ALICE Build Infrastructure

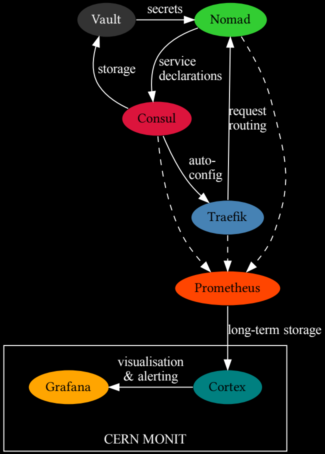
CernVM Workshop, Nikhef

Timo Wilken

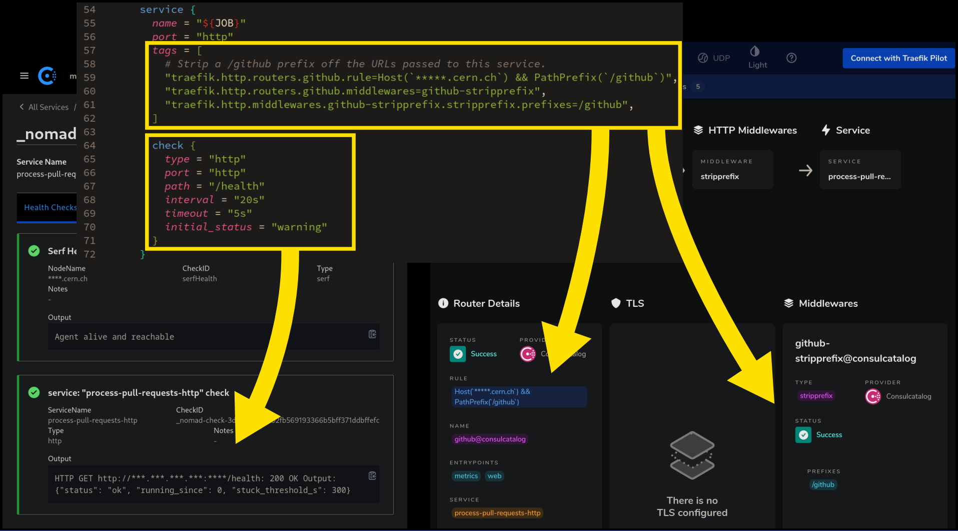
12 September 2022

Architecture overview

- previously: Mesos + Aurora
- Nomad for job scheduling
  - long-running jobs: custom continuous integration builders, Jenkins builders
  - small web services: user account administration, tarball servers
  - cron jobs: software repository maintenance/cleanup
- Consul
  - job discovery: `*.service.consul` DNS
  - Traefik auto-config for web access
  - job monitoring: simple health checks
- Vault stores secrets, using Consul as backend
- Prometheus metrics of the whole cluster monitored and visualised using Grafana



Web services: health checks & Traefik autoconfig



```
service {
  name = "${JOB}"
  port = "http"
  tags = [
    # Strip a /github prefix off the URLs passed to this service.
    "traefik.http.routers.github.rule=Host(`*****.cern.ch`) && PathPrefix(`/github`)",
    "traefik.http.routers.github.middlewares=github-stripprefix",
    "traefik.http.middlewares.github-stripprefix.stripprefix.prefixes=/github",
  ]
  check {
    type = "http"
    port = "http"
    path = "/health"
    interval = "20s"
    timeout = "5s"
    initial_status = "warning"
  }
}
```

Serf He

NodeName
****.cern.ch

CheckID
serfHealth

Type
serf

Notes
-

Output

```
Agent alive and reachable
```

service: "process-pull-requests-http" check

ServiceName
process-pull-requests-http

CheckID
_nomad-check-3c 2fb569193366b5bff371ddbffefc

Type
http

Notes
-

Output

```
HTTP GET http://***.***.***.***:****/health: 200 OK Output:
{"status": "ok", "running_since": 0, "stuck_threshold_s": 300}
```

STATUS
Success

PROVID
Co alog

RULE
Host(`*****.cern.ch`) && PathPrefix(`/github`)

NAME
github@consulcatalog

ENTRYPOINTS
metrics web

SERVICE
process-pull-requests-http

There is no TLS configured

github-stripprefix@consulcatalog

TYPE
stripprefix

PROVIDER
Consulcatalog

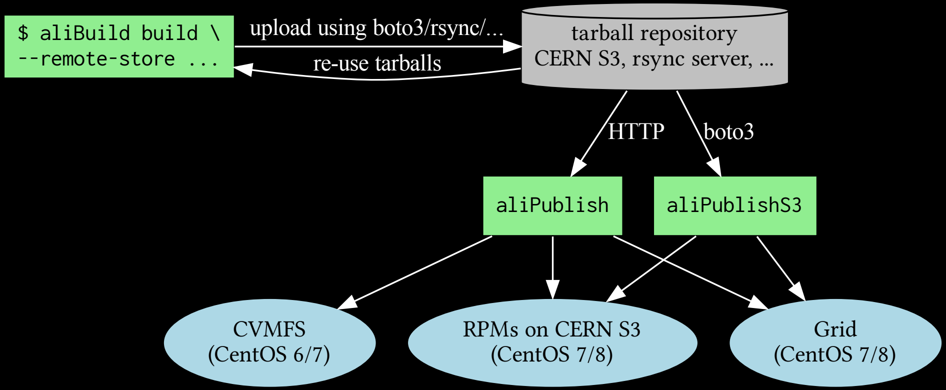
STATUS
Success

PREFIXES
/github

Rough edges

1. Nomad's handling of disk space allocation
   - restarting daemon with non-empty disk confuses Nomad's accounting
   - can cause scheduling issues much further down the line
   - must manually clean up the node and restart the Nomad agent process
2. integration with CERN SSO
   - rely on Nomad/Consul/Vault tokens for authentication
   - could integrate SSO with Vault, which would then issue Nomad/Consul tokens
   - client certificate authentication is supported, so we use that in addition to tokens

Software publishing

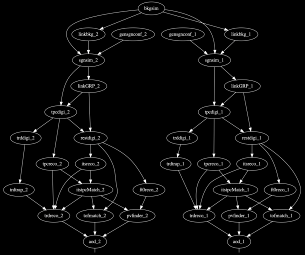


```
$ aliBuild build \
--remote-store ...
```

GRID processing at ALICE (adapted from S. Wenzel)

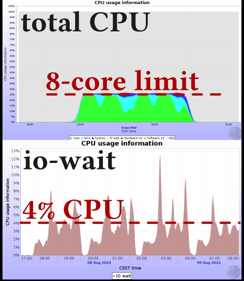
- micro-services (processes) transform streaming data
- offline/GRID: limited memory (2 GB/core)
  - cannot have all services up at the same time
  - need a runtime for scheduling and consecutive steering of processing stages
- runtime based on a directed acyclic graph
  - processes started when input + resources available
    - data communicated via intermediate files
  - multi-core ready: multiple parallel processes + internal multithreading; parallel in stages (sim, reco, etc)

⇒ multiple processes loaded from storage, potentially in many repetitive cycles



Offline multiprocess execution: sample of a graph

CVMFS feedback from GRID processing (adapted from S. Wenzel)



Significant io-wait load when processes are loaded from CVMFS

- generally good multi-core usage and efficiency; very nice from local disk
- we have multiple parallel GRID jobs accessing CVMFS simultaneously
- improvement: increase CVMFS caching size, e.g. 4 ⟶ 50 GB
- interested in further collaboration with CernVM team on adaptation and tuning for our use-case
- in parallel, working on internal solutions to improve performance, such as:
  - structural changes to graph layout
  - optimising process granularity or sequencing (for better cache behaviour)
  - RPath optimisations

Questions?

Links

- `aliBuild`: `https://alisw.github.io/alibuild/`
- CI & ALICE software documentation: `https://alisw.github.io/`

Contact details

- timo.wilken@cern.ch
- GRID processing: sandro.wenzel@cern.ch

Backup: ALICE's Continuous Integration system

- custom build tool (`aliBuild`) and software distribution
  - suited to local development: native incremental build support
- pull requests against software projects are tested
  - built against full software suite
- nightly builds on Jenkins produce reusable tarballs
  - speed up CI, local builds, …
- repack tarballs and publish to different backends
  - CVMFS, RPMs, Grid

Backup: Reusable tarballs

- hashing of build metadata to uniquely identify builds
- synchronise with repository to assign sequential version numbers
  - version numbers used for published packages: packages built later from the same git tag under different conditions get "later" version numbers
- large speed-ups for:
  - continuous integration: no need to rebuild full software stack for each pull request
  - local builds, e.g. for physics analysis: rebuild only what is needed
  - published packages: re-use previous days' dependencies, which rarely change