Utilizing cloud computing resources for Belle-II

Randall Sobie
Institute of Particle Physics of Canada
University of Victoria

On behalf of the Belle-II Collaboration
Belle-II Experiment

High-intensity frontier
Accelerator commissioning in 2016
Data taking starts in 2017

Belle-II computing model
WLCG structure
Shared infrastructure
See talk by Hara-san
Clouds in Belle-II

10-15% of current MC production is cloud-based

- Clouds at Belle-II member sites
- Opportunistic (private and commercial) clouds

Randall Sobie  University of Victoria
Belle-II methods for using clouds

See contribution by R. Grzymkowski
DIRAC Master
KEK

Watch for jobs assigned to the site

SD submits pilot jobs

JS submits pilot job to VM

Production Job retrieved by pilot

IaaS
Private/Commercial

Request VM

Job Scheduler

VM Provisioning
CloudScheduler
Manages remote IaaS clouds with HTCondor
See Gable talk

ATLAS and Belle-II
Canadian astronomy communities

uCernVM images with Shoal/Squid/CVMFS for OS and application software

Glint extension to OpenStack Glance image repository for managing VM images on multiple clouds (see poster)

Operational for 3 years
Many millions of HEP jobs
Panda/DIRAC workload managers
Amazon EC2 used in production

Transparently integrated into ATLAS and Belle-II computing systems
Australia- Belle II Grid site

Dynamic Torque
NECTAR

Multiusers
ATLAS and Belle-II

SL6 VMs configured with Puppet
CVMFS for Belle-II software

See proceedings of CHEP 2014
Belle-II MC production in the cloud

March 2015 MC production

(DIRAC – mainly cloud sites

Cloud – VMDIRAC

Over 300,000 jobs on the cloud

<table>
<thead>
<tr>
<th>LCG</th>
<th>1610926.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIRAC</td>
<td>313647.0</td>
</tr>
<tr>
<td>CLOUD</td>
<td>33003.5</td>
</tr>
<tr>
<td>SSH</td>
<td>12050.4</td>
</tr>
<tr>
<td>ANY</td>
<td>522.2</td>
</tr>
</tbody>
</table>
Commercial cloud use in Belle-II

Running 180 4-core VMs
Oregon EC2 site

Only utilizing 3 of 4-cores

Using CernVM3 (spot instance)

Booted a local squid for only EC2 VMs

Data transferred to North American SE

Any drop in EC2 usage is a result in our exceeding our max-spot-price limit
Summary

• Cloud resources account for significant fraction of the Belle-II computing
  – Plans for further expansion of cloud use
  – Additional work to improve use and data access is underway
  – Testing of multicore jobs will start after CHEP

• The clouds are utilized in a number of ways
  – Primarily to simplify the integration into existing infrastructure
  – Method is selected to meet requirements of non-HEP users

• Clouds are becoming reliable, stable production platforms
  – Adding dynamic or context-aware capabilities to utilize nearest squid caches

• Clouds will continue to improve
  – CernVM, CVMFS, data federations, 100G networks (SDN)