

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



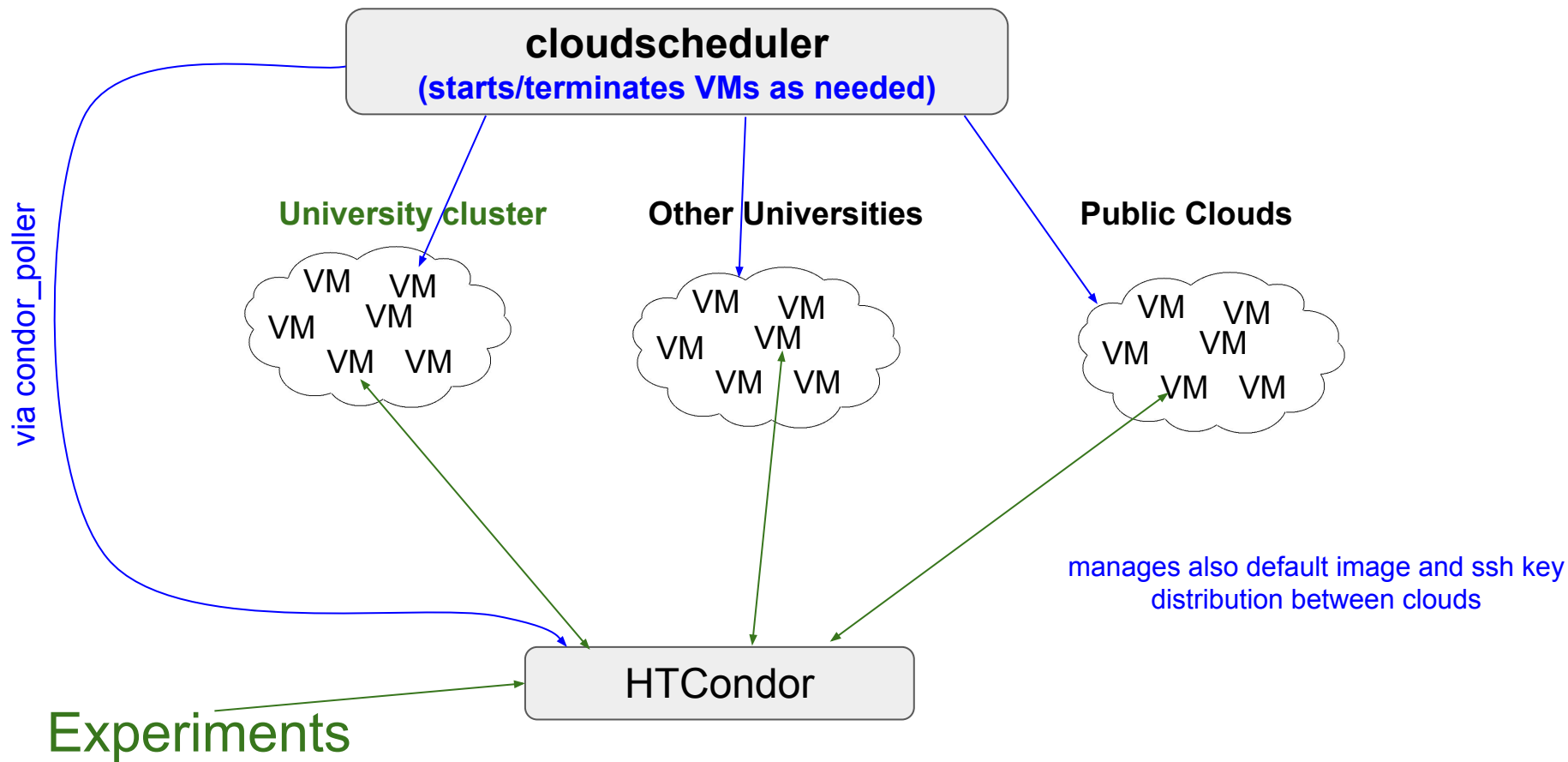
Marcus Ebert
on behalf of the

[HEP-RC group at the University of Victoria, Canada](#)

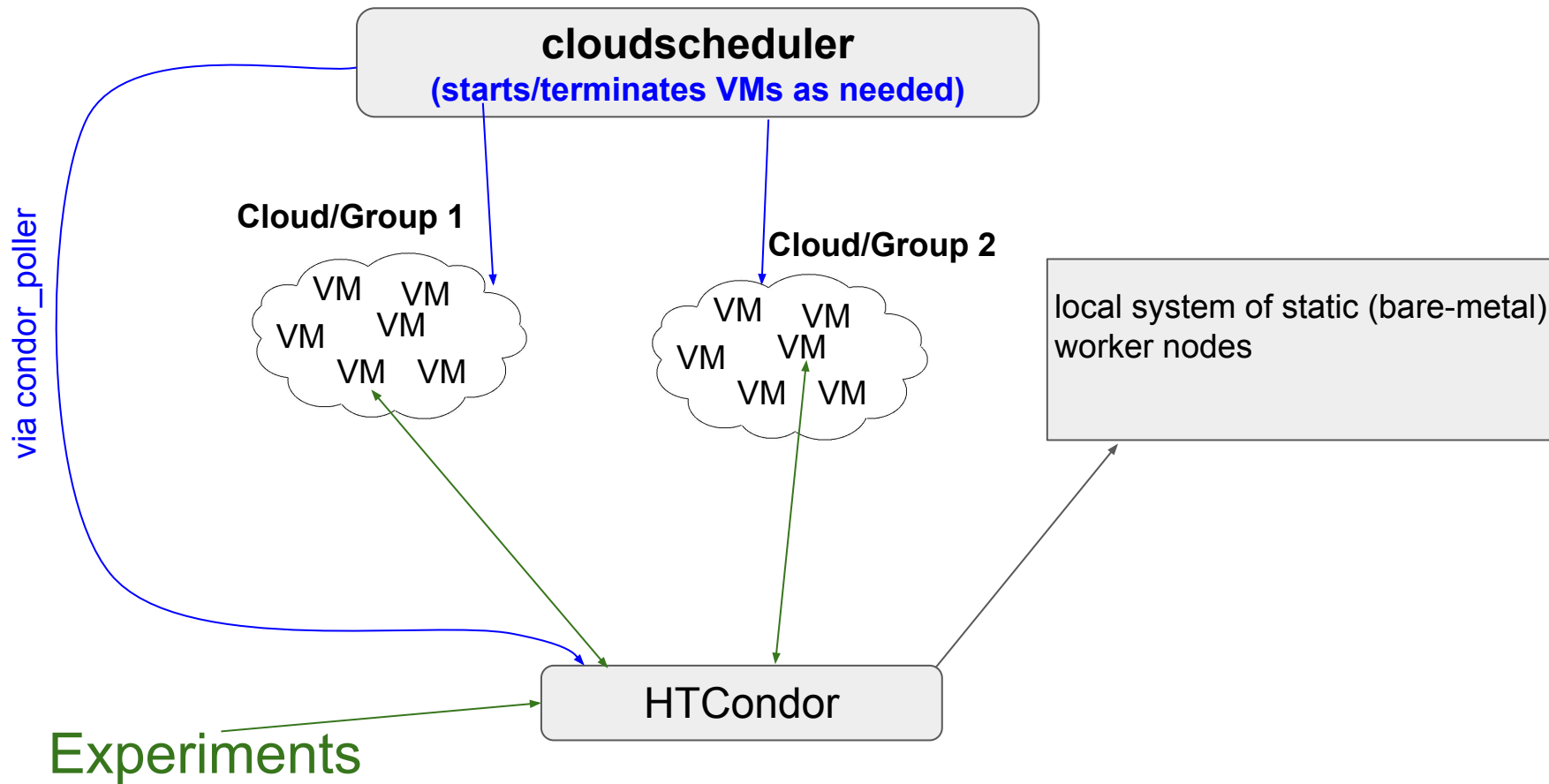
What is cloudscheduler (csv2)

- 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



Multi-cloud batch system with cloudscheduler



How to customize VMs

- csv2 uses cloud init together with [CernVM](#)
 - different set of files for different VO's or even specific to single clouds

group wide configurations
selected group

configurations transferred via cloud-init to the VMs; execution by priority

editor window

```
21 #!f [ -n "$V" ]; then
22 # export PERLSLIB="$PERLSLIB:$V"
23 #!f
24 #!v="$PYTHONPATH"
25 #export PYTHONPATH="$base)/usr/lib64/python2.6/site-packages:$base)/usr/lib/python2.6/site-packages"
26 #!f [ -n "$V" ]; then
27 # export PYTHONPATH="$PYTHONPATH:$V"
28 #!f
29 #export JAVA_HOME="$base)/usr/lib/jvm/jre-1.6.0-openjdk.x86_64"
30 #export LC_LOCALE="$base)/usr"
31 #export CLITE_LOCATION="$base)/usr"
32 #export SWI_PATH="$base)/usr/share/swi"
33 #export GAL_PLUGIN_DIR="$base)/usr/lib64/gfal2-plugins/"
34 #export GAL_CONFIG_DIR="$base)/etc/gfal2.d/"
35 #!v="$base"
36 #!f grep -q chameleon /var/lib/cloud_name; then
37 # export HWDIR=/dev/shm/$USER
38 # mkdir -p $HWDIR
39 #!f
40 owner: root:root
41 path: /etc/profile.d/grid-setup.sh
42 permissions: '0644'
43 content: |
44 #!/bin/bash
45 PATH="/bin:/usr/bin:/usr/local/bin:/sbin:/usr/sbin:/usr/local/sbin
46 cloudtype="unknown"
47 [[ -f /usr/lib/cloud_type ]] && cloudtype=$(cat /usr/lib/cloud_type)
48 logger -t "configure-disks" "Configuring disks on $(hostname) on cloud '$cloudtype'"
49 swap_size=0
50 swap_loc="/mnt/.rw/swap.1"
51 #dd [f=$(dev)zero of=$swap_loc bs=1M count=$swap_size
52 falllocate=$swap_size $swap_loc
53 mkswap $swap_loc
54 swapon $swap_loc
55 summary_rnd$(free)
56 summary_d=$(df -h)
57 logger -t "configure-disks" "Done! $summary_n $summary_d
58 #Setup condor space
59 systemctl stop condor
60 rm -rf /var/lib/condor/execute
61 mkdir -p /mnt/.rw/condor/execute
62 chown -R condor:condor /mnt/.rw/condor
63 ln -s /mnt/.rw/condor/execute /var/lib/condor/execute
64 owner: root:root
65 path: /usr/local/sbin/configure-disks
66 permissions: '0700'
67 #!f
68 local:
69 CVMFS_REPOSITORIES: belle.cern.ch,grid.cern.ch,belle.kek.jp
70 CVMFS_HTTP_PROXY: http://206.12.154.241:3126
71 CVMFS_S_CACHE_BASE: /mnt/.rw/cvmfs:cache
72 shoal:
73 shoal_server_url: http://shoal.heprc.uvic.ca/nearest
74 default_curl_proxy: http://206.12.154.241:3126
75 cron_shoal: '00 01,13 * * * root sleep ${RANDOM%1:2}; /usr/bin/shoal-client -n 8'
76 #!f
77 runCondor
78 echo nameserver 8.8.8.8 >> /etc/resolv.conf
79 echo nameserver 192.168.6.1 >> /etc/resolv.conf
80 [ -f /usr/local/sbin/configure-disks ]
81 [ ! -f /etc/condor/config.d/50cvmn ]
82 [ -f /usr/bin/shoal-client -n 8 -> /tmp/shoal-start.log
83 [ -f /etc/condor/condor/stop ]
84 [ -f /etc/condor/condor/stop ]
```

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

Group	Target Alias	Jobs	Idle	Running	Completed	Held	Other	Foreign	Condor FQDN	Condor Status	Condor Cert	Worker Cert
atlas-cern	None	0	0	0	0	0	0	0	csv2a.heprc.uvic.ca	✓	✓	✓
atlas-cern	cern-extension	179	168	10	1	0	0	0	csv2a.heprc.uvic.ca	✓	✓	✓
atlas-cern	hephy-uibk	38	5	22	9	0	2	0	csv2a.heprc.uvic.ca	✓	✓	✓
atlas-cern	lrz-lmu_cloud	213	204	9	0	0	0	0	csv2a.heprc.uvic.ca	✓	✓	✓
atlas-cern	uki-scotgrid-ecdf_cloud	31	0	30	1	0	0	0	csv2a.heprc.uvic.ca	✓	✓	✓
atlas-uk	None	0	0	0	0	0	0	0	csv2a.heprc.uvic.ca	✓	✓	✓
atlas-uvic	None	0	0	0	0	0	0	0	csv2a.heprc.uvic.ca	✓	✓	✓
atlas-uvic	ca-iaas-t3	230	98	125	7	0	0	0	csv2a.heprc.uvic.ca	✓	✓	✓
australia-belle	None	1326	437	889	0	0	0	0	bellecs.heprc.uvic.ca	✓	-	-
babar	None	0	0	0	0	0	0	0	login.babar.uvic.ca	✓	-	-
belle	None	1843	477	1366	0	0	0	0	bellecs.heprc.uvic.ca	✓	-	-
belle	belle-local-worker	2166	976	1190	0	0	0	0	bellecs.heprc.uvic.ca	✓	-	-
belle-validation	None	0	0	0	0	0	0	0	belle-sd.heprc.uvic.ca	✓	-	-
belle-validation	uvic-worker	5	0	5	0	0	0	0	belle-sd.heprc.uvic.ca	✓	-	-
desy-belle	None	0	0	0	0	0	0	0	bellecs.heprc.uvic.ca	✓	-	-
dune	None	2	0	2	0	0	0	0	dune-condor.heprc.uvic.ca	✓	✓	✓
testing	None	0	0	0	0	0	0	0	csv2a.heprc.uvic.ca	✓	✓	✓

Group	Clouds	RT (µs)	VMs								Slots			Native Cores			Foreign			Global			
			Starting	Unreg.	Idle	Running	Retiring	Manual	Error	Slots	Slot Cores Busy	Slot Cores Idle	Used	Limit	RAM	VMs	Cores	RAM	VMs	Cores	RAM	Volume	
atlas-cern	cern	771	5	0	0	0	5	0	0	0	10	40	0	40	40	0	0	0	5	40	0	0	
	ecdf	801	30	0	0	1	29	0	0	0	30	226	14	240	400	0	1	8	31	248	0	0	
	hephy	633	217	0	0	3	14	200	0	0	36	190	290	1736	1750	0	2	4	219	1740	0	0	
	lrz	637	9	0	0	0	9	0	0	0	9	90	0	90	136	0	3	0	12	90	0	0	
	lrz-pe72te	637	0	0	0	0	0	0	0	0	0	0	0	0	39	0	12	90	12	90	0	0	
Totals			261	0	0	4	57	200	0	0	85	546	304	2106	2326	0	6	12	267	2118	0	0	

Managing multiple VOs

belle		Status	Cloud Config	Aliases	Group Config	Images	Keys	Users	Groups	System Config	User Settings	HTCondor Plugin	Log out								
atlas-uvic	▼	arbutus	▼	728	113	0	0	0	110	3	0	0	114	891	21	904	3000	267	2112	380	3016
		cc-east	▼	571	8	0	0	0	8	0	0	0	8	64	0	64	72	1	4	9	68
		chameleon	▼	719	3	3	0	0	0	0	0	0	0	0	0	24	244	1	4	4	28
		otter	▼	560	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0
		Totals	▼		124	3	0	0	118	3	0	0	122	955	21	992	3316	269	2120	393	3112
australia-belle	▼	melbourne	▼	757	450	0	0	0	440	10	0	0	890	889	11	900	900	5	20	455	920
		Totals	▼		450	0	0	0	440	10	0	0	890	889	11	900	900	5	20	455	920
babar	▼	heprc-cloud	▼	720	0	0	0	0	0	0	0	0	0	0	0	0	608	19	73	19	73
		Totals	▼		0	0	0	0	0	0	0	0	0	0	0	0	608	19	73	19	73
belle	▼	amazon-w	▼	317590	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		arbutus	▼	728	261	0	0	0	260	1	0	0	2081	2078	11	2088	2088	119	928	380	3016
		beaver	▼	720	0	0	0	0	0	0	0	0	0	0	0	0	0	19	73	19	73
		beluga	▼	460	60	0	0	0	60	0	0	0	477	477	4	480	480	1	8	61	488
		cc-east	▼	507	0	0	0	0	0	0	0	0	0	0	0	0	80	9	68	9	68
		chameleon-new	▼	719	0	0	0	0	0	0	0	0	0	0	0	0	160	4	28	4	28
		ecdf-b	▼	655	0	0	0	0	0	0	0	0	0	0	0	0	64	0	0	0	0
		Totals	▼		321	0	0	0	320	1	0	0	2558	2555	15	2568	2568	120	936	441	3504
belle-validation	▼	arbutus	▼	728	0	0	0	0	0	0	0	0	0	0	0	0	5000	380	3016	380	3016
		beluga	▼	730	0	0	0	0	0	0	0	0	0	0	0	0	8	61	488	61	488
		heprc-cloud	▼	720	1	0	0	0	1	0	0	0	2	2	6	8	500	18	65	19	73
		Totals	▼		1	0	0	0	1	0	0	0	2	2	6	8	500	18	65	19	73
desy-belle	▼	desy	▼	833	0	0	0	0	0	0	0	0	0	0	0	0	96	2	5	2	5
		Totals	▼		0	0	0	0	0	0	0	0	0	0	0	0	96	2	5	2	5
dune	▼	dune-axion	▼	570	1	0	0	0	1	0	0	0	1	16	0	16	200	7	55	8	71

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

Running for multiple Grid sites

Group	Clouds	RT (µs)	VMs	Starting	Unreg.	Idle	Running	Retiring	Manual	Error	Slots	Slot Cores Busy	Slot Cores Idle	Native Cores Used	Native Cores Limit	RAM	Foreign VMs	Foreign Cores	Foreign RAM	Global VMs	Global Cores	Global RAM	Volume
atlas-cern	cern	990	5	0	0	0	5	0	0	0	10	40	0	40	40	<div style="width: 100%;"></div>	0	0	<div style="width: 0%;"></div>	5	40	<div style="width: 100%;"></div>	<div style="width: 0%;"></div>
	ecdf	782	32	0	1	1	30	0	0	0	30	233	24	256	400	<div style="width: 64%;"></div>	1	8	<div style="width: 0%;"></div>	33	264	<div style="width: 0%;"></div>	<div style="width: 0%;"></div>
	hephy	700	217	0	0	1	7	209	0	0	25	102	92	1736	1750	<div style="width: 99%;"></div>	2	4	<div style="width: 0%;"></div>	219	1740	<div style="width: 99%;"></div>	<div style="width: 0%;"></div>
	lrz	852	9	0	0	0	9	0	0	0	18	90	0	90	136	<div style="width: 66%;"></div>	3	0	<div style="width: 0%;"></div>	12	90	<div style="width: 0%;"></div>	<div style="width: 100%;"></div>
	lrz-pe72te	852	0	0	0	0	0	0	0	0	0	0	0	0	39	<div style="width: 22%;"></div>	12	90	<div style="width: 0%;"></div>	12	90	<div style="width: 0%;"></div>	<div style="width: 100%;"></div>
Totals			263	0	1	2	51	209	0	0	83	465	116	2122	2326	<div style="width: 91%;"></div>	6	12	<div style="width: 0%;"></div>	269	2134	<div style="width: 0%;"></div>	<div style="width: 0%;"></div>

atlas-cern
Status
Cloud Config
Aliases
Group Config
Images
Keys

cern-extension

hephy-uibk

lrz-lmu_cloud

lrz-pe72te

uki-scotgrid-ecdf_cloud
+

+

Clouds

- cern
- ecdf
- hephy
- lrz
- lrz-pe72te

Update

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

The image shows two side-by-side screenshots of the OpenStack VM configuration interface. The left screenshot is for the 'arbutus' VM on the 'belle' cloud. The 'Cores Softmax' field is set to 4000. The right screenshot is for the 'arbutus' VM on the 'atlas-uvic' cloud. The 'Cores Softmax' field is set to 3000. Red arrows point from the text below to the 'Cores Softmax' input fields in both screenshots.

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

ATLAS can only use up to 3000 cores on that cloud

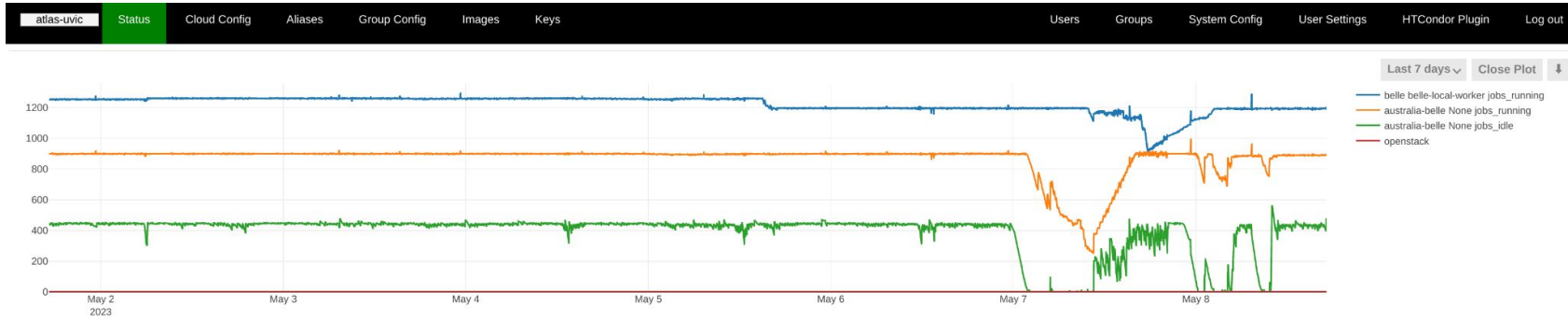
- leaves room for Belle-II to start VMs when needed
- csv2 sees more than 3000 cores used on that cloud
- retires ATLAS VMs, means Belle-II can start more, retires even more ATLAS VMs,....

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



Group	Target Alias	Jobs	Idle	Running	Completed	Held	Other	Foreign	Condor FQDN	Condor Status	Condor Cert	Worker Cert
atlas-cern	None	0	0	0	0	0	0	0	csv2a.heprc.uvic.ca	✓	✓	✓
atlas-cern	cern-extension	47	34	10	1	2	0	0	csv2a.heprc.uvic.ca	✓	✓	✓
atlas-cern	hephy-uibk	41	28	9	4	0	0	0	csv2a.heprc.uvic.ca	✓	✓	✓
atlas-cern	lrz-lmu_cloud	117	108	9	0	0	0	0	csv2a.heprc.uvic.ca	✓	✓	✓
atlas-cern	uki-scotgrid-ecdf_cloud	33	1	28	4	0	0	0	csv2a.heprc.uvic.ca	✓	✓	✓
atlas-uk	None	0	0	0	0	0	0	0	csv2a.heprc.uvic.ca	✓	✓	✓
atlas-uvic	None	0	0	0	0	0	0	0	csv2a.heprc.uvic.ca	✓	✓	✓
atlas-uvic	ca-iaas-t3	220	90	120	10	0	0	0	csv2a.heprc.uvic.ca	✓	✓	✓
australia-belle	None	1337	452	885	0	0	0	0	bellecs.heprc.uvic.ca	✓	-	-
babar	None	0	0	0	0	0	0	0	login.babar.uvic.ca	✓	-	-
belle	None	1934	568	1366	0	0	0	0	bellecs.heprc.uvic.ca	✓	-	-
belle	belle-local-worker	2155	965	1190	0	0	0	0	bellecs.heprc.uvic.ca	✓	-	-
belle-validation	None	0	0	0	0	0	0	0	belle-sd.heprc.uvic.ca	✓	-	-
belle-validation	uvic-worker	1	0	1	0	0	0	0	belle-sd.heprc.uvic.ca	✓	-	-

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/>