DIRAC: Status and future evolution

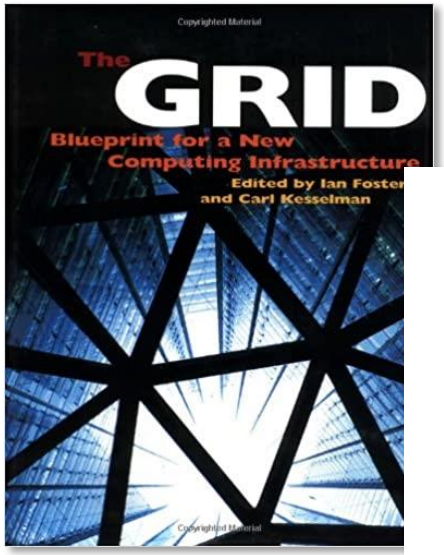
A.Tsaregorodtsev,
CPPM-IN2P3-CNRS, Marseille,
GRID'2025, JINR, Dubna
8 July 2025

DIRAC
THE INTERWARE

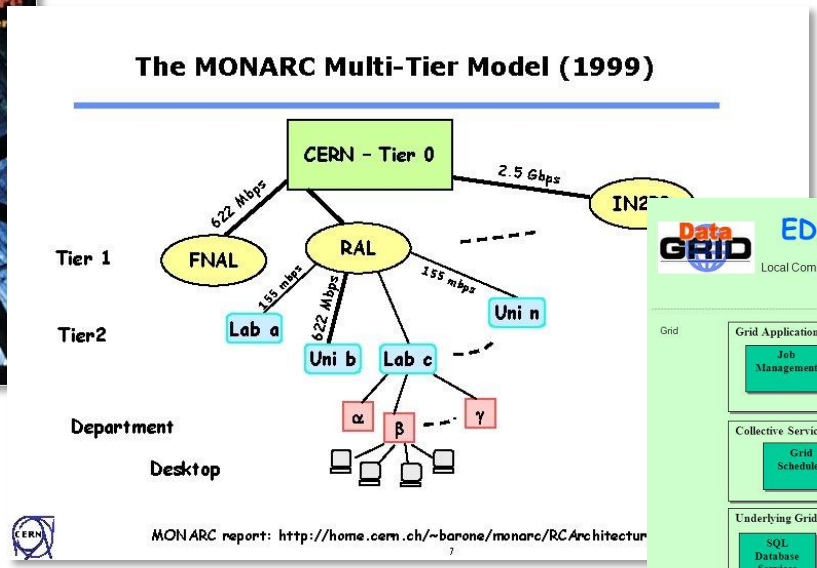
Outline

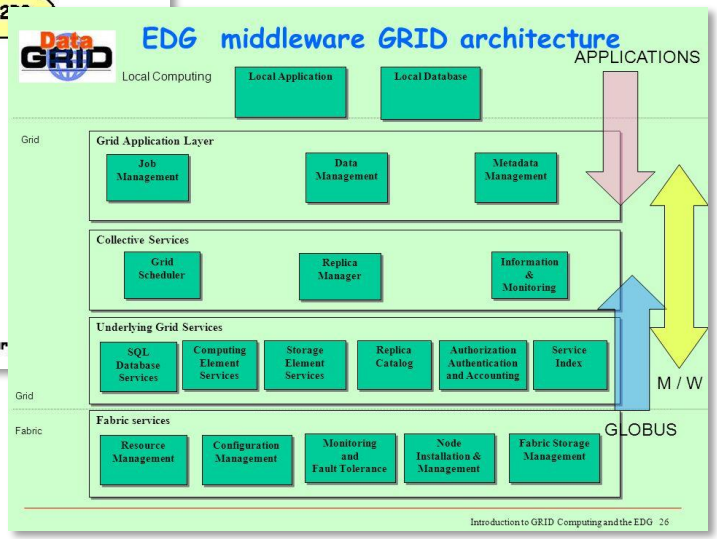
- Retrospective view on the DIRAC evolution
- Current Project Status
- DIRAC Users
- Upgrade to DiracX – the neXt DIRAC incarnation

At the beginning…



Starting with the Grid concept, the MONARC Project proposed an hierarchical model of tier sites

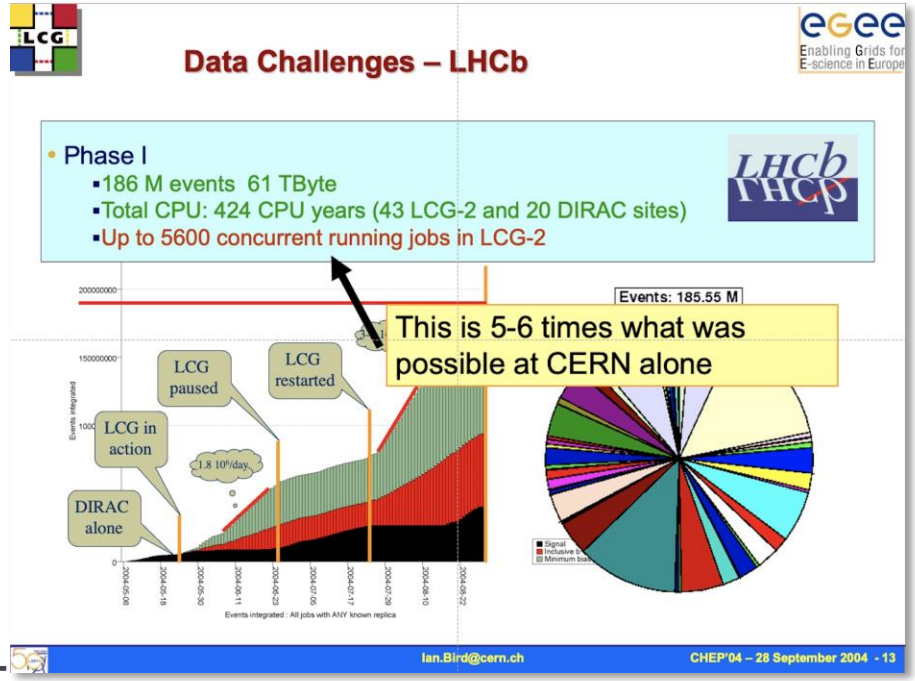




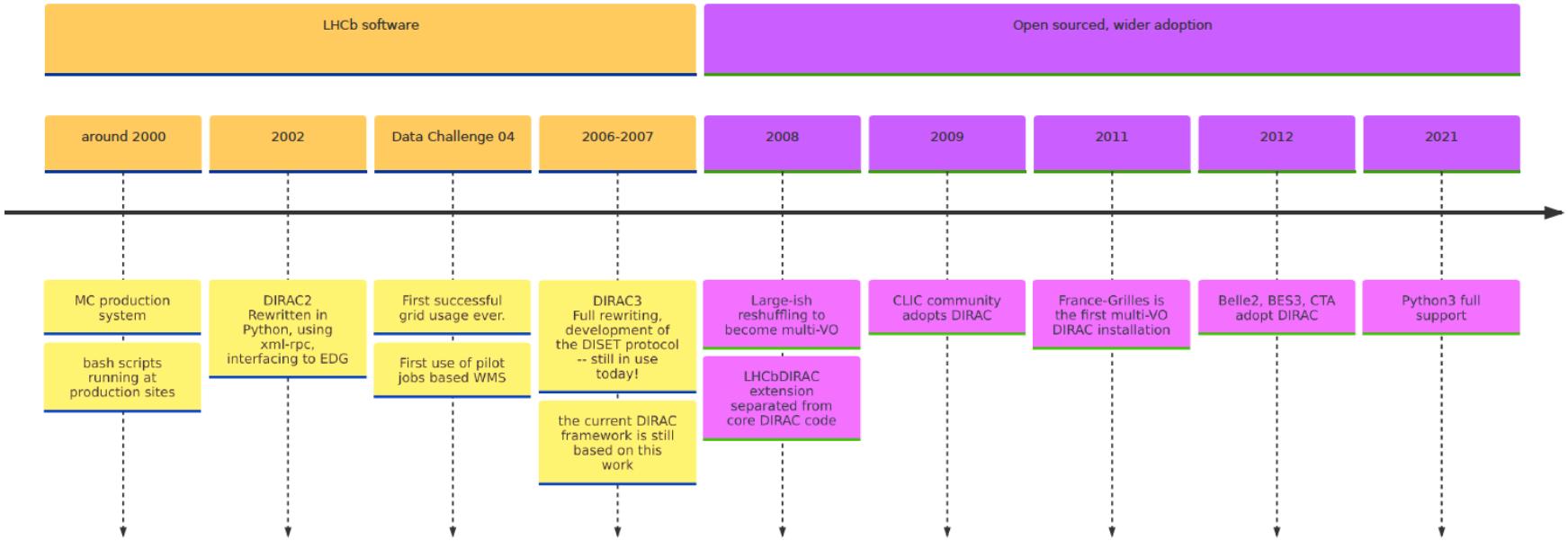
- DataGRID (2001-2003) project came up with the first grid middleware
  - Based on Globus 2 toolkit
  - Did not satisfy production requirements
  - Clear need for an experiment production system on top -> AliEn, DIRAC, PanDA

First successful usage

- DIRAC was the LHCb's response to this challenge
- DIRAC job user efficiency **> 90%**
  - while **~60%** success rate of LCG jobs.
- The first production system to embed in a grid job a script to pull jobs from a central queue
  - Sending agents as regular jobs
  - Now known as pilot jobs
  - The architecture adopted by other L experiments
- The scalability of the system allowed to saturate all available resource of DC'2004
- Record set for the maximum number of running jobs
  - Orders of magnitude less than what we can do now !



Timeline

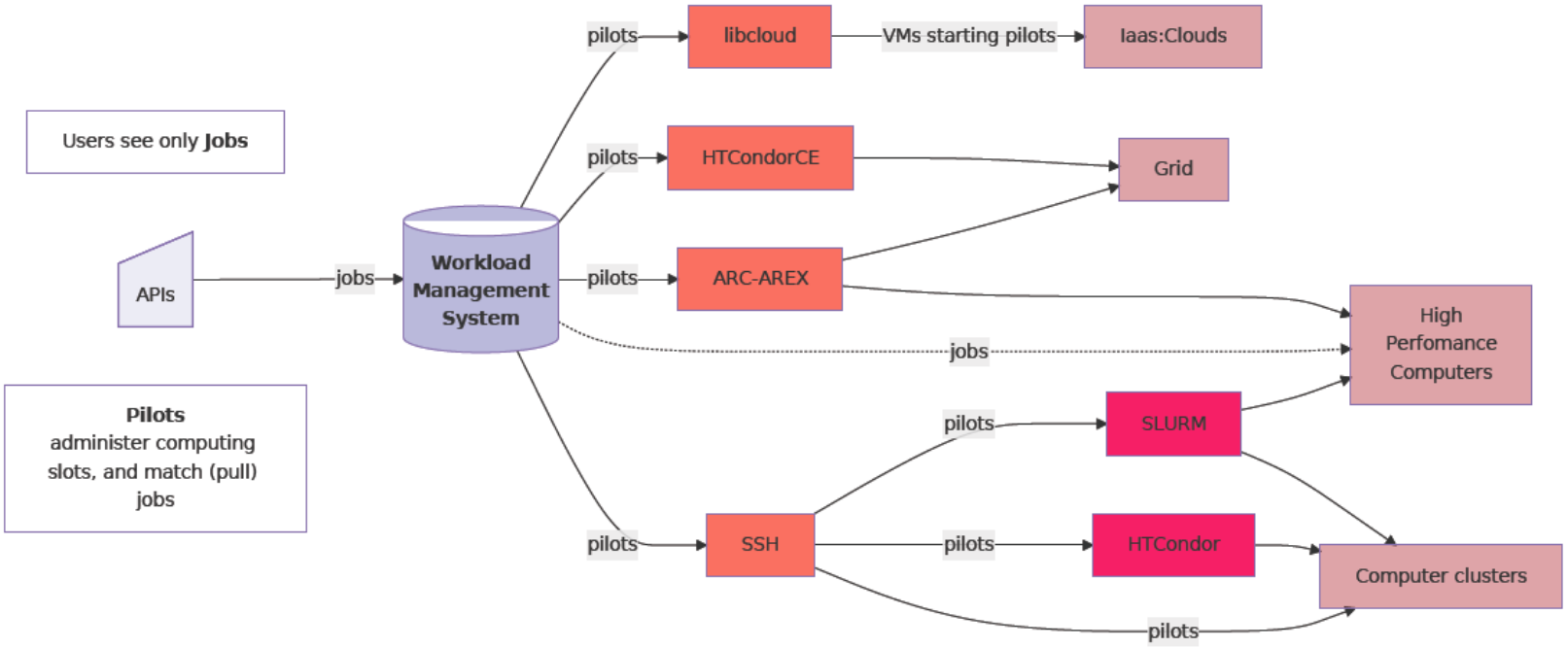


- Evolving from a bunch of shell scripts to a general purpose distributed computing framework
  - DISET secure protocol with data streaming support
  - Multi-VO, extendable
  - Adopted by several HEP communities and grid infrastruture projects

DIRAC
THE INTERWARE

- DIRAC is a complete grid solution for one or multiple user communities that need to exploit distributed heterogeneous resources
- Both computing and storage resources can be handled within the same framework with support for large-scale operations

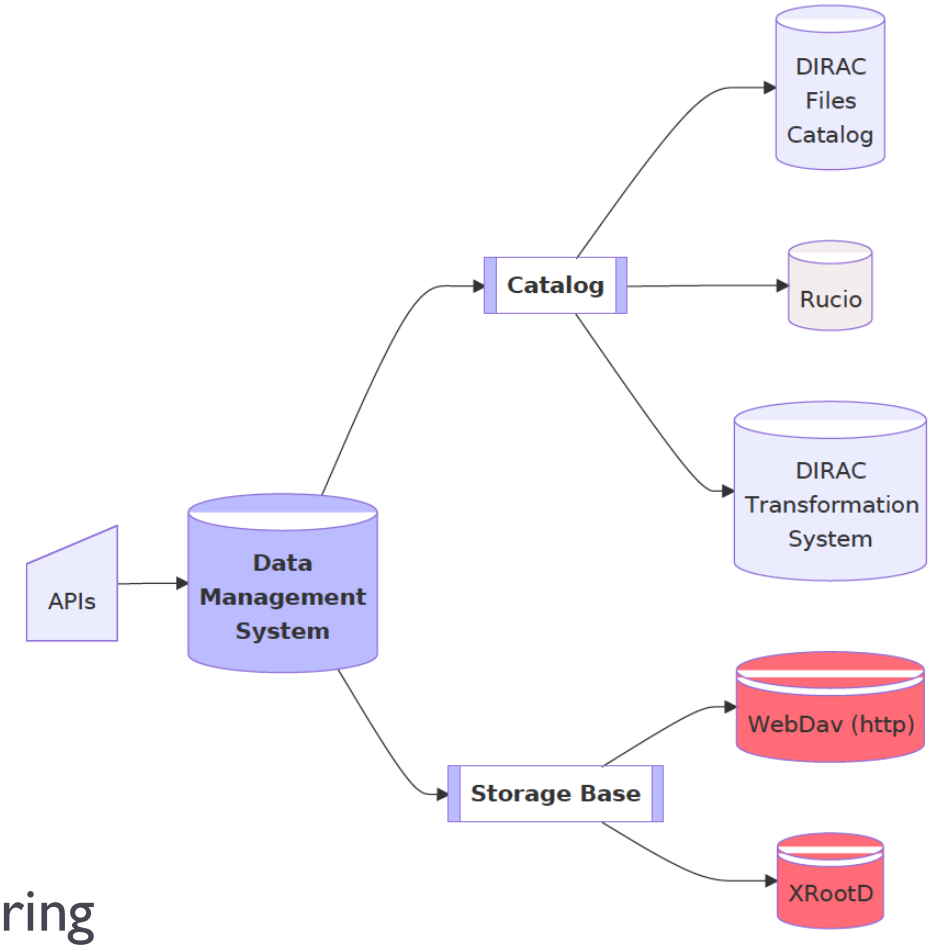
Managing workloads



- « Pull » model with pilot jobs
- « Push » solution for HPC centers that do not allow pilots due to the internet access limitations
- Integrating CWL (Common Workflow Language) for job descriptions in DiracX (*see below*)

Managing data

- Managing files
  - Logical file names - LFNs
  - File Catalogs
    - physical replica locations
    - file metadata
    - DIRAC File Catalog
    - Rucio
  - Storage systems (SEs) with various access protocols
- Support for massive data operations
  - FTS, failure recovery, monitoring

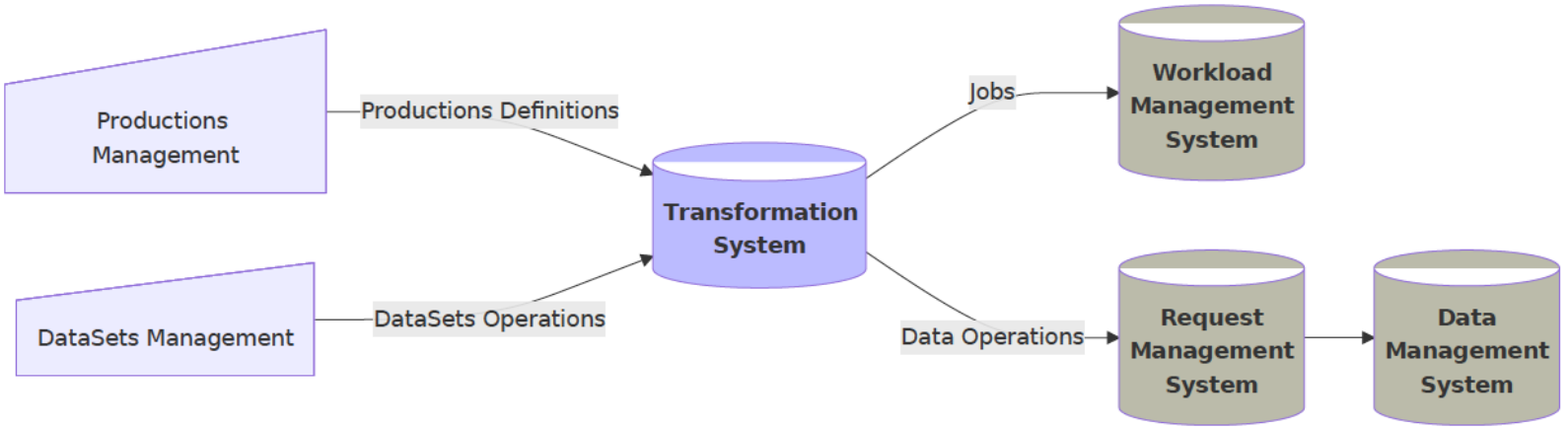


RucioFileCatalog

- Multiple requests to allow DIRAC to work together with the Rucio DMS.

- The RucioFileCatalog plugin developed by the Belle II experiment
  - Developments done both on DIRAC and Rucio sites
  - Implements the DIRAC File Catalog interface
  - Allows access of the DIRAC WMS and DMS components to get the file information from Rucio
    - Data aware job placement
- Now used by the CTAO Collaboration, offered by the GridPP DIRAC service

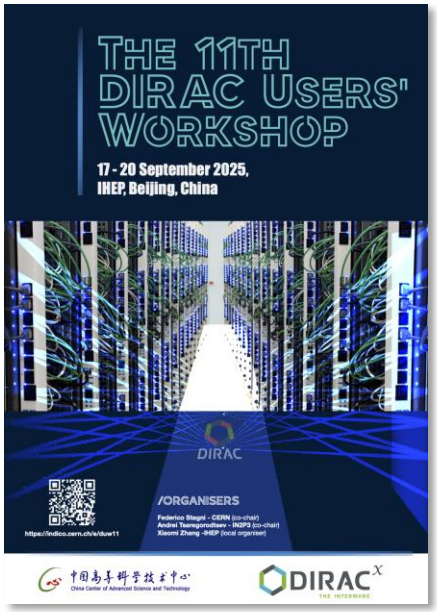
Transformations



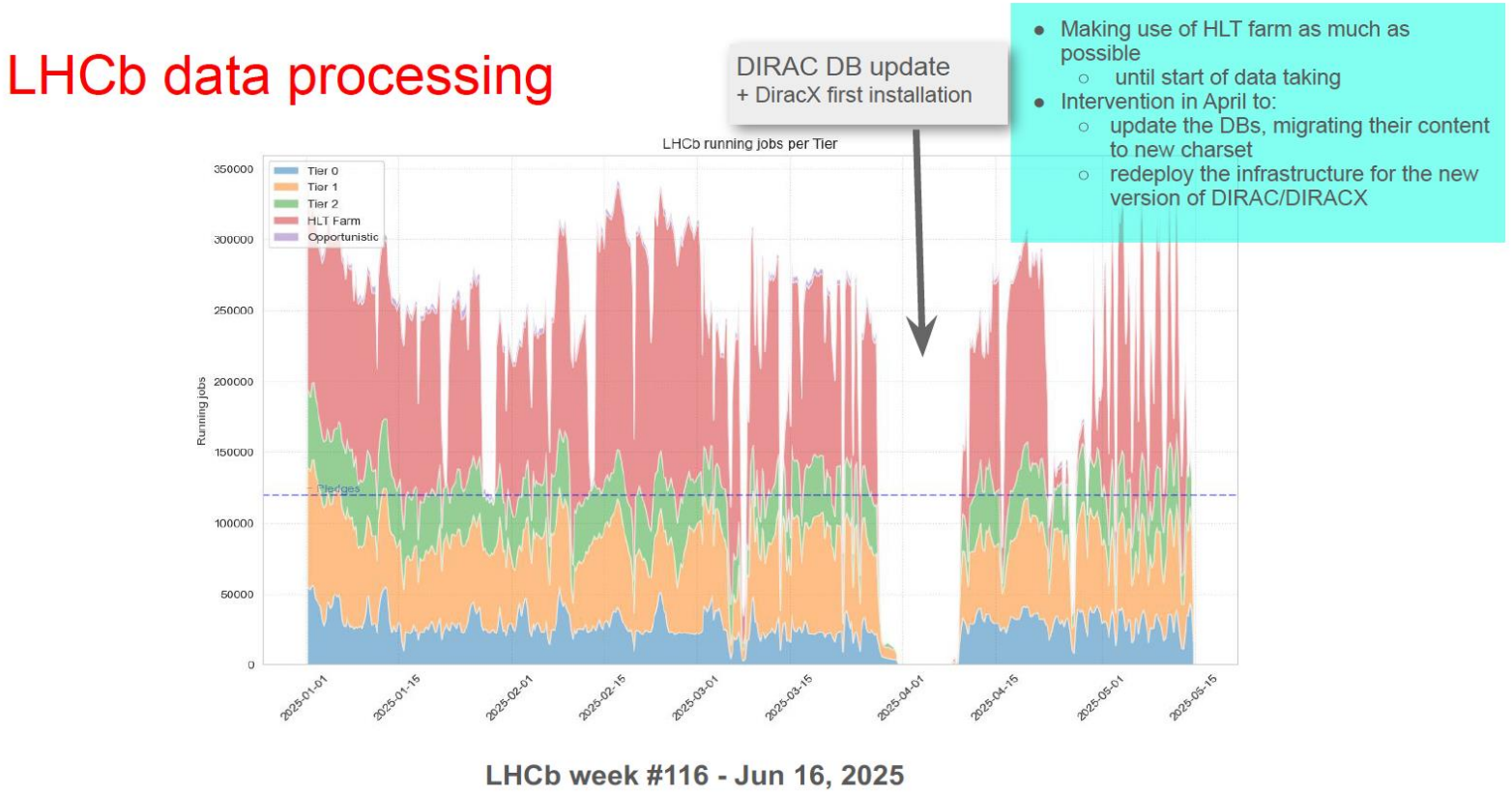
- Transformations System is an engine for managing complex data-driven workflows
  - Automation of common tasks
    - Creation, resubmission, deletion of jobs
    - Replication, registration, removal of data files
  - Handling millions of jobs and files
    - Massive operations validation, failure recovery

Open source project

- Driven by the needs of the user communities
  - With multiple contributions from community developers
  - Example: DIRAC File Catalog (DFC) was developed initially for ILC and BES experiments, it is used now by several experiments including LHCb.
- Public code repository in Github
  - GPL v3 license
  - Automated testing, CI with Github actions
- DIRAC Consortium created in 2017
  - Developing and promoting the DIRAC software
  - Holder of the DIRAC software copyright
  - Current members: CNRS, CERN, IHEP, KEK, Imperial College
  - Organizing DIRAC User's Workshops, tutorials, hackathons
    - The next one is the11th DUW in IHEP, Beijing, 17-20 September ***https://indico.cern.ch/event/1433941/***



Users: LHCb

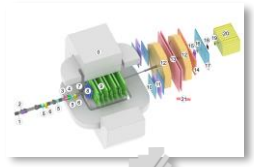


- LHCb is the main DIRAC user
  - Running over 300K concurrent jobs on more than 100 sites
  - Resources: HTC, HPC, clusters, HLT farm

Users

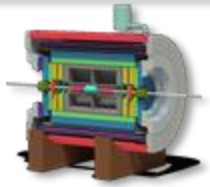
- **Belle II, KEK**
  - Using DIRAC + Rucio combined service for all the production tasks. Developed RucioFileCatalog
- **CTAO** (Cherenkov Telescope Array Observatory)
  - MC production. Develop Production System
- **JUNO** neutrino experiment (*see Xiaomei's talk*)
  - WMS + DMS, Dataset support.
- **ILC/CLIC, CERN**
  - Future accelerator detector modelling, WMS+DMS Developed
- **EGI Workload Manager**
  - Multiple VOs: WeNMR, biomed, Pierre Auger, KM3NeT,…
- **GridPP DIRAC service**
  - Multiple VOs: T2K, NA62, Euclid, …
  - Developed CloudComputingElement, Multi-VO Catalog

Projects at JINR



BM@N

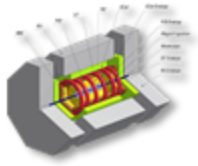
Работает
Монте-Карло,
Анализ,
Обработка данных с детектора
8й сеанс: 436 TB



MPD

Декабрь 2025
Монте-Карло,
Анализ

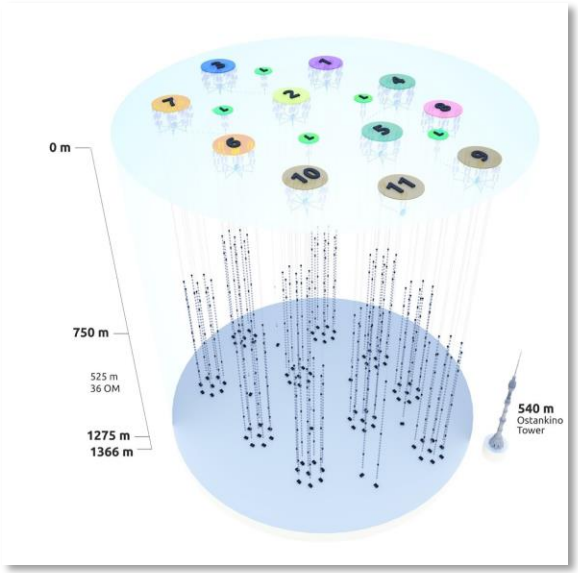
Планируемый поток данных: 7 GB/s



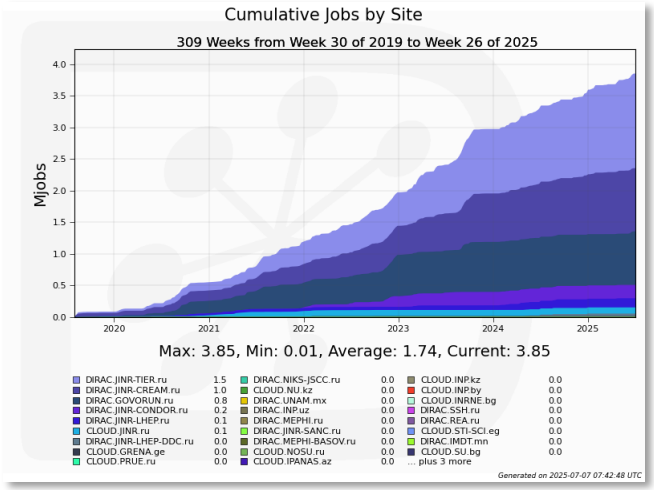
SPD NICA

~ 2028
Монте-Карло,
Анализ

Планируемый поток данных: 20 GB/s



- DIRAC based production systems operations in NICA experiments as well as for Baikal GVD
  - MC production, data analysis
- JINR contributions to DIRAC (*I.Pelevanyuk*):
  - Incorporating OpenNebula cloud sites (from JINR states)
  - Detailed job parameters monitoring system for fine tuning of the site resources usage
  - 14

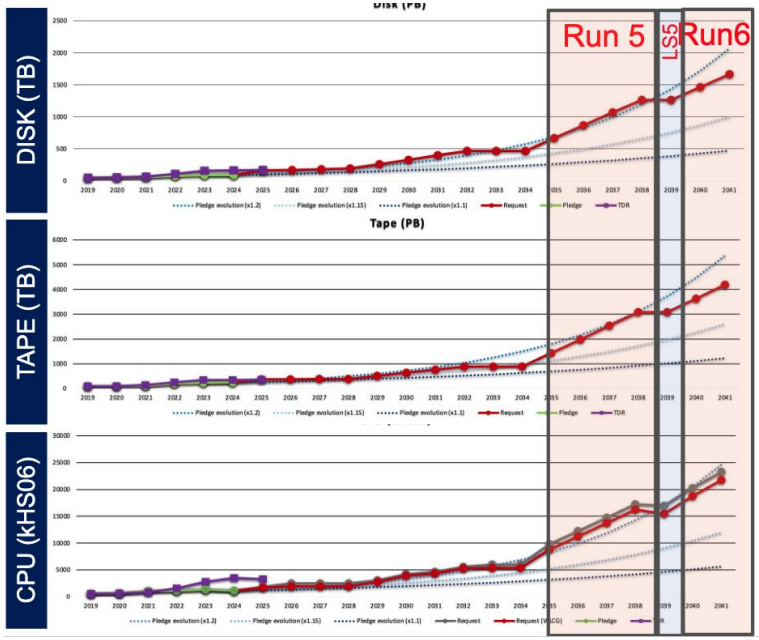


Upgrade: needs

- DIRAC is a successful project
- But:
  - Overly complex development and deployment
  - Late on standards (https, tokens, ad hoc monitoring)
  - Old-fashioned design (RPC vs REST)
  - Multi-VO but not designed from the beginning to do so
  - The list can continue
- It feels like we are the end of the technology cycle
- If we want to keep the project successful we have to make a major upgrade :

Upgrade: requirements

- Multi-VO by the base design, extendable
- Appealing to new communities and developers
- Easy to deploy, single entry point
- Interoperability
- Based on actual standards
- Scalable for LHCb Upgrade II



Still **DIRAC** in terms of functionality

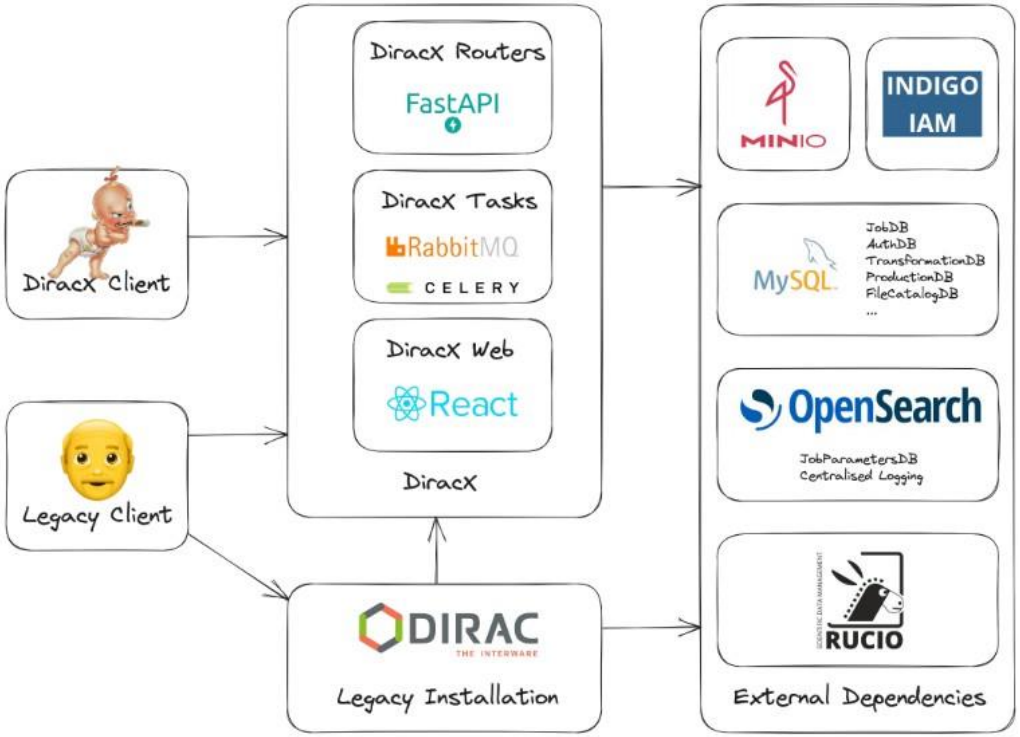
Upgrade: technologies

Major technologies :

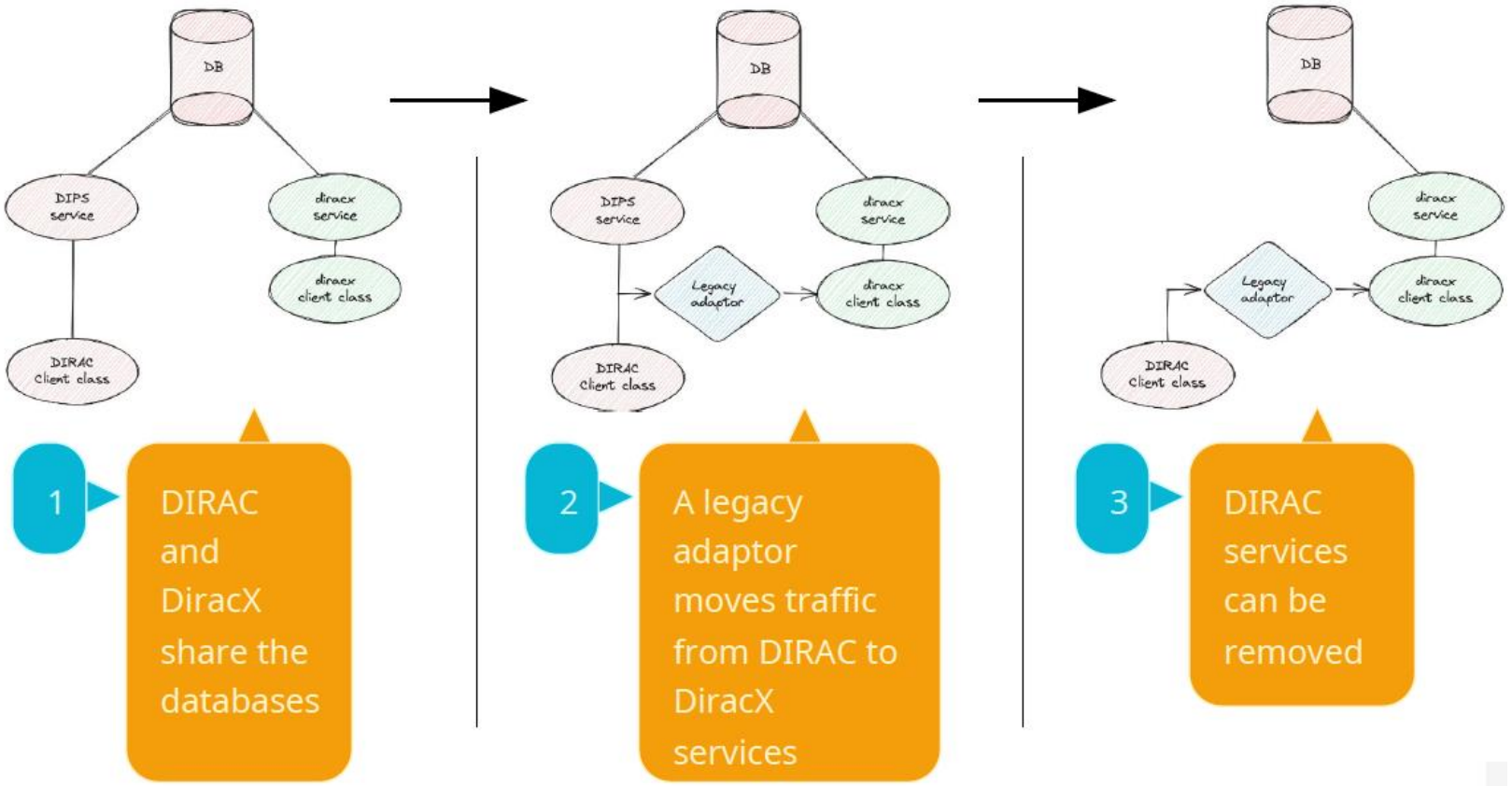
- Services
  - DiracX Web APIs with
  - APIs documented with
  - Following the specification by
- Diracx Web Portal
  - NextJS
  - Material UI
  - TypeScript
- Deployment
  - Kubernetes + Helm

DIRAC/DiracX architecture

- Running in parallel DIRAC and DiracX components
  - To ensure smooth transition of production installations
- Keeping common databases for DIRAC and DiracX components
  - **DIRAC 9.0**
  - **DiracX 0.1**
- Different installations can migrate at different pace
  - **DIRAC 8.0** will be supported up to 2027

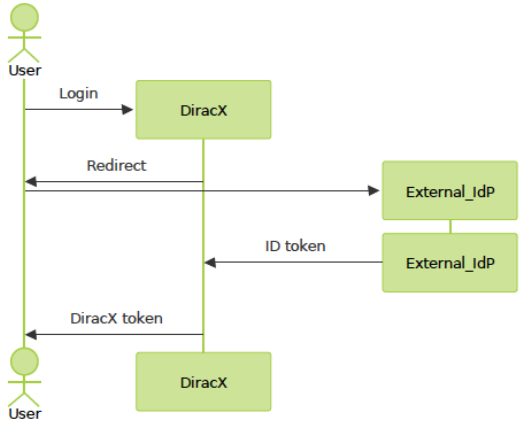


DIRAC->DiracX migration



Tokens

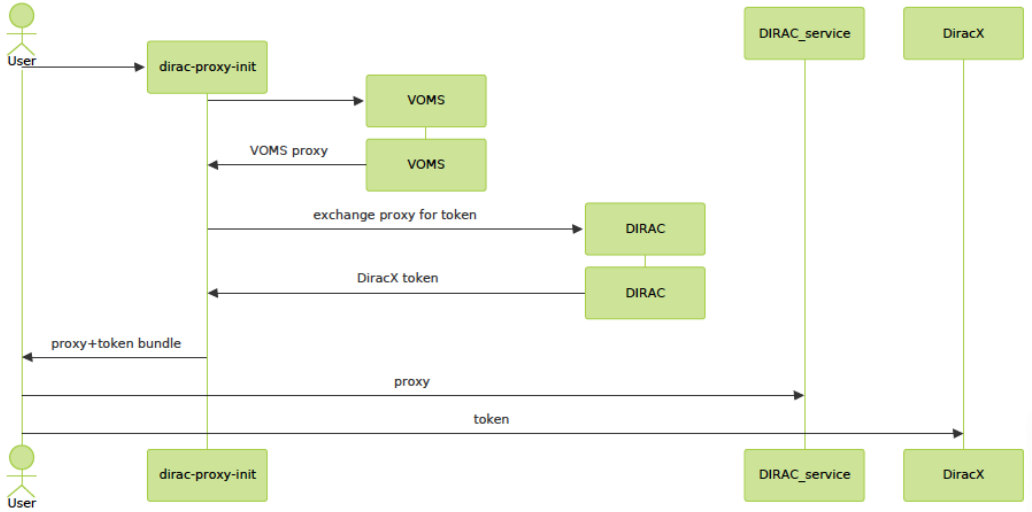
DiracX Authorization with "standard" Authorization Code Flow redirecting to IdP



Note

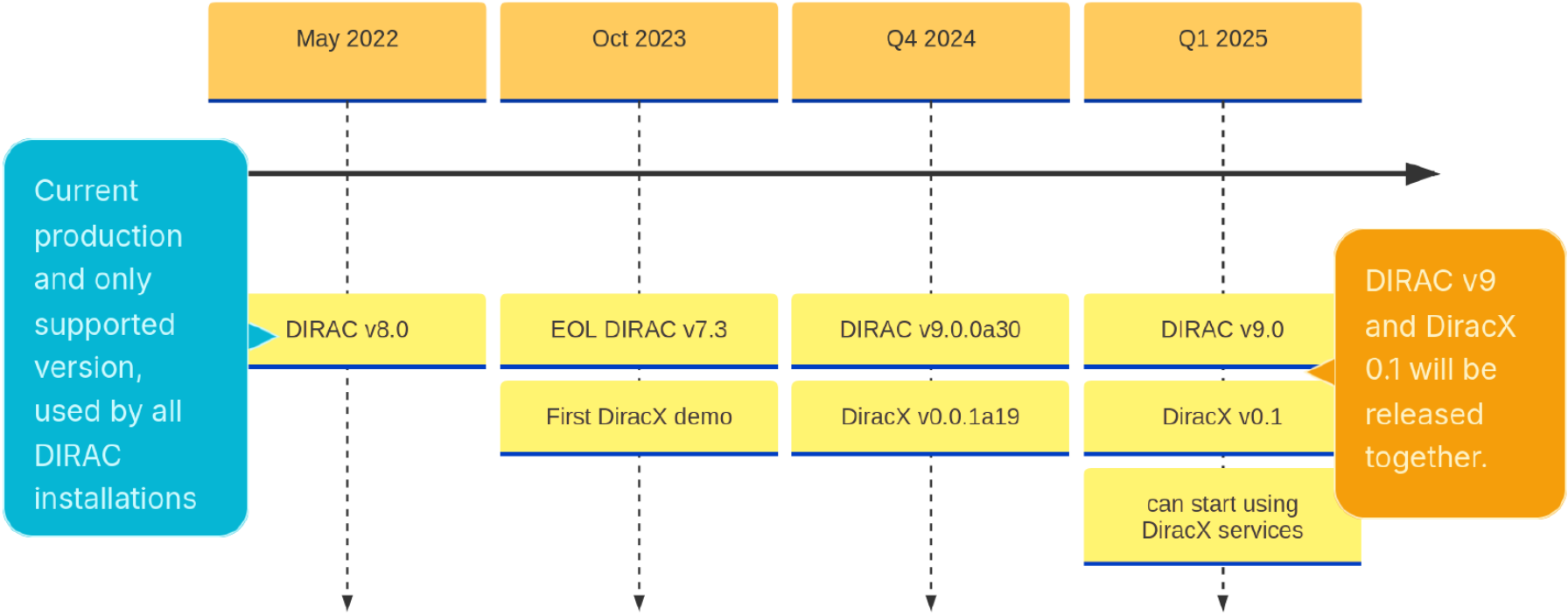
DiracX delivers its own tokens, they are not the same tokens used for the Grid endpoints

DIRAC DiracX: working with proxies and tokens



- Pilot submission with tokens is functional
- Storage access – work in progress
  - Studying possibility for DIRAC FC to mint file access tokens (with high rates) to be accepted by storage providers

Releases



- LHCb deployed DIRAC9/DiracX0.1 in production in April'25
  - During the LHC technical stop, 10 days of the service shutdown
    - Together with several database optimizations
  - Successful restart after several fixes done
- Invaluable experience for other DIRAC installations to follow
- Gradular migration of all the DIRAC subsystems in the next 1-2 years

DIRAC
THE INTERWARE

Conclusion

- DIRAC is an example of a product that evolved from a single experiment development to an open-source project exploited by multiple scientifique communities
- DIRAC offers a complete solution for all the computing and data management tasks for research communities
- It is used successfully by multiple scientific collaborations and it is offered as a service by multiple grid infratsructures
- DiracX is a major upgrade to meet the requirements of large user communities for the years to come
  - New code but it is still DIRAC that you are used to !