



Dirac Grid

DIRAC Users' Workshop

June 19th, 2024

<https://indico.cern.ch/e/duw10>

Federico Stagni

Technical coordinator

On behalf of the DIRAC consortium

Welcome back!



- We were in this same amphitheatre in May 2018 for the 8th DIRAC Users' Workshop, in a larger group wrt today's
- This time, "only" 20-something registered "users"
 - More "Experts" than "Users"
 - Looks like this is going to be an extended BiLD meeting, followed by a hackathon
 - So, we can probably put more "work" in this "shop"
- Why so few? (usually we were around ~40 registrations)
 - (too) close to the previous workshop (Oct 23, KEK)
 - Post-COVID: more selective travel
 - But many other conferences/workshops restarted

Rules



- This is by all means an informal workshop
 - We are here to get things done
- There's a zoom room but we do not attend it attentively
- You can intervene at any time, no microphone needed, just raise your hand and speak up!
 - For questions, comments, critics, even jokes
- We do not have to absolutely stick presentations in the allocated time
- If you want to note something down for later, do the appropriate thing:
 - Make an issue in github, or
 - Make a github discussion, or
 - Add it to this note: <https://cern.ch/duw24note> (login with github)
- Nobody is allowed to eat all the macaroons

Q&A



- Why the "10th" Dirac Users' Workshop?
 - This the 10th in-person workshop, the 9th was in 2019 in London
 - 2020: ~~KEK, Tsukuba~~ --> CANCELED
 - 2021: Virtual WS
 - 2022: 2nd Virtual WS
 - 2023: DIRAC & Rucio WS, KEK, Tsukuba
- Will there be other DIRAC&Rucio Workshops?
 - Possible, but not planned
- Is DIRAC the "Distributed Infrastructure with Remote Agent Control"?
 - Nooooo! Please stop using that!

...at the WS in KEK



DIRAC issues

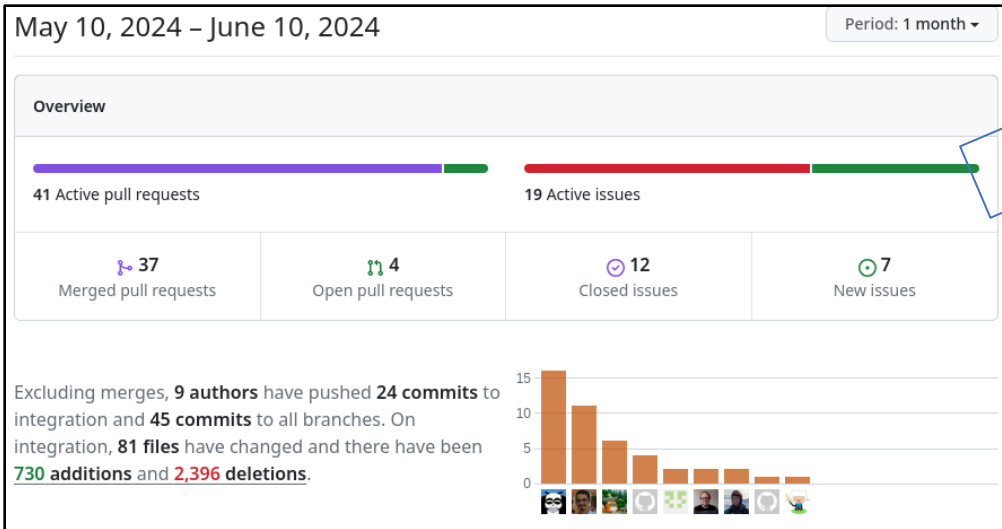
- complex, with high entrance bar
 - got better dropping python2 compatibility
- somewhat cumbersome deployment
 - got better dropping python2 compatibility
- late on “standards”
 - http services
 - tokens
 - monitoring
- “old”-ish design (RPC, “cron” agents...)
- not very developer-friendly
 - rather un-appealing/confusing, especially for new (and young) developers
- multi-VO, but was not designed to do so since the beginning
- no clear interface to a running DIRAC instance



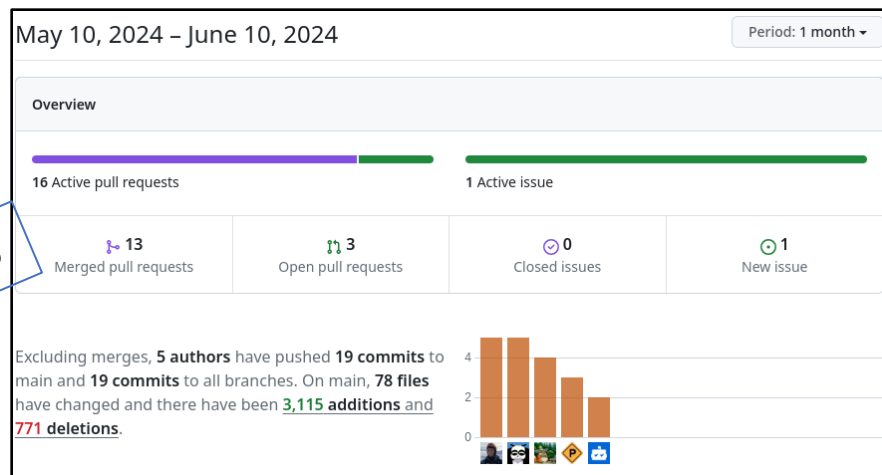
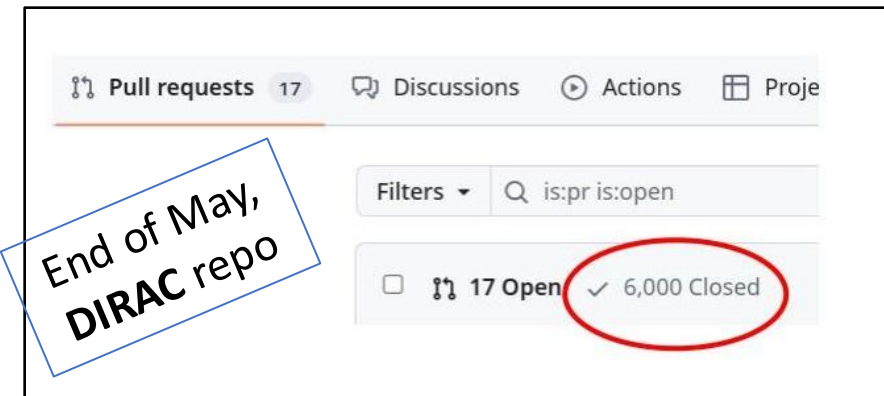
DiracX in just one slide

- The neXt DIRAC incarnation
 - DIRAC functionalities will move there
- A cloud native app
- Multi-VO from the get-go
- Standards-based
- Not a framework

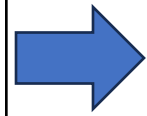
Where we are (visually)



DIRAC repo



DiracX repo



For the actual content: see next presentation

EL7 EOL

EL7 EOL: what is **DONE**



- runsv RPM was rebuilt
- CI container image (e.g. used by DIRAC's integration tests) was moved to Alma9
- Puppet profiles have been updated
- VOMS-admin does not have an officially tested EL9 rpm (also: not compatible with MySQL 8)
 - [PR](#)
 - Christophe will tell you more about it tomorrow morning

EL7 EOL: new DIRAC certification box



- Lbcertifdirac70.cern.ch is CentOS7 and will be retired in few days
 - We can update it, but we take the opportunity to get out of CERN
- New box: ccdirac12.in2p3.fr -- Alma9, provided by cc in2p3
- MySQL also provided by cc in2p3 (MariaDB)
- OpenSearch: not provided by in2p3
 - We can use DiracX K3 kubernetes cluster
- Grafana: as OpenSearch
- Fluentbit can replace LogStash (?)
- No puppet

EL7 EOL: EL9 containers



- SingularityCE by default still points to cernvm4 (EL7), and supposes that CVMFS is available
 - did you update the default? ("[ContainerRoot](#)" option)
- The question here is only which default to use
 - We can provide a default on CVMFS, or point to an existing one
 - There's a [PR](#) with a connected [note](#) attached
 - We hope that `/cvmfs/unpacked.cern.ch` will provide a reference

(not) EL7 EOL: containers – cont



Added "complication": x86_64 (amd64) is not the only platform still on the market (arm64 resources are on the rise)

- A minimal on-the-fly discovery is needed
- The platform variant name has to follow a [specific naming](#)
 - e.g. amd64, not x86_64, arm64 not aarch64
 - This is not what **linux uname** nor **python platform** tells you, so some mangling in the code is needed

(not) EL7 EOL: containers – cont 2



- The hosting of images can be in CVMFS
 - But CVMFS might not be present (HPCs...)
 - And DIRAC should not assume that it is always there
 - In the "old" DIRAC-way we would host them somewhere and download on-the-fly
 - Which, BTW, in the case of HPCs, might not be possible.
 - Somehow we need a pointer to a location where to find containers for all platforms
- What are your (users') case for containers?
 - Expanding on <https://github.com/DIRACGrid/DIRAC/issues/3381> ?

Releases and features

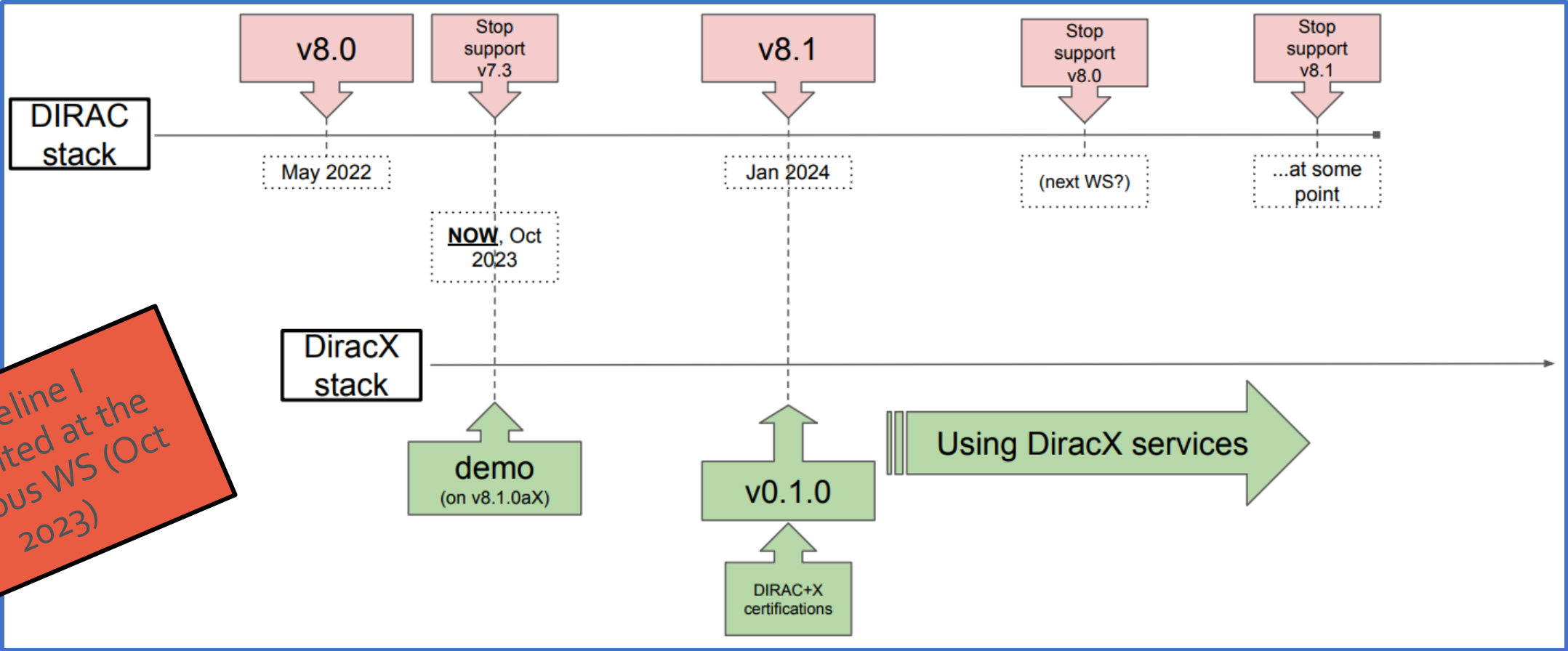
DIRAC versions

long term plan reminders

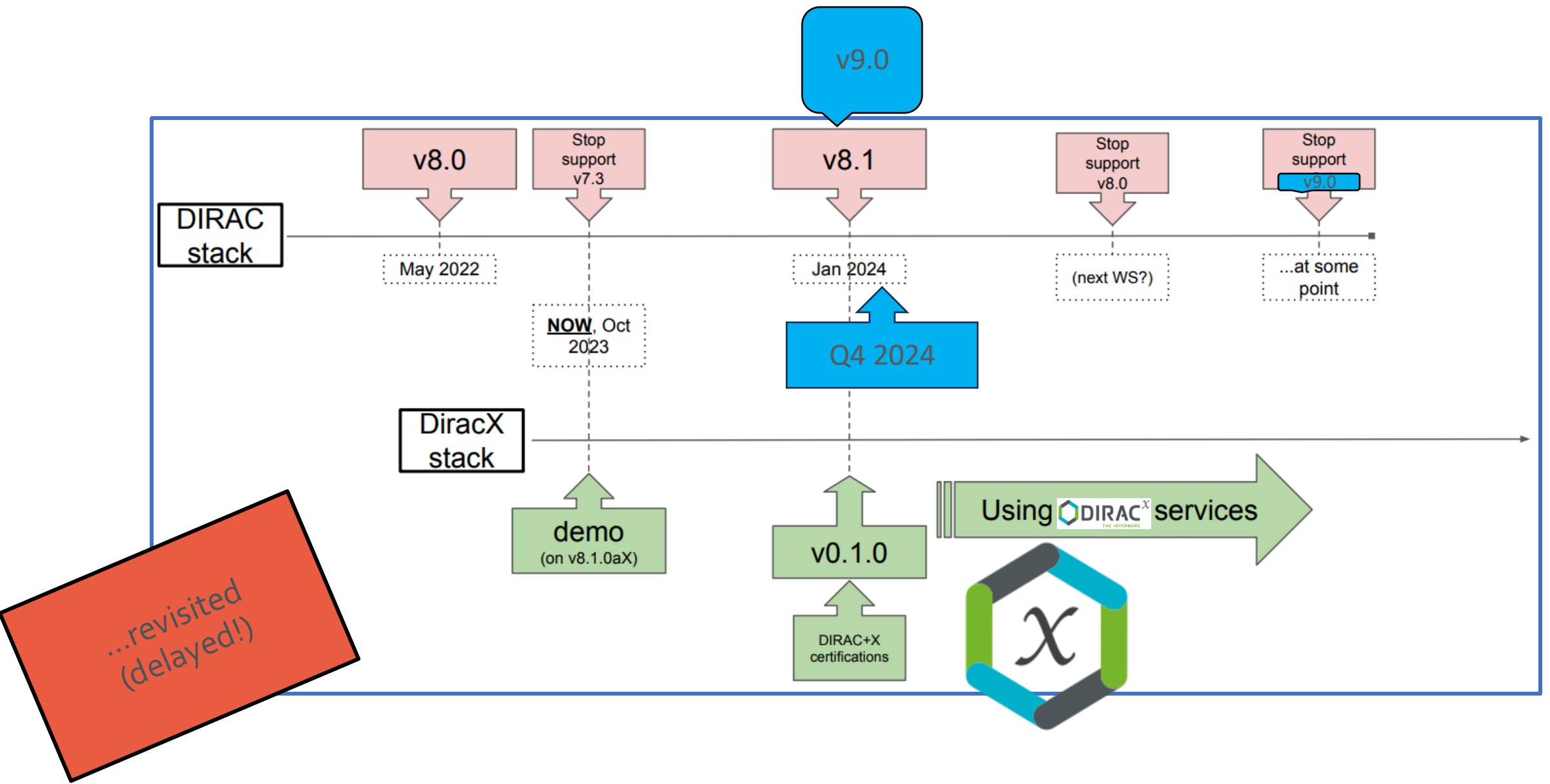


- To our knowledge, all DIRAC installations are using v8 in production
 - Not all: <https://github.com/DIRACGrid/DIRAC/discussions/7673>
- DIRAC v8 is at the moment the only production-level supported release. We will support v8 **for not long** after v9 will be released.
- When v9 will be released (together with DiracX v0.1) we will actively push for a campaign to update existing installations "quickly" (relatively).
 - **We will actively provide help.**

Regarding the DiracX timeline



Timeline 1 presented at the previous WS (Oct 2023)



Why DiracX is delayed?



- The main developers (Chris²) have been considerably busy with LHCb operations (data taking and other activities)
 - Plus Data Challenges and transition to IAM
- Some of the "basics" are still not fully sorted out
- DIRAC v9 requires still a considerable effort
 - Even myself, I did not manage to yet properly contribute to DiracX
- Several workshops/hackathons/holidays in April and May

DIRAC v8 -- features






Features introduced since 8.0.0 (which was releases in Sept 2022):

- ([#7612](#)) VOMS2CSAgent can query IAM (8.0.48)
- ([#7370](#)) add tools for consistency checks (8.0.39)
- ([#7371](#)) pilotWrapper: using CVMFS_locations for discovering the pilot files (8.0.34)
- ([#7336](#)) add an index on Status and Type for the TransformationDB.Transformations table (8.0.33)
- ([#7248](#)) Add support for loading external providers in CloudCE (8.0.31)
- ([#7248](#)) Add OpenNebula6 (XMLRPC) provider for CloudCE/libcloud (8.0.31)
- ([#7113](#)) Implement a finer grained permissions model for Transformations (8.0.27)
- ([#7062](#)) audience support in SiteDirector and CEs (8.0.24)
- ([#6890](#)) Enable PilotManager to manage pilots with tokens (8.0.17)
- ([#6803](#)) HTCondorComputingElement, ARC(6)ComputingElement and AREXComputingElement enabled to for job operations with tokens (8.0.16)
- ([#6682](#)) Introduce getDirectoryDump method in the DFC (8.0.11)

DIRAC v9 -- status



- A large list of changes is in the [wiki](#)
- Many DB changes, requiring manual intervention
 -  We need you for testing the instructions
 -  Basically, let's update your test instance and see what we missed
 -  **WE NEED YOU!** • And I mean to do it these days
- The last hackathon was OK-ish, but more changes are on the way and of course DiracX should be released together with v9.0
- There are still several [issues](#) to sort out
 - Possibly no more MySQL changes
 - Changes to OpenSearch indices will be part of it

AOB

CVMFS

consolidation and organization



NB: DIRAC pilots do not depend on CVMFS, but its presence is of great help

- `/cvmfs/dirac.egi.eu` is the CVMFS repo for DIRAC-related business
 - Managed at RAL, and to be fair, up to now not the most stable nor reliable
 - Might move to CERN
 - DIRAC ready-to-source releases are added here by the CI when created
 - DIRAC pilot code
- `/cvmfs/grid.cern.ch` is the default repo used for CAs and CRLs, voms-related pointers
 - De-facto source of truth for security files
 - vomsdir and vomses can be added here for any VO
- Pilots use it for quickly setting up the releases
 - Extensions can define a different (list of) CVMFS location(s) in `cvmfs_locations`, but the CVMFS structure should be the same

Monitoring (time-series)

I gave a longer talk at last [WS](#), today there are few news



- Historically, we stored time-series data in MySQL (AccountingDB).
- Then, through also ElasticSearch (still bending the arm of ES).
- More thoughts went into this [discussion](#)
- Now, OpenSearch provides [DataStreams](#) for time-series support.
- Plan:
 - Consolidate data on [OpenSearch](#)
 - NB: we will not be able to guarantee ElasticSearch support
 - Use Index Templates (first [PR](#)) and DataStreams
 - Consolidate visualizations through [Grafana](#)
 - Which might become an essential piece of DiracX, and treated "like a DB".

Grafana dashboards



...are nice. Also, fast, easy to setup, and intuitive for the users.

But my experience is only by using CERN's facility

Reminder: Grafana dashboards can already FULLY replace DIRAC's own plotting library and visualizations, plotting also MySQL (Accounting) data.

Qs:

- which installation has already set them up?
- If not, why not?
- If yes, how was the experience?

Ideally, we find time in these days to set them up for the DIRAC certification instance.

DiracX Branding



All logos you'll ever need
in <https://github.com/DIRACGrid/management/tree/master/branding/diracx>



Other Projects (proposals/ideas)



- Anyone wants to do a Dirac chatbot? (e.g. [CERN chatbot](#))
 - Good/funny student project
- Green computing, but not only
 - "select the greener resource", might not be only the "Hype of the day"
 - There are already studies about [carbon cost of processors](#)
 - And ARM processors are at the moment more carbon-cost effective than x86
 - ...and [carbon-cost of storage medias](#)
 - IMHO it looks similar to "select the cheaper cloud provider"
 - At the end, at least for jobs, it is advanced matching
 - We have a description for a PhD program on this topic





Questions?

