HTC Scientific Computing in a Distributed Cloud Environment

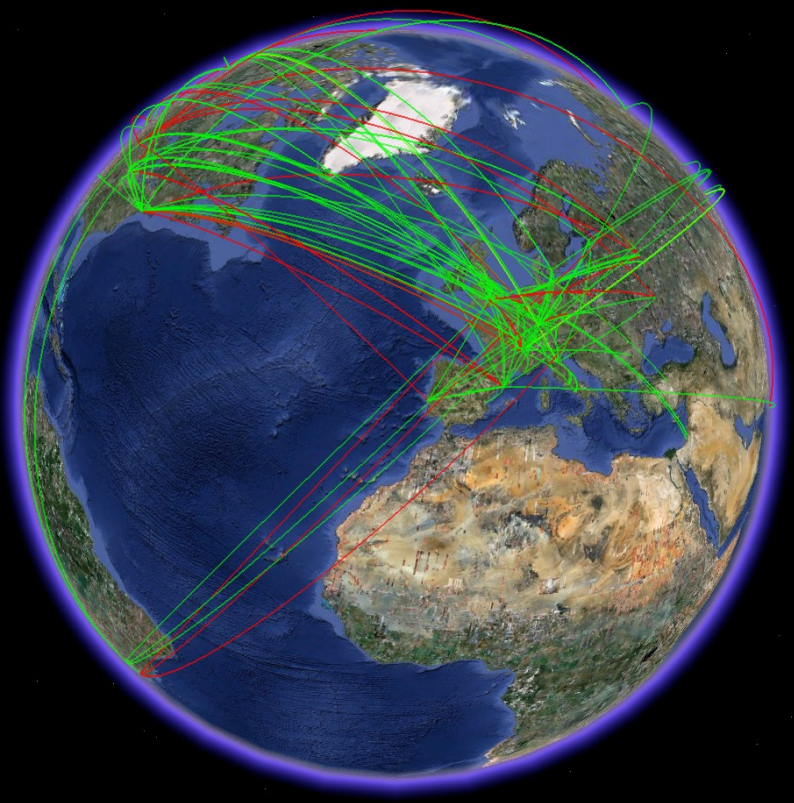
Randall Sobie

A.Agarwal, A.Charbonneau, I.Gable, C.Leavett-Brown, R. Impey

M.Paterson, W. Podiama, R.Taylor

Institute of Particle Physics

University of Victoria and National Research Council of Canada



Research is evolving into "Big Science"

Large international collaborations

Long term projects

Each nation bringing "in-kind" resources

How do we utilize the international computing resources for Big Science projects?

Large Hadron Collider
CERN Laboratory

Worldwide LHC Computing Grid (WLCG)

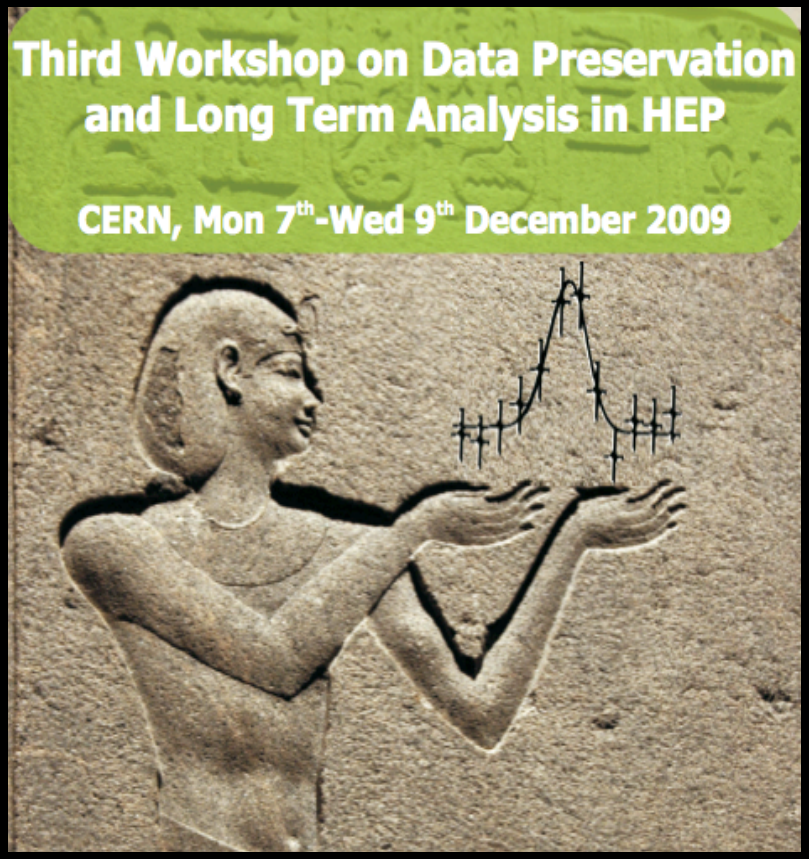
Tightly integrated set of resources
(100 separate sites)

Tiered design based on functionality

Successful operation

Meeting the growing demand will be a challenge

Can we make more use of emerging technologies such as cloud computing?



Virtualization has many benefits for HEP computing

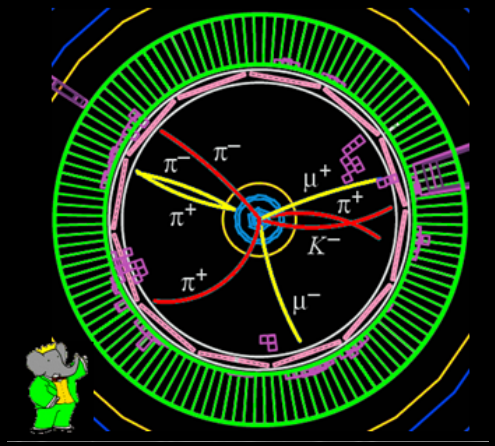
Minimize dependence on local system

Isolate the complex application software

No loss in processing efficiency

BaBar experiment at SLAC (Stanford) ended in 2008

Long Term Data Access (LTDA) system
In-house cloud for preserving the data and software



Many facilities are evolving into clouds

Can we use these sites in an federated way?
And integrate them into our existing system?

Distributed cloud computing system

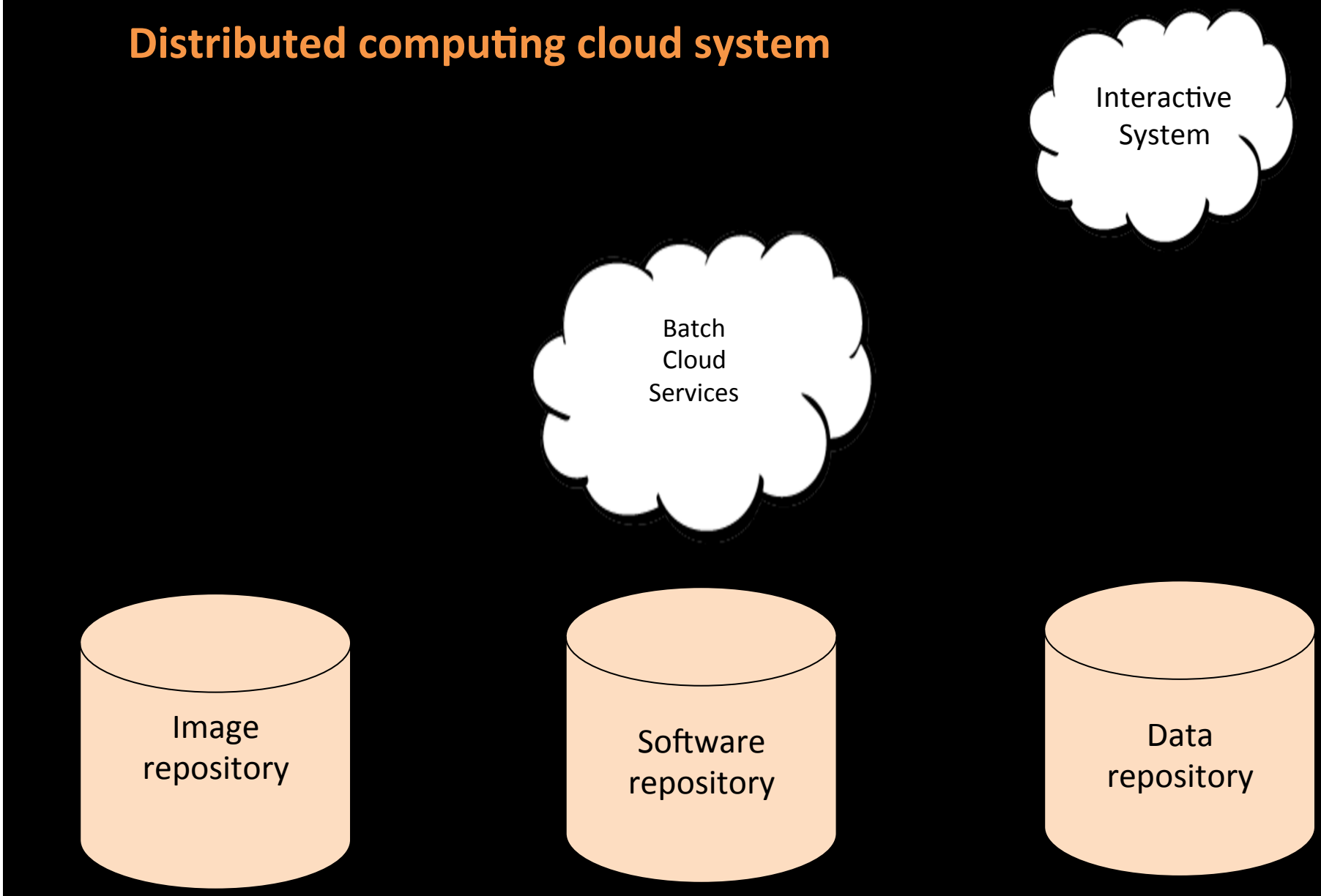
Production system for ATLAS experiment
Integrated into the WLCG
Operating over 3 continents

Grid of Clouds
Sky Computing K.Keahey et al

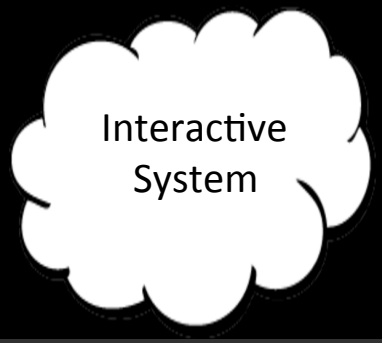
Outline

Design, operation and performance
Issues and challenges
Future plans

Distributed computing cloud system



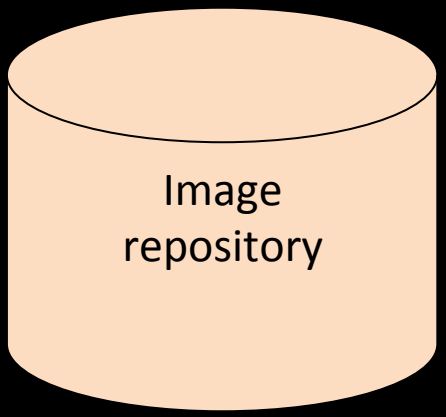
Distributed computing cloud system



Provides platform for developing and saving images

Separate from the batch IaaS clouds

Submission of batch jobs (ATLAS jobs submitted from CERN)



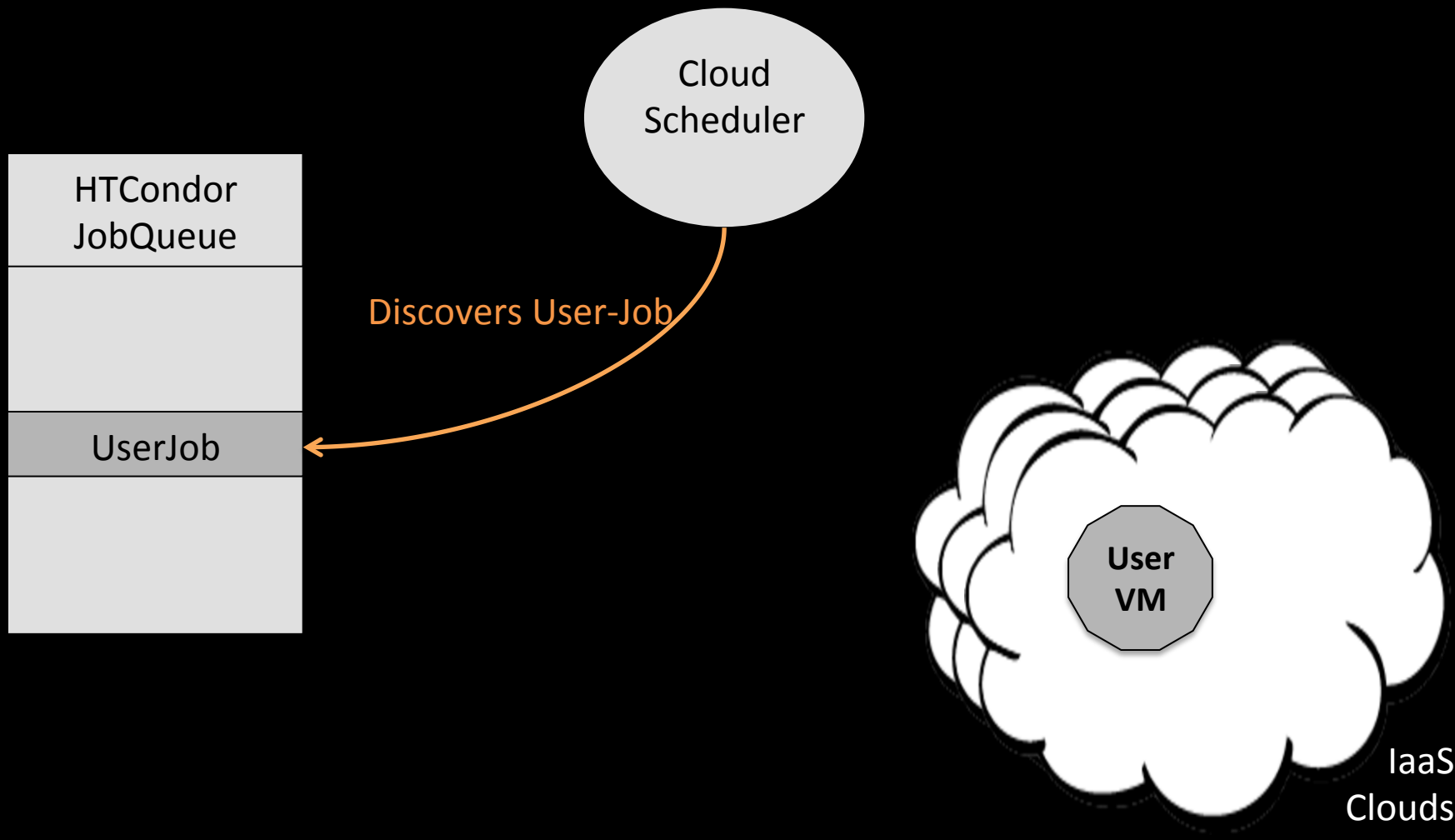
Repoman – custom VM image repository

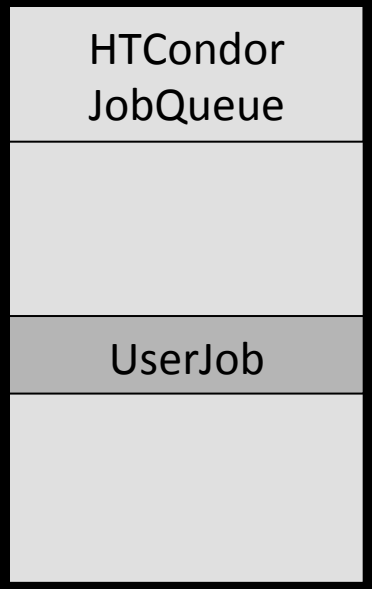
Users can store, retrieve images

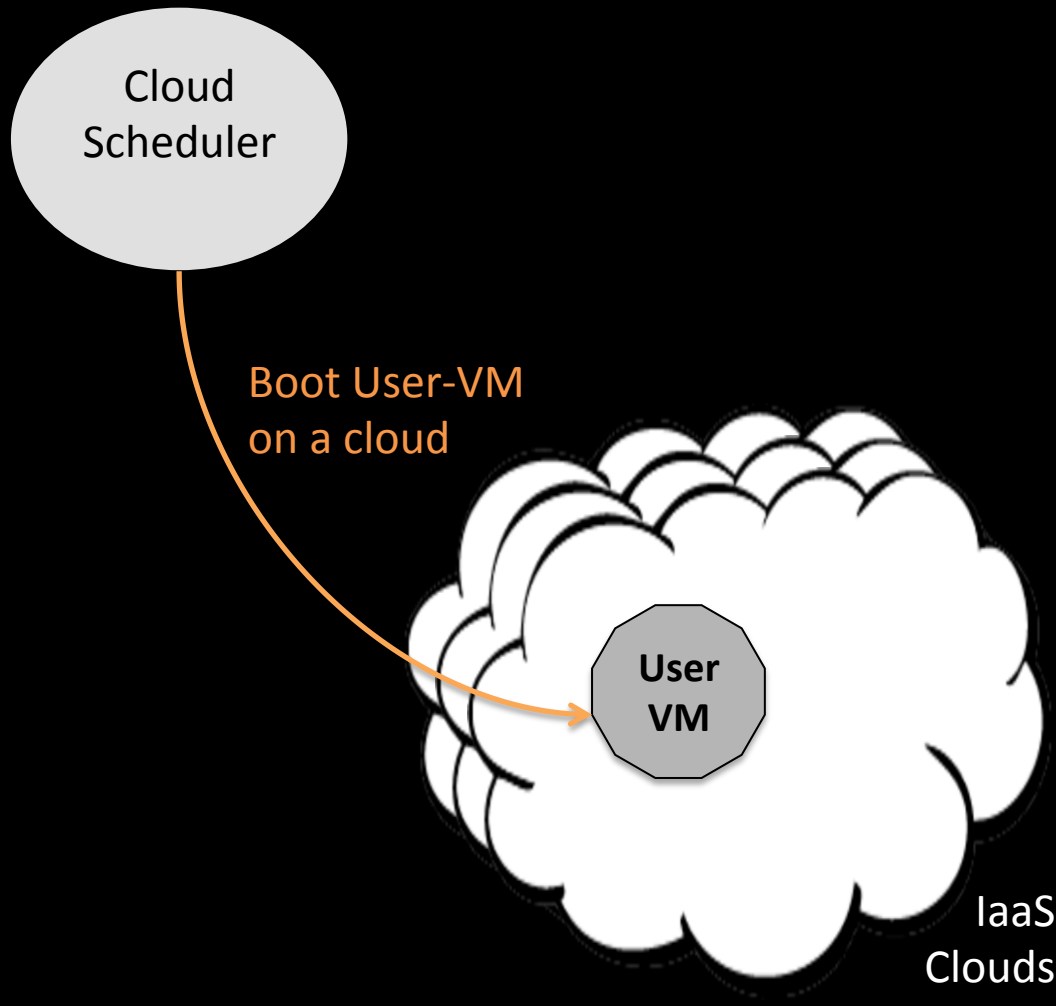
Accessible via HTTP(S)

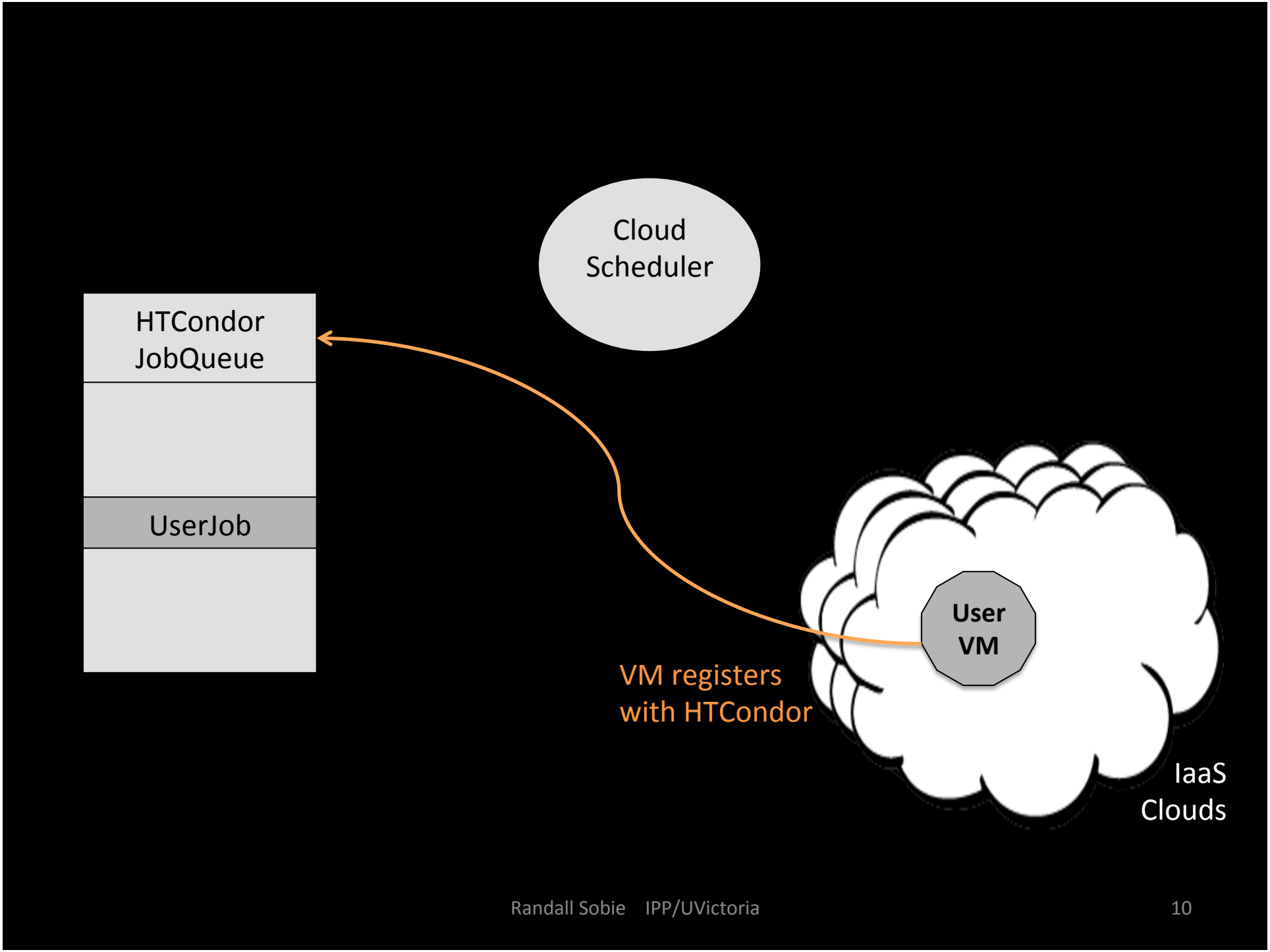
Dual hypervisor capability

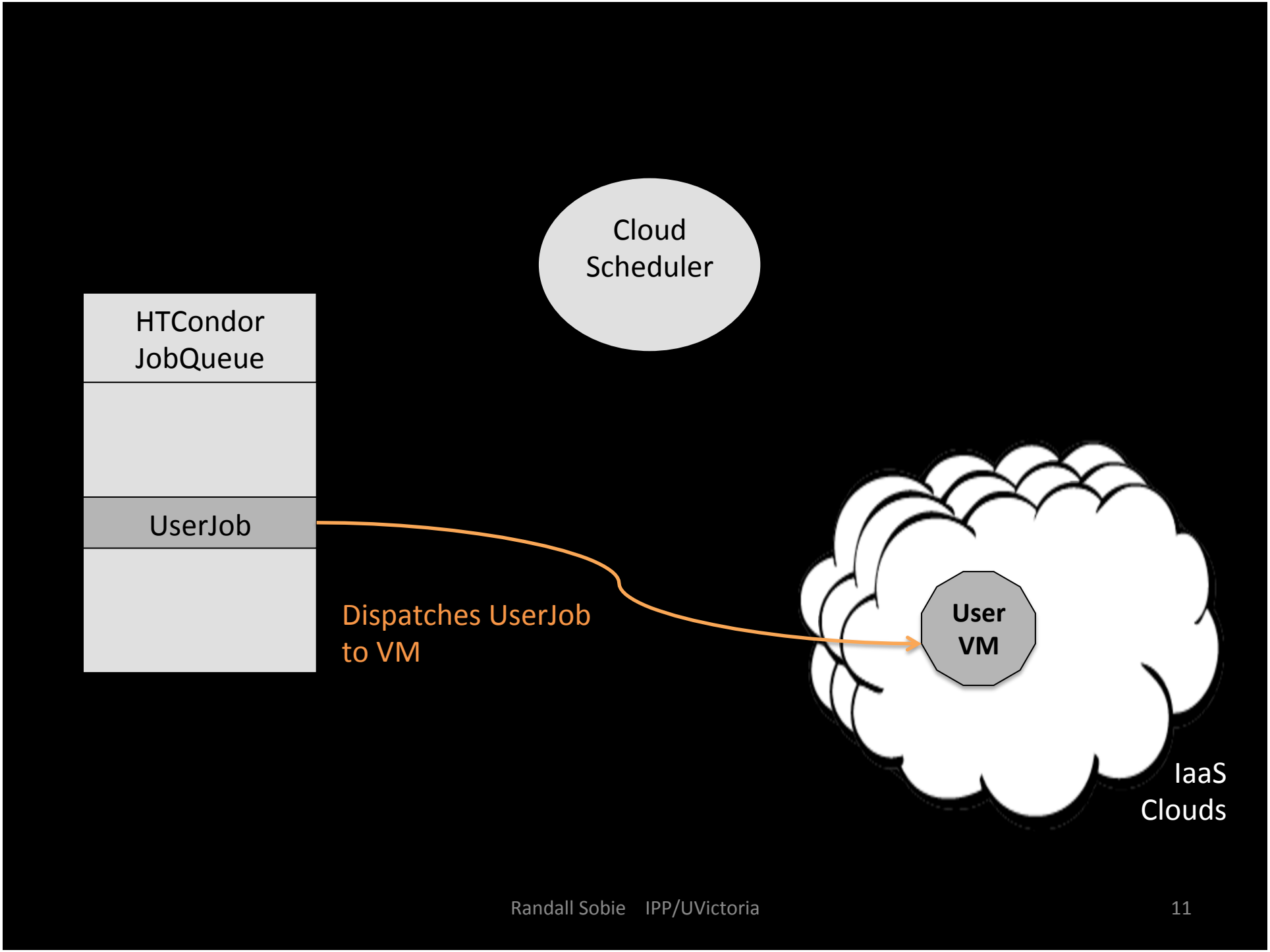
X.509 certificate authentication


Cloud
Scheduler
HTCondor
JobQueue
Discovers User-Job
UserJob
User
VM
IaaS
Clouds

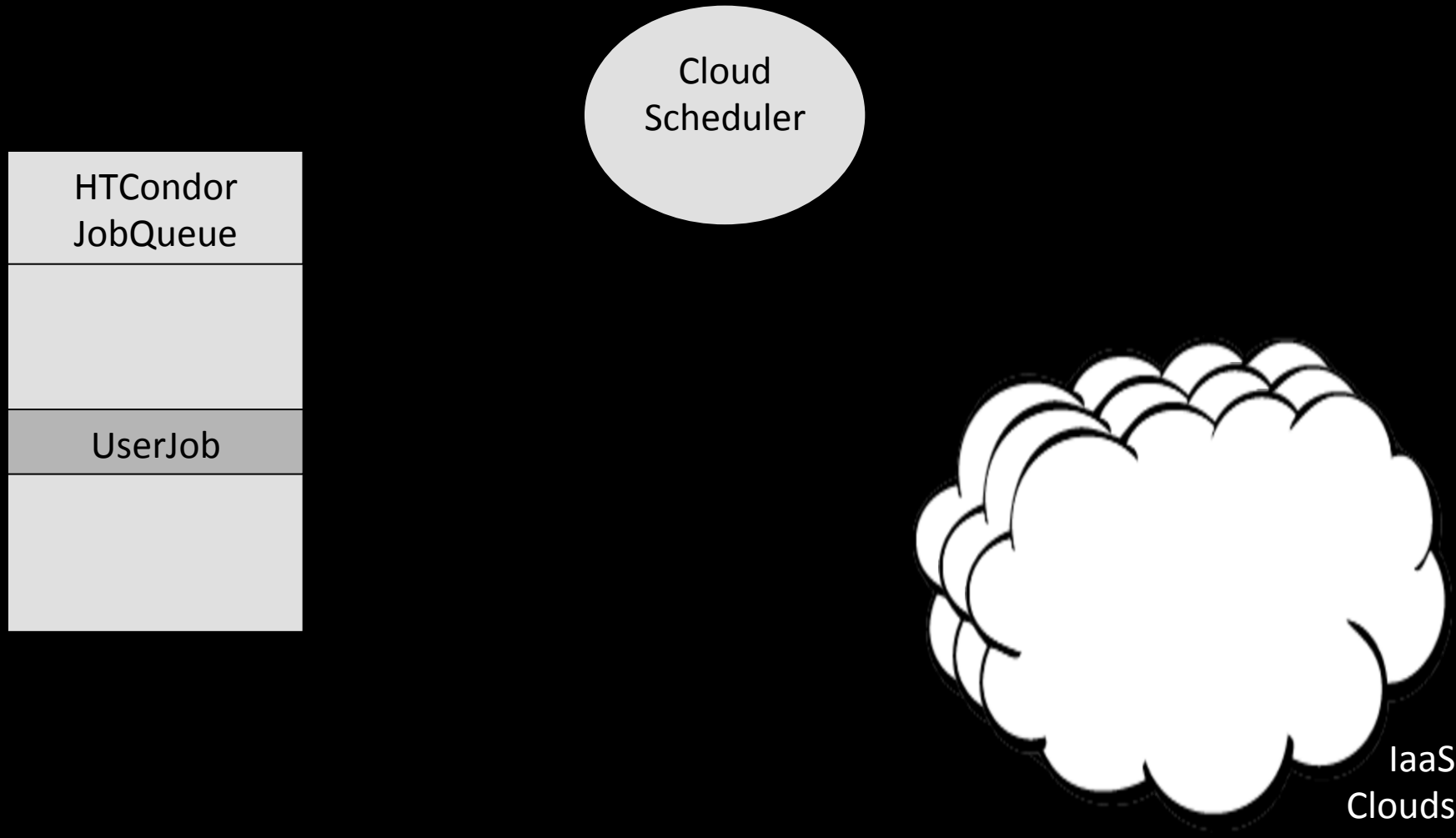

HTCondor
JobQueue
UserJob


Cloud
Scheduler
Boot User-VM
on a cloud
User
VM
IaaS
Clouds

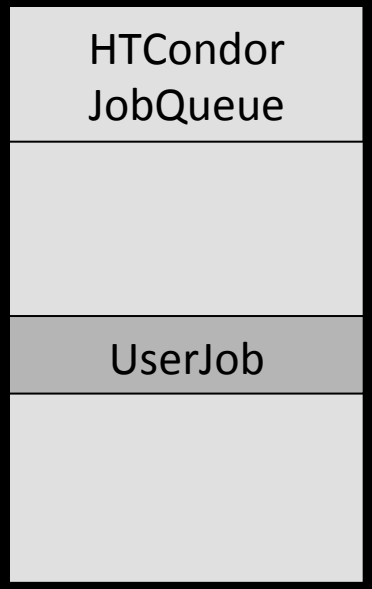

Cloud
Scheduler
HTCondor
JobQueue
UserJob
User
VM
VM registers
with HTCondor
IaaS
Clouds


Cloud
Scheduler
HTCondor
JobQueue
UserJob
Dispatches UserJob
to VM
User
VM
IaaS
Clouds

Overview of the components



Overview of the components



HTCondor

Designed as a cycle scavenger

Ideal as a job scheduler for a dynamic environment

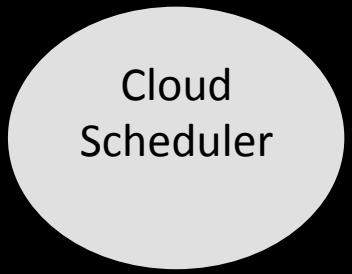
Job description specifies the VM image and requirements for instantiating the VM image on the clouds

Nimbus clouds require URL to VM image, path to users proxy and hostname of the MyProxy server

OpenStack clouds require AMI and instance type

IaaS
Clouds

Overview of the components



Cloud Scheduler

Custom component

Reviews the requirements of jobs in the HTCondor queue

Requests the boot of user-specific images

Monitors the VMs
VMs remain active while user jobs exist
VMs are shutdown if there are no jobs or if VM is in an error state

Overview of the components

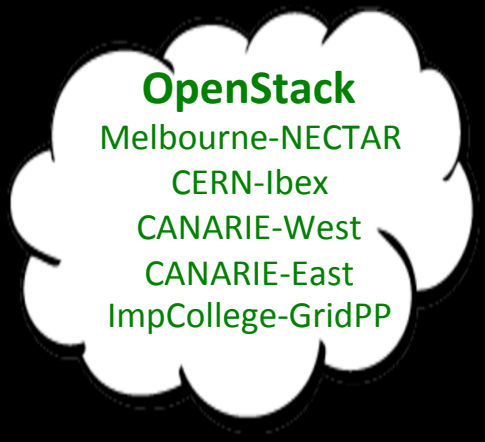
IaaS Clouds

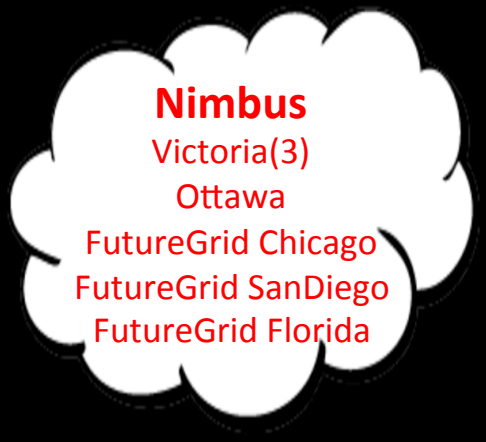
North America, Australia and Europe

Nimbus clouds

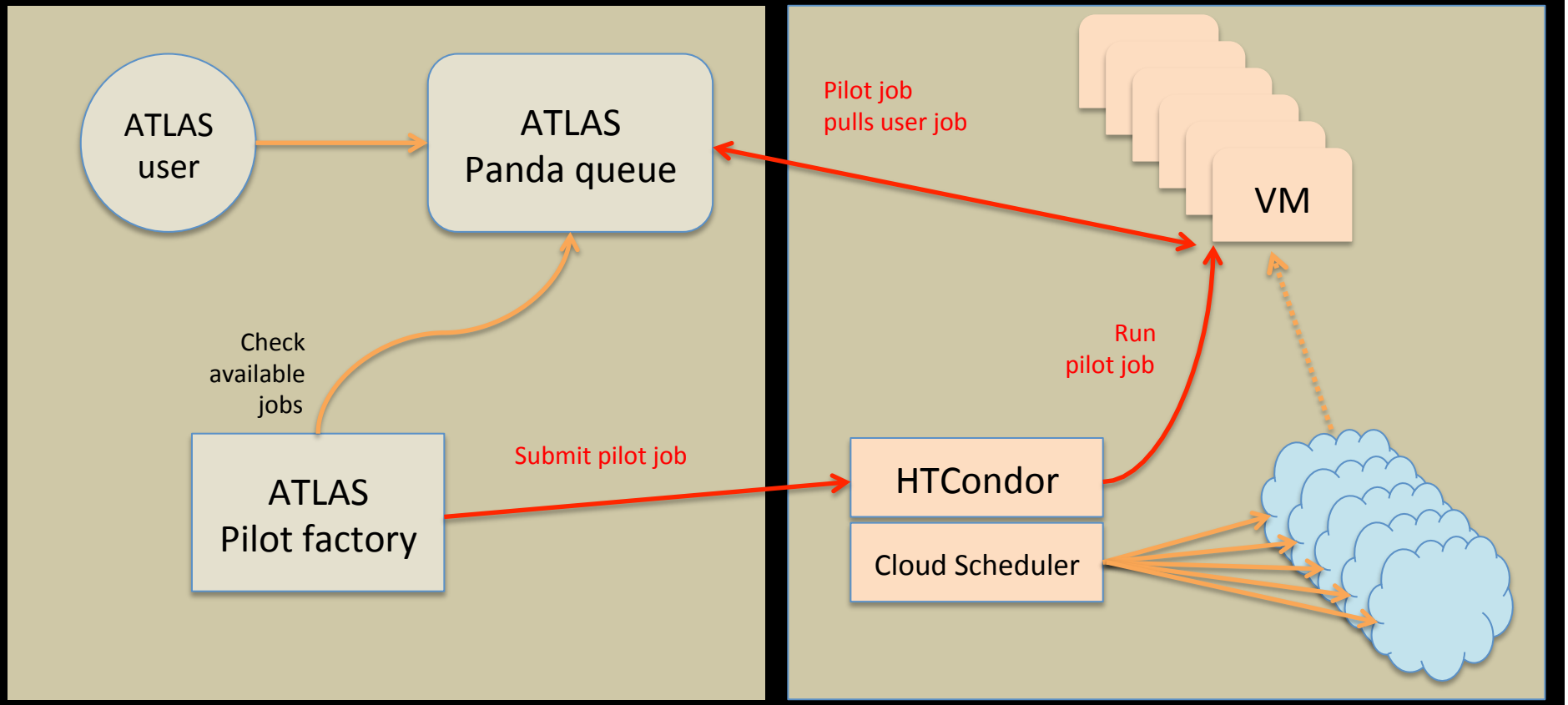
OpenStack clouds

Commercial clouds





ATLAS job submission

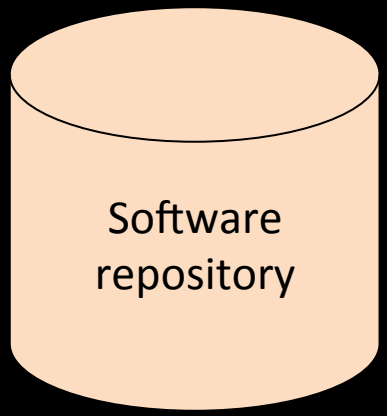


CVMFS

A caching, http based read-only filesystem optimized for delivering experiment software to (virtual) machines.
Originally developed as part of the CernVM project

Ideal for VMs (particularly multi-core VMs)

SQUIDs provide proxies to CVMFS repositories
(Exploring dynamic SQUID creation)

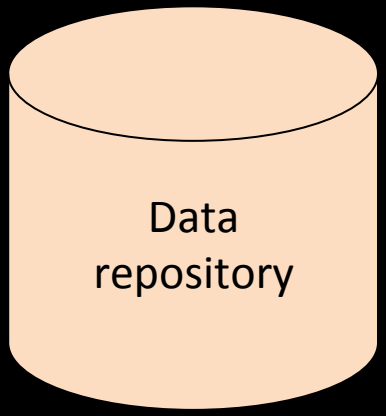


Data repository

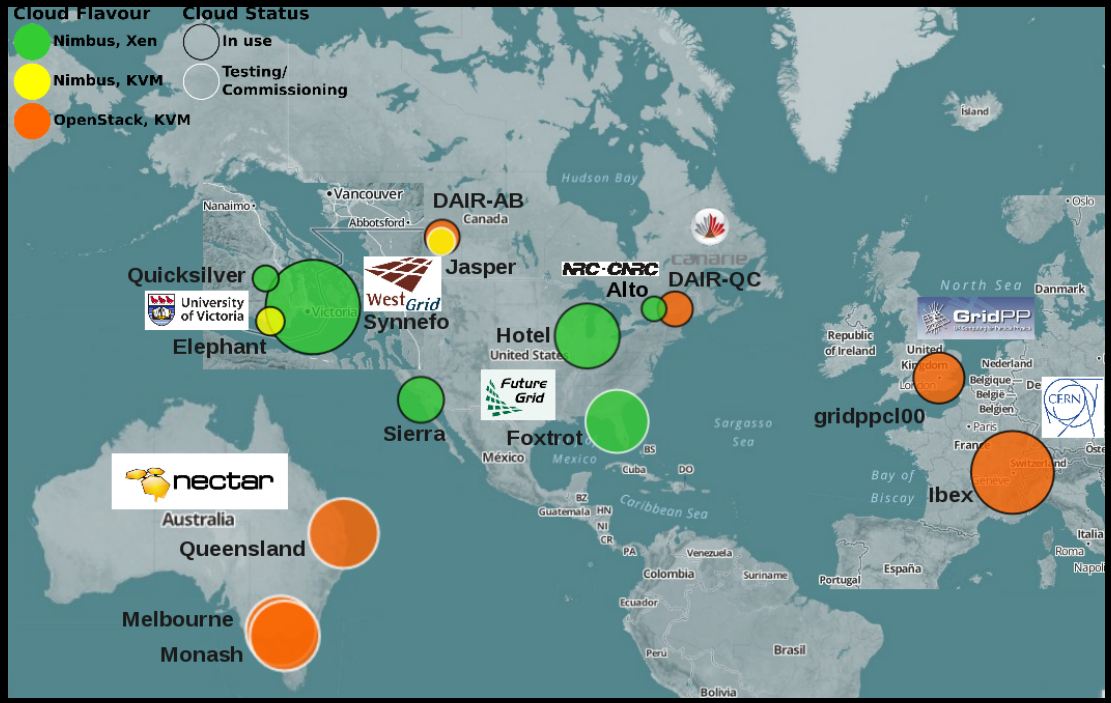
We are limited to applications requiring little input data (and output produced at modest rates)

Webdav has been used for modest I/O workloads

Management of data in the distributed cloud is under study



Clouds

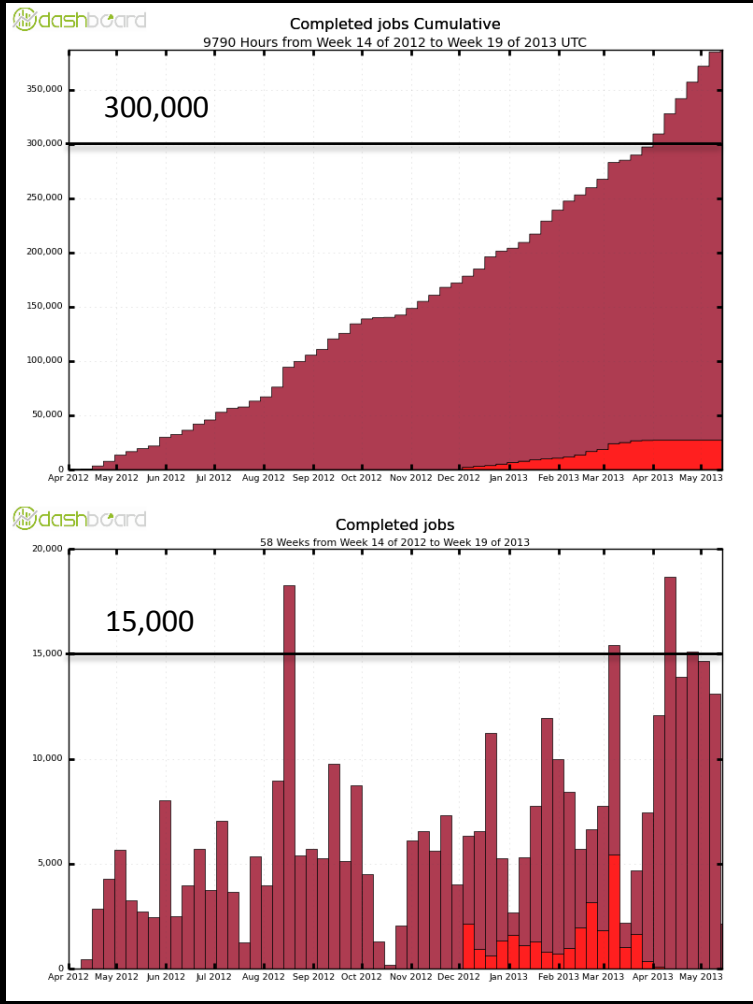


Nimbus

Victoria(3)
Ottawa
FutureGrid Chicago
FutureGrid SanDiego
FutureGrid Florida

OpenStack

NECTAR-Melbourne
NECTAR-Queensland
NECTAR-Monash
CERN-Ibex
CANARIE-West
CANARIE-East
Imperial College-GridPP



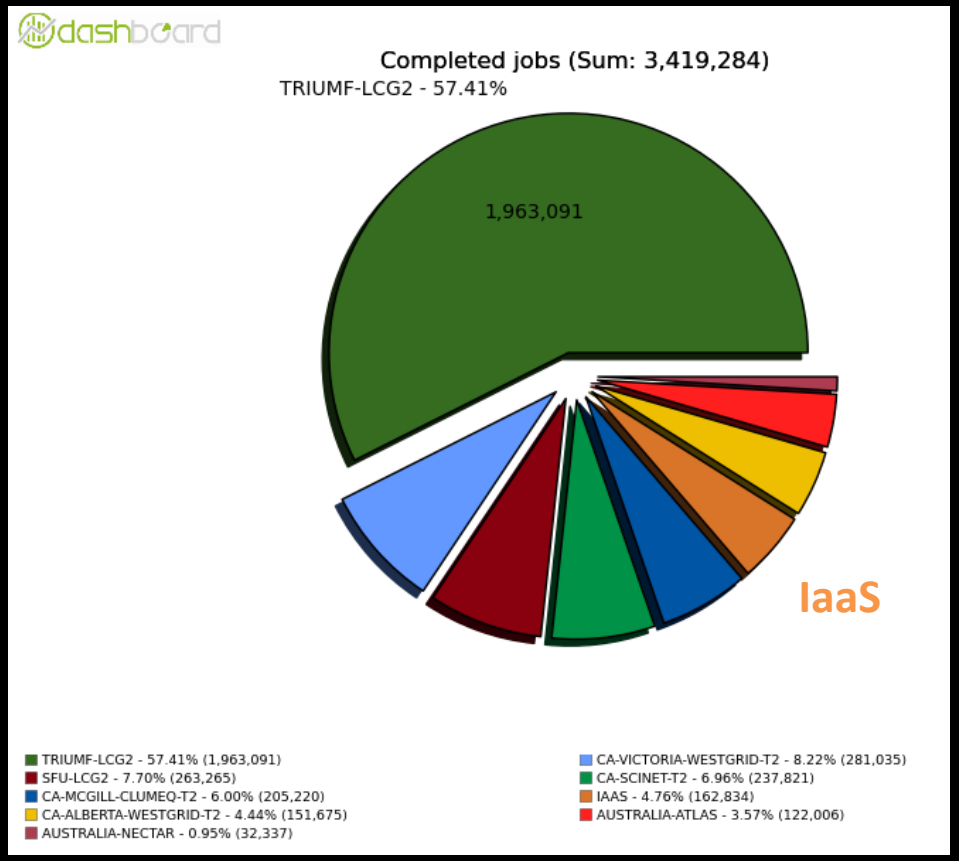
Fully integrated as an ATLAS Grid site
(grid operations, monitoring, …)
April 2012

Integrated number of jobs (380,000)

Weekly jobs (12,500 in 2013)

Peak over 1000 simultaneous jobs
April 2013

Comparison with other ATLAS Canada sites



Results in 2013 (Jan – April)

4% of all CA ATLAS jobs

163,000 **jobs**

12 hour jobs

Virtual Machines in 2013

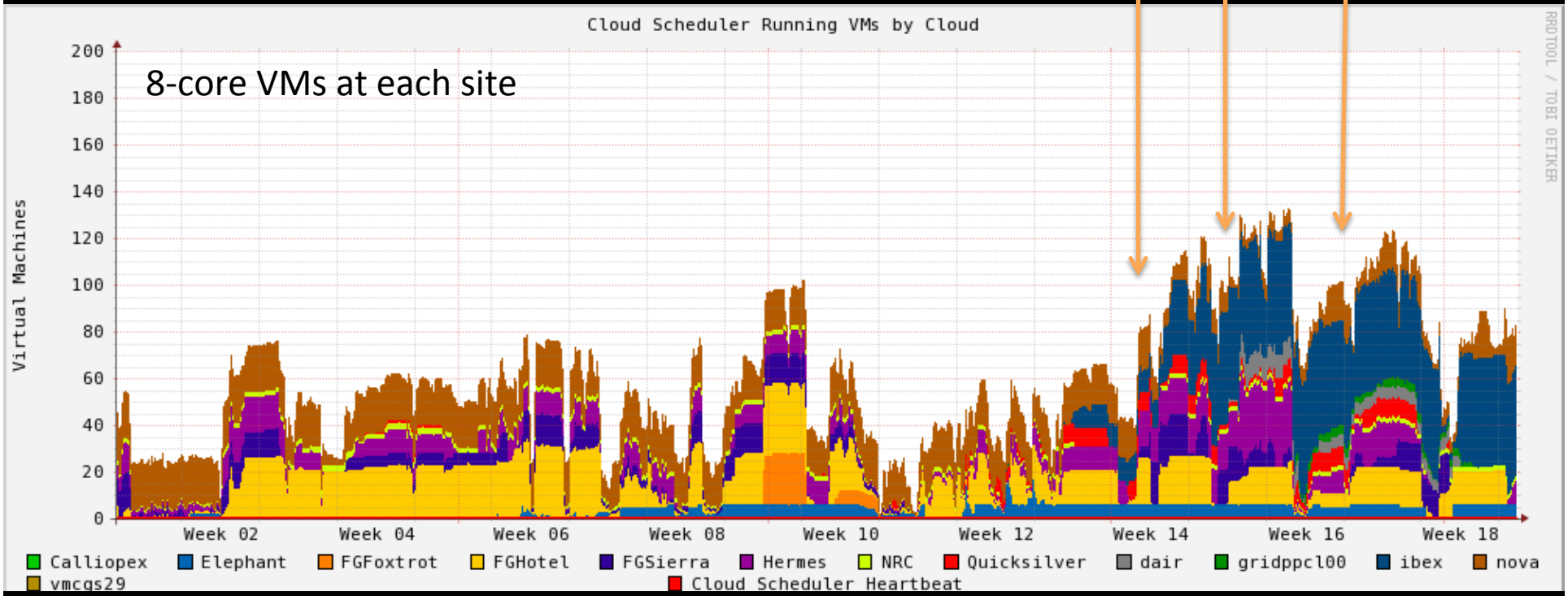
OpenStack Clouds

Nimbus Clouds (and NECTAR)

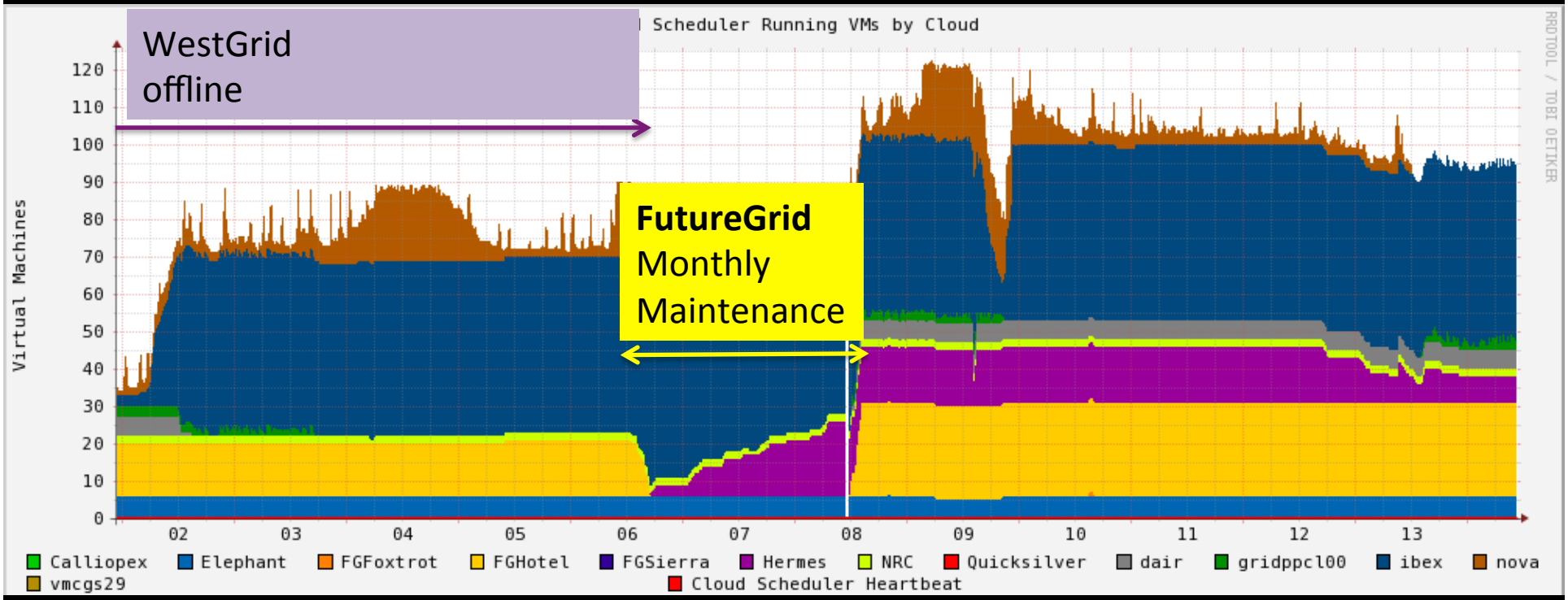
CERN IBex

CANARIE

Imperial College



Cloud evaporation and condensation



Number of 8-core VMs in May 2013

Issues and Challenges

Cloud issues

We are often the first user of the cloud

Clouds are not configured in the same way

(meta-data is often inconsistent)

Clouds use the "default cloud name"

Storage insufficient

(back-end storage not assured)

Grid issues

ID management and authentication

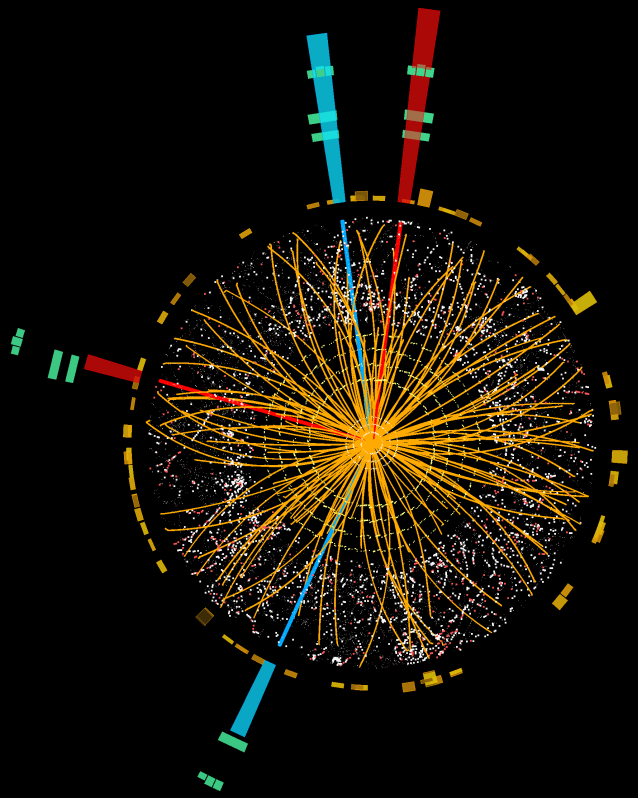
Remote access to repositories

Ability to upload a remote VM image

HEP issues

HEP applications stress the (non-HEP) clouds

Successes



Production system

ATLAS at LHC

CANFAR (astronomy)

Enabling it for Belle2 HEP project (Japan)

Grid of Clouds

Federated system

10+ clouds over 3 continents
(HEP and non-HEP sites)

1000 simultaneous jobs

No HEP-specific services

Dynamic system
(manages cloud variability)

Limited only by the resource availability
(looking for more research/commercial clouds)

Future

Automate integration and testing of new clouds

Standardized cloud configuration

Common authentication and authorization

Improved monitoring

Distributed data management
Dynamic squids
Improved tools to monitor cloud storage

Explore commercial clouds
(Cost effectiveness will likely improve)

Summary

We see ourselves as integrators
rather than developers of cloud technology

Distributed cloud computing
is a viable solution for scientific computing
(Grid of Clouds or Sky Computing)

Optimistic for the future

No limits on the scalability of the system for low I/O applications

Distributed data management is a challenge
but is being addressed by many groups

Acknowledgements

Contact information: rsobie@uvic.ca

Web sites: http://rjs.phys.uvic.ca/
http://heprc.phys.uvic.ca/home/