



Overview of CRAB3: Middleware for submitting CMS users' jobs to the grid

Emilis Antanas Rupeika

CRAB3 integration and release operator

Vilnius University



What is CRAB?

CRAB



Likes:
Cucumbers

Real name:
Bob

About CRAB3

- CMS Remote Analysis Builder version 3
- It's a distributed workflow management tool written in python
- Part of the Data and Workload management project (DMWM) in CMS.
- Software with which CMS users (physicists) can submit their data analysis and monte-carlo simulations to the Worldwide LHC Computing Grid (WLCG).

About CRAB3

- With CRAB, users submit **tasks**:
 - A unit of work corresponding to the processing of an input dataset or simulation.
 - Is split into many jobs to speed up processing.
- Job
 - A small piece of processing of a part of an input dataset or simulation.
 - Generally uses one processor core in the WLCG

Some CRAB3 features

- Simplifies the submission of jobs to the Grid (splits a task into jobs, generates input for the HTCondor batch processing software, checks for sites that are down, input data availability)
- Allows the user to track the status of their jobs via “crab status” command or the task dashboard webpage.
- Allows control of jobs as a single unit (task)
- Resubmits failed jobs
- Asynchronous Stage-Out component manages output data transfers to the user’s specified location
- Publishes output data so that it can be used as input later on)

Job dashboard

Filters Users Help

Found a bug?

REFRESH Disabled



Search

- Nicholas Anthony De Porzio
- Todor Trendafilov Ivanov

Start » [Emilis Antanas Rupeika] » Tasks » Jobs

Task: 161209_105351:erupeika_crab_1612rc6_preprod_test9

NJobTotal: 18 Pending: 0 Running: 0 Unknown: 0 Cancelled: 0 Success: 15 Failed: 0 WNPPostProc: 3 ToRetry: 0

Data Charts Show 25 entries

*All times are in UTC

Search:

Id	Status	ExitCode	Site	Retries	Submit	Start	Finish	Wall Time	Logs	Input Files	FTS File Status
1	finished	0	T3_UK_ScotGrid_GLA	1	09-12 10:54:40	09-12 12:33:33	09-12 12:39:40	00:06:04	Job,Post Job	File Info	N/A
2	finished	0	T2_UK_London_Brunel	1	09-12 10:54:40	09-12 11:00:01	09-12 11:36:56	00:36:55	Job,Post Job	File Info	N/A
3	finished	0	unknown	2	09-12 11:09:42	-	09-12 12:11:34	00:00:00	Job,Post Job	File Info	N/A
4	finished	0	T2_EE_Estonia	1	09-12 11:10:52	09-12 11:10:52	09-12 11:26:40	00:14:51	Not available	File Info	N/A
5	finished	0	T3_HU_Debrecen	1	09-12 10:54:40	09-12 11:00:01	09-12 12:47:03	01:46:19	Job,Post Job	File Info	N/A
6	finished	0	T3_UK_ScotGrid_GLA	1	09-12 10:54:40	09-12 11:17:02	09-12 11:25:14	00:07:40	Job,Post Job	File Info	N/A
7	wnpostproc	0	unknown	1	09-12 10:54:40	-	09-12 11:15:55	00:00:00	Job,Post Job	File Info	N/A
8	finished	0	T1_UK_RAL	1	09-12 10:54:40	09-12 11:18:00	09-12 11:32:25	00:05:15	Job,Post Job	File Info	N/A
9	wnpostproc	0	T2_EE_Estonia	2	09-12 12:40:39	09-12 12:43:02	09-12 12:53:09	00:10:38	Job,Post Job	File Info	N/A
10	finished	0	unknown	1	09-12 10:54:40	-	09-12 11:07:10	00:00:00	Job,Post Job	File Info	N/A
11	finished	0	T1_RU_JINR	1	09-12 11:15:49	09-12 11:15:49	09-12 11:19:52	00:04:08	Not available	File Info	N/A
12	finished	0	T3_US_Colorado	1	09-12 11:00:06	09-12 11:00:06	09-12 11:05:33	00:05:52	Not available	File Info	N/A
13	finished	0	T3_HU_Debrecen	1	09-12 11:00:01	09-12 11:00:01	09-12 12:43:03	01:42:42	Not available	File Info	N/A
14	finished	0	T3_HU_Debrecen	1	09-12 10:54:40	09-12 10:56:39	09-12 12:45:27	01:48:48	Job,Post Job	File Info	N/A
15	finished	0	T2_EE_Estonia	2	09-12 11:37:16	09-12 12:33:29	09-12 12:44:25	00:11:15	Not available	File Info	N/A
16	finished	0	T3_US_UMD	3	09-12 15:53:11	09-12 16:00:49	09-12 16:14:11	00:08:42	Job,Post Job	File Info	N/A
17	finished	0	T1_DE_KIT	2	09-12 12:10:08	09-12 12:11:47	09-12 12:36:51	00:24:39	Job,Post Job	File Info	N/A
18	wnpostproc	0	T2_EE_Estonia	1	09-12 10:54:40	09-12 12:33:29	09-12 12:42:30	00:09:21	Job,Post Job	File Info	N/A

Showing 1 to 18 of 18 entries

First Previous Page 1 of 1 Next Last

Powered by hBrowse framework

Task Monitoring

Example crab status output

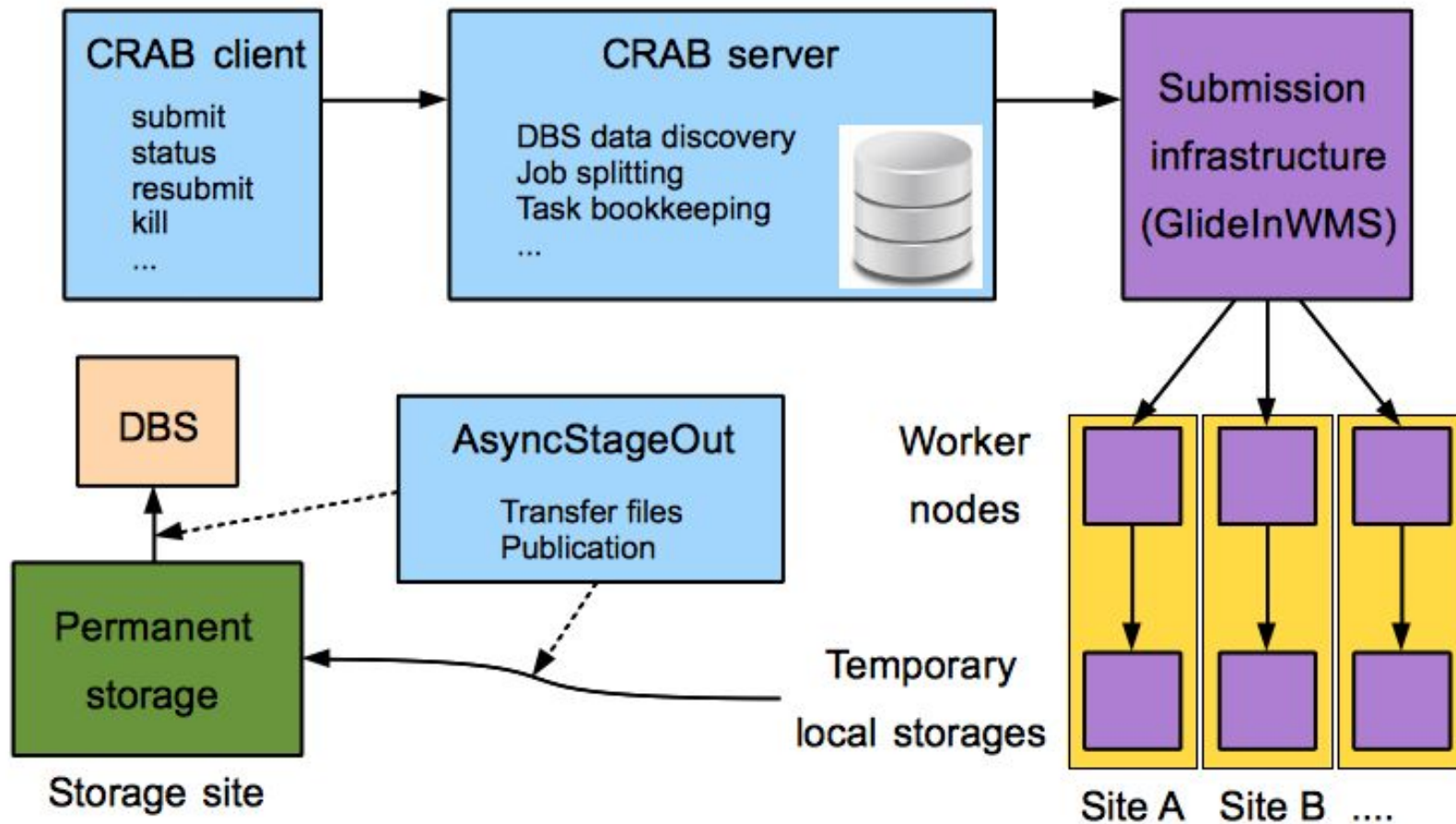
```
[erupeika@lxplus033 crab_projects]$ crab st -d crab_1612rc6_preprod_test9
CRAB project directory:      /afs/cern.ch/user/e/erupeika/crab_projects/crab_1612rc6_preprod_test9
Task name:                   161209_105351:erupeika_crab_1612rc6_preprod_test9
Grid scheduler:              crab3-5@vocms059.cern.ch
Task status:                 COMPLETED
Dashboard monitoring URL:    http://dashb-cms-job.cern.ch/dashboard/templates/task-analysis/#user=erupeika&ta
ble=Mains&pattern=161209_105351%3Aerupeika_crab_1612rc6_preprod_test9

Jobs status:                 finished          100.0% (18/18)

Publication status:         finished          100.0% (18/18)

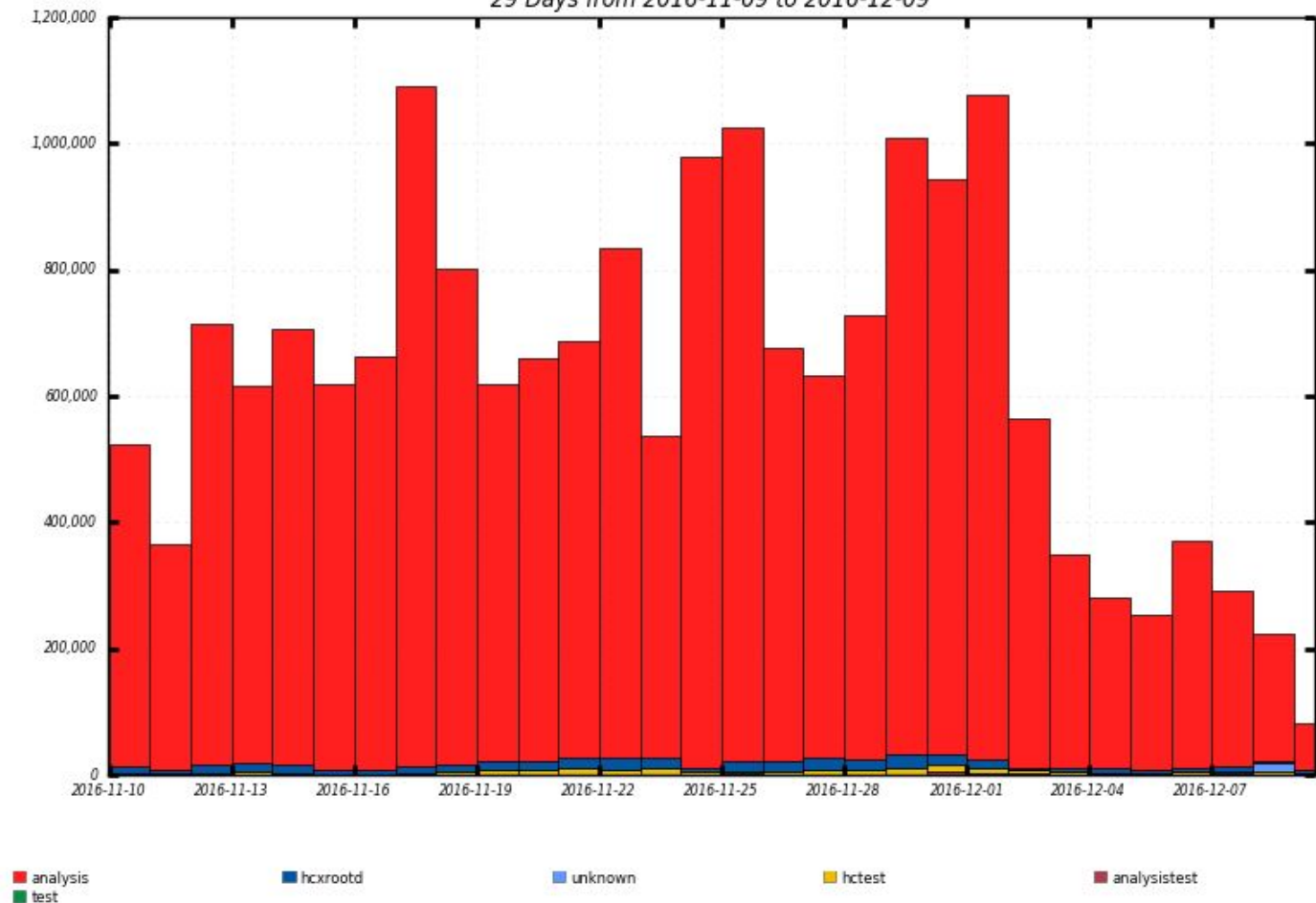
Output dataset:             /GenericTTbar/erupeika-crab_1612rc6_preprod_test9-37773c17ce2994cf16892d5f04945e
41/USER
Output dataset DAS URL:     https://cmsweb.cern.ch/das/request?input=%2FGenericTTbar%2Ferupeika-crab_1612rc6
_preprod_test9-37773c17ce2994cf16892d5f04945e41%2FUSER&instance=prod%2Fphys03
Log file is /afs/cern.ch/user/e/erupeika/crab_projects/crab_1612rc6_preprod_test9/crab.log
[erupeika@lxplus033 crab_projects]$
```


CRAB3 architecture

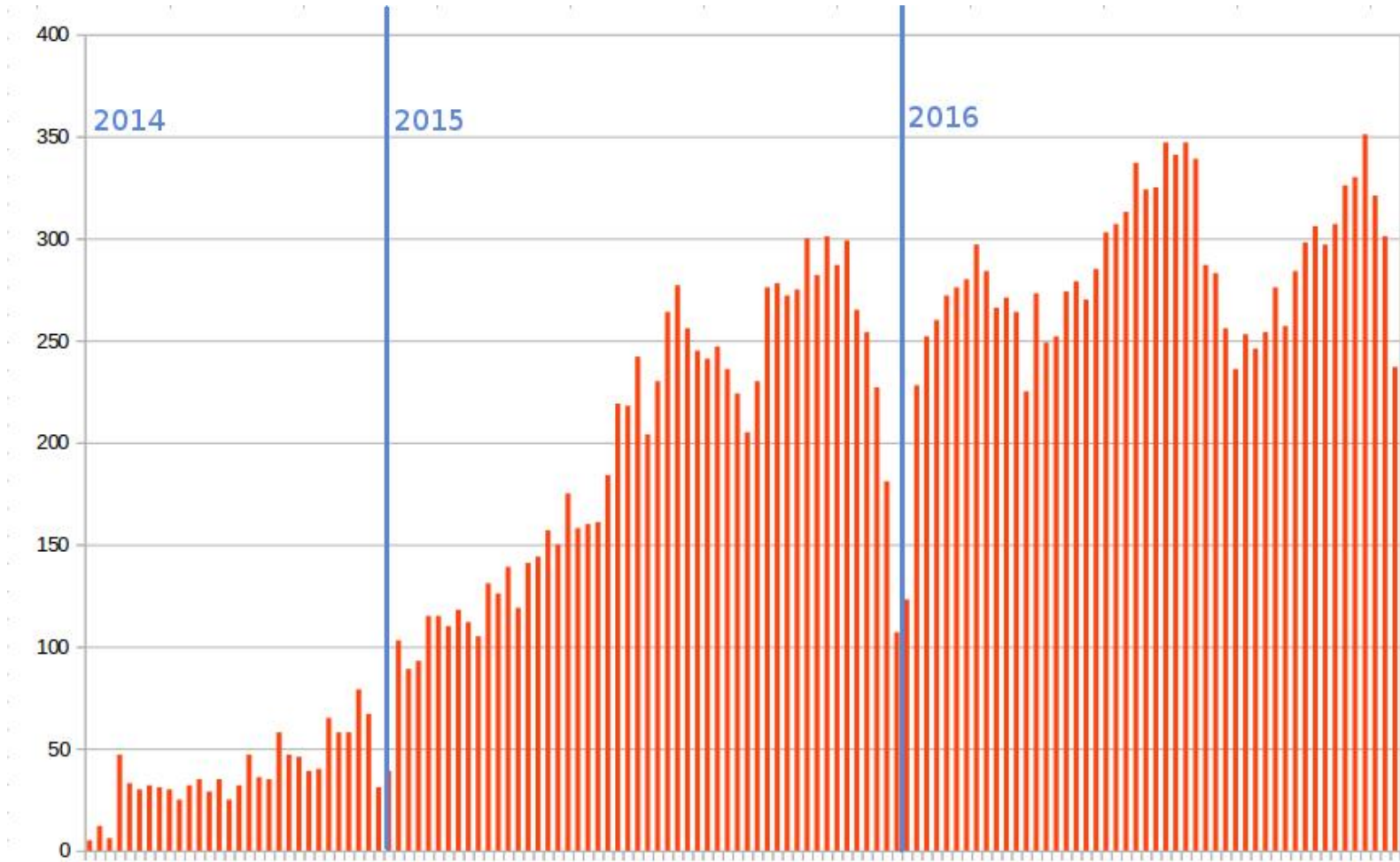


Number of completed jobs last month.

Total ~19 million



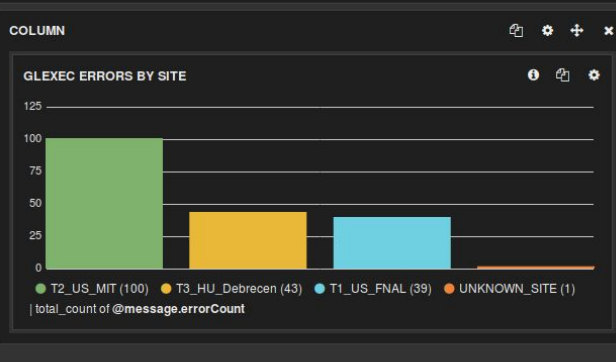
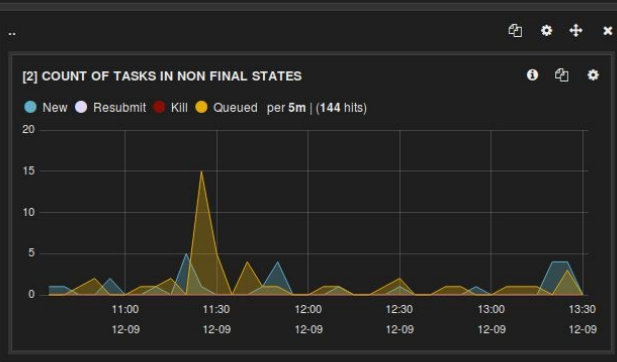
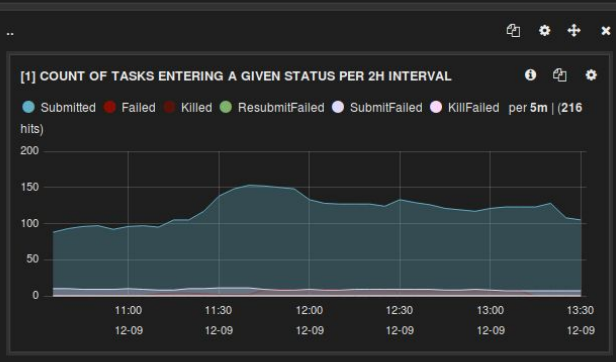
Number of unique users per week since the start of CRAB3



Quickly about what I do

- Maintenance
 - Bugs, improvements of existing features, CRABMon
- Accept pull requests
 - Code changes submitted by other developers
- Keep an eye on the infrastructure
- Integration tests
 - Usually there is a monthly deployment of a new CRAB version
 - All of the components have to be tested working together to make sure the changes don't break something unexpected
- Create packages (rpms) from which our software components are deployed

Kibana monitoring page

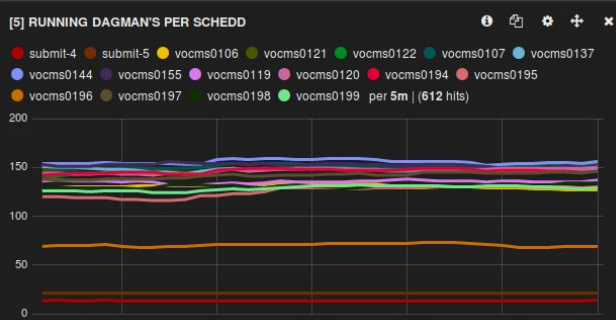
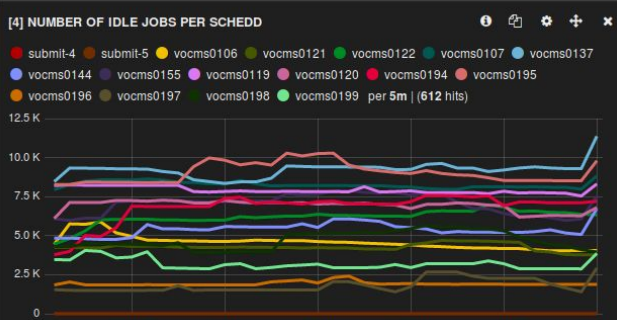
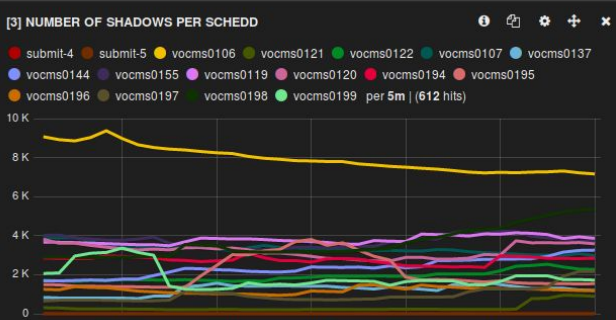


TIPS

If the number of failed tasks increase you can check the last failures reasons here
[Last failed tasks](#)

- ### ALARMS
- Prod TW (2)
 - Pre Prod TW (0)
 - vocms0106 (41)
 - vocms0121 (42)
 - vocms0122 (51)
 - vocms0107 (5)
 - vocms0137 (4)
 - vocms0144 (5)
 - vocms0155 (5)
 - vocms0119 (2)
 - vocms0120 (2)
 - vocms0194 (2)
 - vocms0195 (2)
 - vocms0196 (28)
 - vocms0197 (16)
 - vocms0198 (14)
 - vocms0199 (31)

To see which hosts are showing gLExec errors at a site click [here](#)



CRABMonitor

CRABMonitor

Task info

161205_064427:gurpinar_crab_RSToQQ_GEN-SIM-RAW-v1



prod



Go

Main

[Task Info](#)

[Config](#)

[PSet](#)

[TaskWorker log](#)

[Upload log](#)

[Script Exe](#)

Main

Dashboard	Webdir	DAS Input	DAS Output
-----------	--------	-----------	------------

status SUBMITTED

username gurpinar

jobs [object Object]

submissionTime 1480920269

jobdefErrors

collector cmsgwms-collector-global.cern.ch,cmssrv221.fnal.gov

jobList
idle 240
finished 153
running 1
failed 6
TOTAL 400

publication [object Object]

schedd crab3@vocms0199.cern.ch

taskWarningMsg The CRAB3 server got a non-critical error while checking stageout permissions. Please use checkwrite to check if everything is fine.

jobsPerStatus [object Object]

publicationFailures [object Object]

statusFailureMsg

taskFailureMsg

failedJobdefs 0

command SUBMIT



Thanks!