Status of DiracGrid projects

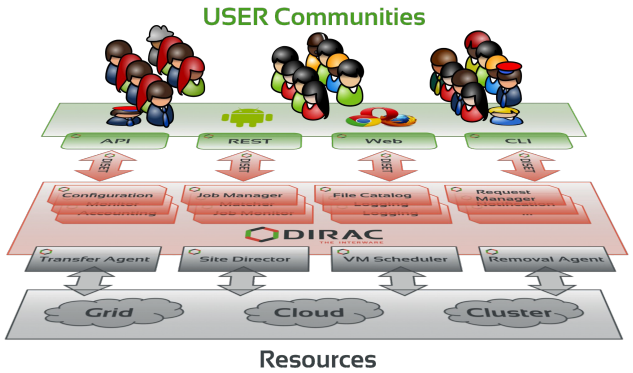
Federico Stagni

DIRAC technical coordinator

DR23, KEK, Tsukuba, 16th October 2023

Slide that I have been presenting for years, with minimal variations

- A software framework for distributed computing
- A **complete** solution to one (or more) user community
- Builds a layer between users and resources



- Developed by communities, for communities
  - Open source (GPL3+), GitHub hosted
  - Python 3
  - Publicly documented, yearly users workshops, open developers meetings and hackathons
  - Deployed mostly via Puppet on VMs (really, not bound to any specific technologies)
- The DIRAC consortium as representing body

DIRAC Public
DIRAC Grid
Python 108 GPL-3.0 167 86 18 Updated 3 minutes ago

diracx-web Public
TypeScript 0 GPL-3.0 2 2 0 Updated 2 hours ago

diracx Public
The neXt DIRAC generation
Python 5 GPL-3.0 14 25 (9 issues need help) 11 Updated 3 hours ago

diracx-charts Public
Helm charts for running DiracX
Shell 0 13 7 2 Update

Pilot Public
DIRAC pilot 3.0
Python 2 GPL-3.0 22 4 3 Updated 3 days ago

diraccfg Public
DIRAC stand-alone cfg files parser
Python 0 GPL-3.0 3 0 0 Updated 5 days ago

WebAppDIRAC Public
DIRAC Web front-end application
JavaScript 10 GPL-3.0 40 9 0 Updated 3 weeks ago

DIRACOS2 Public
Python 0 GPL-3.0 8 5 3 Updated on Aug 7

management Public
management repo for DIRAC
Python GPL-3.0 10 0 3 1 Updated on Feb 22

pr-sweeper Public
Forked from AIDASoft/pr-sweeper
Cheery picks merged PRs to selected branches
Python MIT 3 0 0 0 Updated on Nov 28, 2022

tornado Public
Forked from chaen/tornado
Tornado is a Python web framework and asynchronous networking library, originally developed at FriendFeed.
Python Apache-2.0 5,534 0 0 0 Updated on Aug 10, 2022

COMDIRAC Public archive
Package to supply a comprehensive list of commands for the DIRAC user interface
un 23, 2022

un 16, 2022

RESTDIRAC Public
Web API for DIRAC
Python GPL-3.0 6 3 5 2 Updated on Apr 13, 2022

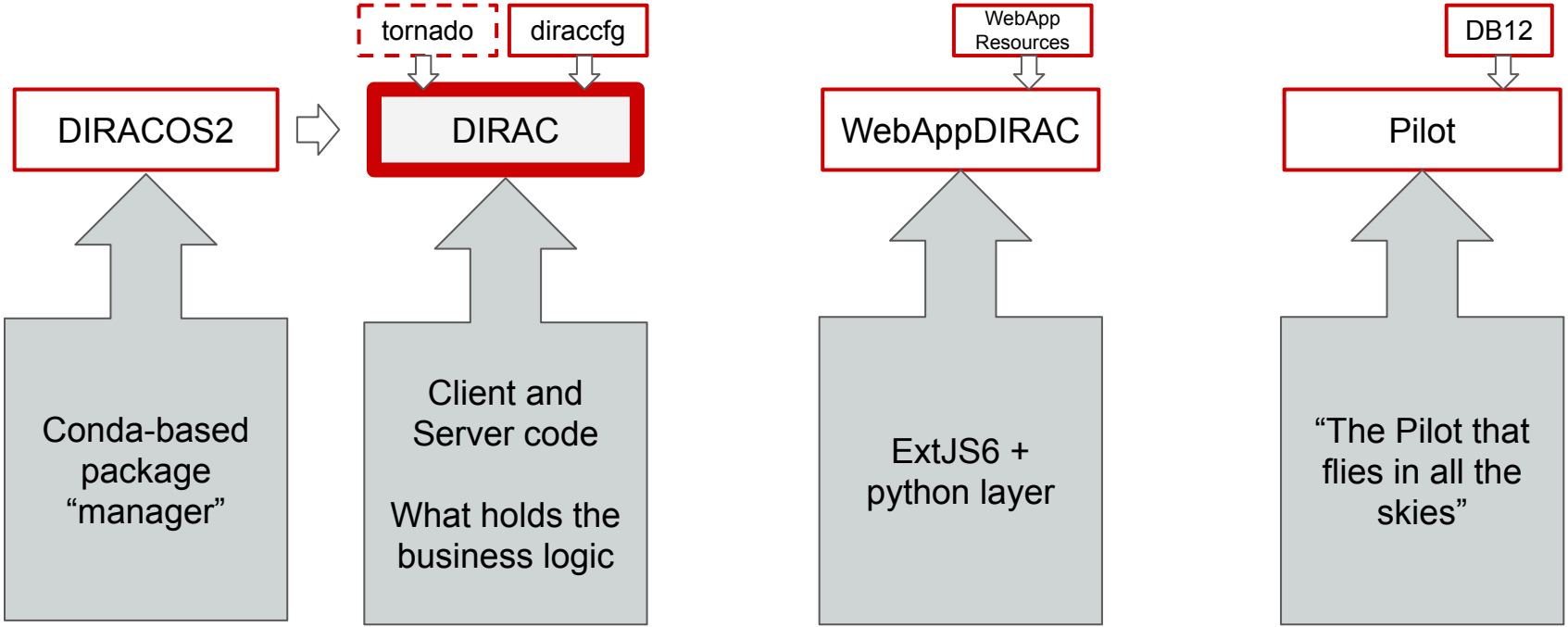
DB12 Public
DIRAC 2012 fast benchmark
Python LGPL-3.0 5 4 2 0 Updated on Feb 28, 2022

DIRACWebAppResources Public
Slowly changing assets for WebAppDIRAC
Python GPL-3.0 0 0 0 0 Updated on Jun 29, 2021

dirac-webapp-packaging Public
Build tools for compiling javascript sources in DIRAC WebApp packages
Python GPL-3.0 0 0 0 0 Updated on Jun 29, 2021

Several active repositories at https://github.com/DIRACGrid

Today's DIRAC (py3) stack



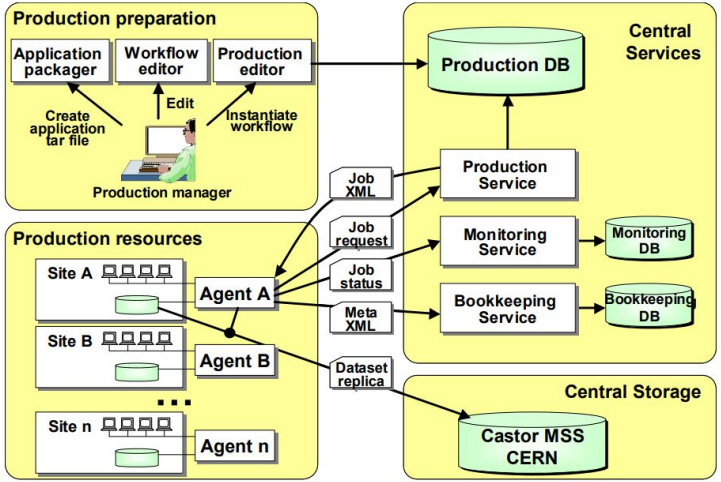
Brief history of DIRAC

Started in LHCb as a MC production system at around 2000

- Not called DIRAC yet
- bash scripts
- Running at production sites

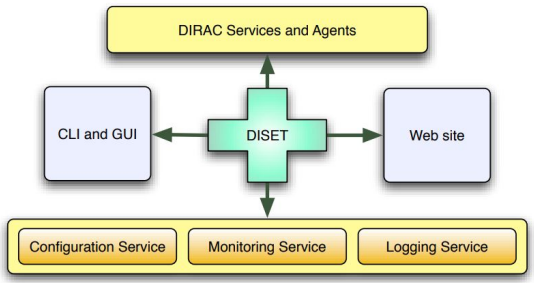
“DIRAC2” - 2003

- Started in 2002
  - Python
  - xml-rpc
  - interfacing to EDG (European DataGRID)
- First successful grid usage ever - Data Challenge’04
  - First use of pilot jobs based WMS
- Inventing DIRAC = “Distributed Infrastructure with Remote Agent Control”
  - a completely made-up name, later abandoned

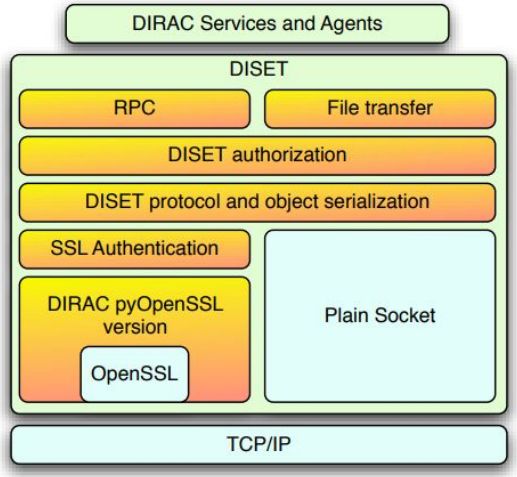


“DIRAC3” - 2006 and 2007

- DISET protocol developed to address shortcomings of python’s xml-rpc
  - provide RPC and file transfer mechanism (“dips://”)
- The CS (Configuration System)
- Accounting System
- …



- **Full** refurbishment
- Planned 6 months, but effectively took 3 years!
- Effectively, what we have still nowadays
- **Needed!** or there would not be a DIRAC today.



Open-sourced, wider adoption

- 2008: large-ish reshuffling to become Multi-VO:
  - DIRAC - the “interware”
  - “vanilla” DIRAC created + extension mechanism
  - LHCbDIRAC extension
    Separated
- DFC
  - LFC/AMGA replacement
- First multi-VO installations:
  - France-Grilles, 2011
  - EGI 2014

… as a Community Grid Solution

DIRAC is now used by multiple High Energy Physics experiments and projects in other domains – ILC, CLIC, Belle II, BES III, CTA, VIP/biomed, etc. The LHCb Collaboration remains the main DIRAC user for all its Distributed Computing tasks.

See, for example [20],[35],[98],[145],[292]

2012

Installations and communities

(that I know)

A *framework* shared by multiple experiments/projects, both inside HEP, astronomy, and life science

Experiment agnostic

Extensible

Flexible

DIRAC
THE INTERWARE

Successful project

- Example of a project evolving from an experiment specific to a general-purpose one
- The pilot based architecture is adopted by all the LHC experiments and multiple grid infrastructures
- Rare example of an efficient complex solution
  - Both WMS and DMS at a scale
- Contributions from more that a hundred developers during 20 years of the project life
  - Plus specific extensions

All good?

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

Some DIRAC developments

- Done: Python 3
  - py3 clients supported since version 7.2 (pip installable)
  - py3 server supported since version 7.3
  - py2 support ended with 8.0 (released last week)
    - with some obvious exceptions of part of pilots code
- Done: ES/kibana/grafana dashboards
- Ongoing/advanced: `dips://` → `https://`
  - `dips`: DIRAC proprietary protocol for RPC calls
  - `http`: based on tornado
  - most DIRAC services already available using HTTP
    - we said that `http` would be the default for all the DIRAC services from version 8.1
- Ongoing: token support, and IdP (IaM, Check-in)
- Ongoing: running on kubernetes (goal: define a *helm* chart)
- Started: using celery and RabbitMQ (retiring executors)

issues with HTTPS in DIRAC

- started (too) long time ago, not yet finished and even less adopted
- the horse (tornado) is feeling old-ish
- there's still an RPC call to be made at the end, no routes
- we are using our own tornado fork because of GSI support

Issues with tokens adoption

- DISET is x509 only, and will stay like that

  DISET is the DIRAC's own protocol

- Adding token support in the current framework is far from easy, and error prone

  (here I am not talking about tokens for submitting to HTCondorCE…)

The list can go on

- the WebApp is highly custom, and somewhat un-maintainable
  - with an intermediate python layer
- runsv is “dead”, we create the RPMs…
- DIRAC’s plotting is “old-ish”
- Moving to JSON serialization quite painful
- Upgrades are not always easy, and sometime scary
- …

- we have been accumulating problems for years
- out there the world evolved in different directions (e.g. REST APIs)

Keeping the project successful

- ➔ It felt like we were at the end of a technology cycle.
- ➔ in order to keep the project successful we are creating the neXt dirac incarnation in what we dubbed project “DiracX”

DiracX in just one slide

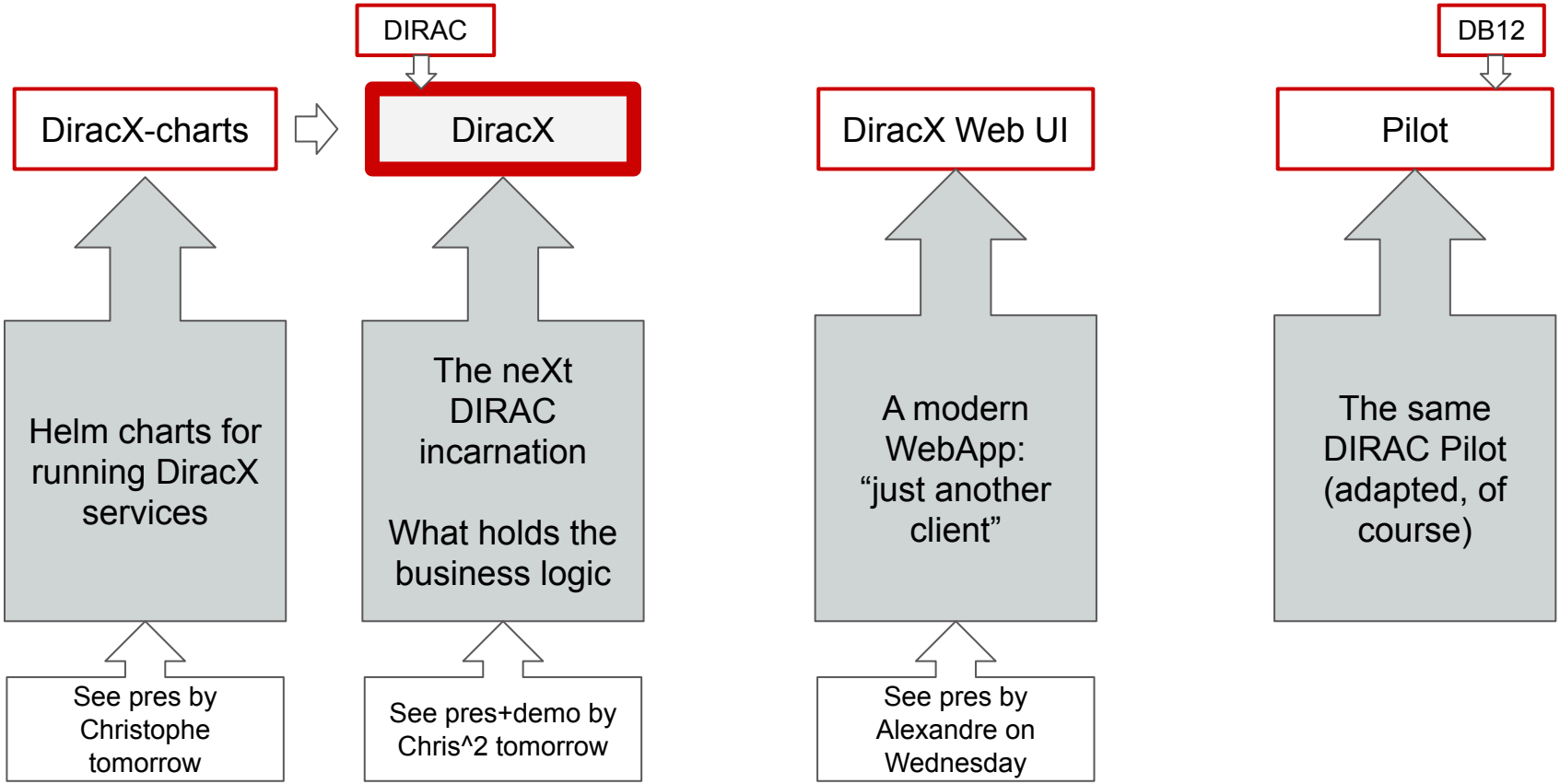
- The neXt DIRAC incarnation
  - DIRAC functionalities will move there
- A cloud native app
- Multi-VO from the get-go
- Standards-based
- Not a framework

DIRAC
THE INTERWARE

Considerations

- The moment is **now**
- What is possible to do nowadays, would have not been possible only few years ago
- Risks
  - **time**! We don’t have the luxury to go slowly

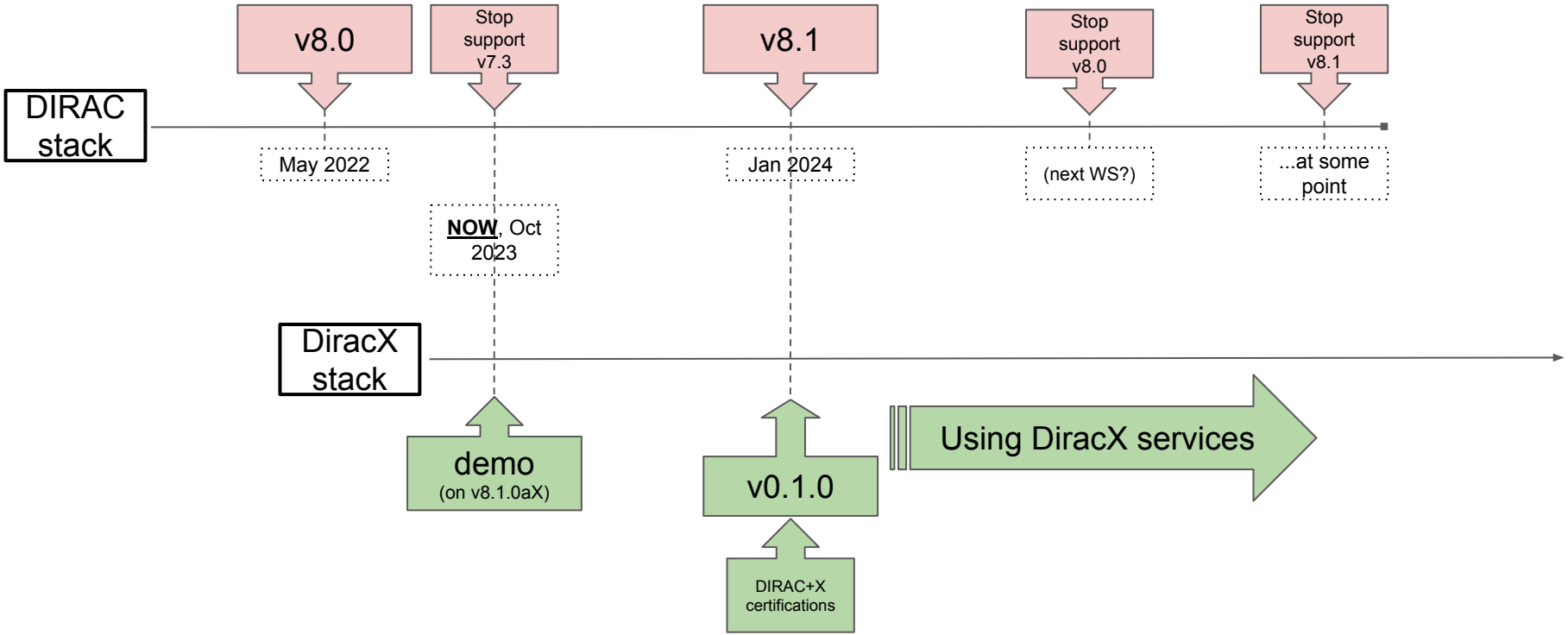
(in dev) DiracX stack



So, right now

- Supporting the current, production-level DIRAC “stack”
- Developing for the neXt stack

Timeline


v8.0
Stop support v7.3
v8.1
Stop support v8.0
Stop support v8.1
DIRAC stack
May 2022
Jan 2024
(next WS?)
...at some point
NOW, Oct 2023
DiracX stack
demo
(on v8.1.0aX)
v0.1.0
Using DiracX services
DIRAC+X certifications

DIRAC releases

DIRAC v8.0 (production)

First major release in a while:

- Abandoned Python 2
- Added support for IdPs (IaM, Check-IN)
  - Can use tokens for submitting pilots to CEs
- Monitoring capabilities expanded

Python3 and PyPI

- DIRAC releases using standard `pip` package manager, found on PyPI
  - extensions had to adapt (already in DIRAC v7.3)
- Deployed in a conda environment created by DIRACOS2 installer
  - which provides Python 3.11
- Support for platforms `ppc64le` and `aarch64` (in addition to the more common `x86_64`) have also been added
  - through conda/mamba

Tokens support

Timeline

Milestone ID	Date	Description	Dependencies	Teams
M.1	Sep 2022	IAM is also in production for ALICE and LHCb.		CERN IT, IAM devs
M.2	Dec 2022	DIRAC versions supporting job submission tokens deployed for concerned VOs (LHCb, Belle-II, …).		DIRAC, LHCb, Belle-II, …
M.3	Feb 2023	VOMS-Admin is switched off for one or more experiments. Prerequisites: • Significant VO admin functionality issues in IAM have been resolved • User registration, group and management have been swi IAM • IAM services are sufficiently • CERN IAM team is sufficientl • Remaining VOMS-Admin use have been moved or will be dropped		IAM devs, CERN IT, experiments
M.4	Mar 2023	HTCondor installations at EGI sites have been upgraded to supported versions > 9.0.x. Prerequisites: • DIRAC versions supporting job submission tokens have been deployed for the concerned VOs (LHCb, Belle-II, EGI catch-all, …) • HTCondor CE supports (adjusted) EGI Check-in tokens • IAM or equivalent in production for ALICE, LHCb, Belle-II, …	M.1 M.2	HTCondor Dev Team, WLCG ops, EGI ops, sites
M.5	Mar 2023	End of HTCondor support for GSI Auth (link).		
M.6	Mar 2023	Some storage endpoints provide support for tokens (at least one per service type).		WLCG ops, storage devs, sites
M.7	Feb 2024	Rucio / DIRAC / FTS have sufficient token support in released versions to perform DC24 using token authorization.	M.6	Rucio, DIRAC, FTS, experiments

https://doi.org/10.5281/zenodo.7014668

Basically: trying to respect the WLCG timeline

Interfacing with IAM and EGI Check-IN IdP

DIRAC v8 adds `client_credentials` flow for submitting pilots

See pres by Andrei tomorrow

FTS only?

DIRAC / dashboards /

fstagni Merge pull request #7039 from EwoudK/MonitoringDashboards

Name

..

AgentMonitoring

DataOperation

ElasticJobParameters

GrafanaDemo

PilotSubmissions

PilotsHistory

RMS

ServiceMonitoring

WMS

diracLogs

More details tomorrow afternoon

Monitoring

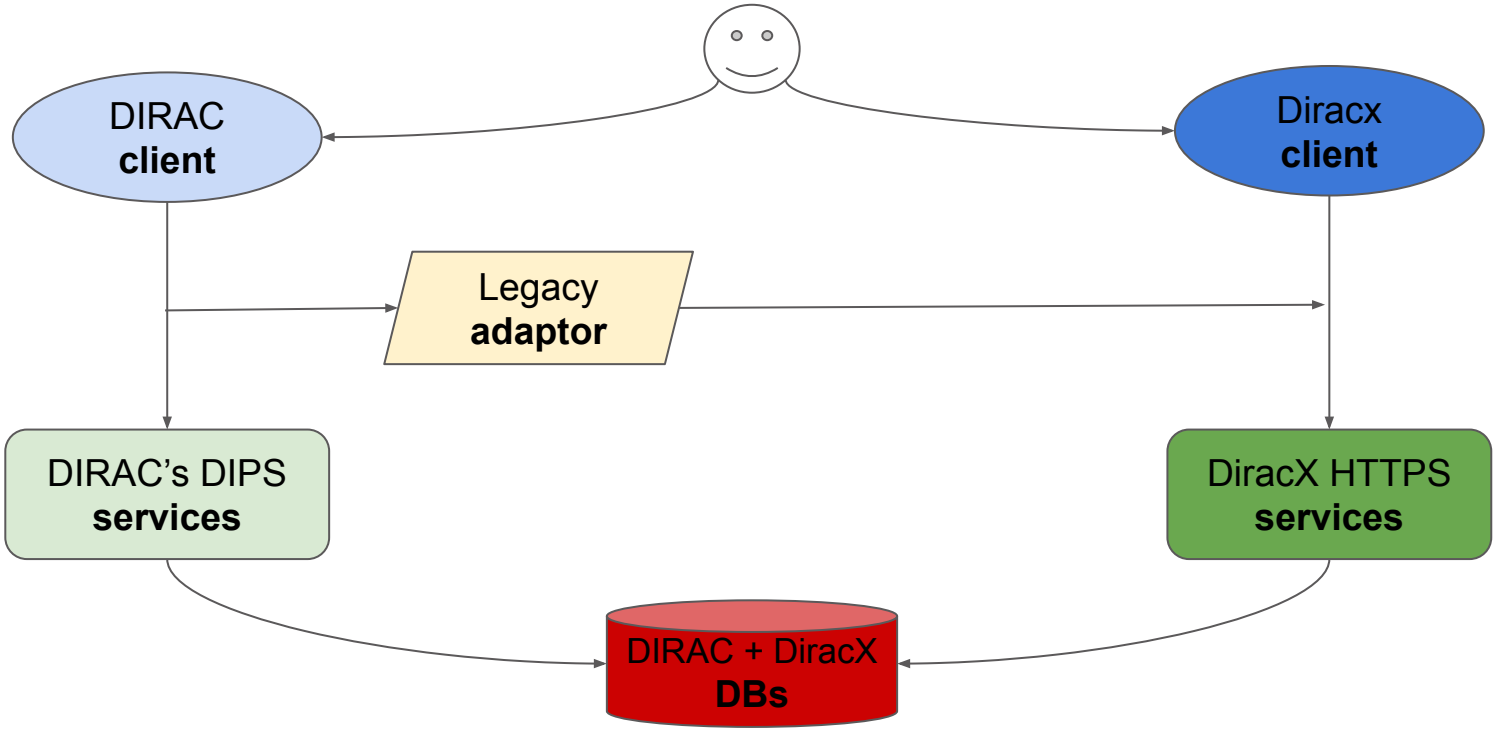
- Added support for OpenSearch (ElasticSearch support was already there), which also becomes the favourite option
  - dropped ES6 support
- Added several OpenSearch indexes that can be filled in
- Added dashboard definitions for Kibana and grafana
- removed `gMonitor` and the `Framework/Monitoring` service (“ActivityMonitoring”)

DIRAC v8.1

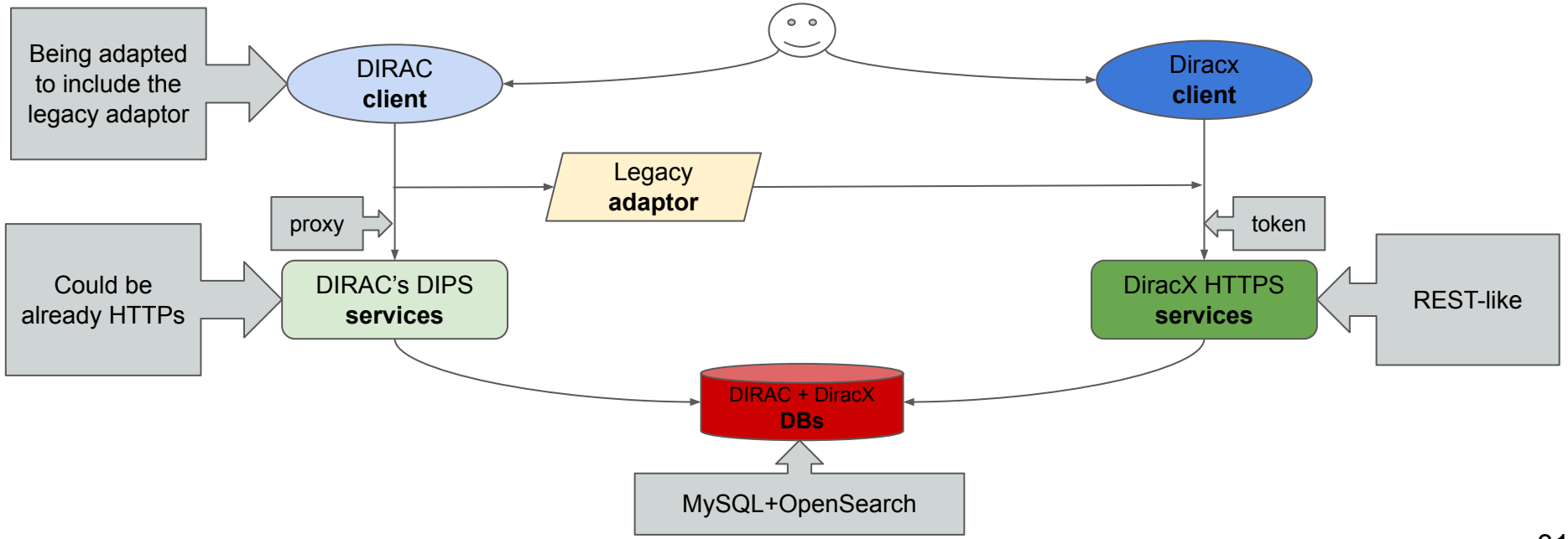
- Postponed to Jan 2024
- Abandoning the concept of “Setup”
  - several changes/simplifications at CS and DB level
  - unfortunately, many manual changes too
- Really, a major release, and the last of DIRAC releases!
  - with all the groundwork for transitioning to DiracX
    - especially the database changes

From DIRAC to DiracX

Transitioning (services)



Transitioning (services)



Transitioning
(agents + executors)

Python **celery** + **RabbitMQ**

https://docs.celeryq.dev/en/stable/getting-started/introduction.html

https://www.rabbitmq.com/

NB: we have not yet started coding for this!

Transitioning from DIRAC agents and executors to DiracX tasks should be easy and straightforward

DIRAC
THE INTERWARE

Q1+Q2
2024

Transitioning stages
(extreme summary)

1. Update to DIRAC v8.1
   a. this, effectively, means also installing DiracX
2. Run few services in DiracX
3. Activate the legacy adaptor
   a. traffic for the selected services will be redirected to DiracX services
   b. proxy → token behind the scene
4. You can now remove the legacy DIRAC services

Notes on updating

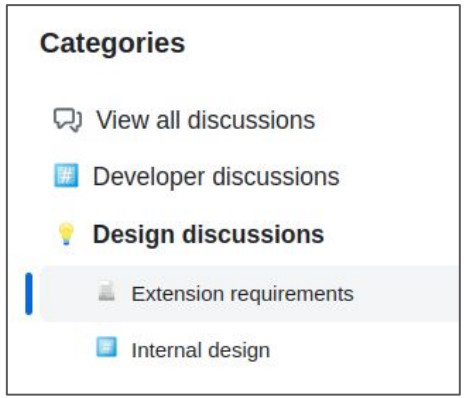
- Transitioning to DIRAC v8.1 and DiracX is, *today*, not trivial. We will make sure that the procedure will be made easier by the time the update will be performed (it won't be "just a wiki page").
  - the update procedure will be tested.
- DIRAC and DiracX will coexist for a while. The shortest, the best.
- We'll be there to help, in various ways.

To conclude

Exciting and busy time

- Rewriting DIRAC
  - WMS functionalities will come first
  - you are very welcome to come onboard
  - your input is needed:

    https://github.com/DIRACGrid/diracx/discussions



- DIRAC v8.1 will be the bridge for getting there
  - We'll try to ensure stability as much as possible

For Devs

BILD meetings:

DIRAC meeting

every 3rd week

Thursday at 10:00 AM CET

LHCb hosted

Clic, Belle2, EGI, GridPP, IHEP, JINR, CTA represented

→ *you want to be invited? Just let me know*

DIRAC Certification hackathons:

every 3rd week

Thursday at 10:00 AM CET

LHCb hosted

Clic, EGI, GridPP represented

→ *you want to be participate? Just let me know*

lhcbdirac.slack.com + Trello

BILDx meetings:

DiracX meeting

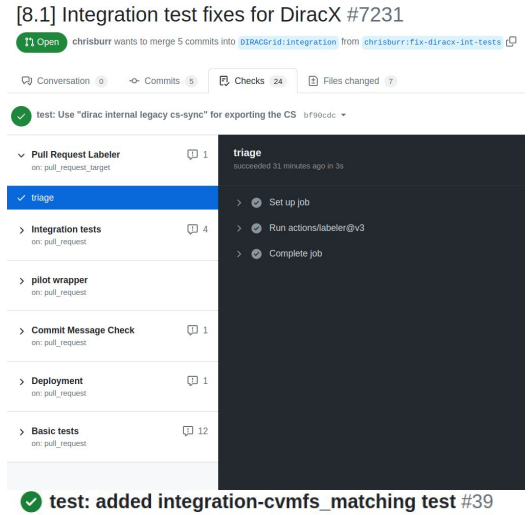
every 3rd week

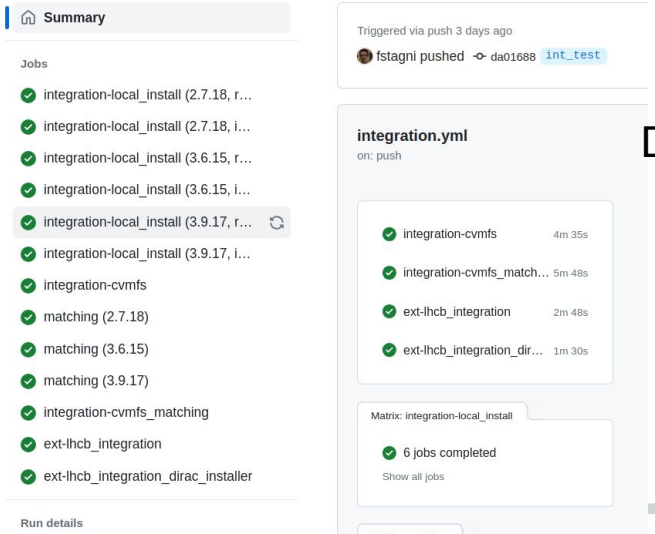
Thursday at 10:00 AM CET

LHCb hosted

Clic, Belle2, EGI, GridPP, IHEP, JINR, CTA represented

→ *you want to be invited? Just let me know*





Development and testing

~6 FTE as core developers, a dozen contributing developers

Tests, certification, integration process is a daily work.

- We use (lots of) GitHub Actions, and Jenkins for some bits
- We run certification hackathons every 3rd week
  - install the latest pre-reason, run several system tests

DiracX hackathons:

- 1st: https://indico.cern.ch/event/1292289/ (4-5 Jul, CERN)
- 2nd: https://indico.cern.ch/event/1304626/ (4-5 Sept, CERN)
- 3rd: here, on Wednesday/Thursday/Friday
- 4th: 22-23 Jan ?? (CERN)

Links

- DIRAC's doc: dirac.readthedocs.io
  - including code documentation
- DiracX's doc
  https://github.com/DIRACGrid/diracx/tree/main/docs
  - We might use RTD also for DiracX
- Dev+Ops+general questions:
  - DIRAC github discussions
  - DiracX github discussions
    - DiracX-Web discussions
    - for speedy communications: https://mattermost.web.cern.ch/diracx/

      do you *really* want also slack?

Plans for this week's Q/A + hackathon

- DiracX developers hackathon
- … what you ask!
  - "help me with update"
  - "I need info on topic xyz"
  - …

Add your comments/questions to https://cernbox.cern.ch/s/DNZcRP5KCsbvVZ7

Questions?

https://github.com/DIRACGrid