Managing remote cloud resources for multiple HEP VO’s with cloudscheduler

Marcus Ebert on behalf of the HEP-RC group at the University of Victoria, Canada
What is cloudscheduler (csv2)

- software that is able to start Virtual Machines (VMs) on clouds
  - clouds can be local or far away
  - concept of groups
    - multiple clouds can be used in each group
    - multiple groups are possible
  - has web interface and CLI

- startup of VMs depends on jobs in an HTCondor queue
  - dynamic process, on demand
  - VMs are started depending on the resources needed by jobs for a specific group
  - VMs are automatically deleted when no more jobs available that can use those resources
Multi-cloud batch system with cloudscheduler

cloudscheduler
\textit{(starts/terminates VMs as needed)}

University cluster

Other Universities

Public Clouds

\textit{via condor\_poller}

manages also default image and ssh key distribution between clouds

HTCondor

Experiments

CHEP 2023, May 11th

Marcus Ebert (mebert@uvic.ca)
Multi-cloud batch system with cloudscheduler

**cloudscheduler**
*(starts/terminates VMs as needed)*

Cloud/Group 1

Cloud/Group 2

local system of static (bare-metal) worker nodes

Experiments

via condor_poller
How to customize VMs

- csv2 uses cloud init together with CernVM
  - different set of files for different VOs or even specific to single clouds
Managing multiple VOs

● one csv2 group per VO
  ○ multiple clouds per VO possible

● different HTCondor systems if configuration is too different between groups
  ○ and for best practice

● we currently run single csv2 instance for
  ○ Belle-II
  ○ ATLAS
  ○ DUNE
  ○ BaBar
## Managing multiple VOs

<table>
<thead>
<tr>
<th>Group</th>
<th>Target Alias</th>
<th>Jobs</th>
<th>Idle</th>
<th>Running</th>
<th>Completed</th>
<th>Held</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>atlas-cern</td>
<td>None</td>
<td>179</td>
<td>168</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>atlas-cern</td>
<td>orcm-extension</td>
<td>213</td>
<td>204</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>atlas-cern</td>
<td>Hephy-silk</td>
<td>31</td>
<td>30</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>atlas-cern</td>
<td>Iro-lms_cloud</td>
<td>90</td>
<td>889</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>atlas-cern</td>
<td>Uki-scotgrid-ecdf_cloud</td>
<td>2166</td>
<td>1190</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>atlas-cern</td>
<td>Co-ias3-t3</td>
<td>96</td>
<td>126</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Australia-Beller</td>
<td>None</td>
<td>1326</td>
<td>437</td>
<td>889</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>babar</td>
<td>None</td>
<td>1343</td>
<td>477</td>
<td>1366</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>belle</td>
<td>None</td>
<td>2166</td>
<td>976</td>
<td>1190</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>belle</td>
<td>Belle-local-worker</td>
<td>None</td>
<td>None</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>belle</td>
<td>Belle-validation</td>
<td>None</td>
<td>None</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>belle</td>
<td>Uvic-worker</td>
<td>None</td>
<td>None</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dasy-belle</td>
<td>None</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dune</td>
<td>None</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>testing</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### Condor QDNN

- **condor-group-1**: 100
- **condor-group-2**: 200
- **condor-group-3**: 300

### Condor Status

- **Total Jobs**: 1000
- **Total Running Jobs**: 500
- **Total Completed Jobs**: 200

### Condor Cert

- **Worker Cert**: 1

### HTCondor Plugin

- **Plugin Name**: MyCondorPlugin
- **Plugin Version**: 1.0.0
## Managing multiple VOs

<table>
<thead>
<tr>
<th>Site</th>
<th>Status</th>
<th>Cloud Config</th>
<th>Alasos</th>
<th>Group Config</th>
<th>Hignas</th>
<th>Keys</th>
<th>Users</th>
<th>Groups</th>
<th>System Config</th>
<th>User Settings</th>
<th>HTCondor Plugin</th>
<th>Log out</th>
</tr>
</thead>
<tbody>
<tr>
<td>allen-crie</td>
<td>online</td>
<td>72B</td>
<td>113</td>
<td>0</td>
<td>110</td>
<td>3</td>
<td>0</td>
<td>114</td>
<td>891</td>
<td>21</td>
<td>954</td>
<td>3000</td>
</tr>
<tr>
<td>alsatron</td>
<td>online</td>
<td>617</td>
<td>8</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>64</td>
<td>0</td>
<td>20</td>
<td>72</td>
</tr>
<tr>
<td>ca-crest</td>
<td>online</td>
<td>729</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>44</td>
<td>24</td>
</tr>
<tr>
<td>charlesdon</td>
<td>online</td>
<td>130</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>otter</td>
<td>offline</td>
<td>560</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td>124</td>
<td>3</td>
<td>0</td>
<td>118</td>
<td>3</td>
<td>0</td>
<td>122</td>
<td>955</td>
<td>992</td>
<td>2316</td>
</tr>
<tr>
<td>australian</td>
<td>online</td>
<td>757</td>
<td>450</td>
<td>0</td>
<td>0</td>
<td>640</td>
<td>10</td>
<td>0</td>
<td>880</td>
<td>889</td>
<td>990</td>
<td>990</td>
</tr>
<tr>
<td>melbourne</td>
<td>online</td>
<td>555</td>
<td>450</td>
<td>0</td>
<td>0</td>
<td>640</td>
<td>10</td>
<td>0</td>
<td>880</td>
<td>889</td>
<td>990</td>
<td>990</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td>900</td>
<td>0</td>
<td>0</td>
<td>1280</td>
<td>20</td>
<td>0</td>
<td>164</td>
<td>959</td>
<td>1992</td>
<td>990</td>
</tr>
<tr>
<td>babar</td>
<td>online</td>
<td>729</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>hep-ex-cred</td>
<td>online</td>
<td>729</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>belle</td>
<td>online</td>
<td>33780</td>
<td>261</td>
<td>0</td>
<td>0</td>
<td>260</td>
<td>1</td>
<td>0</td>
<td>2081</td>
<td>2078</td>
<td>2088</td>
<td>2083</td>
</tr>
<tr>
<td>amazon-er</td>
<td>online</td>
<td>729</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>berlexia</td>
<td>online</td>
<td>480</td>
<td>90</td>
<td>0</td>
<td>0</td>
<td>80</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>co-crest</td>
<td>online</td>
<td>729</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>charlesdon-mine</td>
<td>offline</td>
<td>729</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>exlll-bxi</td>
<td>offline</td>
<td>555</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td>321</td>
<td>0</td>
<td>0</td>
<td>320</td>
<td>1</td>
<td>0</td>
<td>2598</td>
<td>2555</td>
<td>2568</td>
<td>2568</td>
</tr>
<tr>
<td>belle-valdation</td>
<td>online</td>
<td>729</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>alsatron</td>
<td>online</td>
<td>617</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>bellexia</td>
<td>online</td>
<td>729</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>hep-ex-cred</td>
<td>online</td>
<td>729</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>dnc-ex-bxi</td>
<td>online</td>
<td>523</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dune</td>
<td>online</td>
<td>519</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>dune-acienc</td>
<td>online</td>
<td>519</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

[CHEP 2023, May 11th] Marcus Ebert (mebert@uvic.ca)
Running for multiple Grid sites

● if site administrators want to manage cloud resources themselves:
  ○ single csv2 group per Grid site
    ■ Belle-II: Melbourne, DESY
  ○ create csv2 user with access to their own group
    ■ login with certificate or username/password

● otherwise:
  ○ multiple sites can be combined in single csv2 group
    ■ ATLAS: ECDF, LRZ, HEPHY, CERN-cloud
From the experiments, jobs come in with additional Requirements-string:

```
Requirements = ( group_name =?= "atlas-cern" &&
                target_alias =?= "uki-scotgrid-ecdf_cloud" ) && ...
```

“group_name” needs to be there for any csv2 job
Opportunistic usage between VOs

- csv2 has concept of
  - hard max: max number of cores on the cloud that csv2 could use for that group
  - softmax: no more cores than that should be used in total on the cloud
    - default: same as “hard max” (core quota on the cloud)

- opportunistic usage: softmax for main VO larger than for the other VO
  - higher softmax for main VO means it can still start VMs when needed
  - lower softmax for secondary VO means it will automatically retire VMs when main VO starts VMs
    - that way frees up more resource, main VO can start more VMs, secondary VO retires more,....
    - no jobs for main VO: it retires its VMs, secondary VO sees more resources available to start own VMs again

- fully automatic
Opportunistic usage between VOs

Belle-II can use up to 4000 cores on that cloud

- leaves room for Belle-II to start VMs when needed
- csv2 sees more than 3000 cores used on that cloud
- retires ATLAS VMs, means Belle-II can start more, retires even more ATLAS VMs,...

ATLAS can only use up to 3000 cores on that cloud
Opportunistic usage of cloud resources

- it may happen that a cloud has unused resources
- resource usage usually limited by quotas

- we can use opportunistic unused resource that a cloud admin wants to be used:
  - set cloud quotas to max as default (depending on max possible use)
    - instead of normal allocation
  - in csv2 set softmax to what should be used (e.g. normal allocation)
  - cloud admin adds new property to cloud project: “dynamic-cores”
  - via cli and cronjob, set softmax to retrieved “dynamic-cores”
    - cloud admin can change dynamic-cores as needed
    - csv2 will automatically retire and remove VMs when over quota

- we have it in place on two clouds currently
Monitoring

- all properties on the status page can be plotted and have a timeline
Summary

● a single csv2 instance can be used to manage multiple VOs and resources for multiple GRID sides efficiently

● can manage same resources for multiple VOs in an opportunistic way

● different VOs or site resources can be managed by different people
  ○ user accounts can be for a single csv2 group, access via username/password or certificate

● web interface and cli available

● we run single instance for 4 VOs, and as a service for 8 grid sites and for one local non-grid VO - works very well

more information about csv2:
Ansible playbook to install: https://github.com/hep-gc/uvic-heprc-ansible-playbooks
source: https://github.com/hep-gc/cloudscheduler
administration: https://indico.cern.ch/event/1222948/contributions/5321031/
public status page: https://csv2.heprc.uvic.ca/public/