Globus as a national data transfer platform

frederick.lefebvre@calculquebec.ca
GlobusWorld - April 2014

Compute Canada
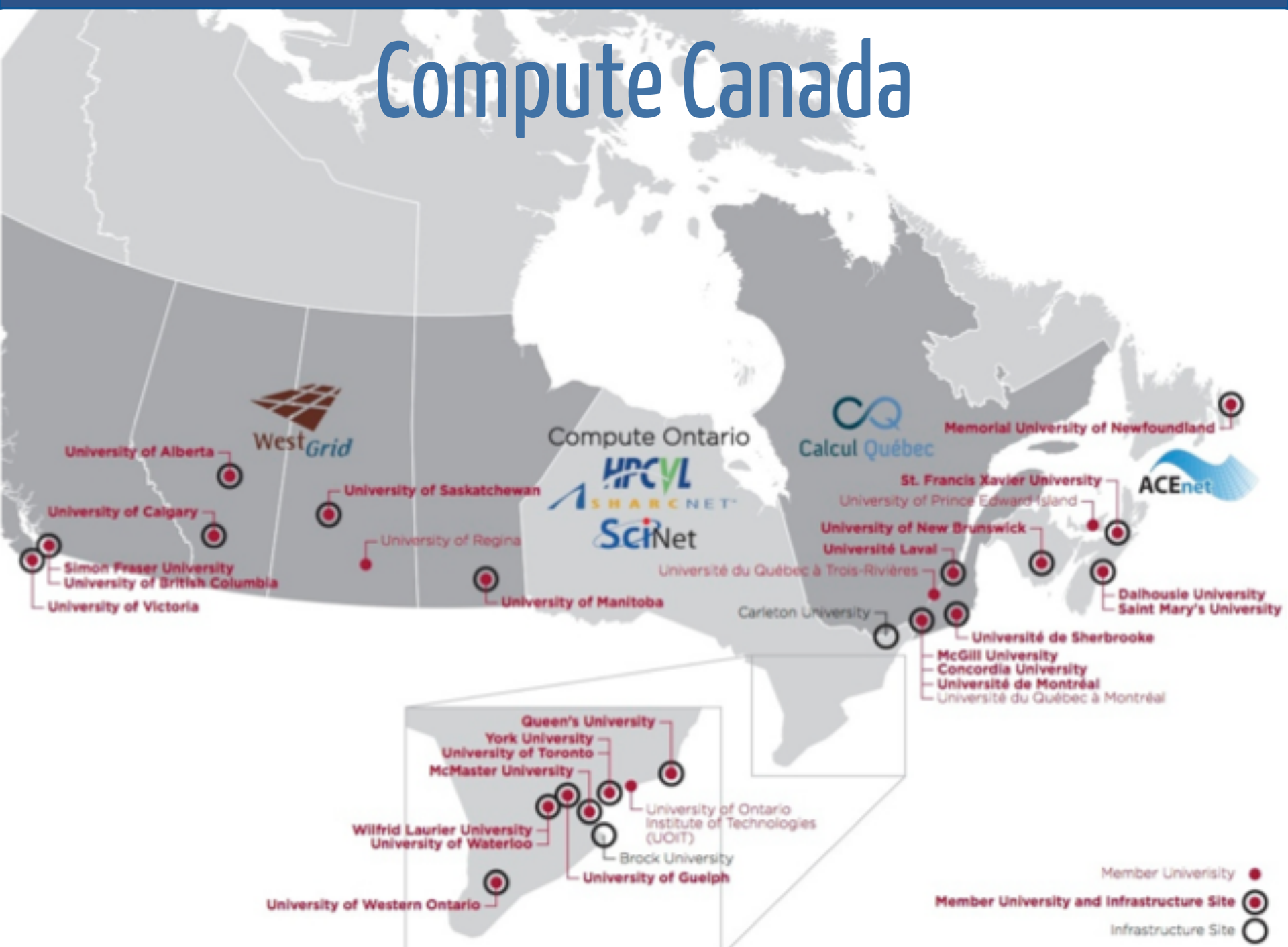
6 regional consortia of Universities

Free access to advanced computing resources for academic research

Geographically dispersed infrastructure

21 sites across 10 provinces and more than 3000 miles

30+ computing systems

Compute Canada

Data

~10 sites with 0.5 to 5 PB of storage

9000+ users

Multiple off-campus sites

i.e. research hospitals, sequencing centers, etc.

Interconnects

10Gbps between most sites through Canarie network

- Research network similar to Internet2 in US
- Transfers between sites don't go through the commercial Internet

Ensure some level of control on all transfer paths

A history of local solutions

Local sites deployed a variety of solutions

SCP, BBCP, GridFTP, HPN-ssh

Often all systems in a region used a prefered solution

Documentation challenge

For users

For science support staff

New challenges

Compute Canada has made it easier for users to use multiple sites

Disjointed solutions make it harder for:

- Users moving to a new system
- Staff having to support users from a remote site
- Negotiating security policies across multiple institutions
- Providing best-practices for optimal transfer speeds

Streamlining data transfers

Better network monitoring & measurements

Use "best current practices" that have been proven effective elsewhere: http://fasterdata.es.net/

Standardize a data transfer tool & document it

Network measurements

perfSONAR

ongoing Compute Canada wide deployment at end sites (18 out of 21) using mesh configuration

15 more perfSONAR nodes at the Canarie/ORAN level

Using Maddash

Identify problem links

Understand the baseline

Globus

Used at a local level since early the 2010s

Today used in 12 sites across 6 provinces

Still very localized config/setup:

- 1 globus namespace per Consortia (ie: westgrid#dtn1, calculquebec#dtn1, etc)
- Half OAuth, half gridFTP certificates
- Some dedicated DTNs, some shared nodes

Globus - short-term plans

Deploy everywhere

We have set up a small team with representation from each region to push Globus in major storage site

Unified namespace for Globus endpoints

computecanada#westgrid-dtn1

Train users and staff

Globus - long-term plans

Unified authentication scheme

Need to agree on how to authenticate users

Mostly an administrative/political issue

Globus-Plus and sharing outside Compute Canada

Globus - issues and questions

Should we create a Globus account for all Compute Canada users ?

Privacy and legal issues

Can the Globus authentication be delegated to Compute Canada ?

Our experience so far

2 motivations for users

- Transfer lots of data
- Ease of use

Surprisingly few questions from users about security

Transfer speeds seems satisfying and often limited by filesystem use

Our experience so far (cont)

Many users are using Globus only to easily transfer data from their laptop to the parallel FS

Using Globus Connect Personal

Thanks

Only a couple of months into this project

A lot more to do

We still haven't considered what to do with the usage data

Expect it will give us insight into data movement between all sites