VERA C. RUBIN OBSERVATORY

Phalanx state of the union

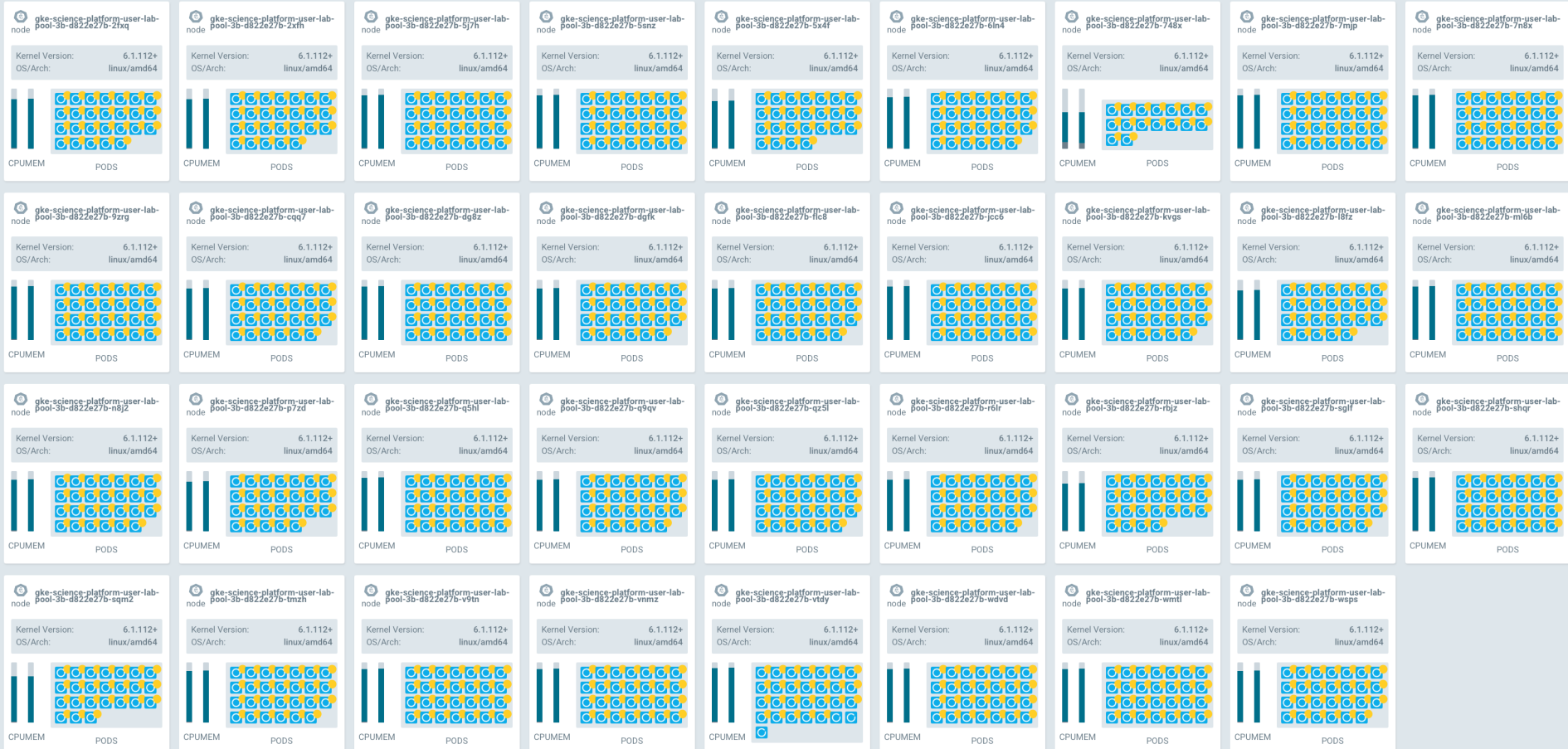
Frossie Economou for SQuaRE

NSF AURA U.S. DEPARTMENT OF ENERGY SLAC CHARLES AND LISA SIMONYI FUND FOR ARTS AND SCIENCES LSST CORPORATION

Scale Friday

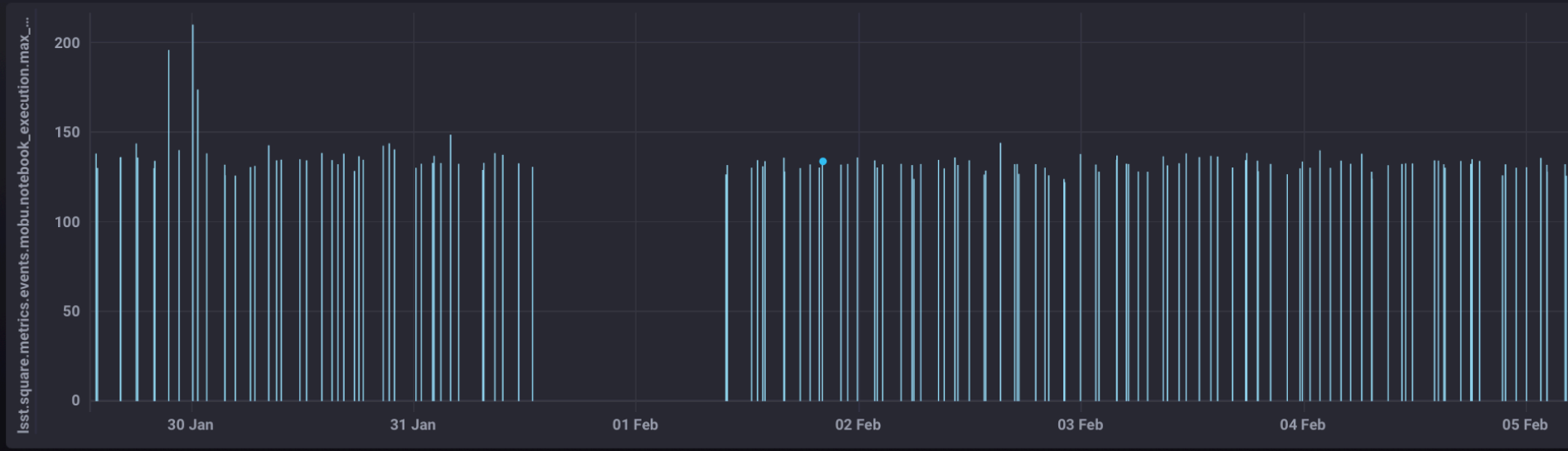
- Every Friday 12:30 PT until it stops being fun (or useful I suppose)
- We're finding lots of concurrent load issues and fixing them
- Highest successful test so far was 3K users (mobu emptyloop), nublado 1K (empty notebook), SIAv2 over Butler C/S 100, Qserv 20
- Results are reported on #rsp-team on Wednesday SupBot
- Running on Cloud
- Indirectly hits USDF
- USDF is not particularly special hybrid-wise, could hit eg. Qserv (even Butler C/S) at UK/FR if network access allows it with a bit of set-up and co-ordination
- You could run on-prem scaling tests using your on-prem mobu but do not recommend (for example mobu users have to be "hand-provisioned" at USDF)

This is what 1,000 users looks like



Metrics infrastructure (Safir/Sasquatch)

- Helpers for publishing service metrics to Sasquatch now in Safir
- Metrics are Not [telemetry, monitoring, logs] (rant on request) - high-value, intentionality, context
- Docs: https://safir.lsst.io/user-guide/metrics/index.html
- Instrumenting services in progress but Gafaelfawr Knows All



Rate-limiting infrastructure

- Did I mention Gafaelfawr Knows All and gives us a “cheap” entry point
- Architecture is https://sqr-073.lsst.io/
- We have a start, docs: https://gafaelfawr.lsst.io/user-guide/quotas.html
- Fairly coarse API quotas right now (eg “user gets max 1000 Datalinker requests in 15 minutes”), might be enough to protect service during DP1 against DDOS levels activity, we’ll see
- Setting API quota to 0 = administrative block (blunt hammer, user gets 403)
- Quota grants per group name
- Dynamic overrides via Gafaelfawr API

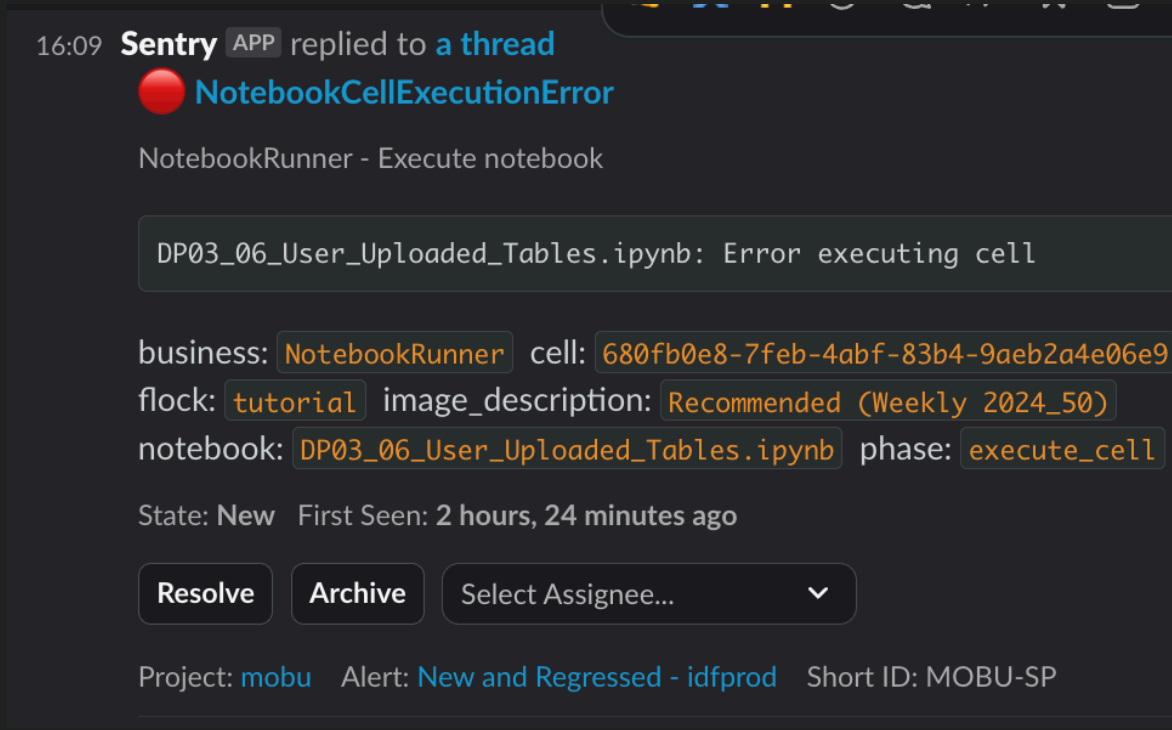
```
config:
  quota:
    default:
      api:
        datalinker: 1000
      notebook:
        cpu: 2.0
        memory: 4.0
```

Kubernetes bump (very soon)

- We do use newer k8s features and are opinionated about what version we run
- Eg we want to use v1.29 for native k8s support for sidecar containers on Cloud and would like sites to target that

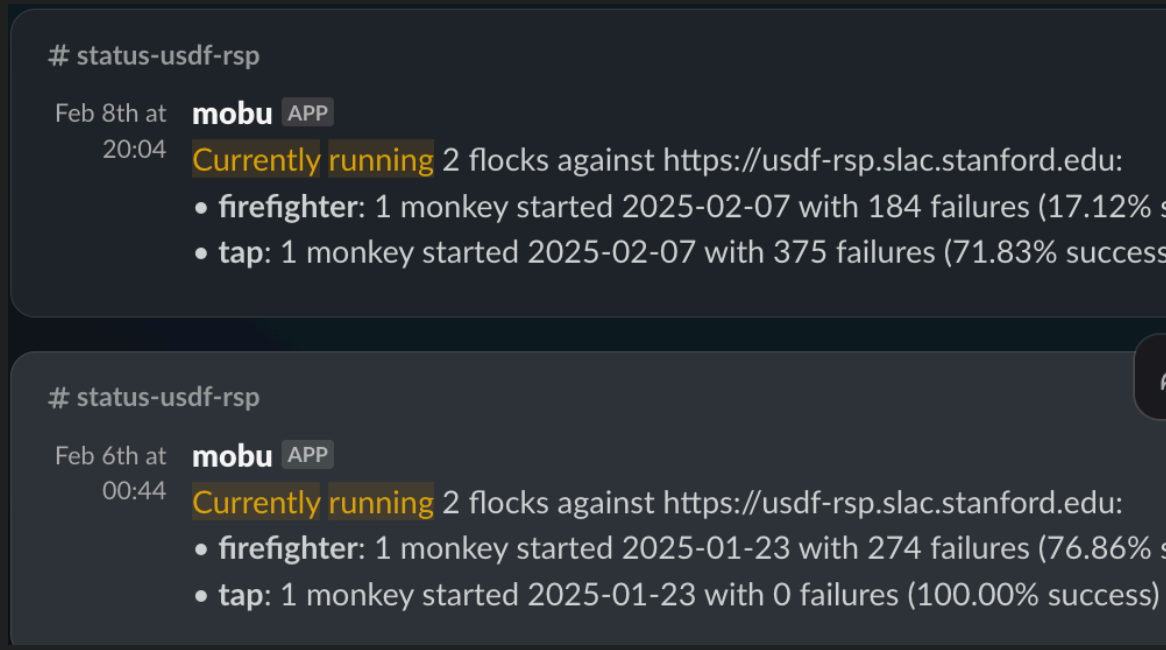
Alert infrastructure

- Sentry for our Cloud clusters (on-prem too but we can quickly exhaust quota)
- Class to abstract alert generation (slack, Sentry, logs etc) to avoid baking in dependance to Sentry and retaining flexibility planned (not imminent)
- Mobu slack alerts on cloud now using Sentry API
- You do need a Sentry account to click on the link but there is enough info for reproducibility



Speaking of mobu

- Mobu under-used on-prem (only USDF runs it, not clear it's being used to monitor services consistently)
- Over-reliance on tutorial notebooks
- More/better system-test coming
- Roadmap at sqr-080.lsst.io (halfway implemented, complete soon)
- Can be noisy (we'll work on non-fatal intermittent errors "mañana") but daily summary very information rich



Separate sub-domains for apps (soon)

- Design in: https://dmtn-193.lsst.io/
- Part of our security posture eg data.lsst.cloud/portal -> portal.data.lsst.cloud
- Timeline for DP1 aggressive, but it’s coming

Service discovery

- Planned but will come
- Not just VO Registry
- Also will allow us to get rid of Segwardes

Wobbly

- We have consolidated all our UWS databases into one
- New service Wobbly (github.com/lsst-sqre/wobbly/) to manage that database via REST API without knowledge of the underlying schema
- Design in sqr-096.lsst.io
- UWS-based services will be increasing

Leveraging DF compute - awooga

- More on this at 4pm
- But basically we (RSP services, Jenkins too?) need to be able to run pipeline workloads at USDF soon or else we can’t have nice things or bulk cutouts etc