My projects

E. Rasche

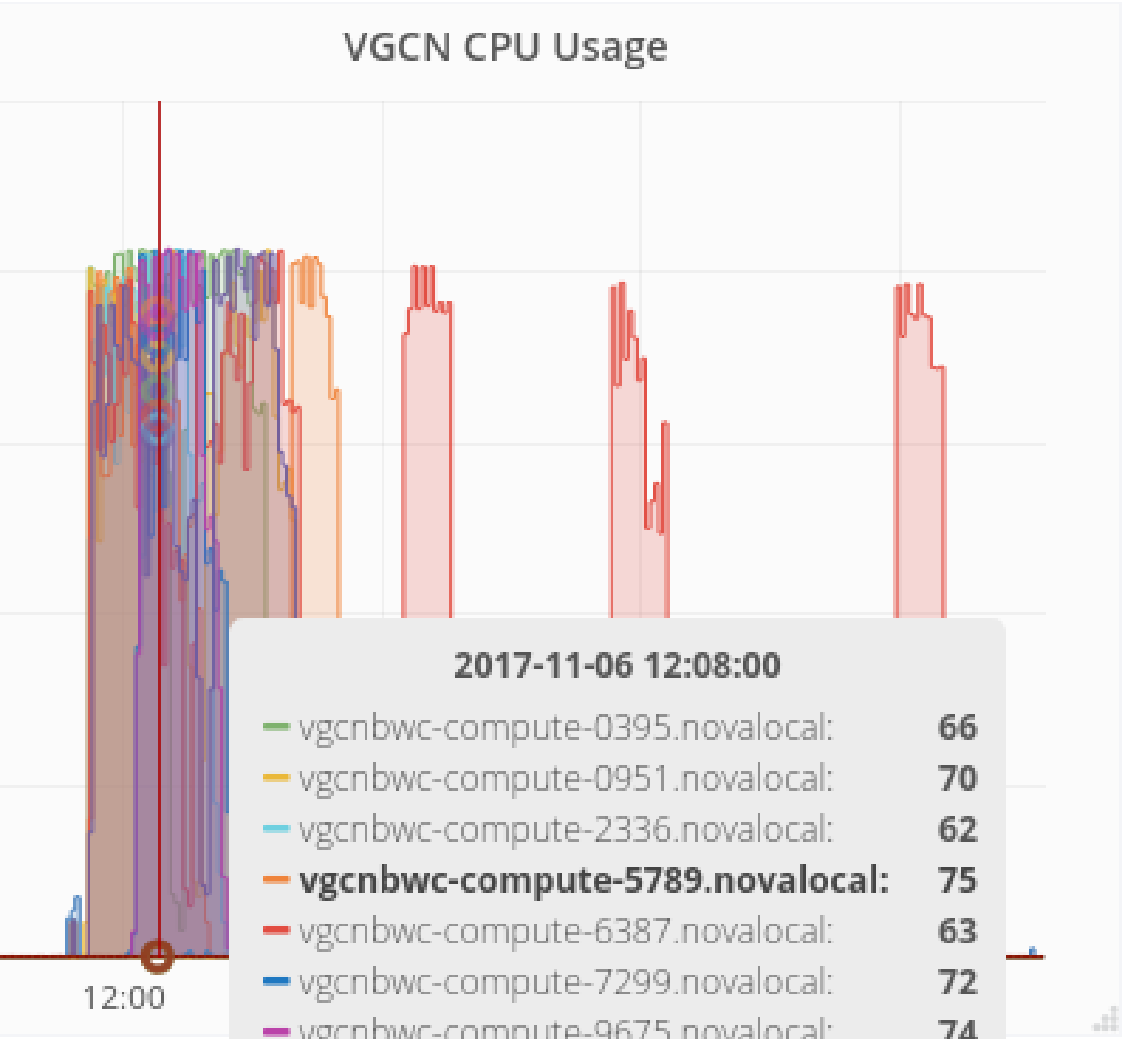
2017-11-09T09:00:00Z

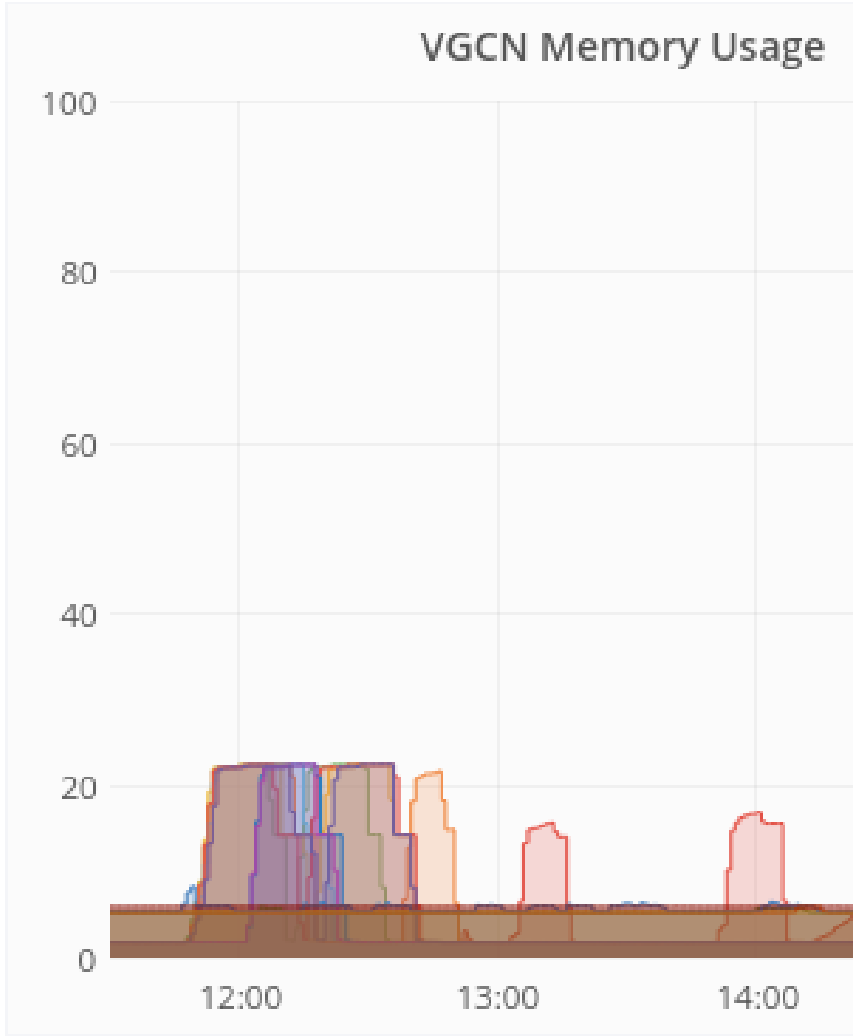
Completed Projects

Initial Monitoring

New Service(s)

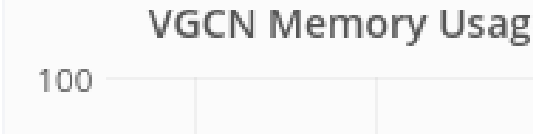
- Telegraf (collect) → InfluxDB (store) → Grafana (view)
- https://grafana.denbi.uni-freiburg.de



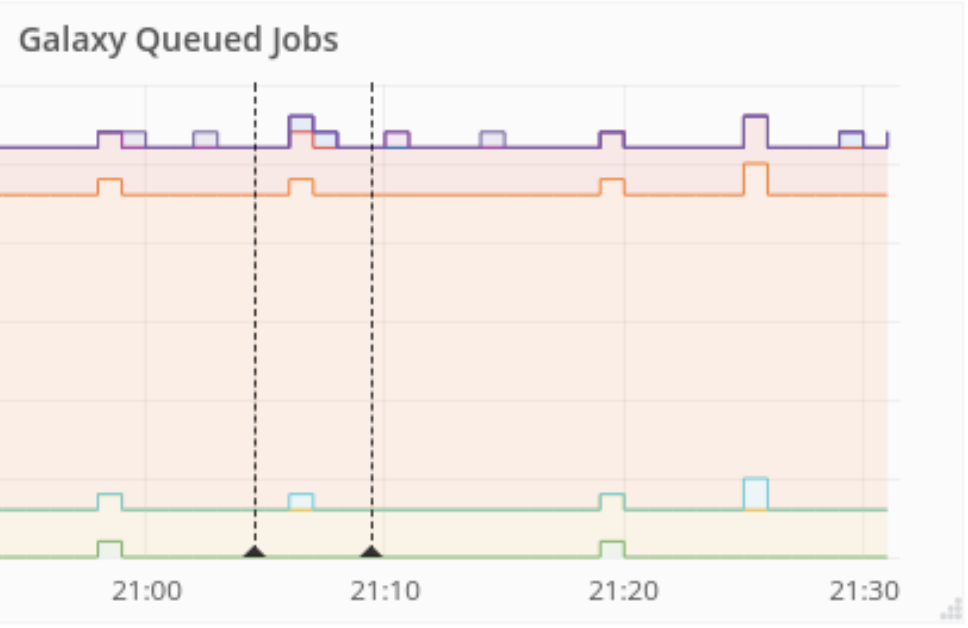


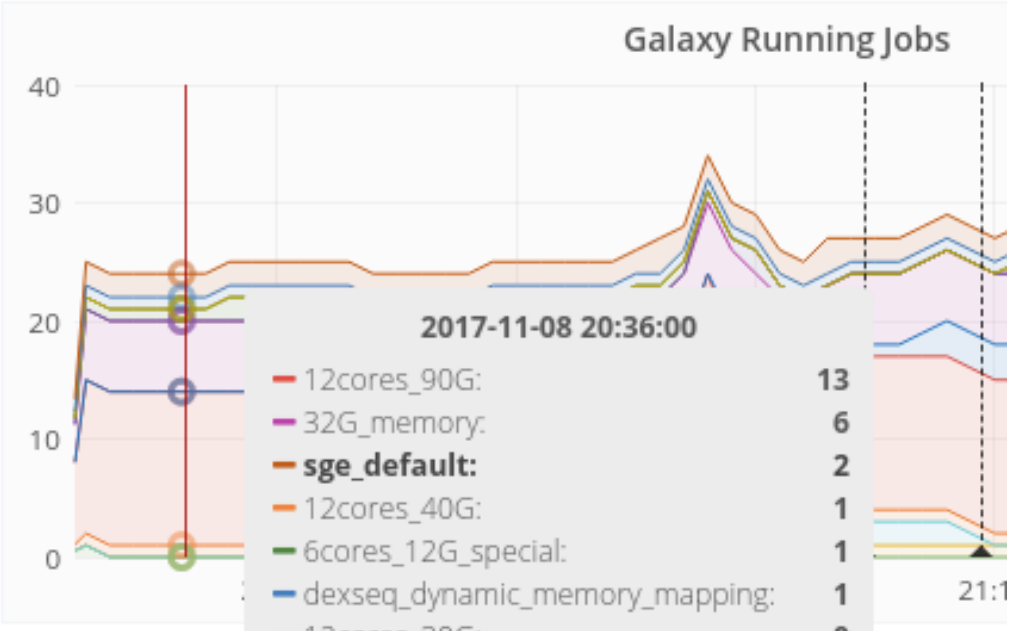
y Usage [Up

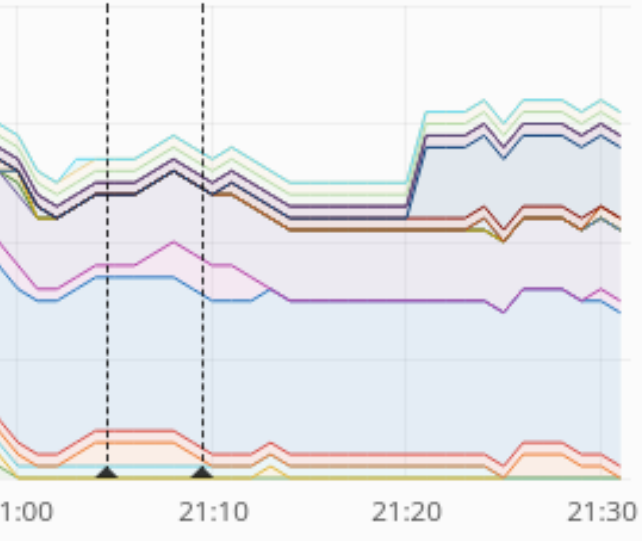
e [Metadata]

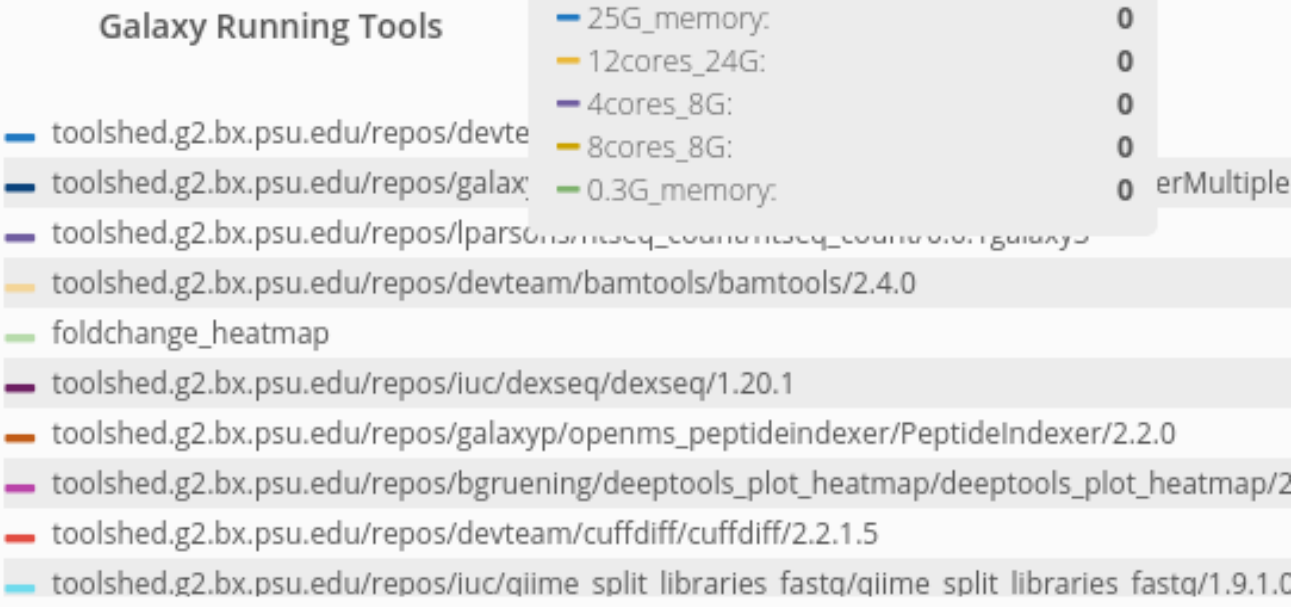


es


Galaxy Queued Jobs
21:00
21:10
21:20
21:30


Galaxy Running Jobs
40
30
20
10
0
2017-11-08 20:36:00
12cores_90G: 13
32G_memory: 6
sge_default: 2
12cores_40G: 1
6cores_12G_special: 1
dexseq_dynamic_memory_mapping: 1
12cores_30G: 0
25G_memory: 0
12cores_24G: 0
4cores_8G: 0
8cores_8G: 0
0.3G_memory: 0
21:1


1:00
21:10
21:20
21:30


Galaxy Running Tools
toolshed.g2.bx.psu.edu/repos/devte
toolshed.g2.bx.psu.edu/repos/galax
erMultiple
toolshed.g2.bx.psu.edu/repos/lparsons/htseq_count/htseq_count/0.6.1galaxy3
toolshed.g2.bx.psu.edu/repos/devteam/bamtools/bamtools/2.4.0
foldchange_heatmap
toolshed.g2.bx.psu.edu/repos/iuc/dexseq/dexseq/1.20.1
toolshed.g2.bx.psu.edu/repos/galaxyp/openms_peptideindexer/PeptideIndexer/2.2.0
toolshed.g2.bx.psu.edu/repos/bgruening/deeptools_plot_heatmap/deeptools_plot_heatmap/2
toolshed.g2.bx.psu.edu/repos/devteam/cuffdiff/cuffdiff/2.2.1.5
toolshed.g2.bx.psu.edu/repos/iuc/qiime_split_libraries_fastq/qiime_split_libraries_fastq/1.9.1.0

Initial Monitoring

. . . but how is it useful to you?

- How is a webservice we develop being used?
- How many requests did we receive?
- Is my service online + responding?
- How is my VM doing on memory/cpu/disk when I ran a new tool overnight?

Sentry

New Service

- Application (Galaxy, webservice, javascript site) exception collection
- Helps catch + fix bugs!
- Galaxy: stack traces, `log.warn`, `log.error`, `log.exception`
- Galaxy Frontend: javascript exceptions + browser info + user identity

D [redacted]

Chrome
Version: 61.0.3163

Windows 7
Version: Unknown

TAGS

browser	Chrome 61.0.3163
browser.name	Chrome
level	error
logger	javascript
os.name	Windows 7
transaction	/static/scripts/bundled/libs.bundled.js in n. iframeChangeHandler
url	https://galaxy.uni-freiburg.de/root
user	[redacted]

EXCEPTION (most recent call first) Full Raw

SecurityError

```
Blocked a frame with origin "https://galaxy.uni-freiburg.de" from accessing a cross-origin frame.
```

JS /static/scripts/bundled/libs.bundled.js in **n._iframeChangeHandler** at line **28:4931**

/static/scripts/bundled/libs.bundled.js in **HTMLIFrameElement.dispatch** at line **25:3616**

/static/scripts/bundled/libs.bundled.js in **HTMLIFrameElement.m.handle** at line **25:318**

/static/scripts/libs/raven.js in **HTMLIFrameElement.d** at line **1:4788**

BREADCRUMBS Search breadcrumbs...

Show 8 collapsed crumbs

ui.click	div.header > div.buttons > a#history-refresh-button > span.fa.fa-refresh	16:59:18
ui.click	div.header > div.buttons > a#history-refresh-button > span.fa.fa-refresh	16:59:18
ui.click	div.header > div.buttons > a#history-options-button > span.fa.fa-cog	16:59:18

FIRST SEEN

When: 2 days ago
October 30 2017 16:59:45 UTC

Release: *not configured*

LAST SEEN

When: 2 days ago
October 30 2017 16:59:45 UTC

Release: *not configured*

Tags

browser	100% Chrome 61.0.3163
browser.name	100% Chrome
level	100% error
logger	100% javascript
os.name	100% Windows 7
transaction	100% /static/scripts/bundle...
url	100% https://galaxy.uni-frei...
user	100% [redacted]@bi...

Notifications

level error | logger galaxy.model.metadata | server_name cn029.bi.uni-freiburg.de

MESSAGE

```
Failed to cleanup MetadataTempFile temp files from /data/dnb01/galaxy_db/job_working_directory/002/021/2021719/metadata_out_HistoryDatasetAssociation_4685734_IdczeA: No JSON object could be decoded
```

EXCEPTION (most recent call first)

Full | Raw

ValueError

```
No JSON object could be decoded
```

json/decoder.py in **raw_decode** at line **384**

```
379.
380.         """
381.         try:
382.             obj, end = self.scan_once(s, idx)
383.         except StopIteration:
384.             raise ValueError("No JSON object could be decoded")
385.         return obj, end
```

idx	0
s	''
self	<json.decoder.JSONDecoder object at 0x195fcd0>

json/decoder.py in **decode** at line **366**

json/__init__.py in **loads** at line **326**

json/__init__.py in **load** at line **278**

galaxy/model/metadata.py in **cleanup_from_JSON_dict_filename** at line **668**

LAST 30 DAYS

FIRST SEEN

When: 2 months ago

September 12 2017 18:00:07 UTC

Release: *not configured*

LAST SEEN

When: 2 days ago

October 30 2017 14:40:10 UTC

Release: *not configured*

Tags

level 100% error

logger 100% galaxy.model.metada...

server_name 100% cn029.bi.uni-freiburg...

1 Participant

Notifications

You're receiving updates because you have changed the status of this issue.

Unsubscribe

Sentry / Freiburg Galaxy

Issues Overview User Feedback Releases

Error /docs/errors.html in ?

IS
FREIBURG-GALAXY-

Script error for "mvc/tours" http://requirejs.org/docs/errors.html#scripterror

javascript

Resolve | Ignore | GitHub

Details Comments 0 User Feedback 0 Tags Related Events

Affected Users

Export to CSV

%	
20.00%	F [redacted]
40.00%	T [redacted]
40.00%	B [redacted]

Note: Percentage of issue is based on events seen in the last 7 days.

Sentry

. . . but how is it useful to you?

- When a javascript exception occurs in a user's browser and no one sees it, did it really happen?
- Do you read your server logs?
- Really? Every line?

Jenkins

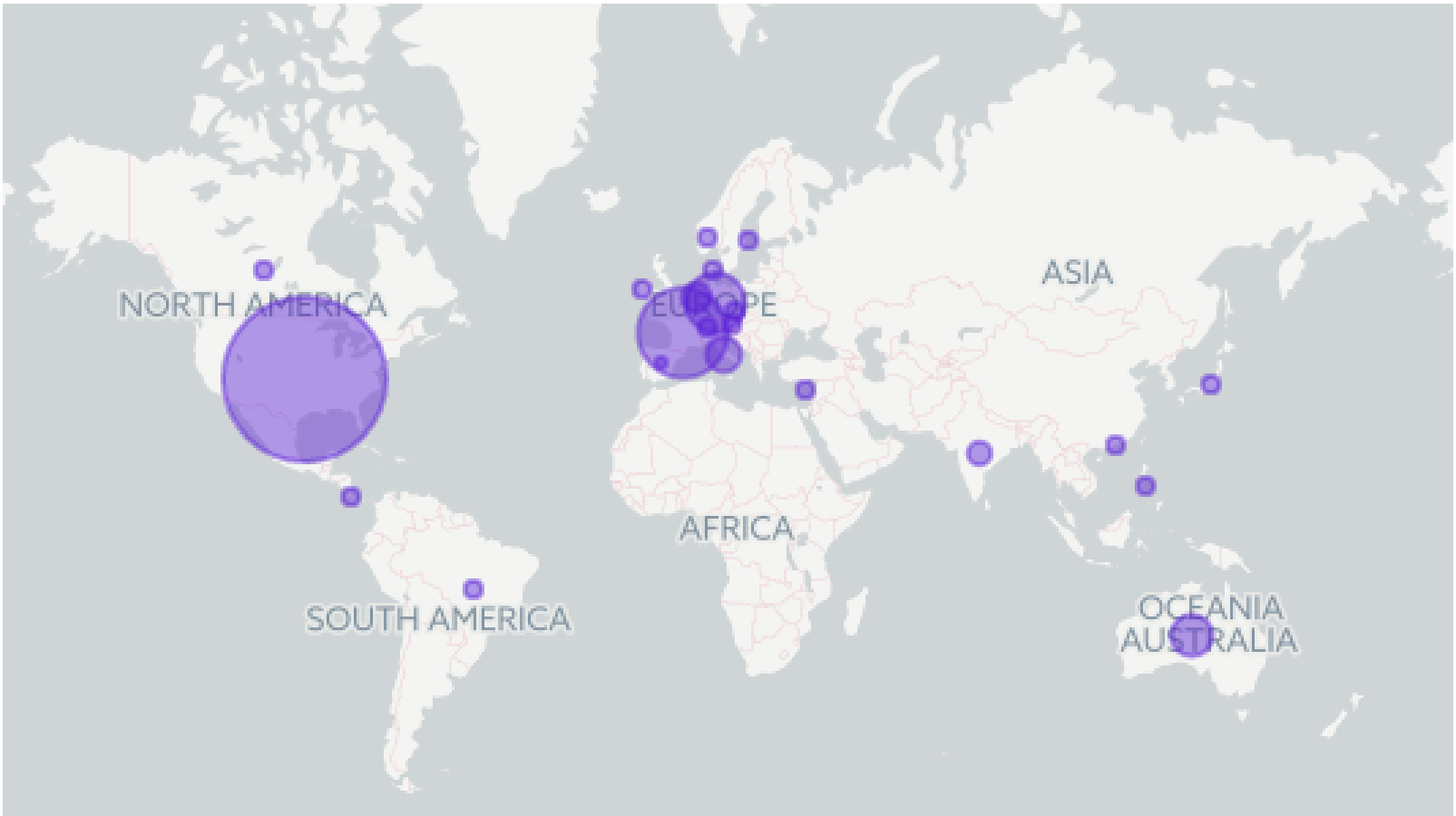
New Service

- Like **Travis-CI** but different
- Runs in the bwCloud
- Useful if:
  - Jobs need to run really long
  - Code is extremely private/sensitive
  - Jobs produce artifacts that must be stored and available to download (e.g. PDF for course slides)

public-galaxy-servers

- CSV file of public galaxy servers
- Simple script to contact server + record configuration
- Uptime badges!

- Cool maps!



Galactic Radio Telescope

- Part of Galaxy 17.09
- Efficiently collects basic metrics (job, state, runtime, input sizes) from participating Galaxies
- A later update will add more metrics
- Currently collecting from UFR Galaxy
- https://telescope.galaxyproject.org

Job Conf Migration

- Job configuration migrated from XML to python
- Job conf now under testing (Jenkins)
- Enables dynamically choosing where to send jobs
- Highlighted use case: Training-as-a-Service. Users in specific roles can be directed to dedicated VMs.

Upload + Metadata VMs

- Leverages our bwCloud Condor Cluster
- "Upload" VMs only handle upload jobs, "metadata" VMs are similar.
- All uncompression of uploaded files is done off of `cn029`
- *Fewer* random slow downs

yaml linting

- pykwalify is great
- Use it.

```
---
type: map
mapping:
  "=":
    type: map
    mapping:
        "name":
            type: str
        "tmp":
            type: str
            enum: ["large"]
        "cores":
            type: int
        "mem":
            type: int
        "runner":
            type: str
            enum: ["condor", "sge"]
        "env":
            type: map
            mapping:
              "=":
                  type: text
        "nativeSpecExtra":
            type: str
        "priority":
            type: int
            range:
                min: 0
```

Ansibilization of Galaxy Configuration

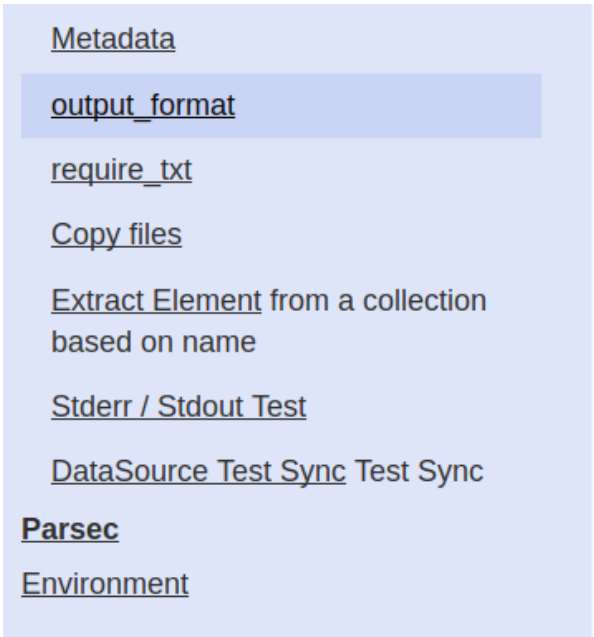
- Major Galaxy configuration files now as some custom playbooks
- Not the entirety of galaxy
- Playbooks automatically run every hour
- GitHub enables reviewing + "auditing" of configuration changes
- *Nearly* no more SSHing

Bug Reporting Plugins

- Bug reporting as “plugins”
- Plugins for: email, sentry, GitHub, influxdb, logfile
- Easy to develop new ones
- Control over automated submission without user interaction (e.g. sentry vs email)
- galaxyproject/galaxy#4305

Bigger Click Targets

- galaxyproject/galaxy#4470
- I hate small click targets
- Be the change you want to see in the world



Zerg Mode

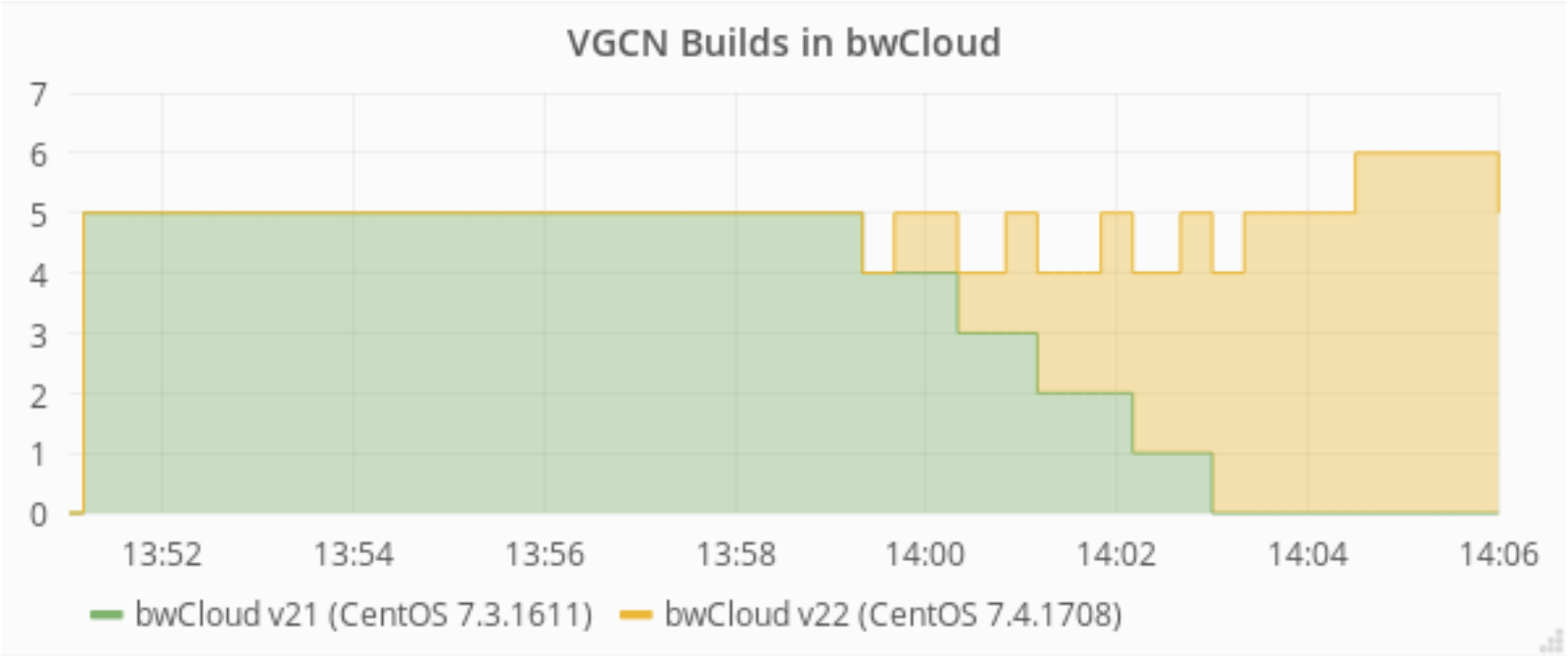
- *Zero-downtime restarts of Galaxy*
- Configuration worked out by Nate Coraor (@natefoo)
- It is super effective!
- We annotate the restart events in Grafana if you are curious when it happens

VGCN Images

- *Cluster in the Cloud*
- usegalaxy-eu/vgcn
- Pre-configure Condor Master + Executor VM images
- Converted existing playbooks to reusable roles
- Integrated work from RZ for simplified build process
- Now regularly building + deploying images

VGCN Infrastructure

- usegalaxy-eu/vgcn-infrastructure
- Regularly deploying new images sucked
- Rolling restarts are necessary
- Infrastructure-as-code (e.g. Terraform/CloudFormation/Heat)



bwCloud

- Temporary maintainer
- Full time employee replaces me (yay!)
- No major improvements made, some networking issues triaged and bandaids applied around libvirtd

Securing Galaxy

- Tracking security issues in emails is not scalable
- We now have a git repo for this
- Testing setup is WIP

Galaxy subprocess hardening

- galaxyproject/galaxy#4604
- **BAD**

```
subprocess.call("ls /some/path", shell=True)
```

- **GOOD**

```
subprocess.call(["ls", "/some/path"])
```

- Missed some :(

Reaching the promised LAN

- galaxyproject/galaxy#4289
- Galaxy's upload lets you paste in URLs
- Even URLs inside your super secure network (e.g. `http://192.168.0.1:8080`)
- Galaxy can be used to probe infrastructure and possibly download restricted material
- Now administrators must whitelist local network resources which should be accessible.

Ongoing Projects

Projects

E. Rasche

Completed
- Monit/Infra
- Freiburg Galaxy
- Galaxy Project
- Performance
- Cloud
- Security

Ongoing
- Freiburg Galaxy
- Galaxy Project
- Performance
- Cloud
- Security

Upcoming
- Monitoring
- Freiburg Galaxy
- Galaxy Project
- Performance
- Cloud
- Security

usegalaxy.eu rollout

Done:

- Many uptime and performance enhancements in place
- Monitoring continues to help identify issues

TODO:

- Logo
- Branding things (UFR downplayed, usegalaxy.eu everywhere)
- English and grammar checks of all written materials
- Terms of Service (SLA, SLO) + Privacy Policy
- Communications Materials
- User-facing mailing list, plan out mails

Apollo

Background:

- We'll be deploying the Apollo Genome Browser alongside `usegalaxy.eu`
- Apollo has horrifying performance in some parts
- UI is miserable
- Underlying code is not easily fixed

TODO:

- Working on splitting apart frontend + backend
- Collaborating with some other "big" apollo-galaxy sites

JBrowse Chat Plugin

Background:

- Apollo is great for collaborative, real-time editing
- . . . but it feels very lonely since you cannot communicate

Done:

- Proof-of-concept implementation works OK

TODO:

- Needs more authentication options
- Needs archiving / chat log database

Production Services

Background:

- Increasing number services that become slowly critical
- SLAs? SLOs?

Done:

- Ansible roles for deploying infrastructure (Grafana, InfluxDB, Sentry, public-galaxy-servers)

TODO:

- SLO: *Must* be able to be redeployed quickly ($<$10 min)
- Backups

Advanced Job Conf

Background:

- "New" job conf is nice
- But can only be reconfigured by restarting handlers and waiting 5 min to 10 min

TODO:

- Force it into a secondary process which can be re-deployed with zero downtime
- Need to choose a method + test (unix sockets, load balanced web service, `SO_REUSEPORT`, etc.)
- SLO: $\leq \frac{1}{100.000}$ jobs dropped in networked version of the job destination service

User Filters

Background:

- Too many tools in the toolbox
- Too few are relevant to specific users
- With user toolbox filters users can "build" their own toolbox

TODO:

- Currently broken, will be fixed real soon now™

pg_tools

Background:

- (Re)discovered script `pglite`
- Will provide *real* PostgreSQL databases during tool execution
- Useful for chemiinformatics + genome annotation

Done:

- First set of tools being built
- Datatypes in Galaxy

TODO:

- GMOD tool suite
- chado-schema-builder brought back to life

Mirroring Galaxy Repos

Background:

- Some devs worry: "What if GitHub goes away?"
- Plan is to mirror all of galaxyproject's repos

Done:

- Simple mirror of Galaxy codebase

TODO:

- Mirror all other repos
- Update mirrors regularly (i.e. Jenkins job)
- Recently discovered `git-appraise`. Should permit archiving PRs + discussion (+ issues?) all within portable git repo.

deNBI Cloud

- Adding Elixir AAI authentication to deNBI Cloud @ Freiburg
- Unlike any other cloud, so we have special issues

Have had this conversation numerous times:

Them: So... why isn't it working yet? Why are you so slow at this?

Me: [Explains setup]

Them: Oh. Ohh yeah... ok. *[Nervous laughter]* Good luck.

Automating VGCN Image Building

Completed
Monit/Infra
Freiburg Galaxy
Galaxy Project
Performance
Cloud
Security

Ongoing
Freiburg Galaxy
Galaxy Project
Performance
Cloud
Security

Upcoming
Monitoring
Freiburg Galaxy
Galaxy Project
Performance
Cloud
Security

Background:

- Building VGCN images by hand is time consuming
- Builds are quite reliable, unlikely to produce a broken VM
- So let's make Jenkins build them on VGCN playbook changes

TODO:

- Builds in bwCloud time out when Jenkins runs them
- But they'll build fine in SJ's "imager" cloud VM
- ???

Content Security Policy (CSP)

Background:

- CSP helps prevent huge swaths of security issues
- Can prevent browsers from executing unexpected scripts, CSS, media, frames, etc.

Done:

- Currently in "reporting" mode, collecting any infractions

TODO:

- Need to analyze infractions, identify real issues (vs noise)
- Need to cleanup galaxy, removing inlined styles, scripts
- Will move to "enforcing" mode after issues are patched

Cron jobs logged to Jenkins

Background:

- Ansibilization lets us discover all manner of hidden jobs
- Let's push the logs to a central location for auditing
- Jenkins can "monitor" external jobs
- Public visibility $\Rightarrow$ forced accountability

Done:

- All cron tasks from `galaxy@cn029`

TODO:

- Identify cron jobs on other hosts, VMs, etc.

Upcoming Projects

Still in planning phase

Testing Workflow execution

Question: Is Galaxy working as expected?

Proposed Solution:

- Regular, automated testing of our Galaxy installation
- Commonly executed workflows will give good idea
- Can cover jobs going to condor + sge + odd configuration we might have

More monitoring

Question: Why is everything still slow?

Proposed Solution:

- Upload jobs should be timed (time to upload, job execution, green history element)
- More job timings available in the database
- Need to track all of this
- In general, are things getting better or worse for users?
- Is user-facing infra more or less reliable?

Caching Framework

Question: How can we keep Galaxy from re-building the same tools, the same JSON blobs *ad nauseam*?

Proposed Solution:

- Provide a generic interface in codebase for caching data
- Provide methods for expiring the cache as needed
- Will support in-memory, on-disk, and networked caches (memcached)
- Will convert subset of slow computations to use new cache

Galaxy OAuth2

Question: How can I grant partial access to my account to a third party without violating every security policy ever?

Proposed Solution:

- Implement *OAuth2* over the Galaxy API
- API keys with *scopes* that can be granted
- Third parties can build things on top of Galaxy, new UIs, new (limited) tools
- Users feel safe granting privileges

User Generated Content

Question: How can admins be 100% absolutely sure that user generated content is safe?

Proposed Solution:

- Host user-content on a second domain (i.e. HTML render of datasets)
- Just like GitHub (`githubusercontent.com`)
- Will allow us to ignore / disable whitelisting of "safe" tools
- Token-based access to datasets instead of cookie-based

Secrets Server

Question: How can we securely store user-supplied secrets in the safest way possible?

Proposed Solution:

- Galaxy should not be responsible for storing super secret secrets (e.g. API keys to third party services)
- Hashicorp's Secrets Server is *literally built for this and only this*
- Integration paves way for tools with easier access to remote services (i.e. Apollo suite)

Medical / DFARS

Question: How can users with extreme data security requirements use Galaxy?

Proposed Solution:

- Allow users to mark datasets, libraries, entire histories as "secure"
- Any datasets produced from these are also tagged "secure" and treated accordingly.
- For datasets marked as "secure":
  - Must securely wipe (DoD standard 32-pass) datasets and temporary files
  - Data must be encrypted in transit + at rest (requires Secrets Server)