GEMC Ongoing Development
Interactive and “farm” Containers
CLAS12 Simulations Jobs using off-site resources

GEMC 4.3.2

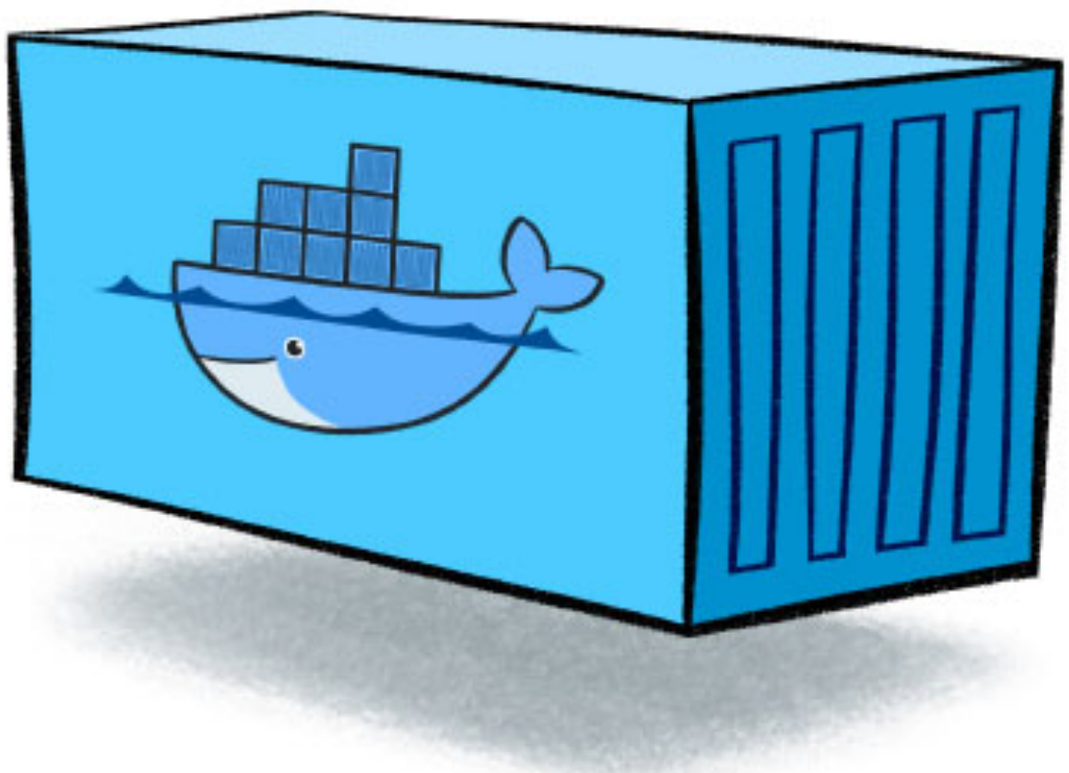
In development:

- 4.3.2:
  - FILTER_HADRONS option to write out events that have hit from specific hadrons in them
  - Rich sector 4 passive materials
  - Background merging memory check ➡ SOON
  - Hipo 4 output ➡ SOON
  - arbitrary number of sequential rotations in the detector definition ➡ SOON
  - BMT digitization with global coordinates instead of locals ➡ SOON
  - TOFs resolutions pars from CCDB ➡ SOON
  - Move LUND vertex based on gcard entry ➡ SOON
  - Time propagation in DC digitization ➡ SOON
  - Rich sector 4 digitization ➡ SOON
  - 3D Cylindrical map field ➡ SOON
  - Detector time signal shift to match data ➡ SOON

Docker: Interactive and Production Containers

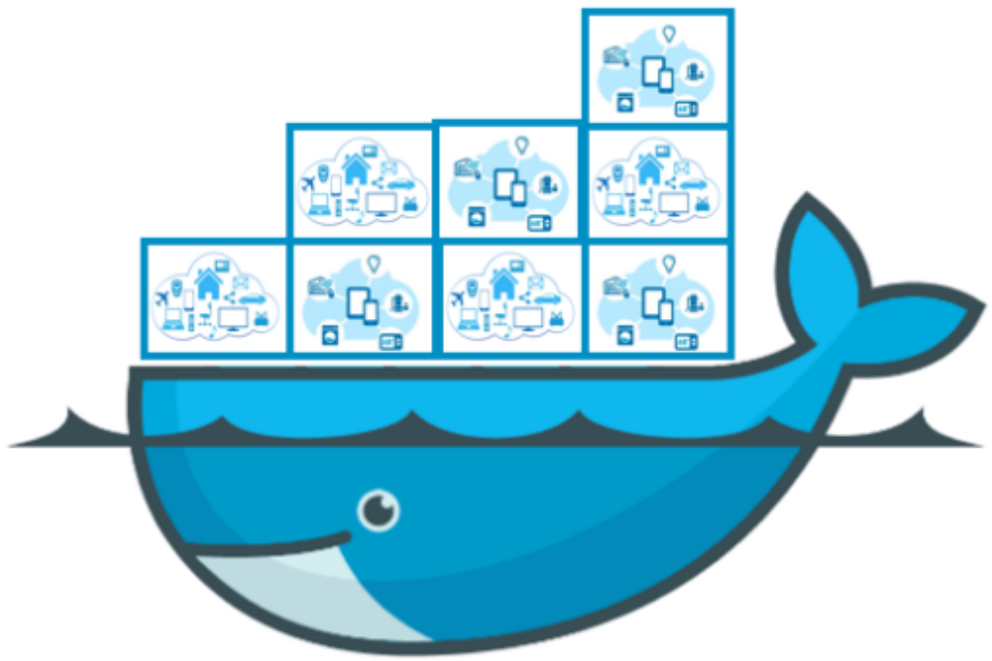
Contain necessary software to perform any of the
generator → GEMC → reconstruction → DST → CED → analysis

`clas12simulations:iprod`
`:1.2`
`:1.1`



- To be used **interactively** (GUI included) by users

`clas12simulations:production`



- Specialized container to be used on computer farms

Interactive Container: “iprod”

Meant to be run on user desktops / laptops. Includes GUI capability

NO REQUIREMENTS (other than docker)

Contains **all the necessary software** to run simulations > reconstruction chain:

- geant4
- GEMC
- CLAS12 geometry
- CLARA, multithreads
- COATJAVA
- CED

Versioning: done with image tag: clas12simulations:iprod (1.3, 1.2, etc)

gemc: version 4.3.1,
CLARA: 4.3.10
Coatjava: 6.3.1
ced 1.4.04

Not feasible for offsite farms, where we will use only 2 lightweight images

Docker images created and released along with our software versioning

Interactive Container: “iprod”

Quickstart: Full chain using docker

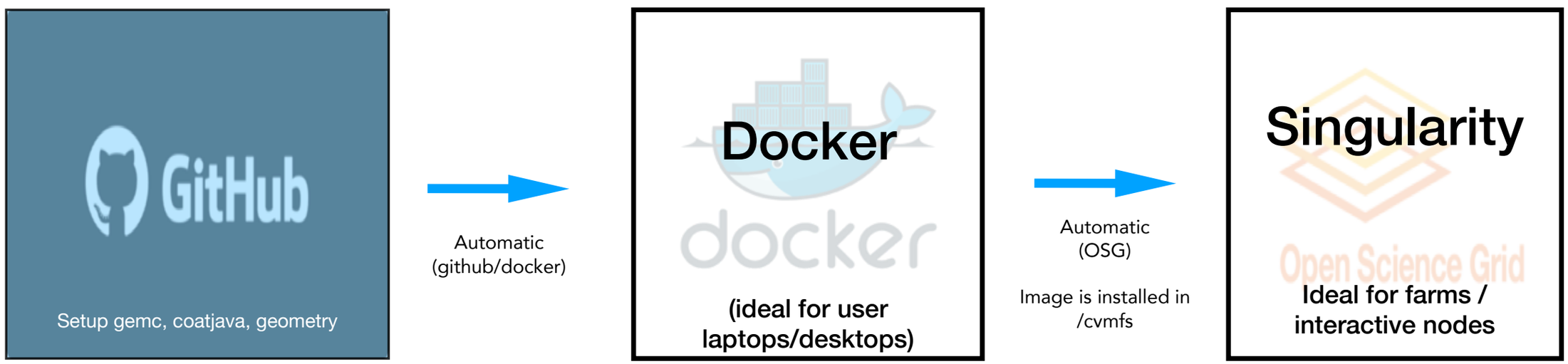
Use the following command to run the clas12 software image using a “~/mywork” local directory, and run clasdis events through the GEMC/COATJAVA chain:

```
mkdir -p ~/mywork
docker run -it --rm  -v ~/mywork:/jlab/work/mywork jeffersonlab/clas12simulations:iprod bash
cd mywork
clasdis --trig 1000 --docker --t 25 35
gemc -USE_GUI=0 -N=100 -INPUT_GEN_FILE="lund, sidis.dat" /jlab/clas12Tags/gcards/rga-spring2018.gcard
evio2hipo -r 11 -t -1.0 -s -1.0 -i out.ev -o gemc.hipo
createClaraCook.csh gemc.hipo rga-spring2018 1
clara-shell cook.clara
```

We already know how to do this.
Will not be shown on Friday.

Farm “production” container

- Currently very similar to “iprod” but the vision is:
- Lightweight images (production and development)
- Will NOT contain all libraries dependencies: we will use CVMS for that: geant4, magnetic fields, etc. These comes with versioning.
- Will contain MANY (if not all) versions of the software, organized by tags that match the “module load” model.
- Distributed automatically to CVMFS



Off-site CLAS12 Simulations

Goals

- About 80 million CPU hours / year, corresponding to about 8,000 cores.
- Run CLAS12 jobs using container that includes the full chain:
    A. Generators and user LUND files;
    B. GEMC (optional: background from experimental data), decoder;
    C. Reconstruction;
    D. DST.
- Submit jobs from a web interface

OSG Opportunistic and Offsite Resources

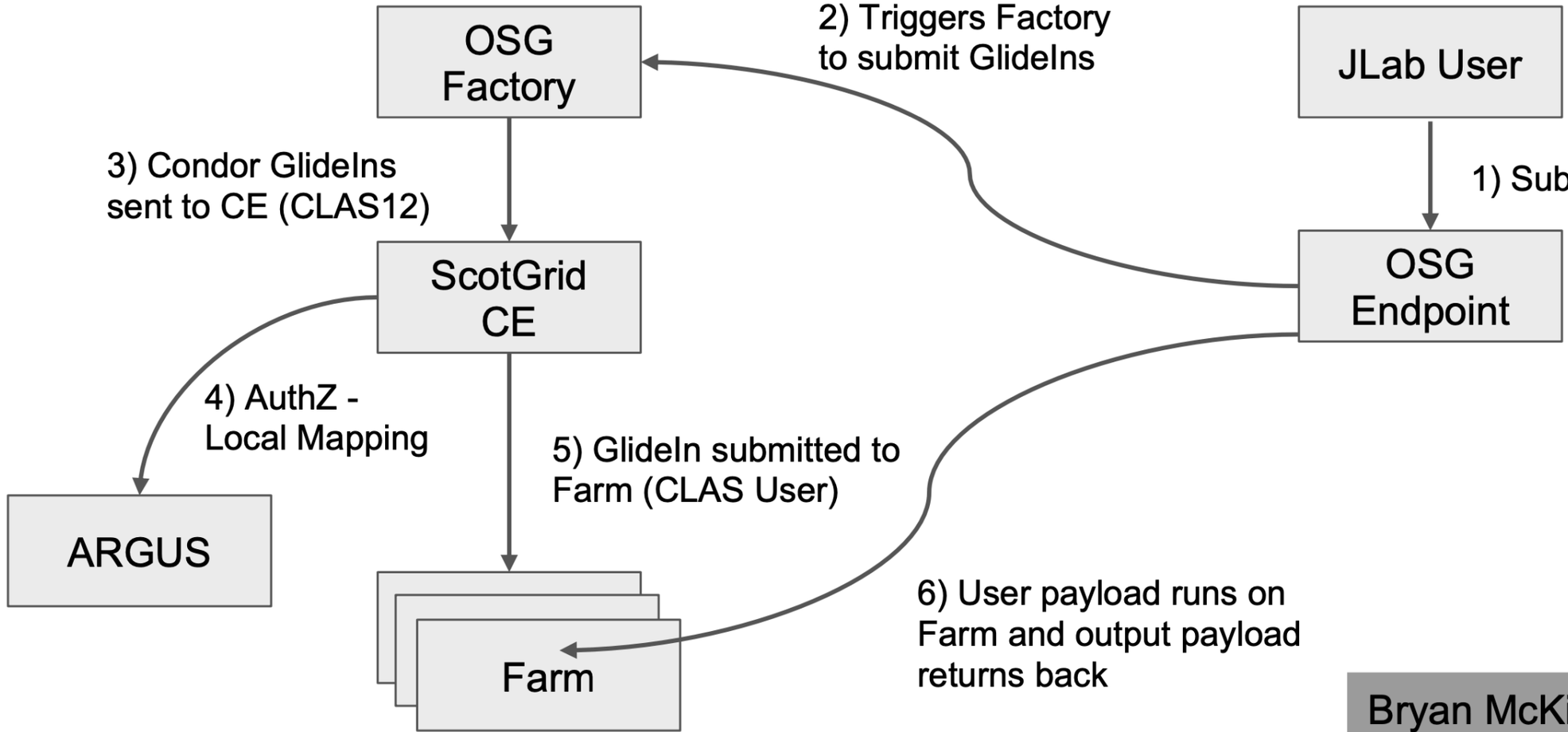
- JLAB node setup to submit jobs to OSG “Opportunistic” CPU time: “free” CPU cycles.

- To use other resources, we need to “authenticate” our CLAS12 submissions: we cannot use the GLUEX Virtual organization anymore.

- OSG + JLAB Computer Center: A **JLAB Virtual Organization** has been created.

- “CLAS12” OSG Project has been created. Jobs submitted under CLAS12 use the JLAB VO.

ScotGrid



Bryan McKinnon (NHPR)
Gareth Roy (ScotGrid/GridPP)

ScotGrid, IRIS and GridPP

Resource:
- Access to our 10% share* (development nodes) for other VOs now (~400 cores)

-- Access to IRIS resources online March 2020 (1250 cores)
(we will have a %share* of 1650 total) - BMK

---Access to IRIS resources based on formalised application - BMK
(Total IRIS is predicted to be 12,000 cores in 2020)

Current cluster snapshot:
Available – 4232
Busy – 2325
Idle** – 1523

* handled by fairshare
** a) not enough memory available to run jobs
b) draining to allow 8 core jobs to get in
c) draining for upgrades…

Bryan McKinnon (NHPR)
Gareth Roy (ScotGrid/GridPP)

ScotGrid, IRIS and GridPP

Resource:
- Access to our 10% share* (development nodes) for other VOs now (~400 cores)

-- Access to IRIS resources online March 2020 (1250 cores)
(we will have a %share* of 1650 total) - BMK

---Access to IRIS resources based on formalised application - BMK
(Total IRIS is predicted to be 12,000 cores in 2020)

Current cluster snapshot:
Available – 4232
Busy – 2325
Idle** – 1523

10% of these resources

* handled by fairshare
** a) not enough memory available to run jobs
b) draining to allow 8 core jobs to get in
c) draining for upgrades…

Bryan McKinnon (NHPR)
Gareth Roy (ScotGrid/GridPP)

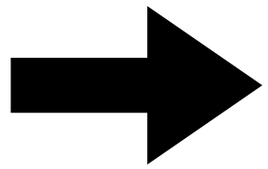
Off-site CLAS12 Simulations

Web Interface

GEMC Server:
gemc.jlab.org

JLAB User Authentication

Prepare
Submit
Monitor



SubMit MYSQL SERVER

User name, ID, IP
Job Information Package

Off-site CLAS12 Simulations

Web Interface

GEMC Server:
gemc.jlab.org

JLAB User Authentication

Prepare
Submit
Monitor

SubMit MYSQL SERVER

User name, ID, IP
Job Information Package
Job Submission Infos

JLAB OSG NODE
Submit

Off-site CLAS12 Simulations

Web Interface

GEMC Server:
gemc.jlab.org

JLAB User Authentication

Prepare
Submit
Monitor

SubMit MYSQL SERVER

User name, ID, IP
Job Information Package
Job Submission Infos

JLAB OSG NODE
Submit

RUN
OSG Opportunistic
ScotGrid

Off-site CLAS12 Simulations

Web Interface

GEMC Server:
gemc.jlab.org

JLAB User Authentication

Prepare
Submit
Monitor

SubMit MYSQL SERVER

User name, ID, IP
Job Information Package
Job Submission Infos

JLAB OSG NODE
Submit

RUN
OSG Opportunistic
ScotGrid

Output @ JLAB
(currently
/volatile/clas12/osg/username)

Off-site CLAS12 Simulations

`https://gemc.jlab.org/web_interface/`

"live" on Oct 24 for alpha release
Beta release announced to the collaboration on Nov. 6.

We invite you to use it, either for your analysis or just to get familiar with the job submission process.

CLAS12 Monte-Carlo Job Submission Portal

Logged in as ungaro

updated on 2019-11-09 15:30:06

Farm Name	Total Available Cores	Busy Cores	Idle Cores
osg	17225	16616	609

username	job_id	submitted	total	done	running	idle	hold	osg_id
avakian	80	11/6 12:50	1000	999	0	0	1	1422770
fxgirod	81	11/6 14:22	3000	2999	0	0	1	1423098
avakian	84	11/7 07:38	660	659	1	0	0	1427085
avakian	85	11/7 08:20	240	239	1	0	0	1427587
fxgirod	87	11/7 16:50	5000	4996	0	0	0	1430324
fxgirod	89	11/8 06:57	5000	4949	51	0	0	1432249
fxgirod	91	11/9 12:49	5000	12	4779	10	199	1438158

19900 jobs, 14853 completed, 10 idle, 4832 running, 201 held, updated on 2019-11-09 15:46:02

Type 1

- Container CLAS12 gcard
- Container or gemc internal generator
- Arbitrary number of jobs
- Arbitrary number of events for each job (max 10,000)

Type 2

- Container CLAS12 gcard
- Use LUND files from a web location or directory in /volatile
- One job per LUND file

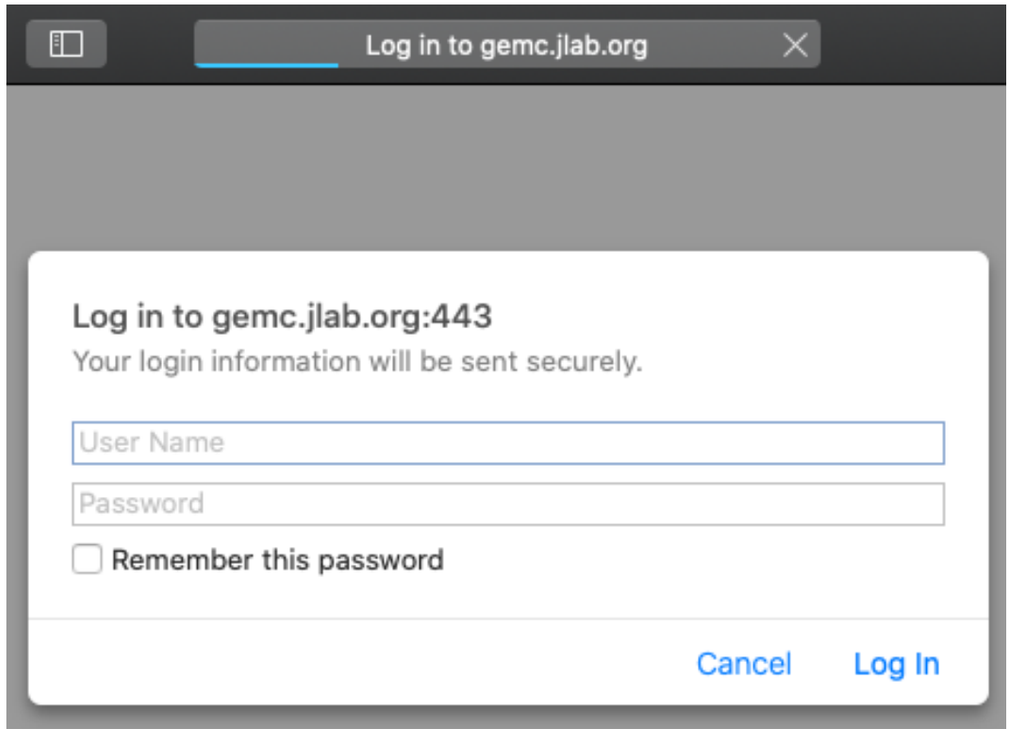
Type 3 (coming soon)

- Use gcards from web location.
- One set of jobs per gcard.
- Container generator
- Arbitrary number of jobs
- Arbitrary number of events for each job

Type 4 (coming soon)

- Use gcards from web location.
- One set of jobs per gcard.
- Use LUND files from a web location
- One job per LUND file

Log-in with JLab Authentication



- This is the JLAB Computer Center authentication mechanism
- Used on the GEMC server gemc.jlab.org

- First step for users using the web interface
- Information collected allow for:
  - organizing output directories
  - collecting users infos (statistical purpose, perhaps expand to priority, etc)

UserID	User
1	robertej
2	ungaro
3	baltzell
4	devita
5	guillaum
6	zwzhao
7	dmriser
8	avakian
9	fxgirod
10	sangbaek
11	sdiehl
12	hattawy

GEMC SERVER

JLAB Authentication

Apache Server
gemc.jlab.org

I/O to submit mySQL server:
submit.jlab.org

Scripting and **PYTHON**:

Create farm statistics
Create job monitoring

Main Page: **PHP**
Read and display farm statistics
Read and display job monitoring
Provide type 1,2 (3,4 coming soon)

Types: **HTML, PHP**
Collect job information
Prepare job “package”

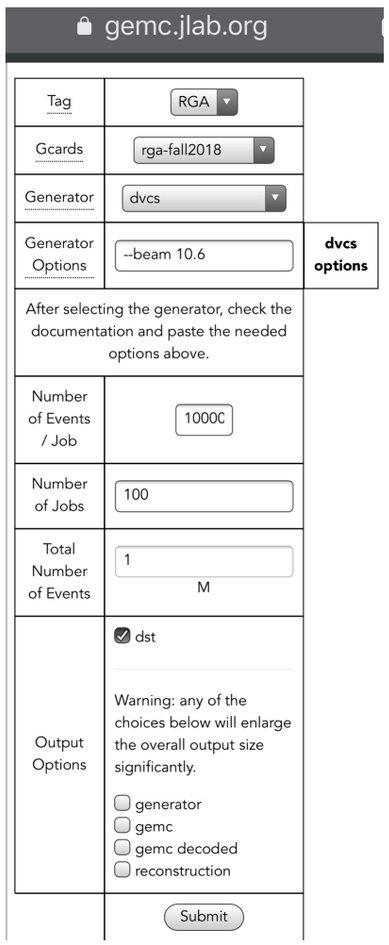
JLAB OSG NODE
"gemc" user: submit

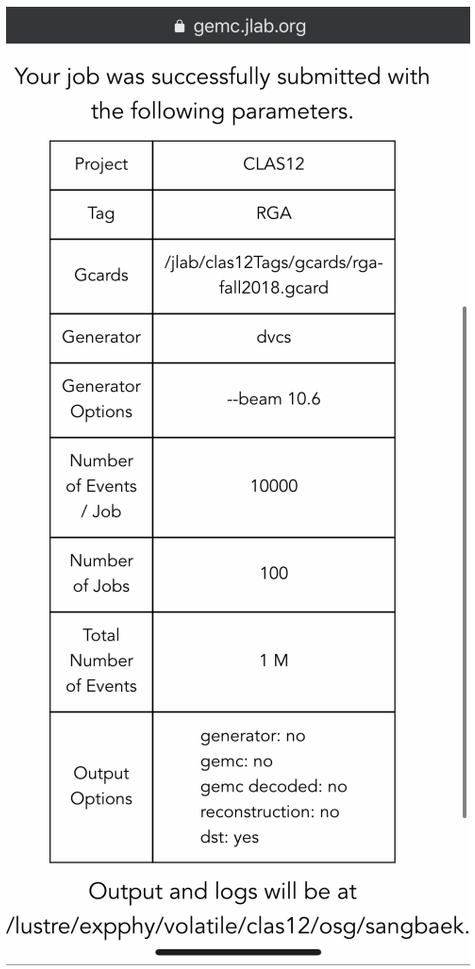
PHP, PYTHON:

Submit job package

CSS Technology: browser independence

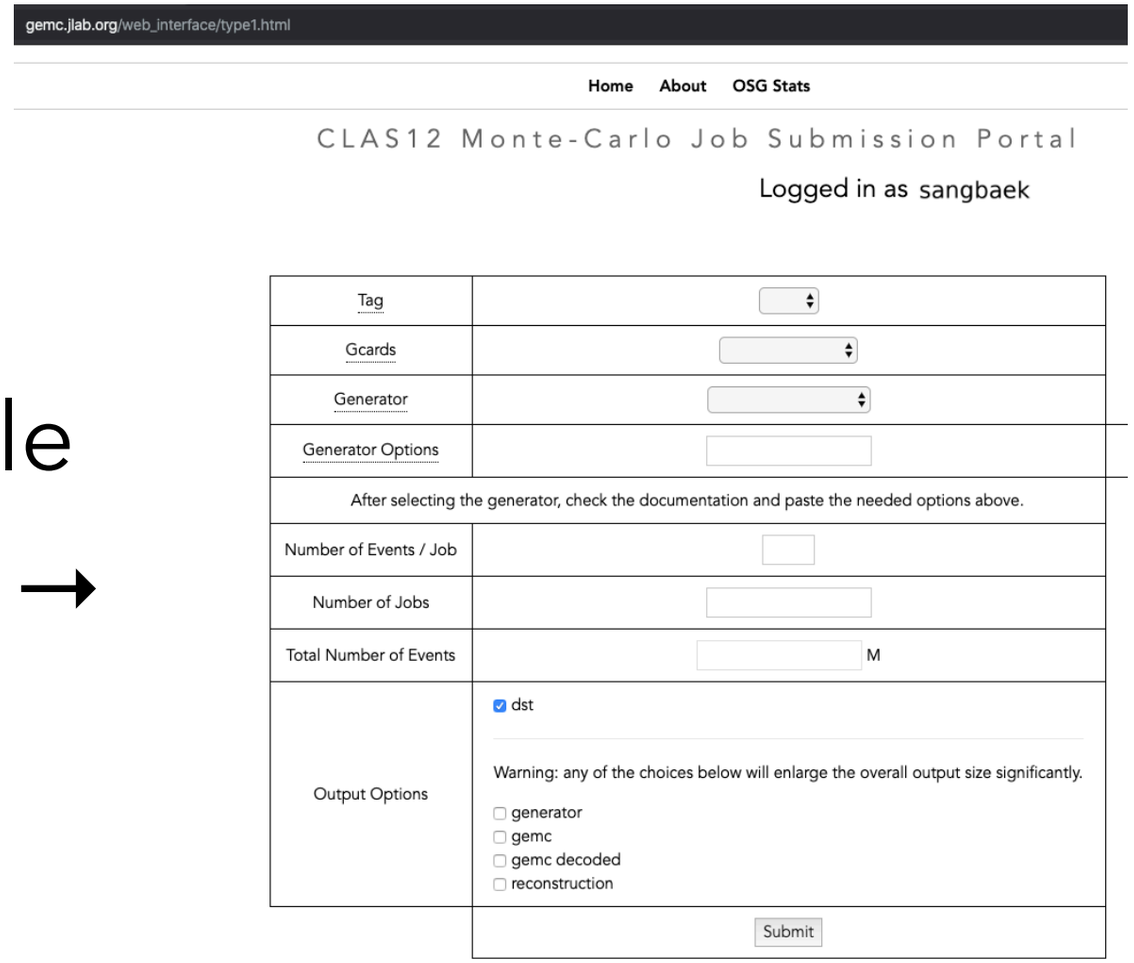
- Pages rendered w.r.t. device widths (responsive design w/ css)
- Can submit jobs from Mobile





← Mobile

Desktop →



We will show this during the Friday session

Monitoring Available Off-site Farms

Home **About** **OSG Stats**

CLAS12 Monte-Carlo Job Submission Portal

Logged in as ungaro

updated on 2019-11-11 13:40:05

Farm Name	Total Available Cores	Busy Cores	Idle Cores
osg	37571	36288	1283

- Farms statistics inquired regularly
- Saved in json format
- Display total / busy / idle cores
- Glasgow will appear shortly
- Framework expandable to other offsite farms as well

Tag	RGA
Gcards	rga-fall2018
Generator	clasdis
Generator Options	
After selecting the generator, check the documentation and paste the needed options above.	
Number of Events / Job	10000
Number of Jobs	1000
Total Number of Events	10 M
Output Options	dst Warning: any of the choices below will enlarge the overall output size significantly. generator gemc gemc decoded reconstruction
	Submit

clasdis options

Stats Purposes

Type 1 and 2 corresponds to conditions of CLAS12 experiments that already ran:

clas12-default: nominal CLAS12 geometry
rga-spring2018: RGA spring 2018 configuration
rga-fall2018: RGA fall 2018 configuration
rgk-fall2018: RGK fall 2018 configuration (torus out-bending)
rga-spring2019: RGA spring 2019
rgb-spring2019: RGB spring 2019

Type 1 submission

Tag	RGA
Gcards	rga-fall2018
Generator	clasdis
Generator Options	
After selecting the generator, check the documentation and paste the needed options above.	
Number of Events / Job	10000
Number of Jobs	1000
Total Number of Events	10 M
Output Options	☑ dst Warning: any of the choices below will enlarge the overall output size significantly. ☐ generator ☐ gemc ☐ gemc decoded ☐ reconstruction
	Submit

clasdis options

Stats Purposes

Type 1 and 2 corresponds to conditions of CLAS12 experiments that already ran:

clas12-default: nominal CLAS12 geometry
rga-spring2018: RGA spring 2018 configuration
rga-fall2018: RGA fall 2018 configuration
rgk-fall2018: RGK fall 2018 configuration (torus out-bending)
rga-spring2019: RGA spring 2019
rgb-spring2019: RGB spring 2019

gemc
✓ clasdis
dvcs
disrad
genKYandOnePion

Tag	RGA
Gcards	rga-fall2018
Generator	clasdis
Generator Options	
After selecting the generator, check the documentation and paste the needed options above.	
Number of Events / Job	10000
Number of Jobs	1000
Total Number of Events	10 M
Output Options	☑ dst Warning: any of the choices below will enlarge the overall output size significantly. ☐ generator ☐ gemc ☐ gemc decoded ☐ reconstruction
	Submit

clasdis options

Stats Purposes

Type 1 and 2 corresponds to conditions of CLAS12 experiments that already ran:

clas12-default: nominal CLAS12 geometry
rga-spring2018: RGA spring 2018 configuration
rga-fall2018: RGA fall 2018 configuration
rgk-fall2018: RGK fall 2018 configuration (torus out-bending)
rga-spring2019: RGA spring 2019
rgb-spring2019: RGB spring 2019

- gemc
- ✓ clasdis
- dvcs
- disrad
- genKYandOnePion

Select N. events / job (max 10K)
Select Number of Jobs

Tag	RGA
Gcards	rga-fall2018
Generator	clasdis
Generator Options	
After selecting the generator, check the documentation and paste the needed options above.	
Number of Events / Job	10000
Number of Jobs	1000
Total Number of Events	10 M
Output Options	dst Warning: any of the choices below will enlarge the overall output size significantly. generator gemc gemc decoded reconstruction
	Submit

clasdis options

Stats Purposes

Type 1 and 2 corresponds to conditions of CLAS12 experiments that already ran:

clas12-default: nominal CLAS12 geometry
rga-spring2018: RGA spring 2018 configuration
rga-fall2018: RGA fall 2018 configuration
rgk-fall2018: RGK fall 2018 configuration (torus out-bending)
rga-spring2019: RGA spring 2019
rgb-spring2019: RGB spring 2019

gemc
✓ clasdis
dvcs
disrad
genKYandOnePion

Select N. events / job (max 10K)
Select Number of Jobs

Select Output
Any non-DST will limit NJOBS to 100 max

Type 2: user provided events in Lund format.

Tag	RGA
Gcards	rgb-spring2019
LUND File(s) Location	
If the LUND File(s) location is a directory, please make sure to end the address with a "/".	

One job will be produced per input file.

- Lund file(s) hosted on the web → Download file(s) to submit node
- Lund file(s) from /volatile → Copy file(s) to submit node

Note: Files copied or downloaded to **/volatile/clas12/osg/username/job_id** just before job is submitted.

Example: Submitting a file/directory of files

LUND File(s) Location	

`https://userweb.jlab.org/~ungaro/lund/`

Name	Last modified	Size	Description
Parent Directory		-	
dis_1.txt	21-May-2019 11:45	231K	
dis_2.txt	21-May-2019 11:45	231K	
dis_3.txt	21-May-2019 11:45	231K	
dis_4.txt	21-May-2019 11:45	231K	
dis_5.txt	21-May-2019 11:45	231K	
dis_6.txt	21-May-2019 11:45	231K	
dis_7.txt	21-May-2019 11:45	231K	

Output Organization

```
output/
   simu_1/
   simu_2/
   simu_3/
   simu_4/
   simu_5/
   simu_6/
   simu_7/
```

LUND File(s) Location	

`https://userweb.jlab.org/~ungaro/lund/dis_1.txt`

Creates one job

```
output/
   simu_1/
```

Off-site CLAS12 Simulations

Output

The output is saved on

`/volatile/clas12/osg/"username"/job_ID`

where "**username**" is your jlab account name and **ID** is the submission ID.

Individual "subjobs" are organized in subdirs named **simu_SJID**, where SJID is the subjob id.

Priority

There is no user priority. Jobs are submitted first come first serve.

Versions

The software versions are:
gemc: 4.3.1
Coatjava: 6b.3.2

We plan to support these and the upcoming versions.

If you submit 500 jobs, SJID will run from 0 to 499.

If the job id was 87:

job_87/output/

simu_0
simu_1
.
.
.
simu_499

You will see the submission file and logs within a couple of minutes of submission:

job_87/log

Web Submission Alpha (mini-stress) Tests

username	job_id	submitted	total	done	running	idle	hold	osg_id
avakian	80	11/6 12:50	1000	999	0	0	1	1422770
fxgirod	81	11/6 14:22	3000	2999	0	0	1	1423098
avakian	84	11/7 07:38	660	659	1	0	0	1427085
avakian	85	11/7 08:20	240	239	1	0	0	1427587
fxgirod	87	11/7 16:50	5000	4996	0	0	0	1430324
fxgirod	89	11/8 06:57	5000	4949	51	0	0	1432249
fxgirod	91	11/9 12:49	5000	12	4779	10	199	1438158

19900 jobs, 14853 completed, 10 idle, 4832 running, 201 held, updated on 2019-11-09 15:46:02

Examples:

Harut 84: `/volatile/clas12/avakian/mcdata/lund/eventfiles/`
FX 91: `/volatile/clas12/fxgirod/dvcs/gemc/lund2/`

Current record: > 4800 jobs running at the same time

Home About OSG Stats

CLAS12 Monte-Carlo Job Submission Portal

Logged in as ungaro

Project Summary

Zoom Out Last 10 days UTC

Bin size auto Project CLAS12 Filter +

? Batch vs Payload Jobs

Total Wall Hours

190 K

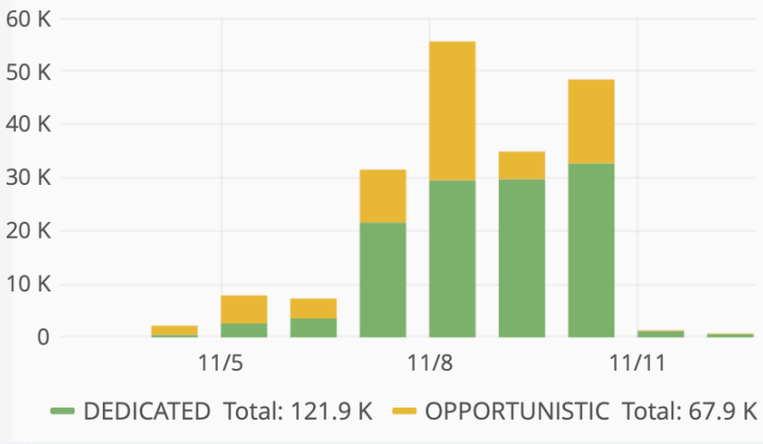
Total CPU Hours

126.9 K

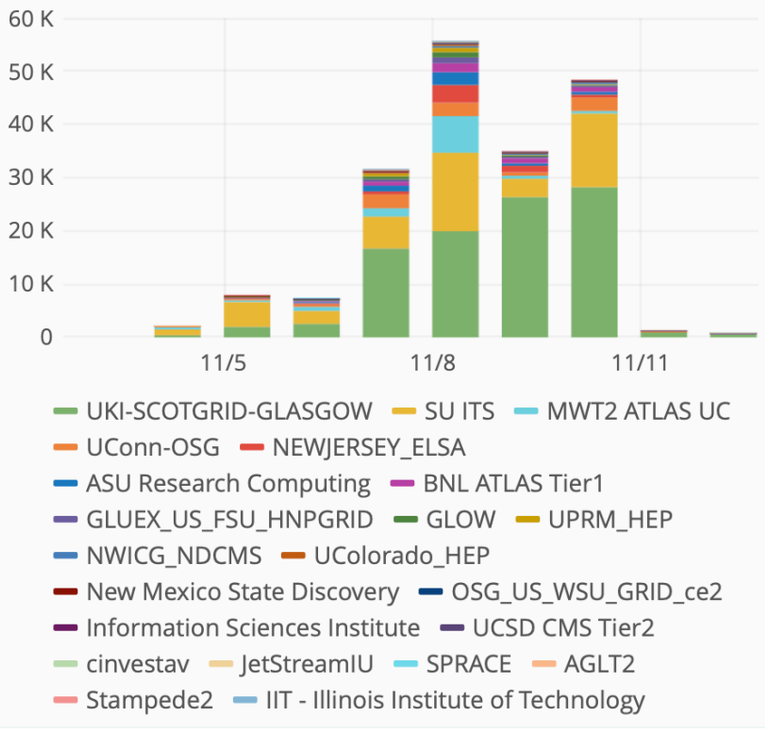
Total Jobs

29.3 K

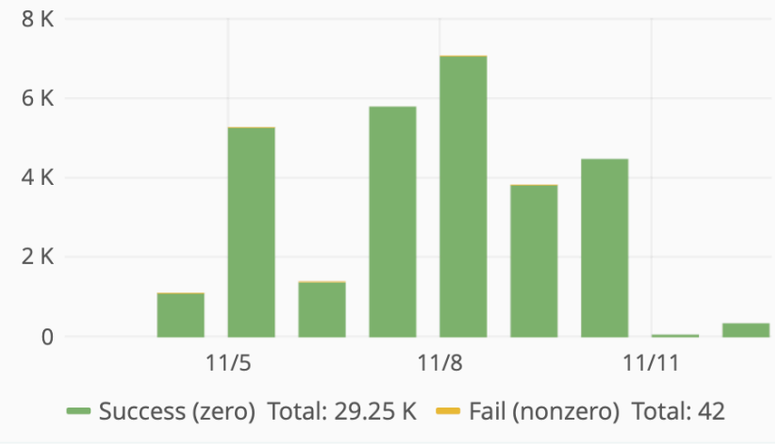
Wall Hours by Usage Model by 1d



Wall Hours by Site per 1d



Jobs by Exit Code per 1d



0.15% Fail:
No output came back

40% of Total CPU Hours went to Scotgrid

Upcoming

- Type 3 and 4: User custom gcards from web or /volatile

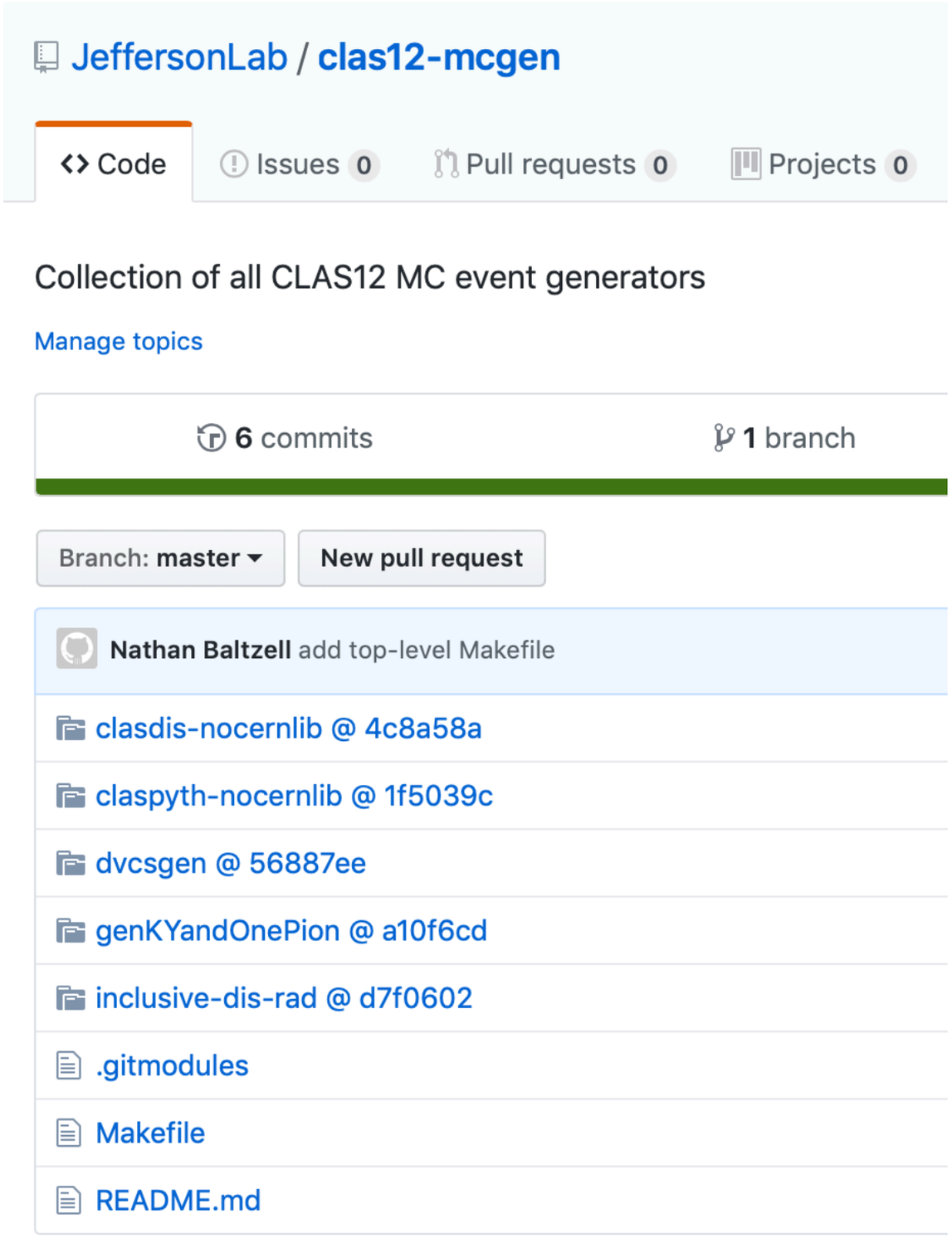
  This will allow custom configurations:
  - magnets currents
  - change detector configuration or
  - new detectors
  - etc

- Ability to cancel a job

- All types: add background from experimental data

- Email when job is completed

Generators in the container



Collected on github at

JeffersonLab/clas12-mcgen

Must:

1. compile in the docker container
2. be executable from any location

Can: make use of env variable to point to data.

Want:

1. Good Summary of Physics And Usage

2. HTML Table with the options, including documentation and default value.

Adding Institutions Computing Resources to this Pool

ScotGrid:
JLAB VO added (once and for all)
Mechanism to add Grid Institution resources to the CLAS12 pool exercised

Next Grid Resources:

- MIT computing farms (Tier2, possibly Tier3 and more)
- Worldwide LHC Computing Grid (WLCG), for example INFN

Phil Cole: adding Lamar University 40 CPUs (32 cores each).
Process / Requirements seems simple, even for farms not on the grid:

1. Install HTCondorCE (to make it look like a grid site)
2. Ssh key based access (one user)
3. Outgoing network access from the compute nodes
4. CVMFS installed in the compute nodes and in the head node

OSG Group (Edgar) + Farm Admins

If your institution wants to contribute please contact us!

Summary and Outlook

- Gemc 4.3.2 development ongoing
- Containers for interactive and offsite farms
- Portal For CLAS12 Simulation Job Submissions went alpha on Oct. 24, beta on Nov. 6. Ongoing development.
  - ~30K jobs submitted, 0.15% failure rate
  - We invite CLAS12 collaborators to start submitting jobs
  - Friday we'll submit jobs together

`https://gemc.jlab.org/web_interface`