

Distributed clouds for particle physics and astronomy research computing

Randall Sobie and Ian Gable
University of Victoria

Motivation :

Sophisticated user communities in physical sciences

Non-GUI users Batch computing environments

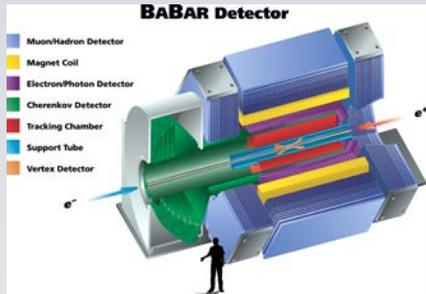
Complex software packages and demanding system requirements

Specific OS system Specific application libraries and compilers

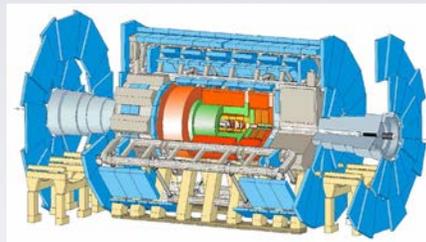
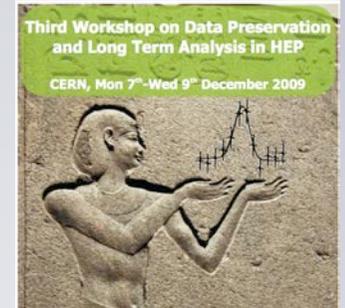
Medium-scale data sets (100s TBs)

Data accessed (on-demand) from remote or local repositories

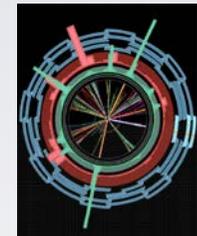
Applications



BaBar experiment (SLAC)
Stopped recording particle collisions in 2008
VMs and clouds seen as a way to preserve access and analysis capability

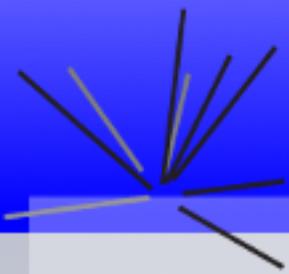


ATLAS Tier 3
User analysis of smaller data sets (1-10TB) recorded by the ATLAS

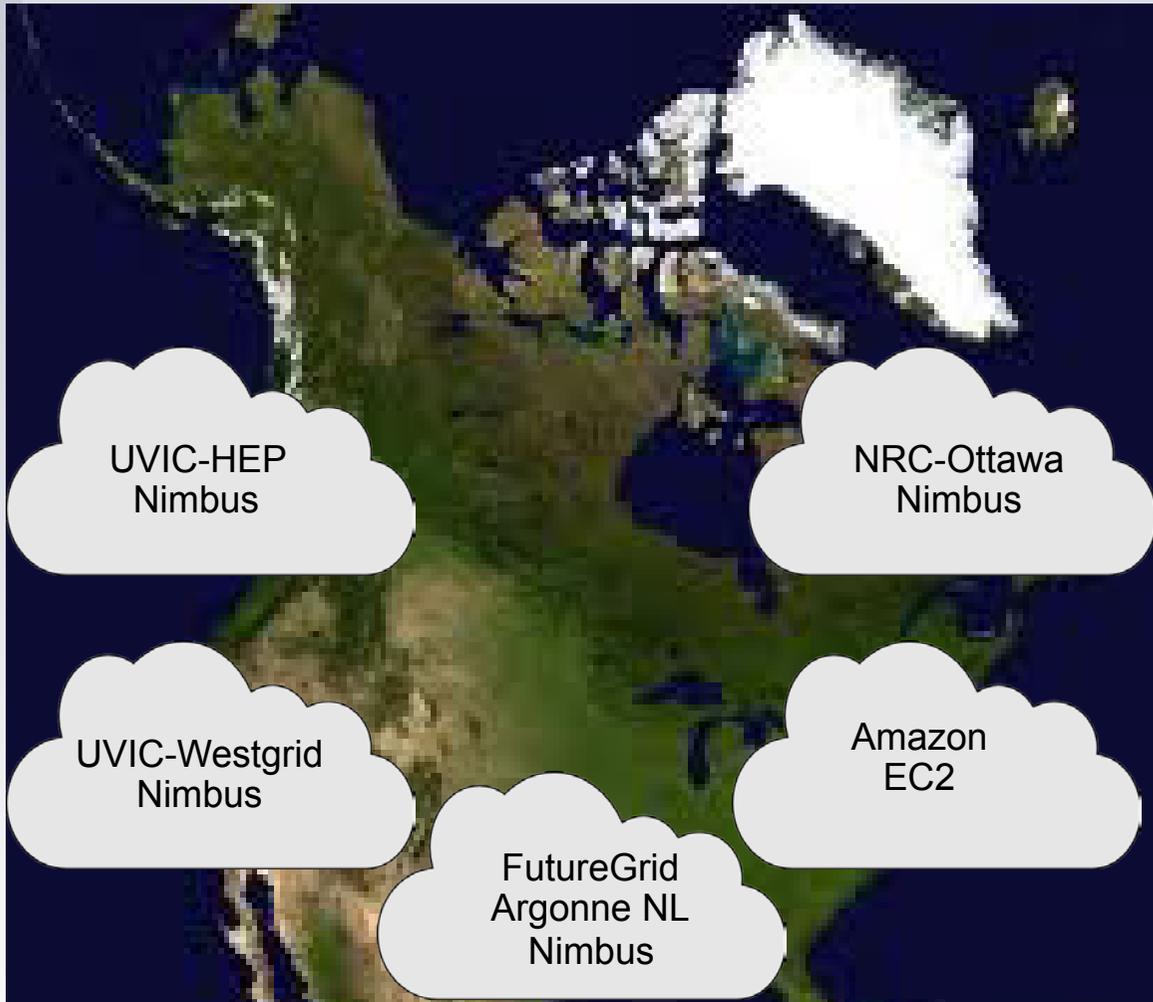


CANFAR Project
(NRC Herzberg Institute for Astrophysics Victoria)
Provides for the delivery, processing, storage, analysis, and distribution of astronomical survey data





Distributed compute cloud



Virtual machines and clouds

Provide users with base-VMs to customize and store in Image-Repository

Distributed compute cloud

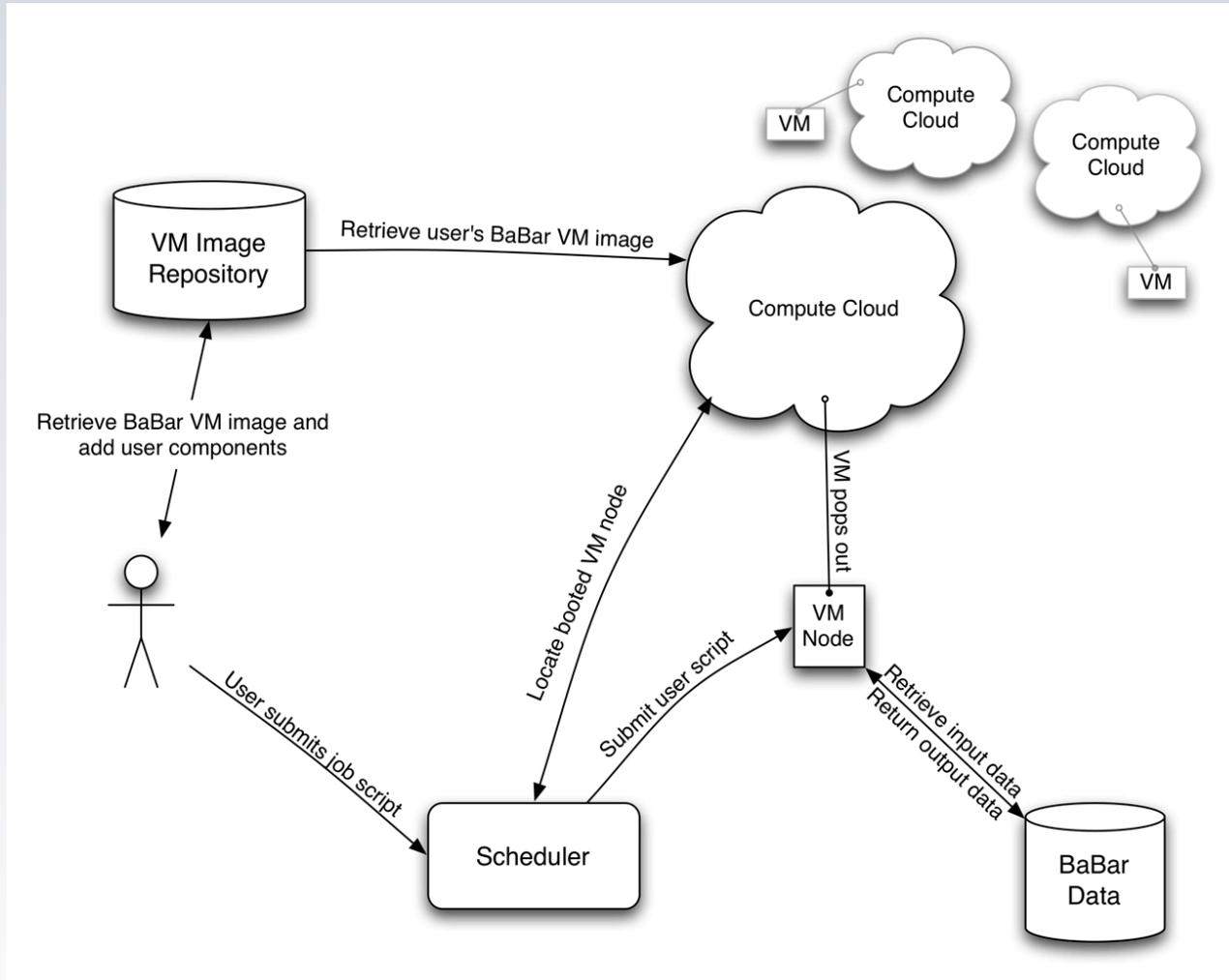
We have access to multiple computing centres configured as clouds

Our goal is to utilize these clouds as a single resource for batch jobs

Remove need for sysadmins being “application specialists”

X509 certificates used for authentication

Overview of system



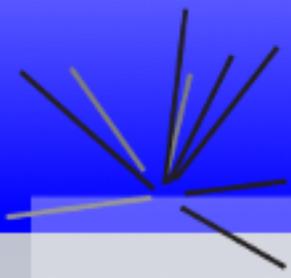
Familiar environment

Users build their code within an application-specific VM

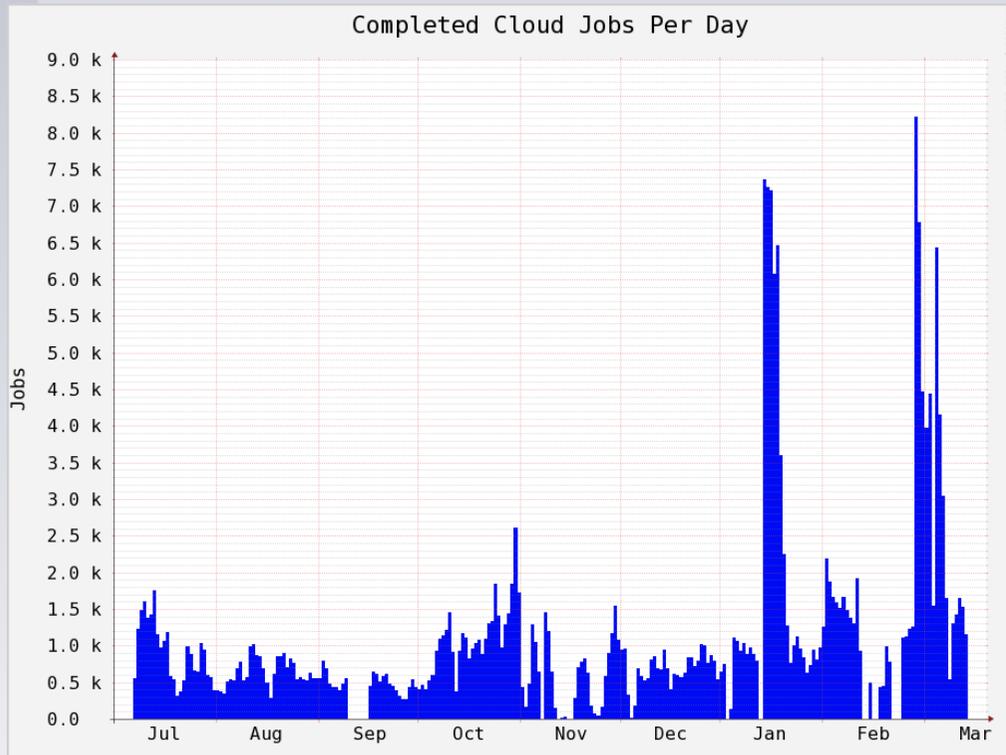
Submit to a batch system

System boots the user-customized VM on one of the clouds

Retrieves data from a local or remote source



Current status



Operational for 1 year

Tens of thousands of VMs booted and jobs processed

Using Nimbus, EC2 clouds

Particle Physics (BaBar, ATLAS T3)
Astronomy (CANFAR)

Main challenges

Security and authentication
Data I/O

Support provided by
CANARIE, NSERC, NRC, Amazon, Google, FutureGrid (NSF)