Constraints

- ▶ Observation Date
- ▶ Integration Time
- ▶ Time Span

Spectral Constraints

- ▶ Spectral Coverage
- ▶ Spectral Sampling
- ▶ Bandpass Width
- ▶ Rest-frame Spectral Coverage
- Do Spectral Cutout

Additional Constraints

Band	Collection	Instrument	Filter	Calibration Level	Data Type	Observation Type
All (8)	DAOPLATES	All (9)	All (584)	All (3)	All (2)	All (1)
Gamma-ray	FUSE	ACS	182NM_MBP	(1) Raw Standard	image	object
Infrared	HST	FOC	191NM_NBP_(CIII)	(2) Calibrated	spectrum	
Millimeter	HSTHLA	FOS	270NM_MBP	(3) Product		
Optical	IRIS	HRS	280NM_NBP(MG1)			
Radio	JCHT	NICMOS	Blank			
UV	MACHO	STIS	CLEAR_FOC/96			
X-ray	OMM	WFC3	CLEAR_HRC			
Unknown	UKIRT	WFPC	CLEAR_NIC1			
	VGPS	WFPC2	CLEAR_NIC2			

Date modified: 2014-05-01

Terms and conditions | Transparency

News

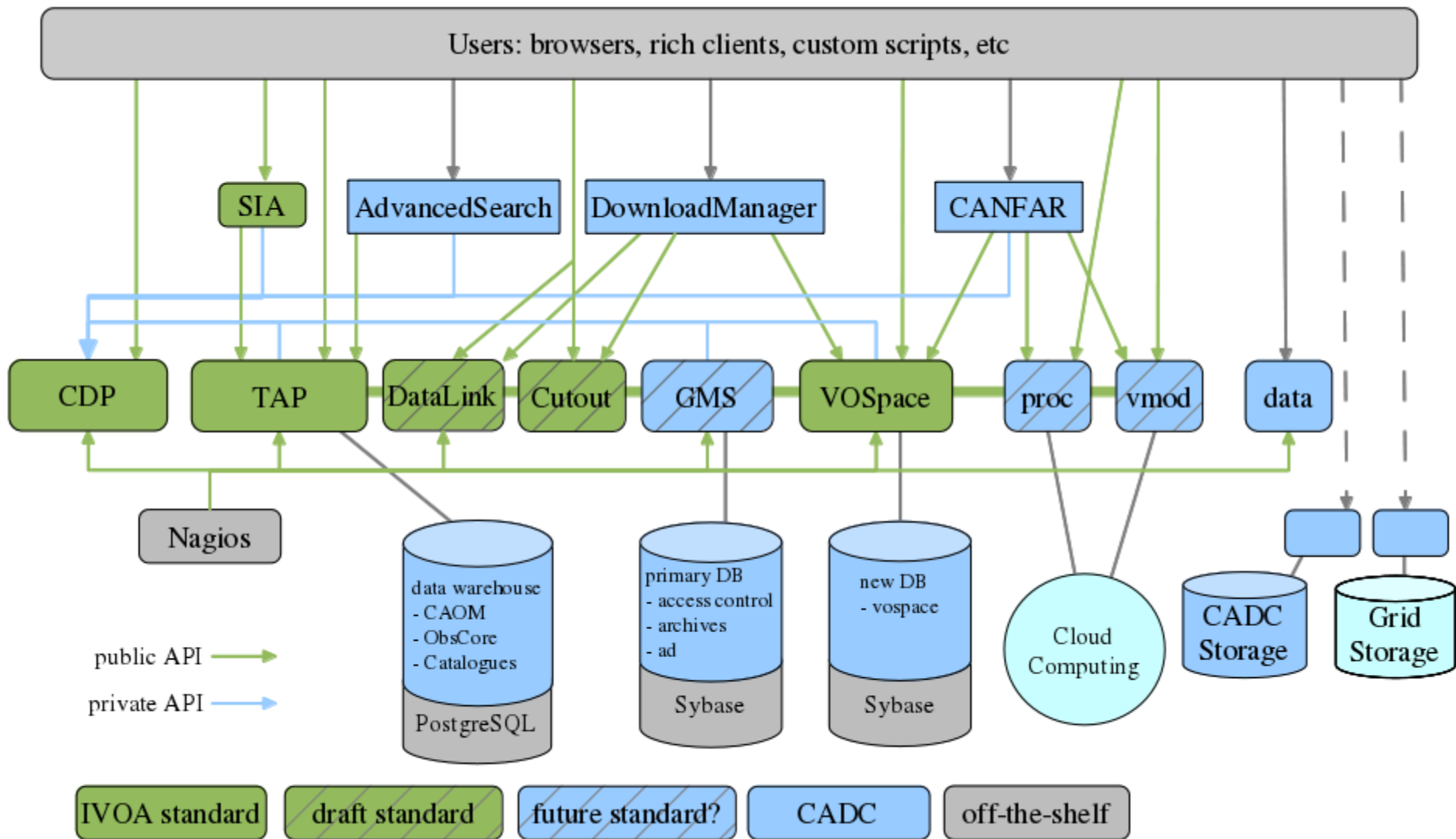
About us
Our mandate
Acknowledgements

Contact us
Email
Address

Virtual Observatory

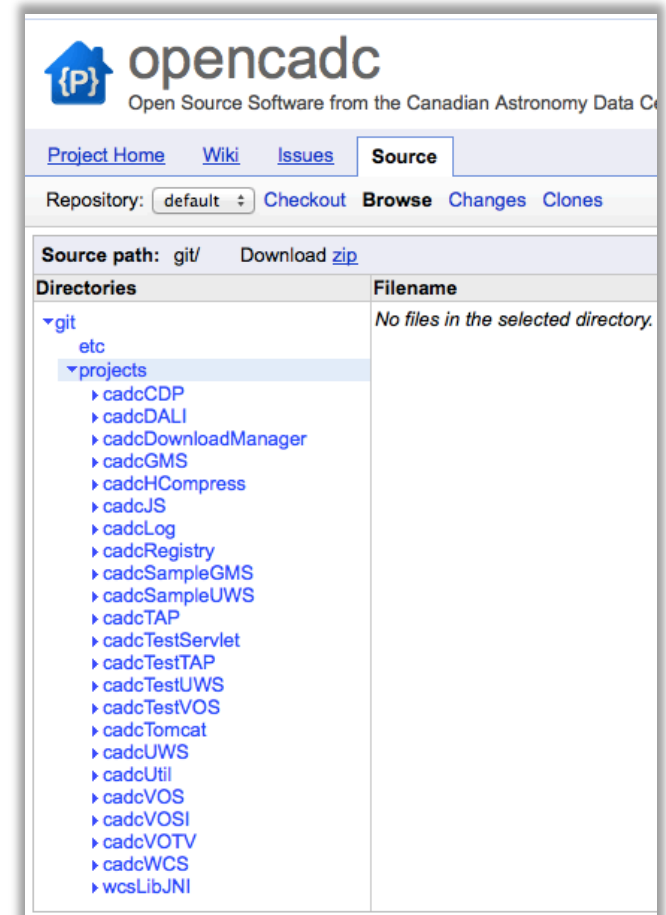
- New perspective on VO standards
- Services
 - Data Models: **ObsCore**, **SIA** (Simple Image) versions 1 and 2
 - Queries: **ADQL**, **TAP** (Table Access) and **VOSpace**
 - Data access: **DataLink**
 - Authorisation: **CDP** (Credential Delegation)
 - **UWS** (Universal Worker Service)
 - Notifications: RSS capability on **TAP** and **VOSpace**
 - Monitoring: **VOSI** (Standard Interface)
 - Storage: **VOSpace 2.0**
- If you use CADC services, you are using VO services





Development Environment

- Linux Fedora
- Core development:
 - Java, Python and C
- Browser app:
 - Javascript, WET
- GIT
- Agile with 3 week iterations
 - Continuous integration (Jenkins)
- Open source repository:
 - VO-related modules
 - CAOM data model tools



Operational Environment

- Linux: RHE and SL
- Multi-Gb LAN
- 1 Gb WAN → 10 Gb (real soon!)
- Service oriented architecture
 - Ecosystem of web applications and interacting RESTful web services
- Web Servers:
 - Apache/Tomcat
 - 4 master (round-robin); 4 slave
 - 1 beta; 2 development
- Databases:
 - Sybase: 2 operational; 1 hot-spare; 1 development
 - Postgres: 1 operational; 1 development with operational mirror
- Storage and Processing
 - On-site and off-site clusters integrated in a unified infrastructure

Current Projects – Advanced Search

Canadian Astronomy Data Centre

Canada

Telescope Data Products | Advanced Data Products | Services | Advanced Search | Login

CADC Home > Advanced Search

Advanced Search

Search | Results | Error | ADQL | Help

Search | Reset

Observation Constraints

- ▶ Observation ID
- ▶ P.I. Name
- ▶ Proposal ID
- ▶ Proposal Title
- ▶ Proposal Keywords

Science and Calibration data

Spatial Constraints

- ▶ Target
- ▶ Pixel Scale
- Do Spatial Cutout

Temporal Constraints

- ▶ Observation Date
- ▶ Integration Time
- ▶ Time Span

Spectral Constraints

- ▶ Spectral Coverage
- ▶ Spectral Sampling
- ▶ Bandpass Width
- ▶ Rest-frame Spectral Coverage
- Do Spectral Cutout

Additional Constraints

Band	Collection	Instrument	Filter	Calibration Level	Data Type	Observation Type
All (8)	DAOPLATES	All (9)	All (584)	All (3)	All (2)	All (1)
Gamma-ray	FUSE	ACS	182NM_MBP	(1) Raw Standard	image	object
Infrared	HST	FOC	191NM_NBP_(CIII)	(2) Calibrated	spectrum	
Millimeter	HST/HLA	FOS	270NM_NBP	(3) Product		
Optical	IRIS	HRS	280NM_NBP(MG11)			
Radio	JCMT	NICMOS	Blank			
UV	MACHO	STIS	CLEAR_FOC/96			
X-ray	OMM	WFPC3	CLEAR_HRC			
Unknown	UKIRT	WFPC	CLEAR_NIC1			
	VGPS	WFPC2	CLEAR_NIC2			

Date modified: 2014-05-01

Terms and conditions | Transparency

News

Contact us
Email
Address

About us
Our mandate
Acknowledgements

- Improving metadata
- Adding functionality
 - Drill down
 - Single-click download
 - Auto-completion
 - ADQL editing and submission
- DataLink interface
- Improving query performance
- Many other ideas!

Current Projects – ALMA

- ALMA Science Archive Query Interface
 - Maintenance
 - Development (VO interfaces)

ALMA Science Archive Query

Query Form | Results Table

Search | Reset [Query Help](#)

Position Source name (Sesame) Source name (ALMA) RA Dec	Energy Frequency Bandwidth Spectral resolution Band	Time Observation date Integration time	Polarisation Polarisation type
Observation Water vapour	Project Project code Project title PI name		Options View: <input checked="" type="radio"/> raw data <input type="radio"/> project <input checked="" type="checkbox"/> public data only <input checked="" type="checkbox"/> science observations only

Current Projects – SKA

- Science Data Processor Work Package
 - DELIV and DATA sub-elements
- Moving data and processing to regional centres
- CANFAR-like infrastructure
- Pre-construction phase
- Contributing to data models and user facing services
 - VO is an important part



Current Projects – MOST

- Microvariability & Oscillations of Stars Telescope
 - 11 year mission ending later this year
 - First time-series collection into CAOM



Canadian Advanced Network for Astronomical Research

A Cloud Ecosystem for Data Intensive Astronomy

- A user facing service
 - Create and interactively configure a VM
 - Store and share a VM
 - Run batches with a VM
 - Store and share data
- Using research cloud resources
 - Compute Canada
 - CADC



University
of Victoria



University of
British Columbia

canarie

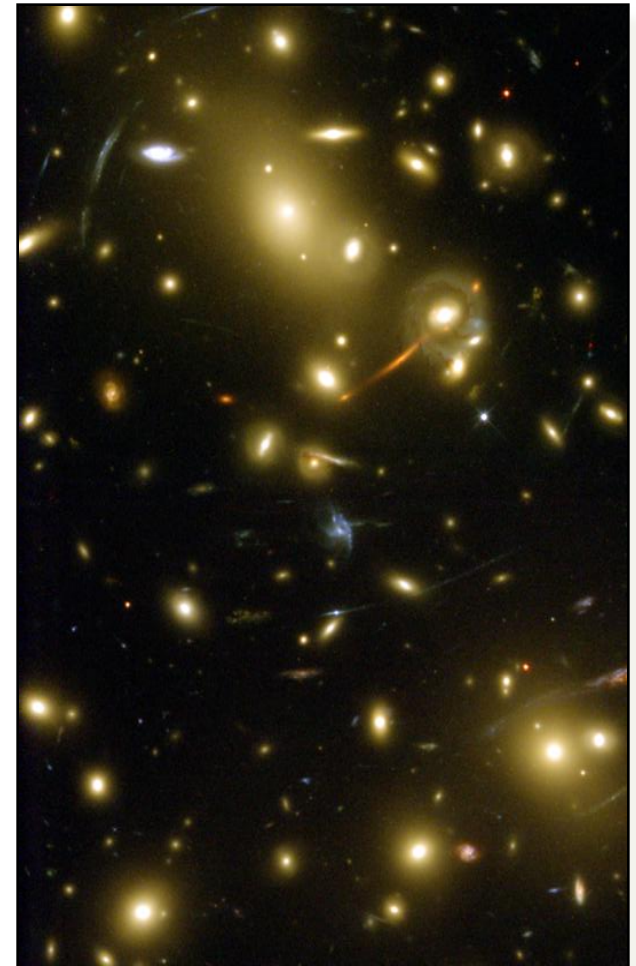


compute + calcul
CANADA



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



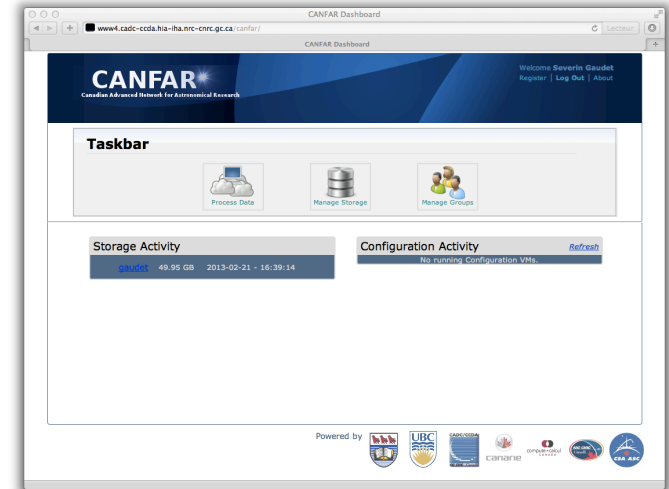
Context: Compute Canada

- Large national computing infrastructure
 - Agencies pushing researchers to use it
- Limited success in data-intensive astronomy
 - Users must adapt to local OS, software and policies
 - Conflicting demands
 - Limited mobility



Solution: CANFAR

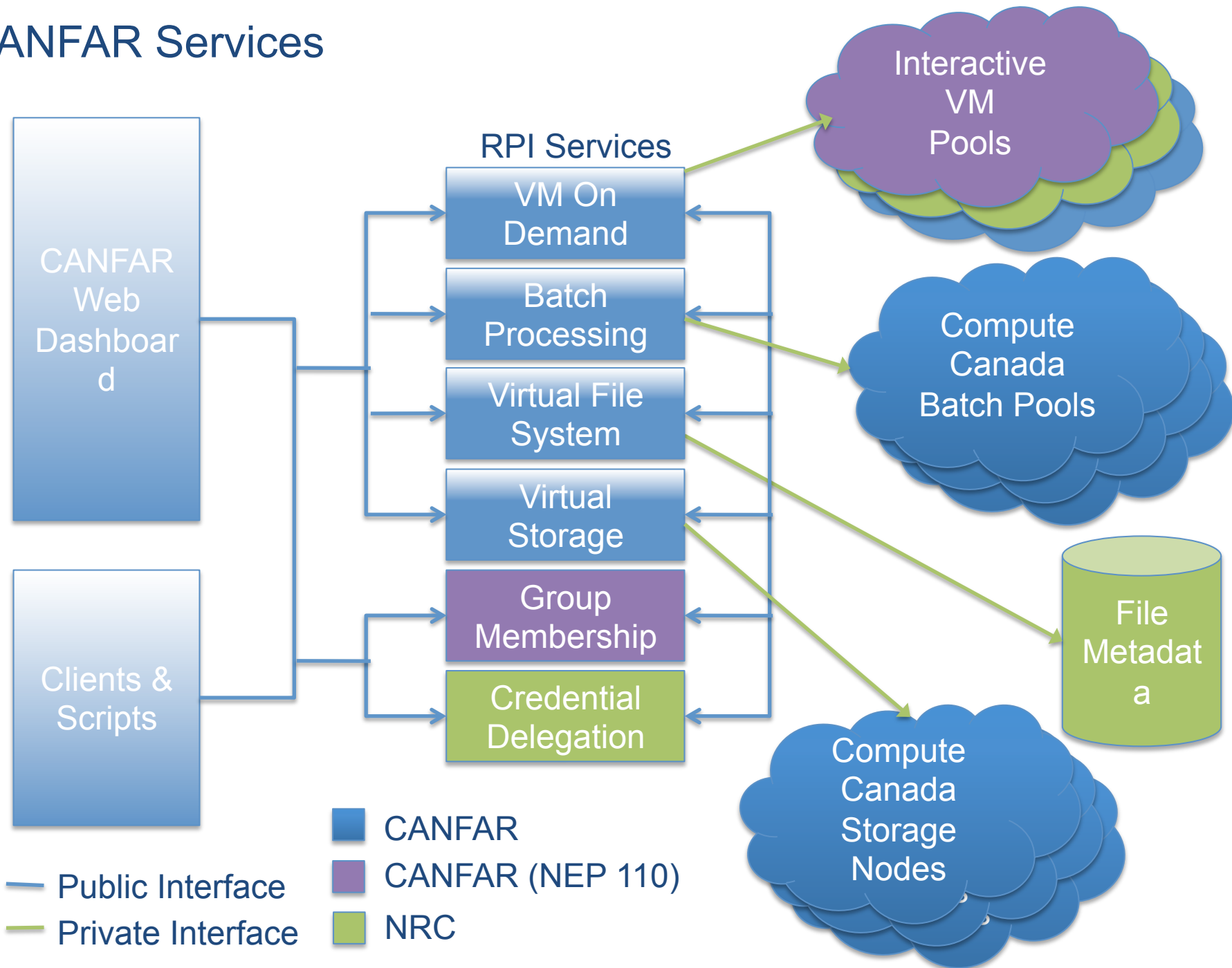
- A new model for supporting community projects
 - Platform for collaborative teams on distributed computing resources
 - Extensible on Compute Canada
- Elements
 - Virtual Storage (VOSpace)
 - Virtual machines
 - Virtual Clusters (Cloud)
 - Group management
 - VO Standards
- Developed and operated by CADC
- If you use CADC services, you are using CANFAR services



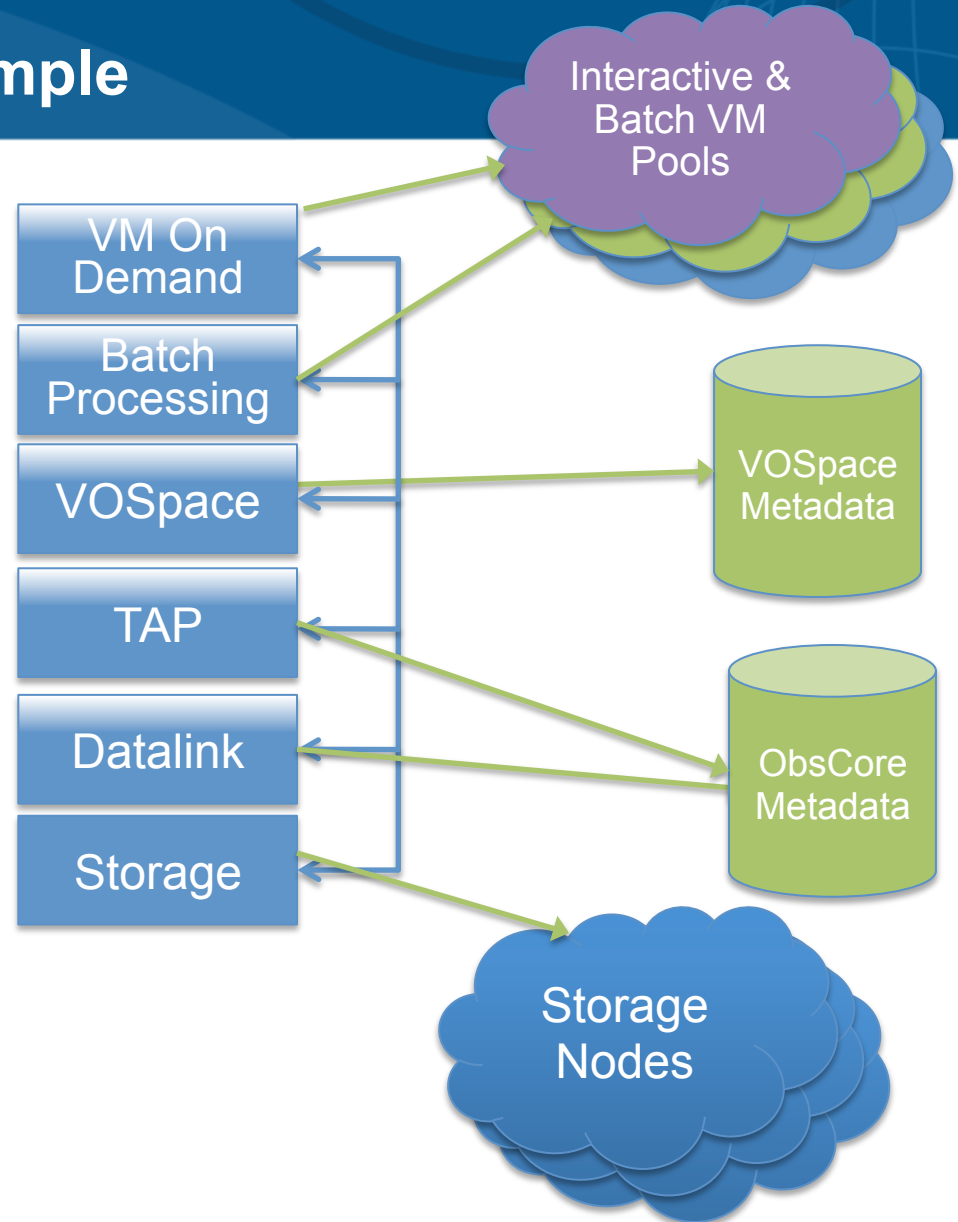
Users

- AAVSO Photometric All-Sky Survey
- ANDROIDS
- CFHT Ecliptic Plane Survey
- CFHT Legacy Survey
- CFHT Megapipe
- CFHTLens
- Disc Emission via a Bias-free Reconnaissance
- HST Processing
- JCMT All-Sky Survey
- JCMT Cosmology Legacy Survey
- JCMT Gould Belt Survey
- Kuiper Belt objects simulations
- L-Band Local Group Legacy Survey
- MACHO Reprocessing
- Millennium dark matter simulations studies
- New Horizons Target Search
- Next Generation Virgo Cluster Survey
- NuGrid stellar evolution
- Outer Solar System Origins Survey
- Pan-Andromeda Archaeological Survey
- Supernova simulations

CANFAR Services

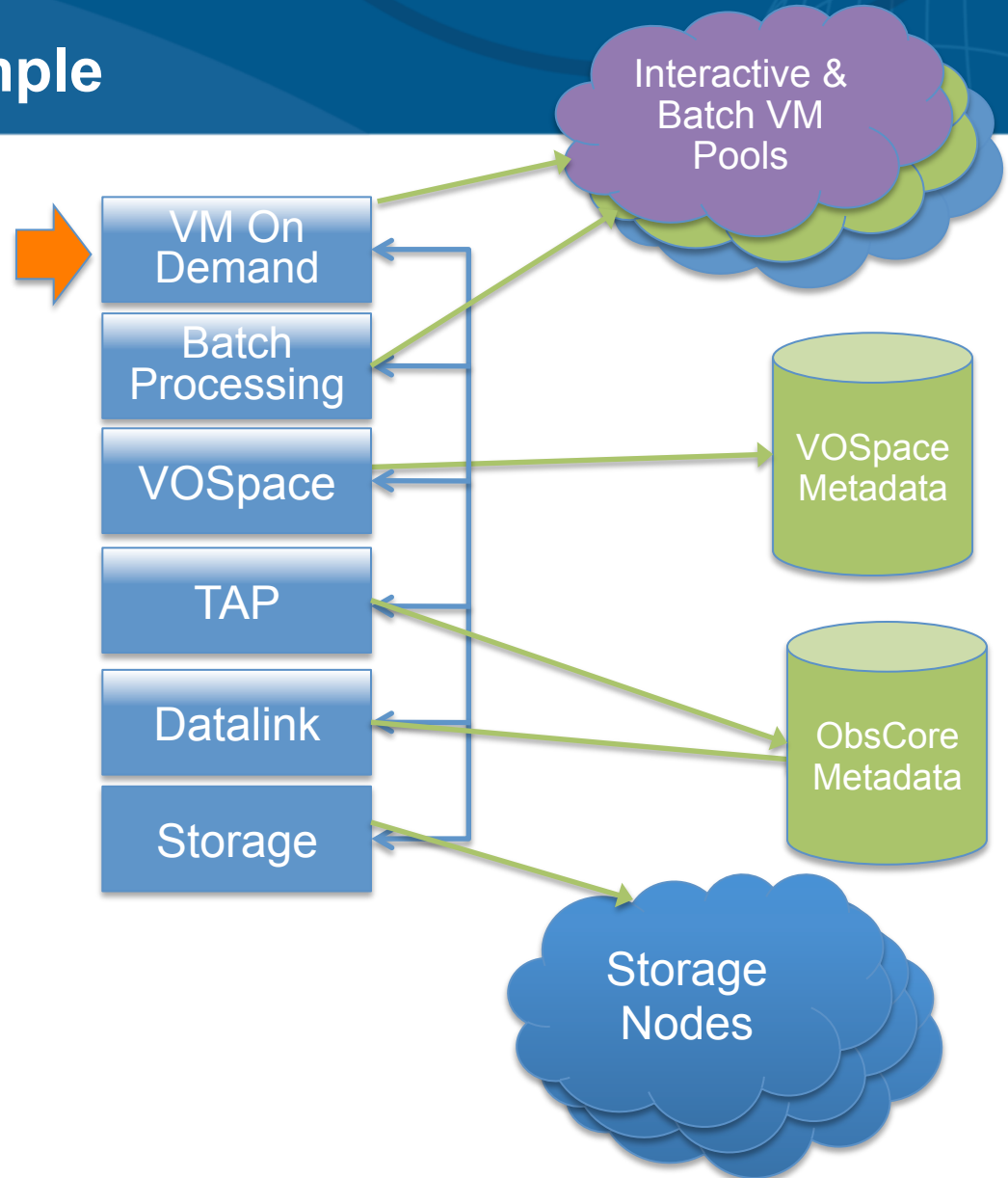


CANFAR VO Science Example



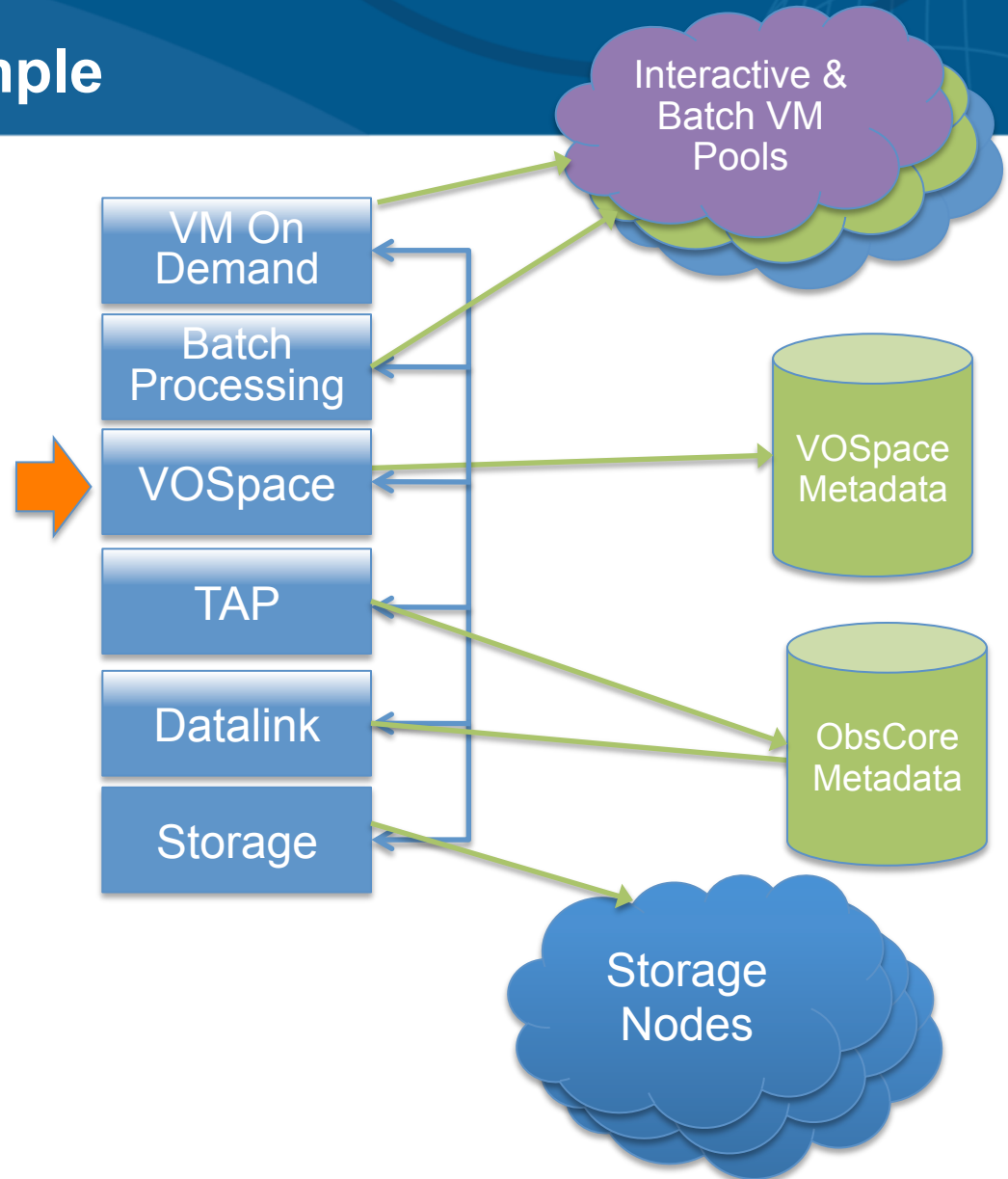
CANFAR Science Example

① User creates and configures VM



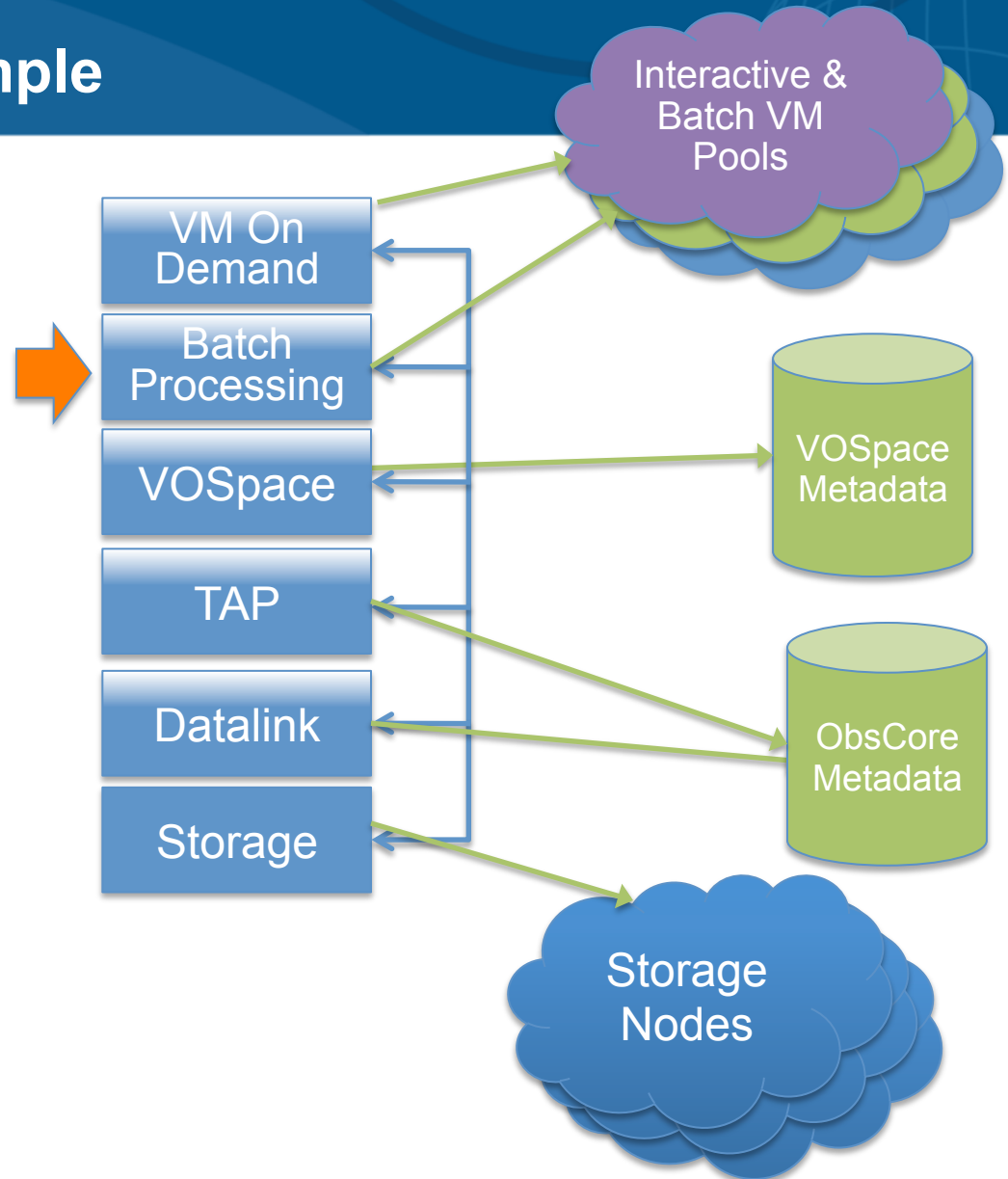
CANFAR Science Example

- ① User creates and configures VM
- ② User saves VM img in VOSpace



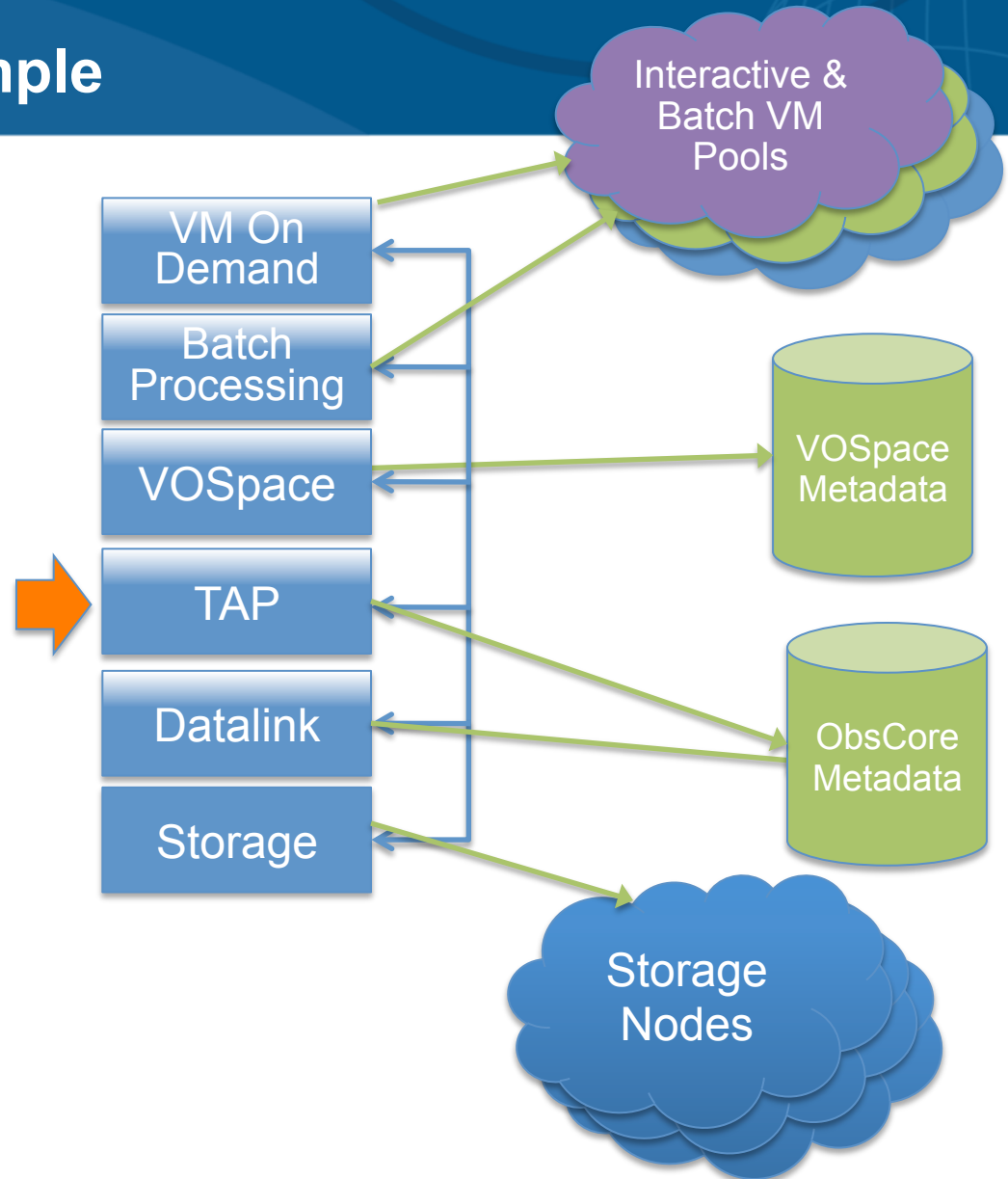
CANFAR Science Example

- ① User creates and configures VM
- ② User saves VM img in VOSpace
- ③ User launches X instances of image in batch processing



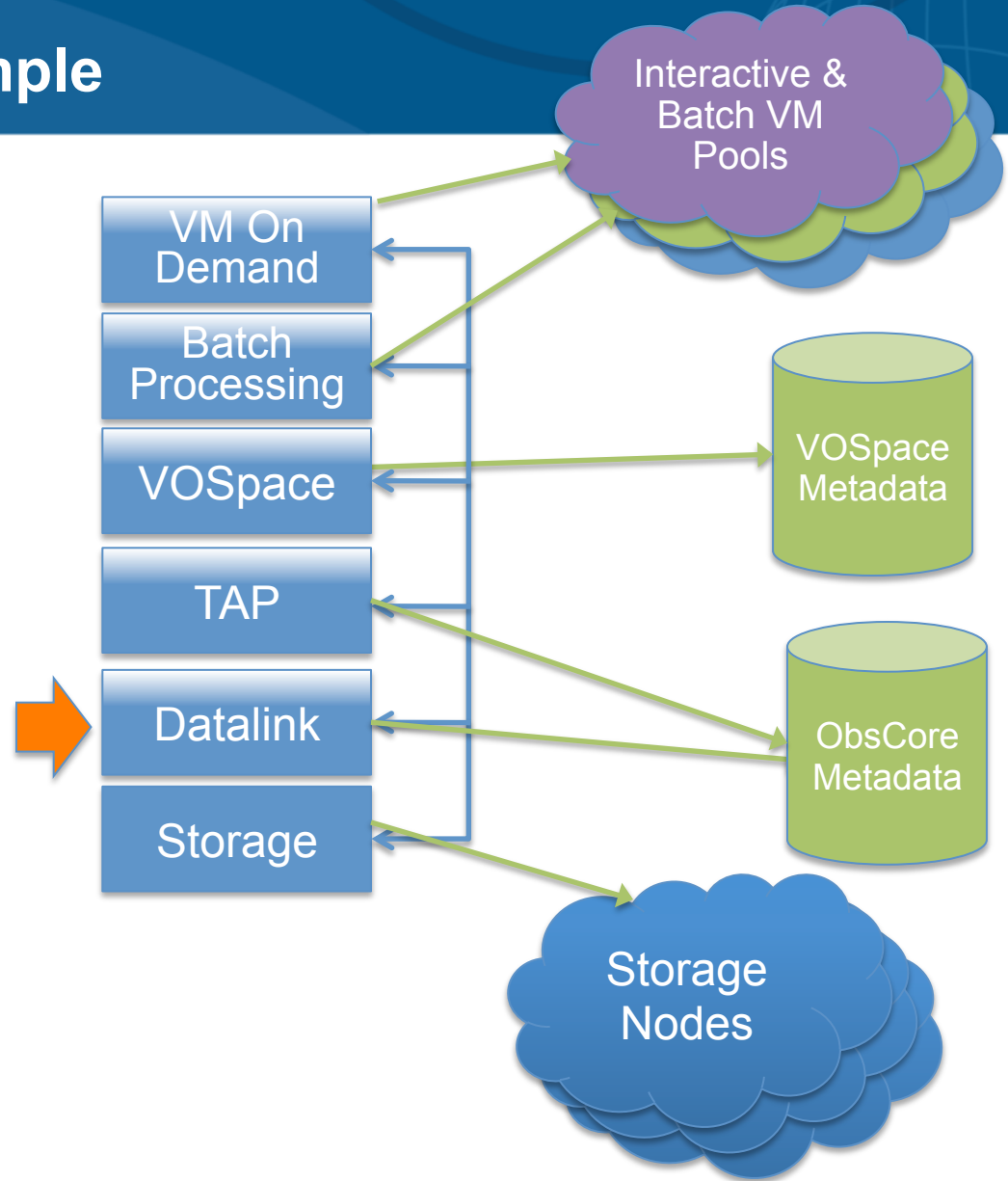
CANFAR Science Example

- ① User creates and configures VM
- ② User saves VM img in VOSpace
- ③ User launches X instances of image in batch processing
- ④ VMs use TAP to find data from ObsCore



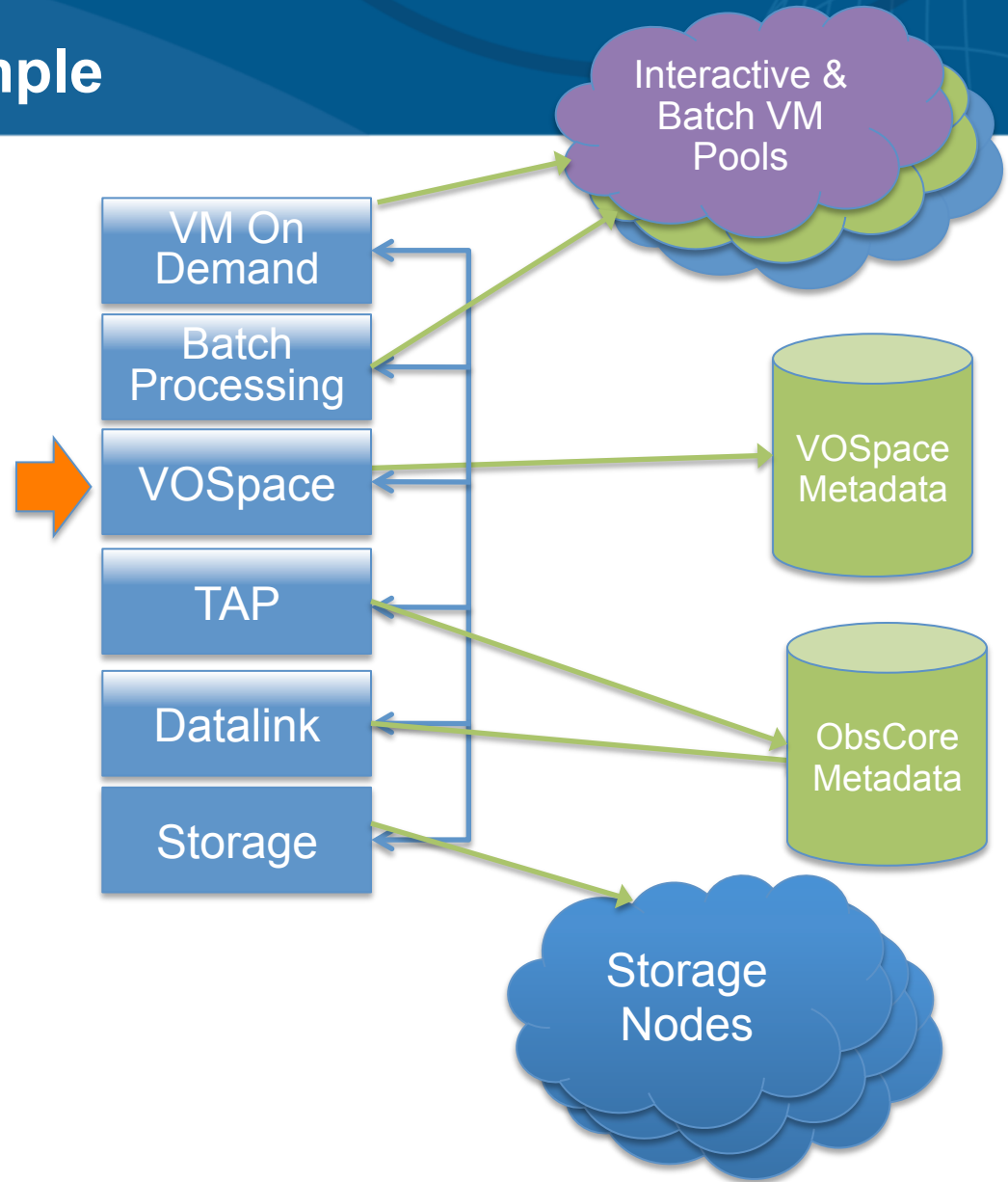
CANFAR Science Example

- ① User creates and configures VM
- ② User saves VM img in VOSpace
- ③ User launches X instances of image in batch processing
- ④ VMs use TAP to find data from ObsCore
- ⑤ VMs use Datalink to access data



CANFAR Science Example

- ① User creates and configures VM
- ② User saves VM img in VOSpace
- ③ User launches X instances of image in batch processing
- ④ VMs use TAP to find data from ObsCore
- ⑤ VMs use Datalink to access data
- ⑥ VMs save science results in VOSpace



Sites

Processing

- University of Victoria
 - 2 cores/job; 2 GB/core
 - 720 core years
- University of Calgary
 - 24 cores/job; 8 GB/core
 - 58 cores years
- CADDC
 - 2 cores/job; 2 GB/core
 - 176 core years

Storage

- University of Victoria
 - 655 TB
 - Near-line backup
- University of Saskatchewan
 - 655 TB
- CADDC
 - 1.4 PB

Virtual Machine on Demand

The screenshot shows the CANFAR web interface. At the top left is the CANFAR logo with the text "Canadian Advanced Network for Astronomical Research". At the top right, it says "Welcome Sebastien Fabbro" with links for "Register", "Log Out", and "About". Below the header is a "Processing" section with a navigation bar containing buttons for "New VM", "Boot VM", "Running VMs", "Submit Job(s)", and "Running Jobs". The "Boot VM" button is selected. The main content area is titled "Boot Virtual Machine" and contains a form for "Virtual Machine Location". The form has two input fields: "VM Name:" with the value "adass" and "Location:". Below these fields is a "- OR -" separator. Further down are three dropdown menus: "Processor Cores:" set to "2", "Memory (GB):" set to "12", and "Staging Disk space (GB):" set to "500". At the bottom left of the form are "Boot" and "Reset" buttons.

CANFAR Rest API



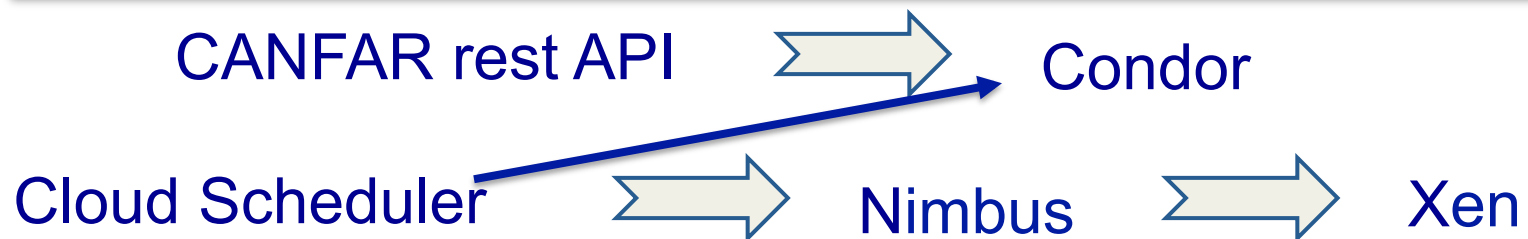
Nimbus



Xen

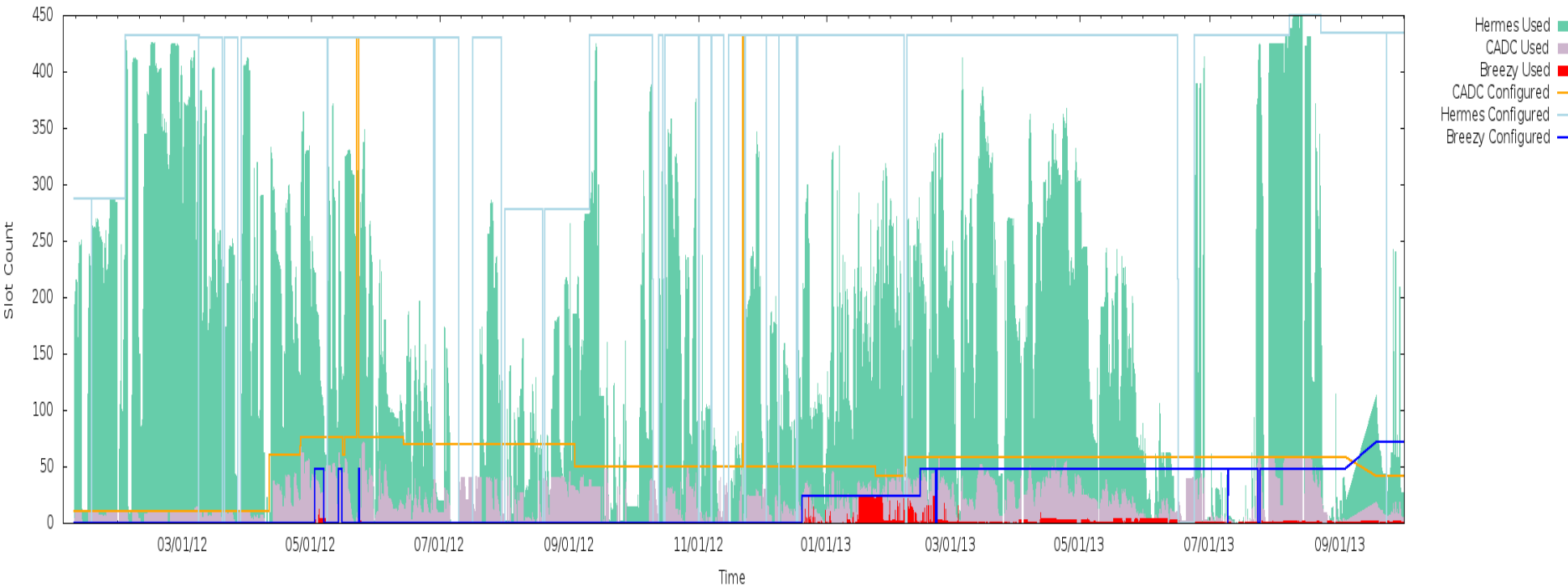
Multi-cluster Batch Processing

The screenshot shows the CANFAR web interface. At the top left is the CANFAR logo with the text "Canadian Advanced Network for Astronomical Research". At the top right, it says "Welcome Severin Gaudet" with links for "Register", "Log Out", and "About". Below the header is a "Processing" section with a navigation bar containing buttons for "New VM", "Boot VM", "Running VMs", "Submit Job(s)", and "Running Jobs". The "Submit Job(s)" button is highlighted. Below this is the "Job Submission" form. It has a "Virtual Machine Location" section with a "VM Name:" field containing ".img.gz" and a "VOSpace:" dropdown menu with four options: "vos://cadc.nrc.calvospace/ShaimaaAli/vmstore", "vos://cadc.nrc.calvospace/adriand/vmstore", "vos://cadc.nrc.calvospace/aikema/vmstore", and "vos://cadc.nrc.calvospace/chiggs/vmstore". Below the VOSpace options is a "- OR -" separator and a "URI/URL:" field. At the bottom of the form are three dropdown menus for "Processor Cores:" (set to 1), "Memory (GB):" (set to 3), and "Staging Disk space (GB):" (set to 50).

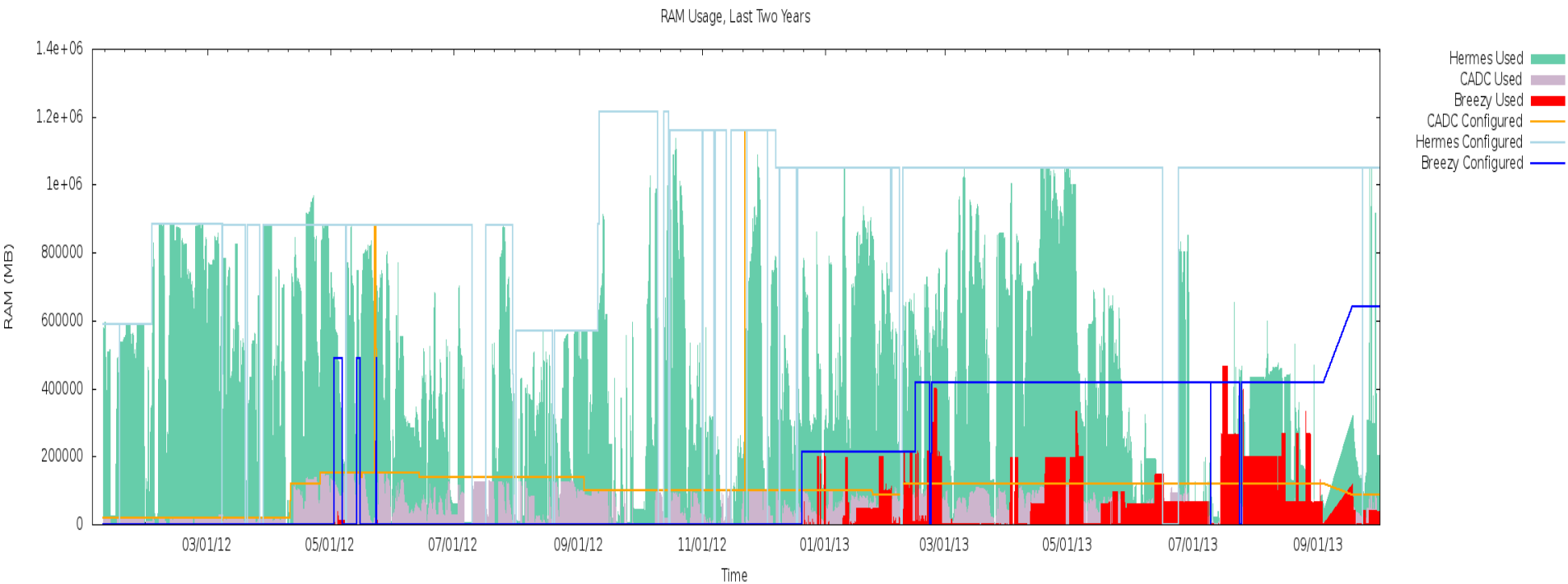


Usage: CPU

Slot Count Usage, Last Two Years

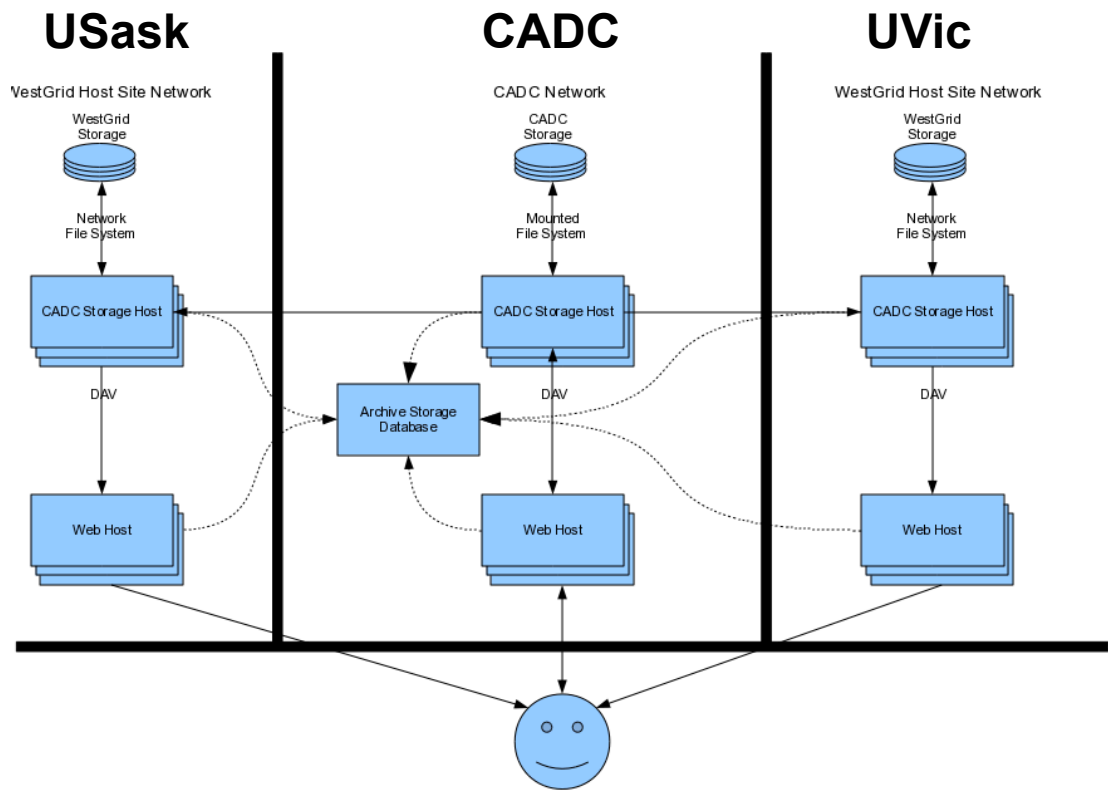


Usage: RAM



Virtual Object Store

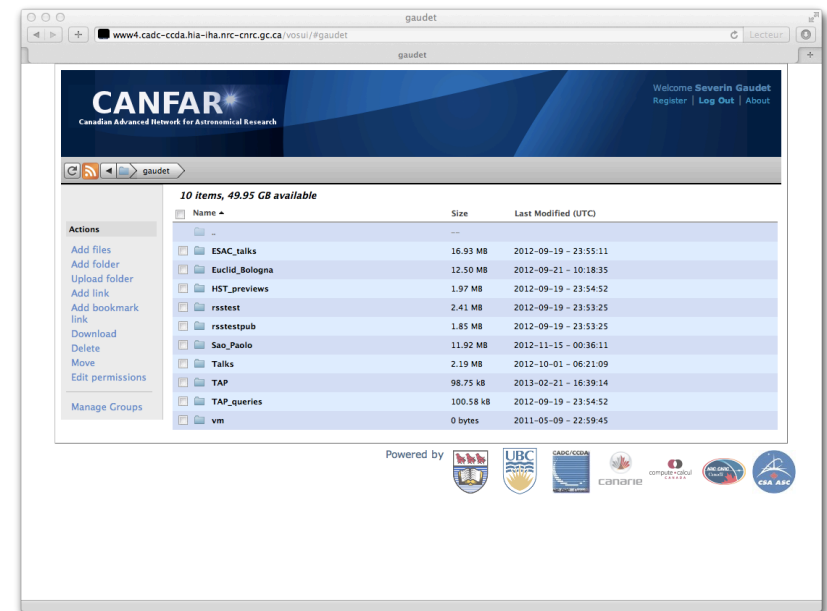
- Storage web services using several distributed storage resources
- Optimization and QoS strategies not user nor provider dependent
- Same system for both archive and users
- File in/File out




Virtual File System: VOSpace


- User storage
- Virtual machine image repository
- Cloud processing file persistence
 - Asynchronous query input and output
- Programmatic access
 - curl, wget, vosClient
- Browser UI
- Mount-able file system (vofs)

- Full access control
- Notifications via RSS feeds



Virtual File System: VOSpace

Welcome **Severin Gaudet**
[Register](#) | [Log Out](#) | [About](#)



10 items, 49.95 GB available

<input type="checkbox"/> Name ▲	Size	Last Modified (UTC)
..	---	
<input type="checkbox"/> ESAC_talks	16.93 MB	2012-09-19 - 23:55:11
<input type="checkbox"/> Euclid_Bologna	12.50 MB	2012-09-21 - 10:18:35
<input type="checkbox"/> HST_previews	1.97 MB	2012-09-19 - 23:54:52
<input type="checkbox"/> rsstest	2.41 MB	2012-09-19 - 23:53:25
<input type="checkbox"/> rsstestpub	1.85 MB	2012-09-19 - 23:53:25
<input type="checkbox"/> Sao_Paolo	11.92 MB	2012-11-15 - 00:36:11
<input type="checkbox"/> Talks	2.19 MB	2012-10-01 - 06:21:09
<input type="checkbox"/> TAP	98.75 kB	2013-02-21 - 16:39:14
<input type="checkbox"/> TAP_queries	100.58 kB	2012-09-19 - 23:54:52
<input type="checkbox"/> 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](#)

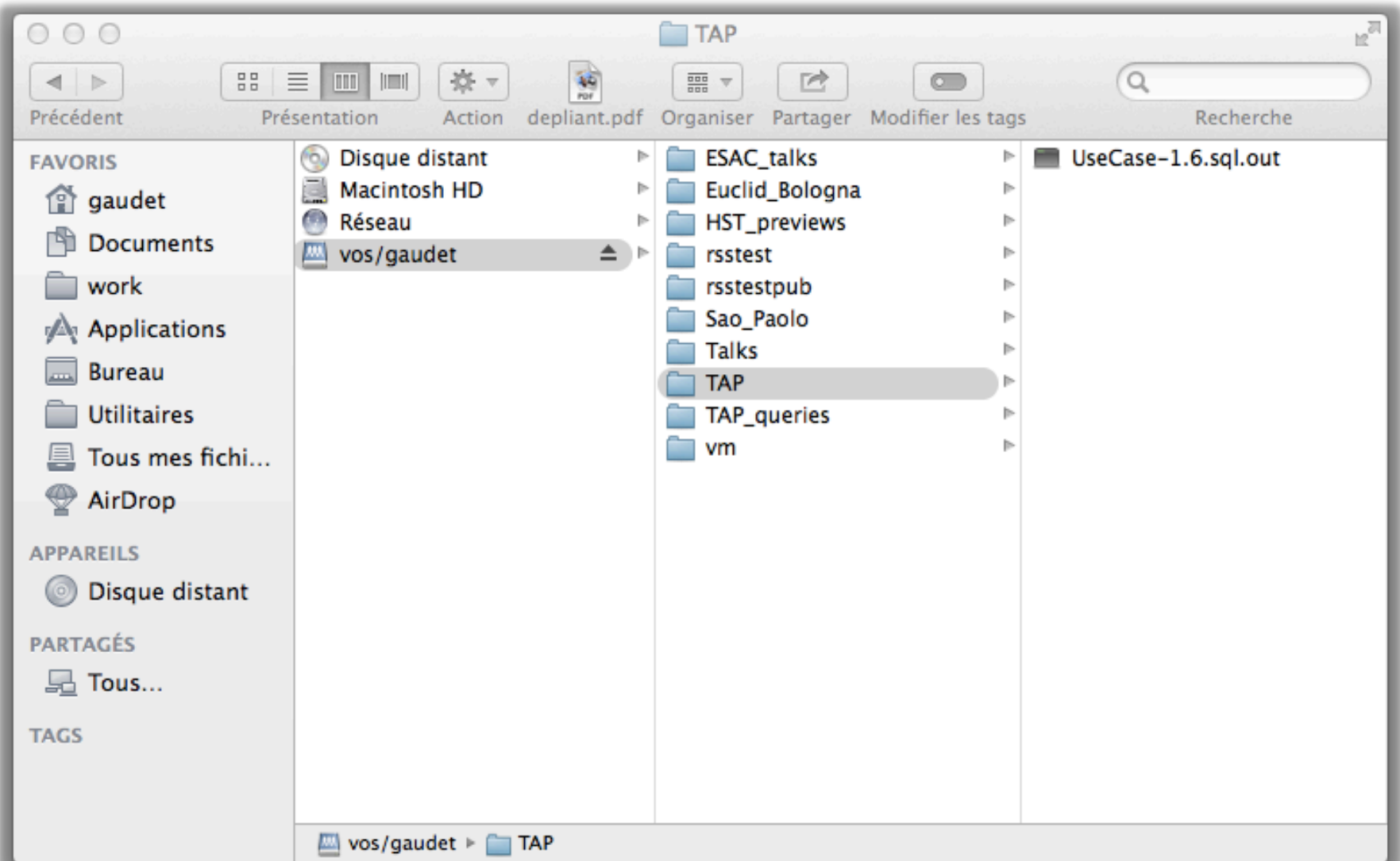
Powered by



Virtual File System: VOSpace

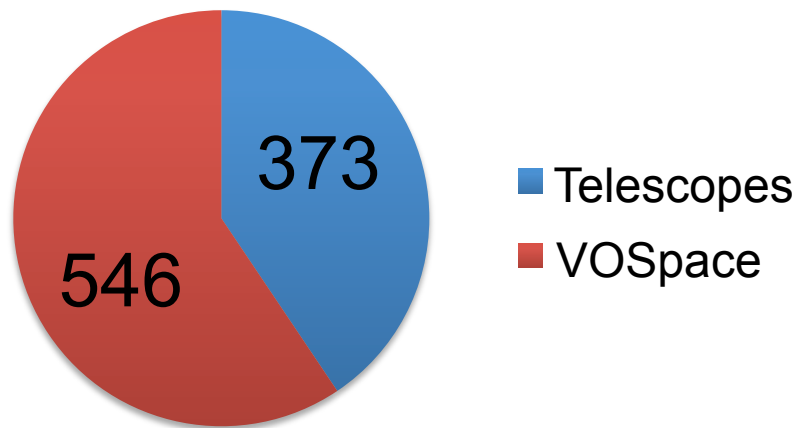
```
Terminal — tcsh — ttys006
[daxka-2:~] gaudet% mountvofs --vospace vos:gaudet --mountpoint /Volumes/gaudet
[daxka-2:~] gaudet% cd /Volumes/gaudet
[daxka-2:/Volumes/gaudet] gaudet% ls
ESAC_talks/   HST_previews/  TAP/           Talks/         rsstestpub/
Euclid_Bologna/ Sao_Paolo/     TAP_queries/   rsstest/       vm/
[daxka-2:/Volumes/gaudet] gaudet% █
```

Virtual File System: VOSpace



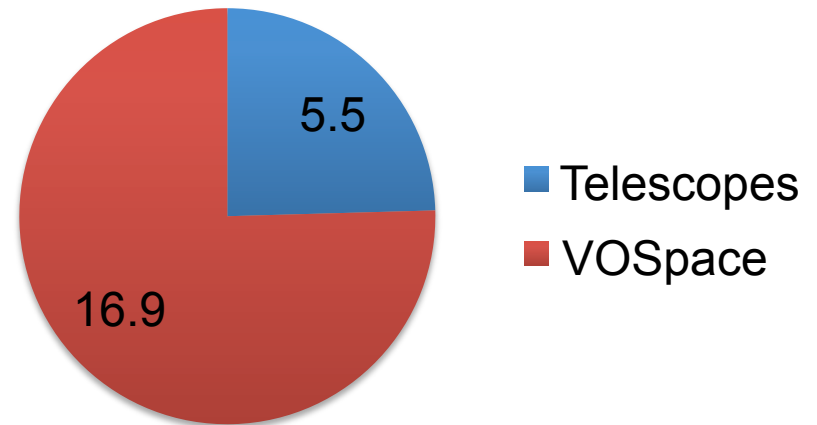
VOSpace Usage in 2013

Bytes (TB)



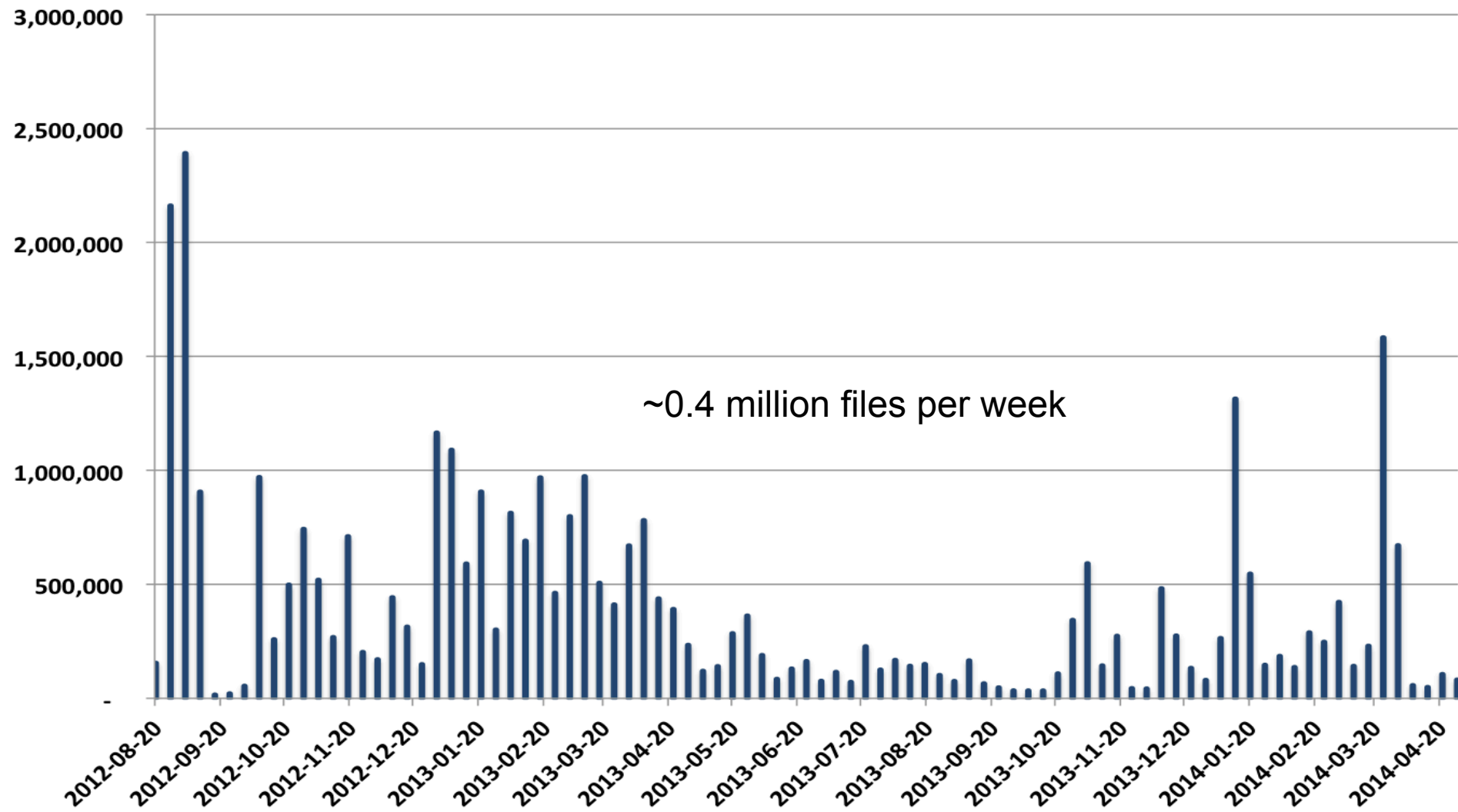
60% VOSpace

Files (millions)

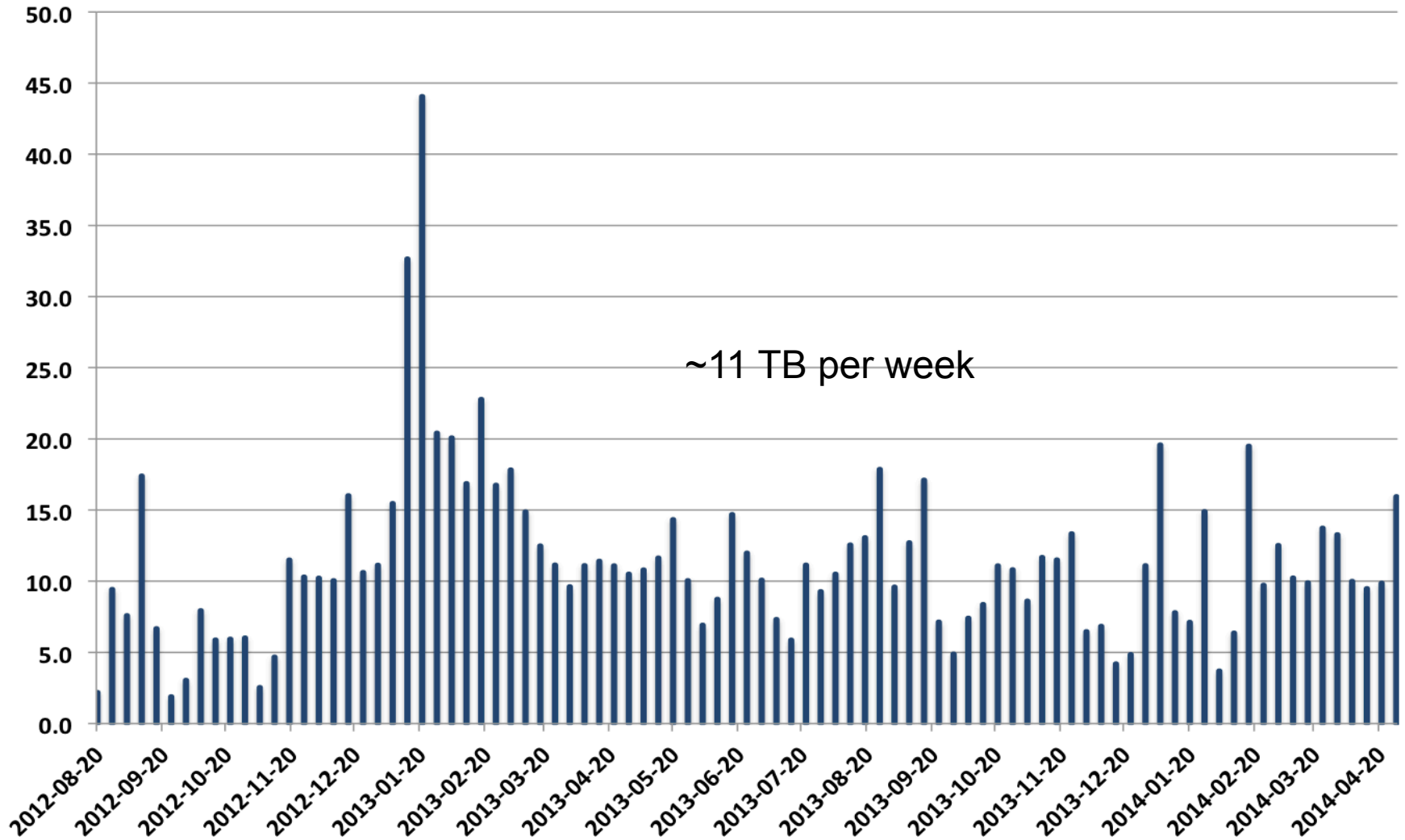


75% VOSpace

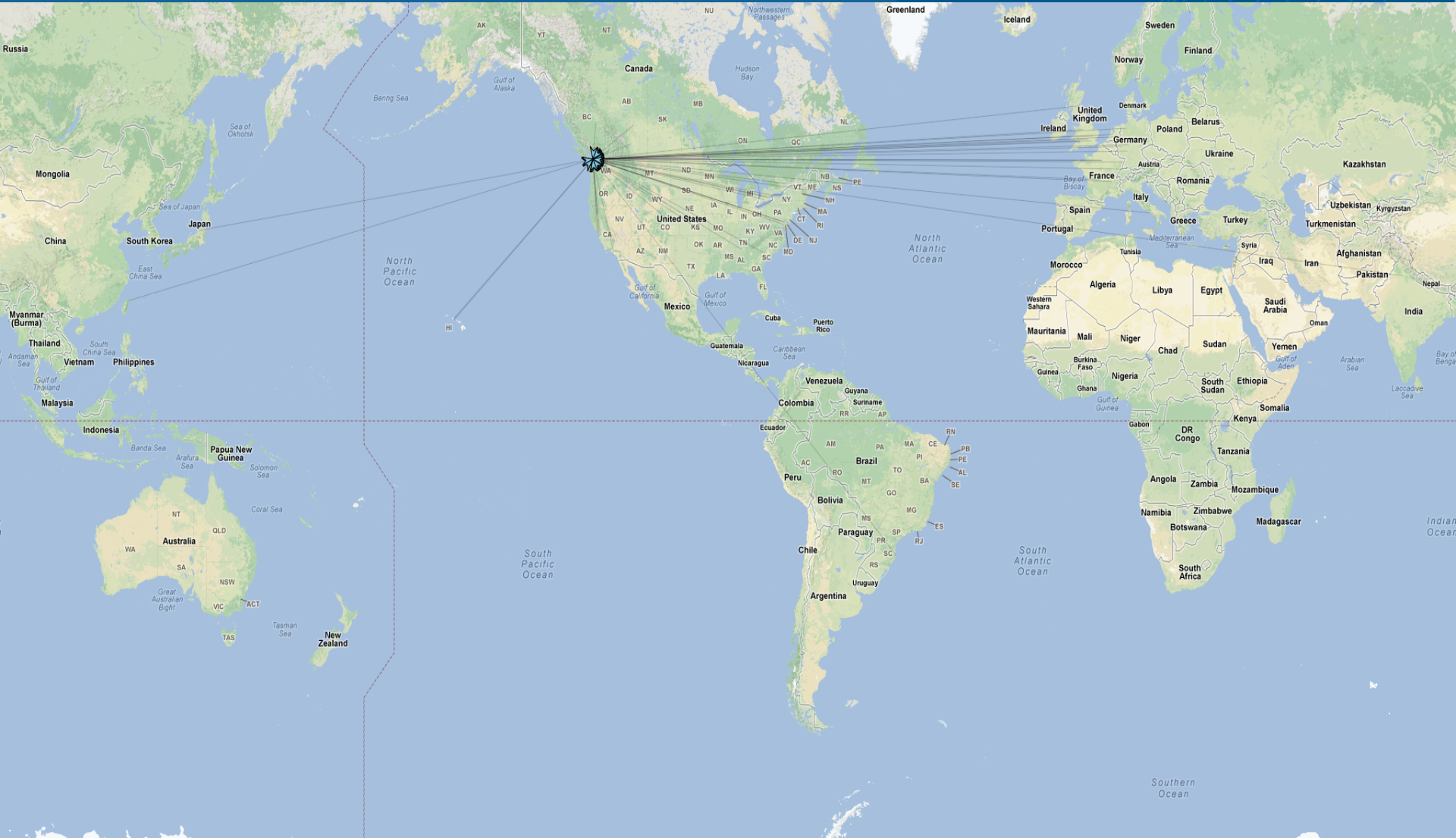
VOSpace GETs: Files per week (Aug 2012 – Apr 2014)



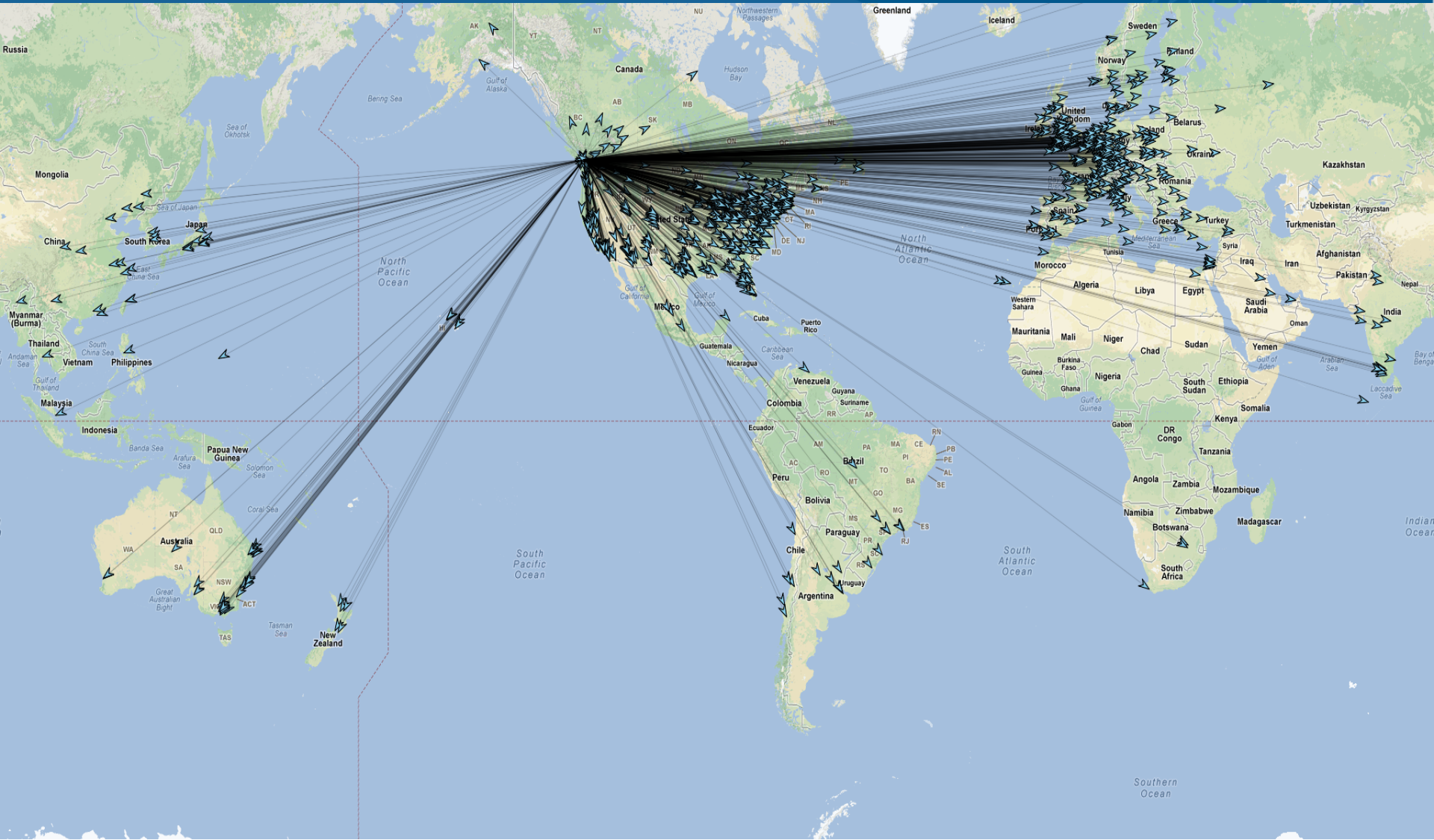
VOSpace GETs: TB per week (Aug 2012 – Apr 2014)



Geography of VOSpace PUTs



Geography of VOSpace GETs



The Power of VOspace Views

- VOspace Views: an powerful optimization
- “Move the code to the data, not the data to the code”
- Views allow one to define a set of operations that usually reduce the number of bytes that need to be transferred.
- Views in use at the CANFAR:
 - On FITS files (data nodes):
 - Cutout view, WCS view, FITS header (fhead)
 - On Container nodes:
 - Manifest view, RSS view

Growing Pains and Lessons Learned...

- As VOSpace usage grew, we had to adjust to meet demand
- Made and learned from mistakes along the way
- The bottleneck kept shifting: fixing one vulnerability would expose the next
- Examples:
 - Tuning database transactions, locking
 - Authorization techniques
 - Contention on root nodes in DB
 - Resource pooling
 - Recognizing our system limits, “try again later” rather than fail
 - Building smart clients, identifying problem ones

Access Control

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

The screenshot shows a web interface for 'Group Management' with the following elements:

- Breadcrumbs:** CADC Home > Group Management > Update Group
- Navigation:** A sidebar with 'Group Management' (active), 'Groups', 'New Group', and 'My Groups'.
- Form Fields:**
 - Owner Name:** Severin Gaudet
 - Name:** Input field containing 'MAP'
 - Description:** Textarea containing 'Multi-Archive Project'
 - Members:** A list with one entry: Severin Gaudet (with a checkbox).
 - User ID:** Input field with a help icon and placeholder text 'CADC Username or X.509 Distinguished Name'.
- Buttons:** 'Delete Selected Members' (yellow), 'Add member' (grey), 'Update' (blue), 'Delete' (blue), and 'Cancel' (grey).

Access Control

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

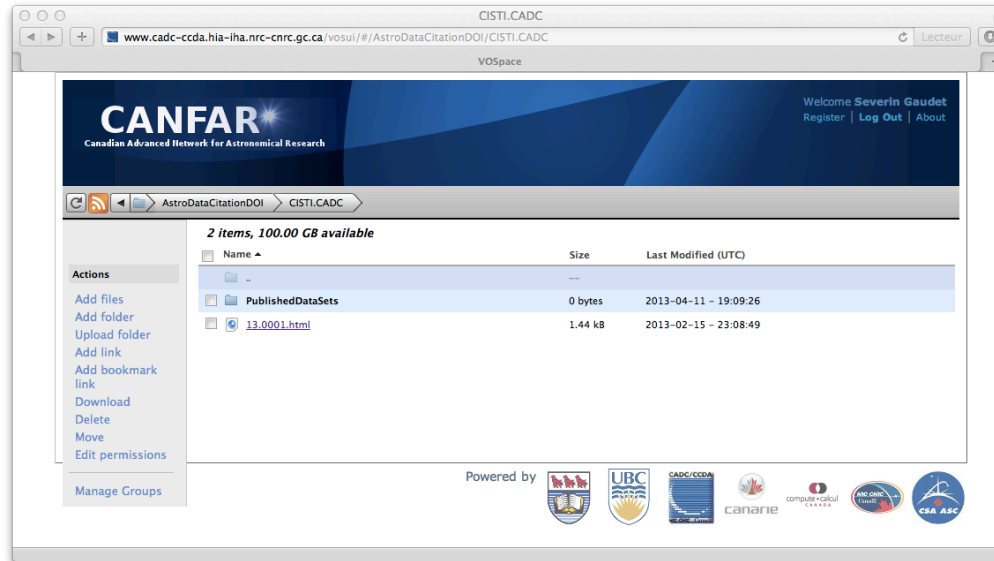
The screenshot shows a web interface for 'Group Management' with the following elements:

- Breadcrumbs:** CADC Home > Group Management > Update Group
- Navigation:** A sidebar with 'Group Management' (selected), 'Groups', 'New Group', and 'My Groups'.
- Page Title:** Group Management
- Section Title:** Update Group
- Form Fields:**
 - Owner Name: Severin Gaudet
 - Name: MAP
 - Description: Multi-Archive Project
 - Members: Severin Gaudet
 - User ID: CADC Username or X.509 Distinguished Name
- Buttons:** 'Delete Selected Members' (yellow), 'Add member' (grey), 'Update' (blue), 'Delete' (blue), and 'Cancel' (grey).

CANFAR Current Projects

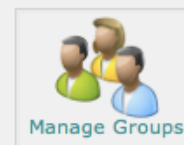
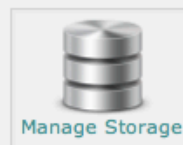
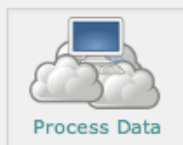
- Moving from Nimbus to OpenStack
 - Integrated identity and authorization
 - Virtual machine image mirroring
 - Incremental deployment
- VOSpace improvements
 - Random access
 - Scalability and robustness
 - Decentralized persistence
- Virtual machine on demand (VMOD)
 - Sharing virtual machines and vm images
 - Software as a service virtual machines
 - Ipython/julia notebook portal
 - Server-side visualization of large datasets
 - Long-lived virtual machines
 - Access control moving to LDAP

CANFAR Current Projects – Data Publishing



- Data Object Identifiers (Data publishing using VOSpace)
 - Science teams use VOSpace as working space
 - User creates a publishable directory
 - Transfer publish-able directory to CADAC control
 - Support refereeing, revision, DOI issuing and finally publication

Taskbar



Storage Activity

[gaudet](#) 49.95 GB 2013-02-21 - 16:39:14

Configuration Activity

[Refresh](#)

No running Configuration VMs.

Powered by


















[CADC Home](#)

[Advanced Search](#)





Telescope Data Products

 Gemini	 CFHT	 JCMT
 HST	 BLAST	 MOST
 DAO	 MACHO	 OMM
 FUSE	 UKIRT	

Advanced Data Products

 MegaPipe	 HLA
 IRIS	 CGPS
 CFHTLS	 WIRwolf

Services

 Meetings	 Community
 SSOIS	 CANFAR

Date modified: 2014-04-28

