From DIRAC to DiracX

A.Tsaregorodtsev,
CPPM-IN2P3-CNRS, Marseille,
CEPC International Workshop, Guangzhou
6-10 November 2025

What is it ? Ask Deepseek

Deepseek: role of DIRAC in the CEPC Project ?

"… The role of DIRAC in the CEPC project is **fundamental**. It is the critical software backbone that transforms a collection of independent computing centers into a powerful, integrated, and efficient data processing factory…"

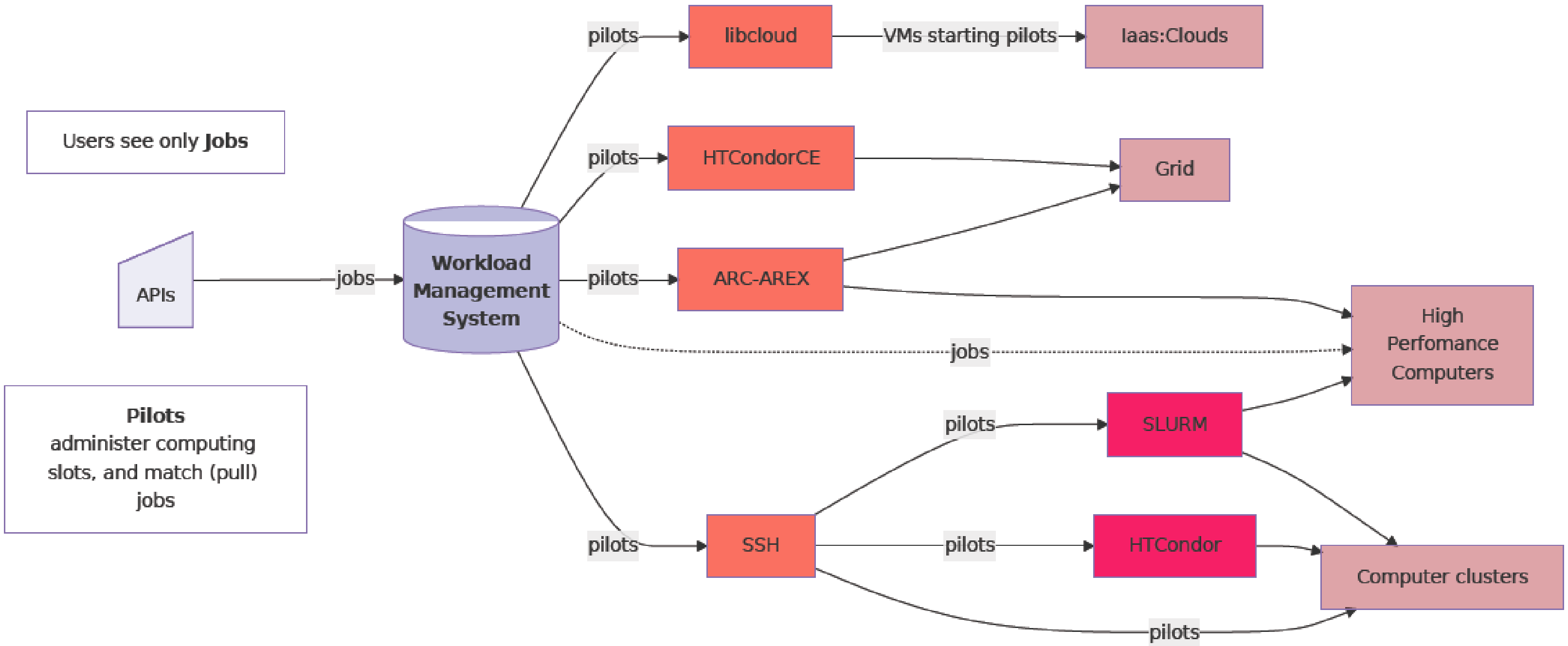
In a nutshell

- 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

JUNO
Belle II
cta
cherenkov telescope array
FUTURE CIRCULAR COLLIDER
ilc
international linear collider
LHCb
BESIII
GridPP
UK Computing for Particle Physics
PIERRE AUGER OBSERVATORY
egi
NA62
T2K
we-nmr
HYPERK
LZ
Euclid consortium
CEPC
NICA

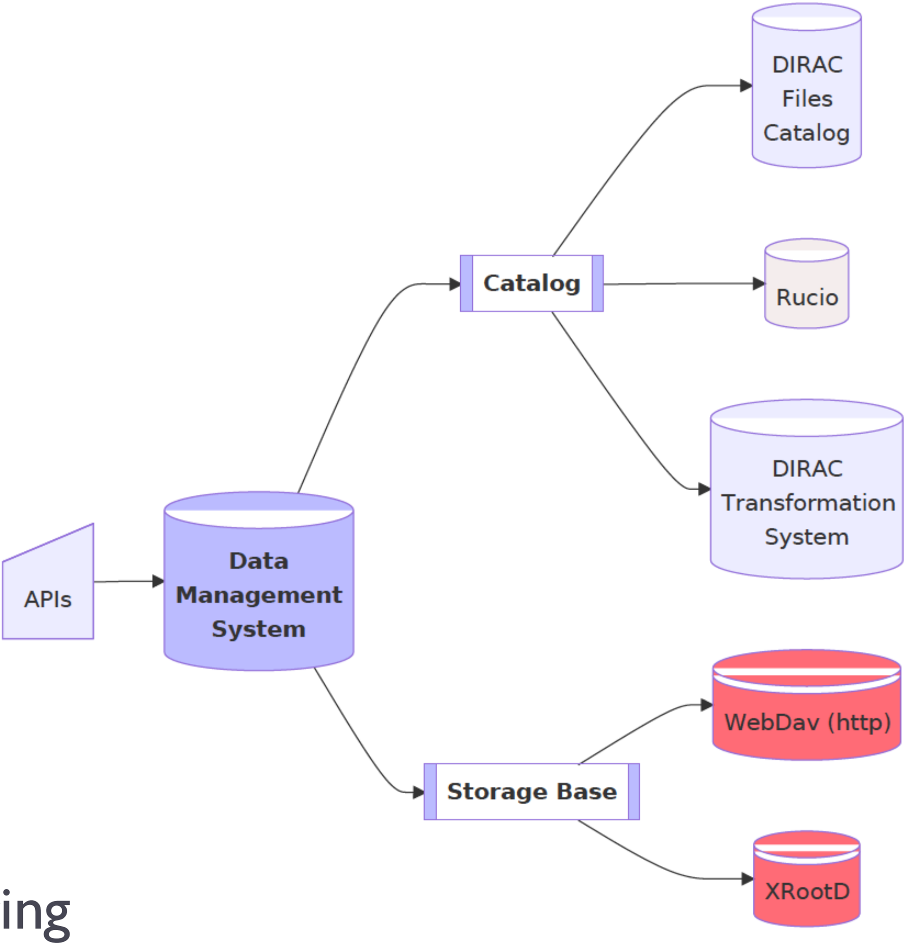
Managing workloads



- « Pull » model with pilot jobs (HTC, Cloud reources)
- « 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

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



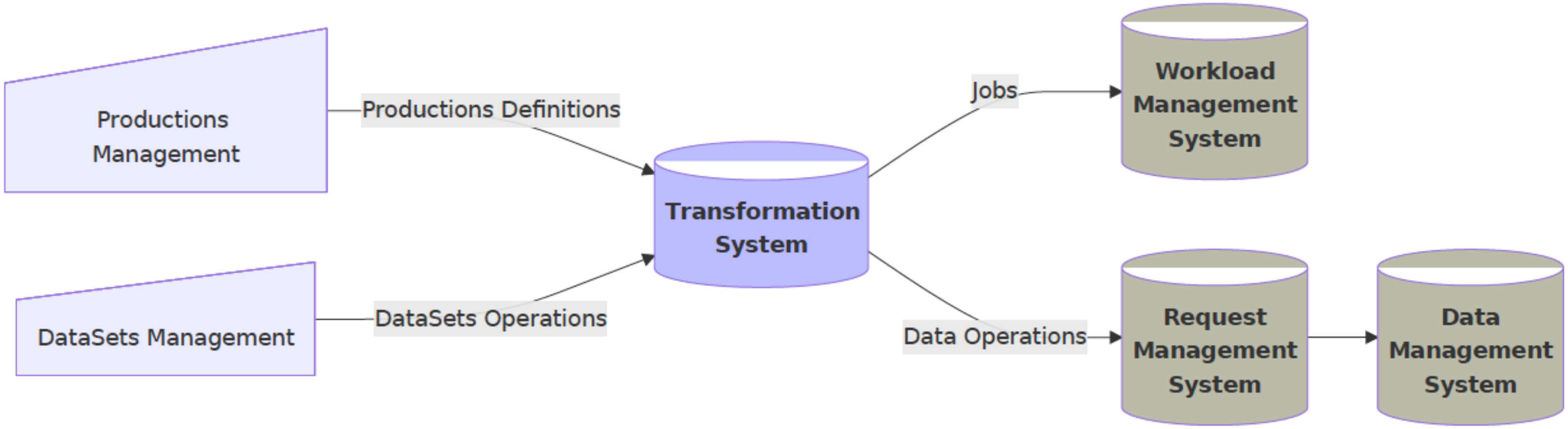
all in 1 solution

- DIRAC is a complete solution including both **workload** and **data** management subsystems
  - Allows to stay in the same software framework for all the tasks of a large experiment
  - Requires a single team of developers and administrators of a distributed computing project
    - Sharing professional experience for different tasks
    - Optimal for managing manpower of the project

- Possibility to use Rucio DMS along with the DIRAC WMS
  - RucioFileCatalog developed by the Belle II experiment
  - Popular solution explored by several communities, e.g. CTAO

- Easily customizable to specific needs of particular communities with the Extensions mechanism

Transformations



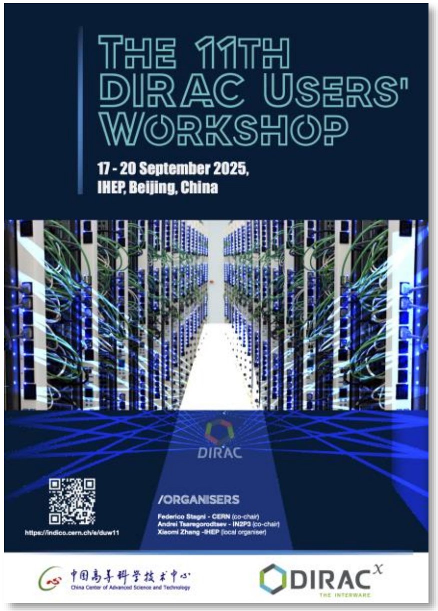
- Transformations System is an engine for managing complex **data-driven** workflows
  - Automation of common tasks
    - Creation, resubmission, deletion of jobs
    - Replication, registration, removal of data files
  - Handling millions of jobs and files
    - Massive operations validation, failure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

Open source project

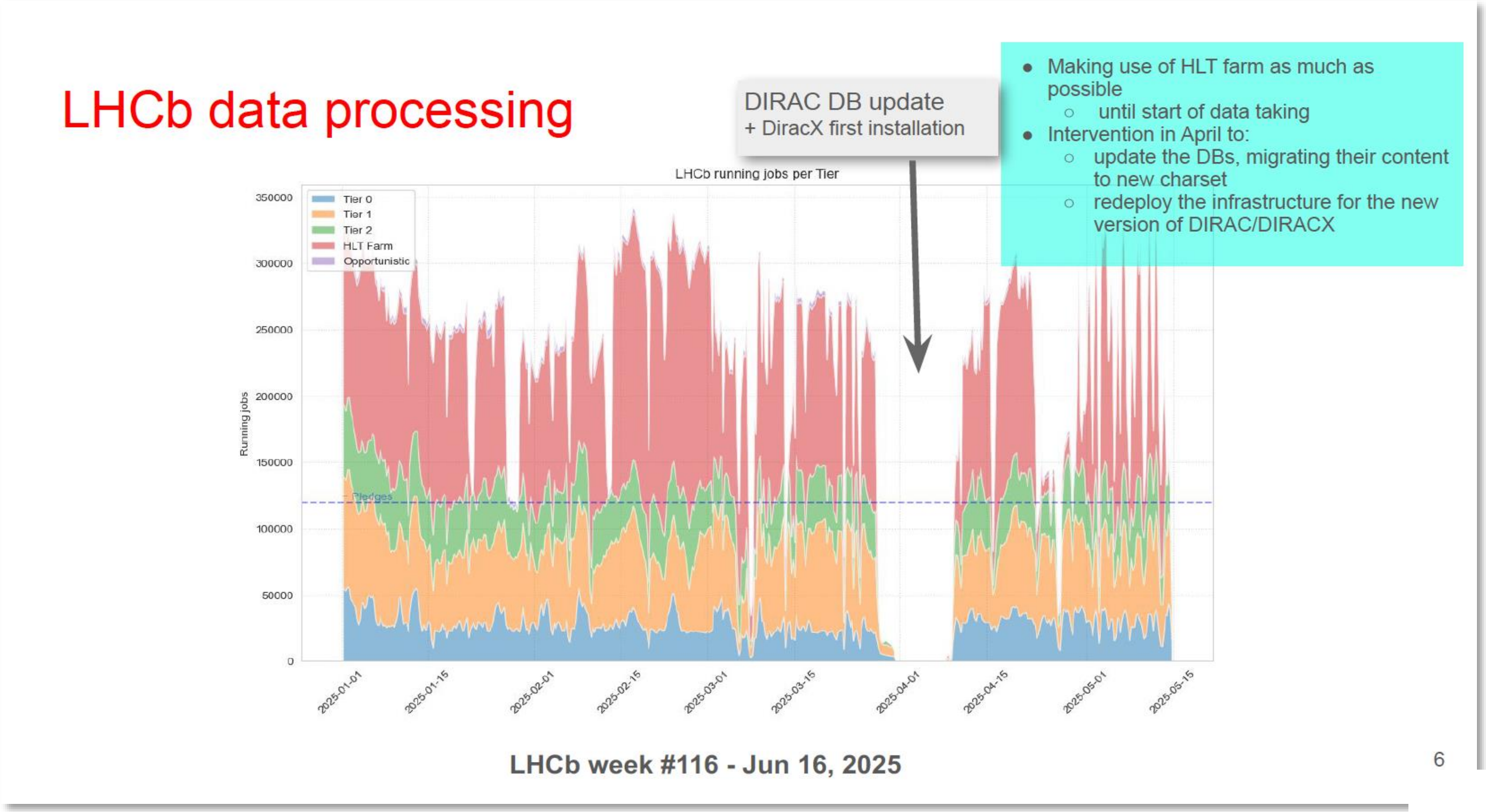
- Driven by the needs of the user communities
  - With multiple contributions from community developers
  - Example: DIRAC File Catalog (DFC) was developed initially for ILC and BES III experiments, it is used now by several experiments including LHCb.

- 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 last one is the11th DUW in IHEP, Beijing, 17-20 September ***https://indico.cern.ch/event/1433941/***


THE 11TH DIRAC USERS' WORKSHOP
17 - 20 September 2025,
IHEP, Beijing, China
DIRAC
https://indico.cern.ch/e/duw11
/ORGANISERS
Federico Stagni - CERN (co-chair)
Andrei Tsaregorodtsev - IN2P3 (co-chair)
Xiaomi Zhang -IHEP (local organiser)
中国高等科学技术中心
China Center of Advanced Science and Technology
DIRAC x
THE INTERWARE

Users: LHCb



LHCb data processing

DIRAC DB update
+ DiracX first installation

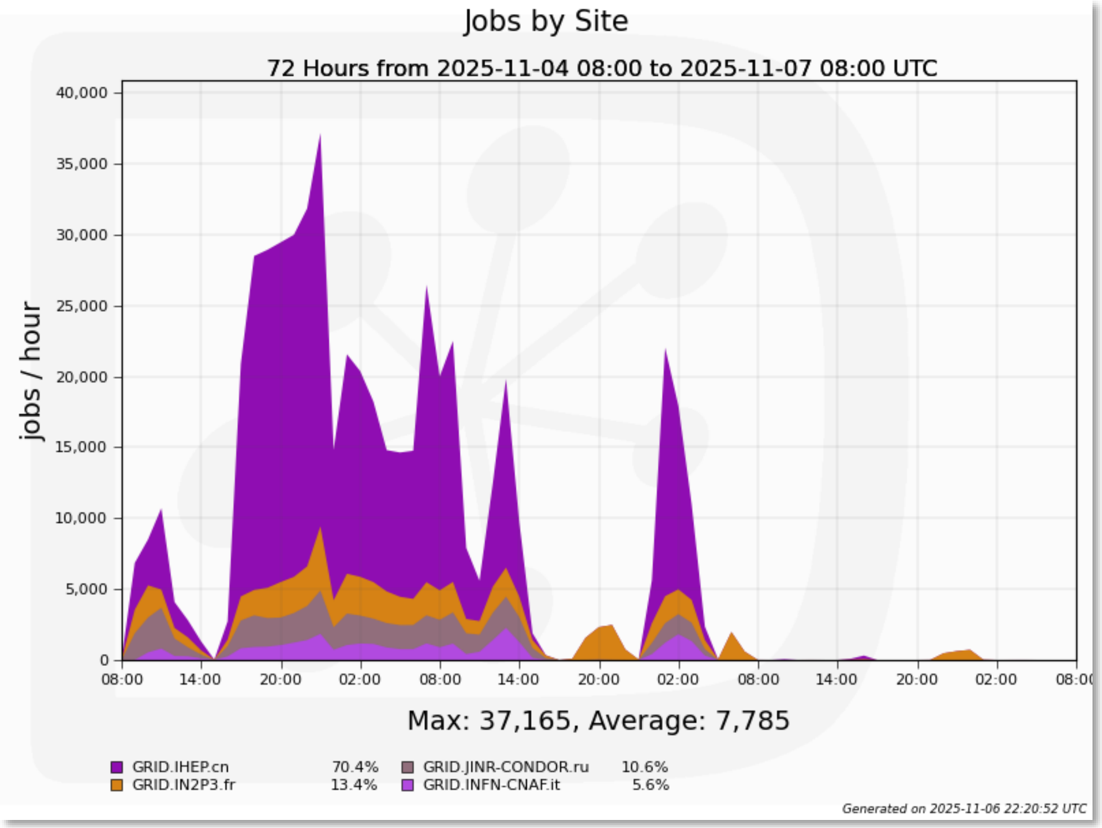
- Making use of HLT farm as much as possible
  - until start of data taking
- Intervention in April to:
  - update the DBs, migrating their content to new charset
  - redeploy the infrastructure for the new version of DIRAC/DIRACX

LHCb week #116 - Jun 16, 2025

6

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

Users: JUNO



- DIRAC service at IHEP for JUNO, BES III, CEPC, …
  - Example: JUNO production data processing in November 2025:
    - Up to 10Hz rate of executed jobs
    - **Two times more than the LHCb's average value !**

Users

- **Belle II, KEK**
  - Using DIRAC + Rucio combined service for all the production tasks. Developed RucioFileCatalog
- **CTAO** (Cherenkov Telescope Array Observatory)
  - MC production. Developed DIRAC Production System
- **ILC/CLIC, CERN**
  - Future accelerator detector modelling, WMS+DMS Developed service supervision tools
- **EGI Workload Manager**
  - Multiple VOs: WeNMR, biomed, Pierre Auger, KM3NeT,…
- **GridPP DIRAC service**
  - Multiple VOs: T2K, NA62, Euclid, …
  - Developed CloudComputingElement, Multi-VO Catalog
- **JINR DIRAC service**
  - Multiple JINR experiments: NICA, BM@N, MPD, Baikal
  - Developed support for OpenNebula clouds

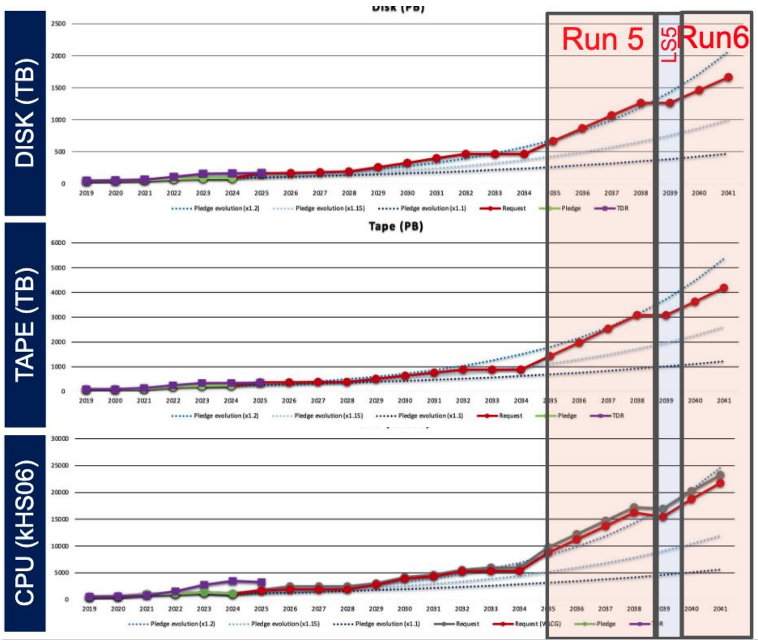
Upgrade: needs

- 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 :

DIRAC^{x} THE INTERWARE

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

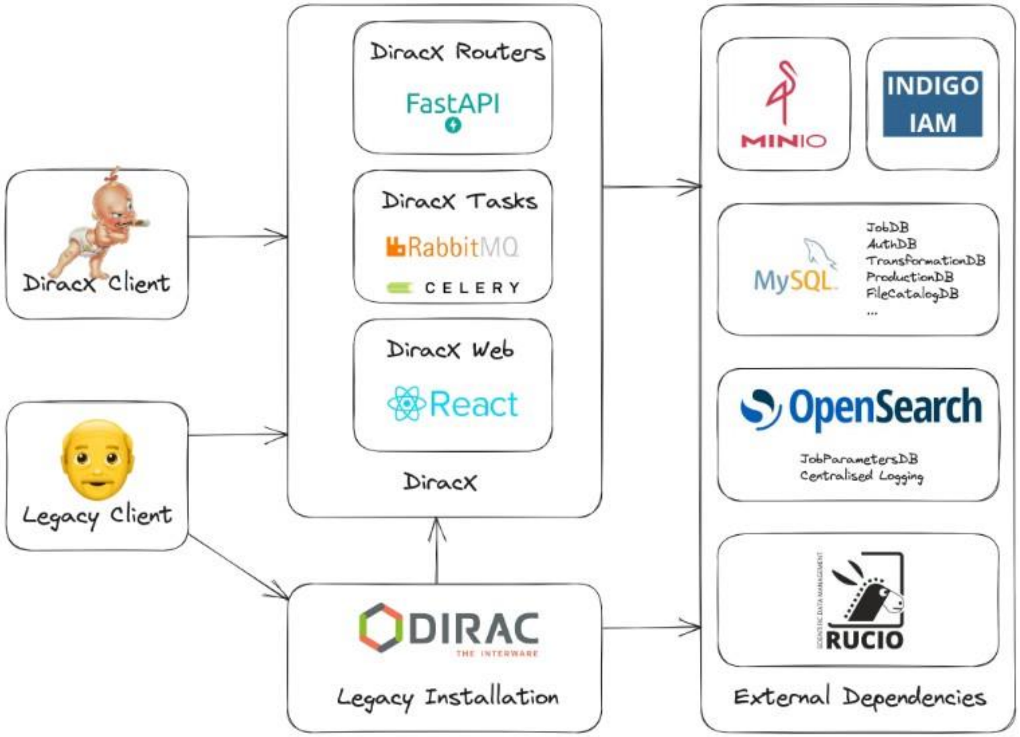
Major technologies :

- Services
  - DiracX Web APIs with
  - APIs documented with
  - Following the specification by
- Diracx Web Portal
  - NextJS
  - Material UI
  - TypeScript
- Deployment
  - Kubernetes + Helm
- Security
  - OAuth2/OIDC tokens

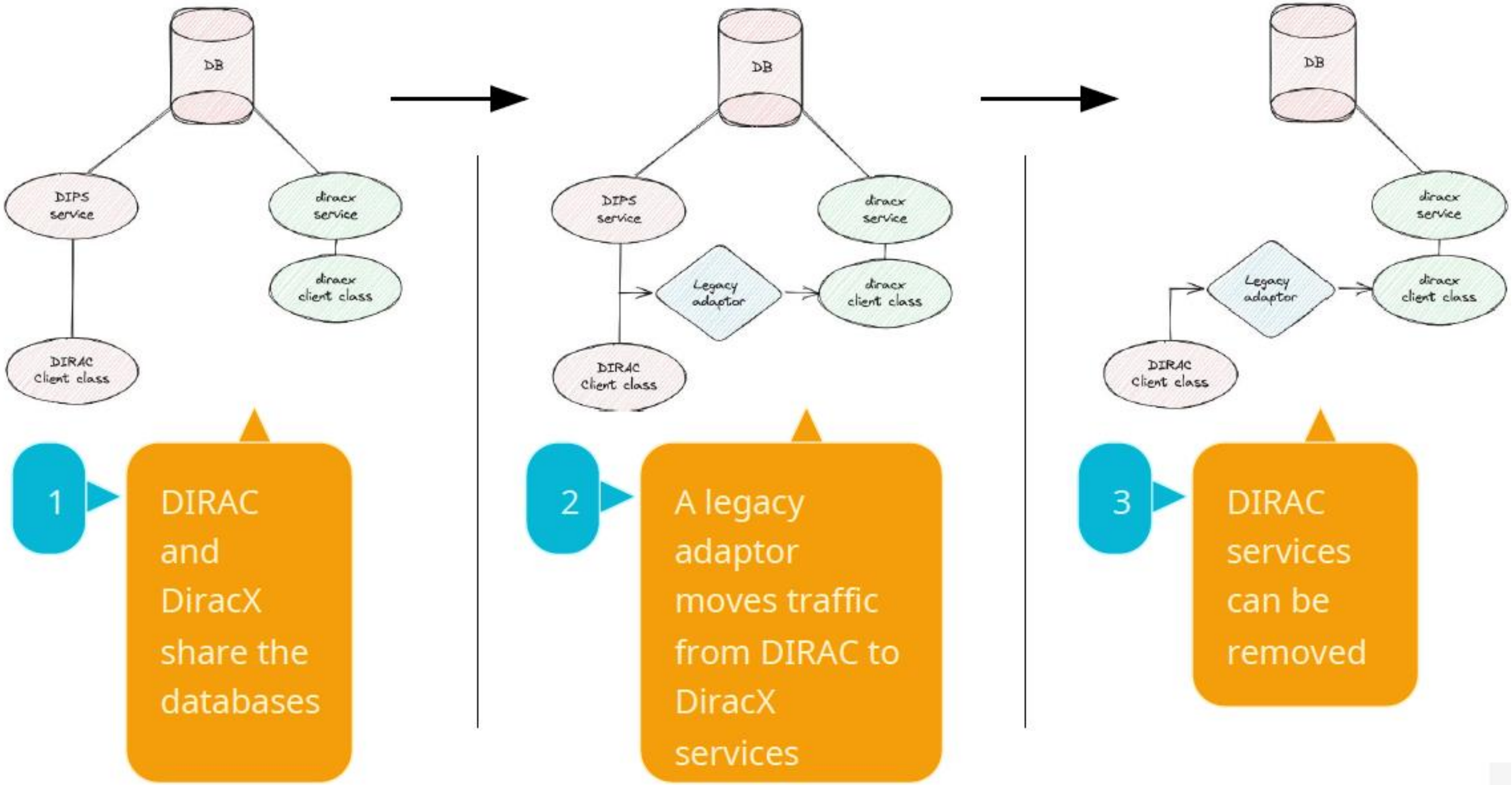
FastAPI
Swagger
OPEN API
INITIATIVE
NEXT.JS
TS
HELM

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

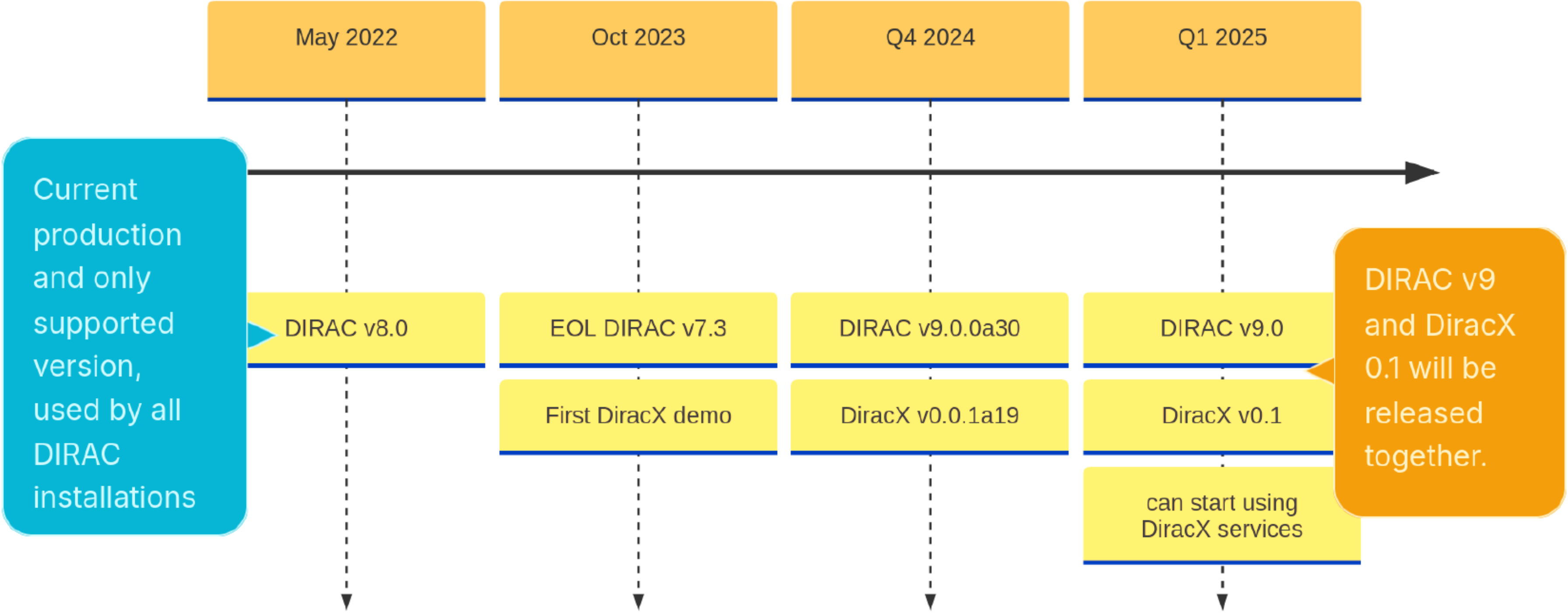


DIRAC and DiracX share the databases

A legacy adaptor moves traffic from DIRAC to DiracX services

DIRAC services can be removed

Releases

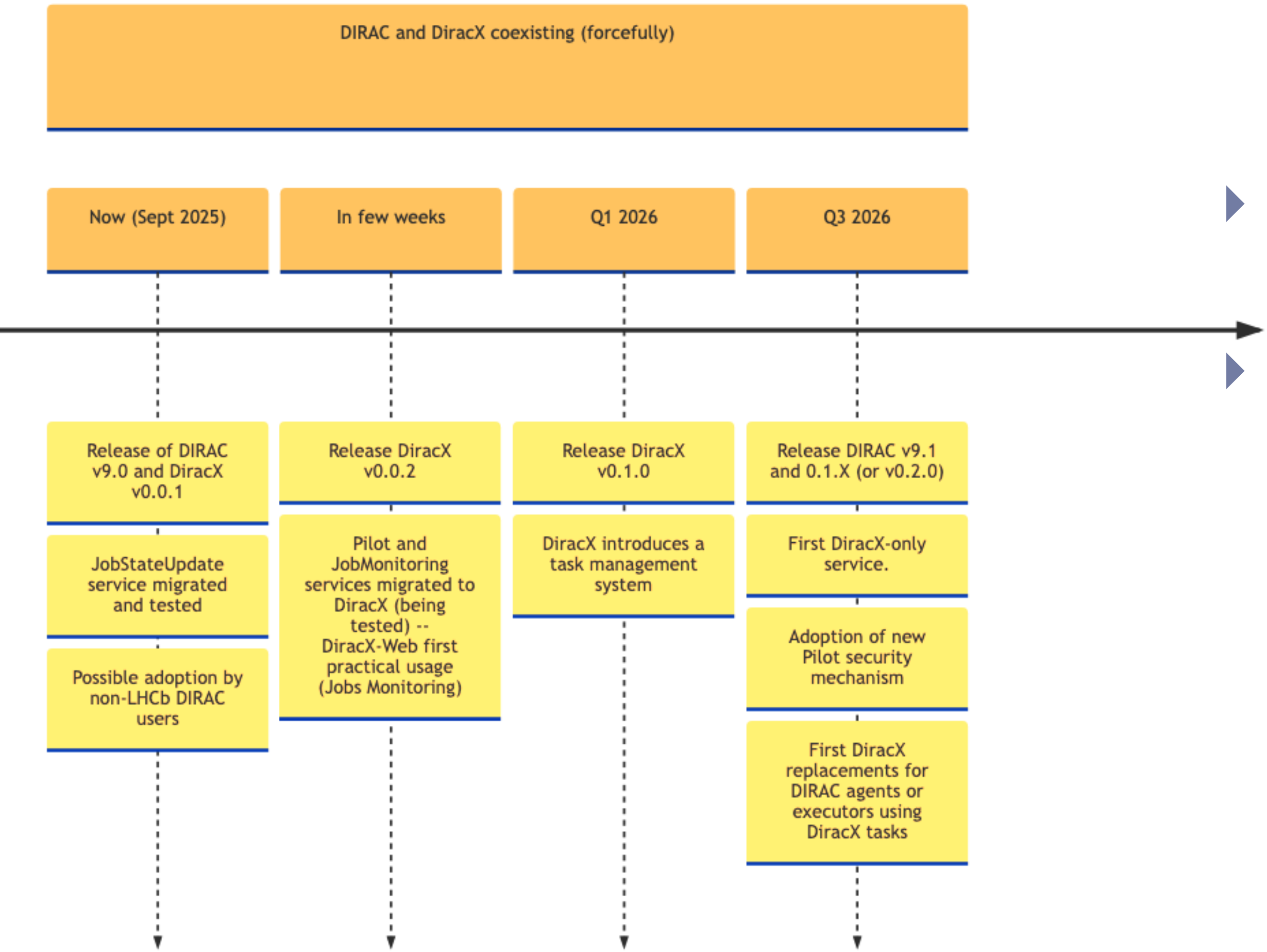


May 2022
Oct 2023
Q4 2024
Q1 2025
Current production and only supported version, used by all DIRAC installations
DIRAC v8.0
EOL DIRAC v7.3
First DiracX demo
DIRAC v9.0.0a30
DiracX v0.0.1a19
DIRAC v9.0
DiracX v0.1
can start using DiracX services
DIRAC v9 and DiracX 0.1 will be released together.

- 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
  - First Release DIRAC v9.0 + DiracX v0.1 in September'26

Roadmap

2026 Developments



DIRAC and DiracX coexisting (forcefully)

Now (Sept 2025)	In few weeks	Q1 2026	Q3 2026
Release of DIRAC v9.0 and DiracX v0.0.1	Release DiracX v0.0.2	Release DiracX v0.1.0	Release DIRAC v9.1 and 0.1.X (or v0.2.0)
JobStateUpdate service migrated and tested	Pilot and JobMonitoring services migrated to DiracX (being tested) -- DiracX-Web first practical usage (Jobs Monitoring)	DiracX introduces a task management system	First DiracX-only service.
Possible adoption by non-LHCb DIRAC users			Adoption of new Pilot security mechanism
			First DiracX replacements for DIRAC agents or executors using DiracX tasks

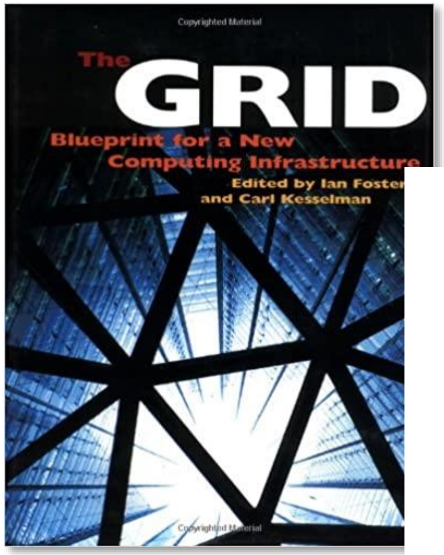
- Q4’2026 EOL
  - DIRAC v8
- Q4’2027
  - DiracX v1.0
  - Full replacement of DIRAC v9
  - Very indicative guess

Conclusion

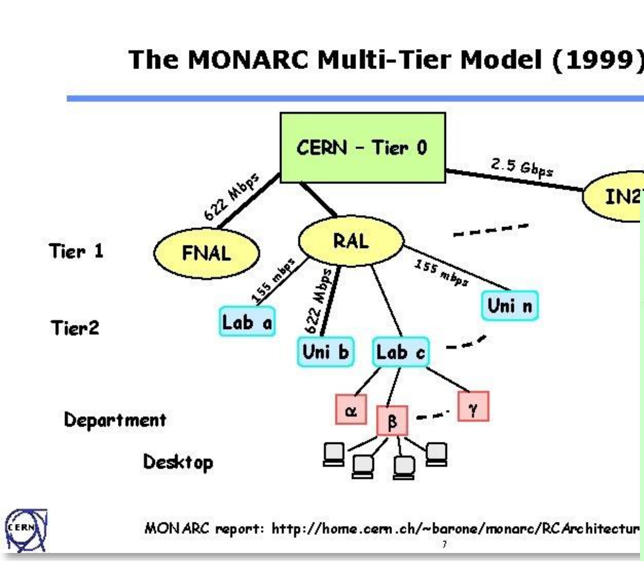
- DIRAC is an example of a product that evolved from a single experiment development to an open-source project exploited by multiple scientifique communities
- DIRAC offers a complete solution for all the computing and data management tasks for research communities
- It is used successfully by multiple 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 !

Back-up slides

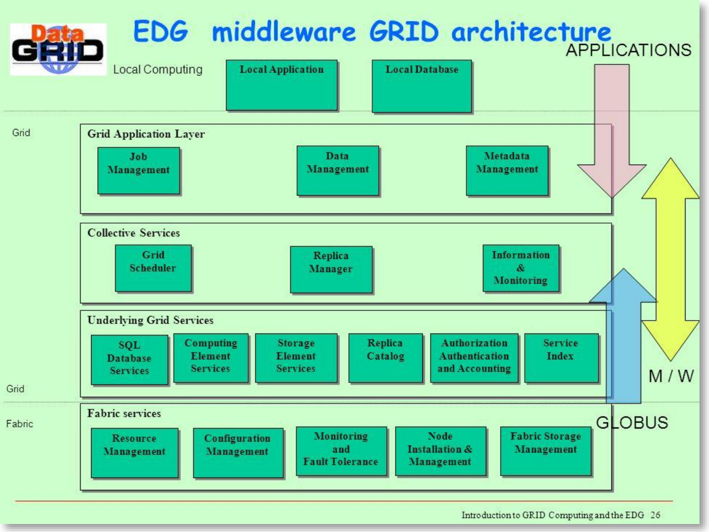
At the beginning…


Copyrighted Material
The GRID
Blueprint for a New Computing Infrastructure
Edited by Ian Foster and Carl Kesselman

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


The MONARC Multi-Tier Model (1999)

MONARC report: http://home.cern.ch/~barone/monarc/RCArchitectur


EDG middleware GRID architecture

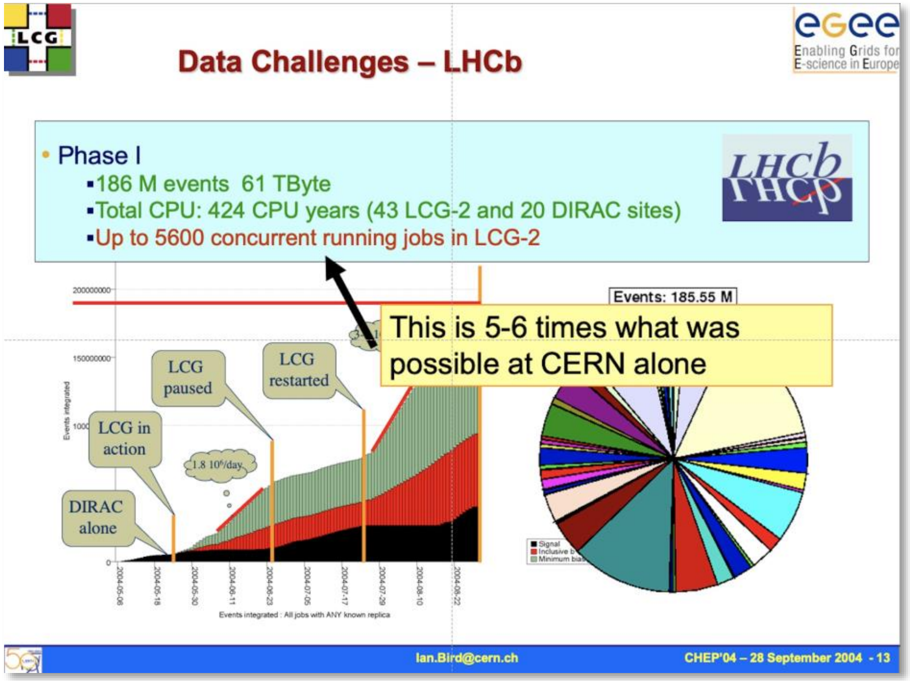
Introduction to GRID Computing and the EDG 26

- 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 the LHC Computing 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 LHC experiments



Data Challenges – LHCb

- Phase I
  - 186 M events 61 TByte
  - Total CPU: 424 CPU years (43 LCG-2 and 20 DIRAC sites)
  - Up to 5600 concurrent running jobs in LCG-2

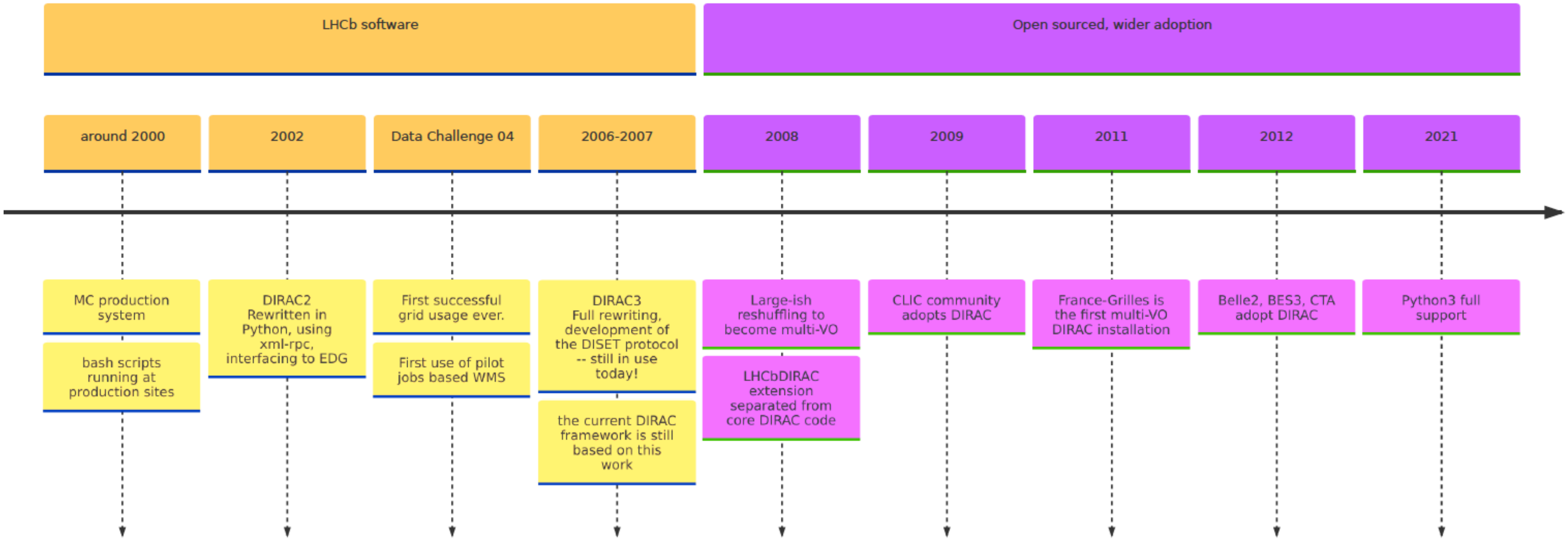
This is 5-6 times what was possible at CERN alone

Ian.Bird@cern.ch

CHEP'04 – 28 September 2004 - 13

- 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



LHCb software

Open sourced, wider adoption

around 2000
MC production system
bash scripts running at production sites

2002
DIRAC2 Rewritten in Python, using xml-rpc, interfacing to EDG

Data Challenge 04
First successful grid usage ever.
First use of pilot jobs based WMS

2006-2007
DIRAC3 Full rewriting, development of the DISET protocol -- still in use today!
the current DIRAC framework is still based on this work

2008
Large-ish reshuffling to become multi-VO
LHCbDIRAC extension separated from core DIRAC code

2009
CLIC community adopts DIRAC

2011
France-Grilles is the first multi-VO DIRAC installation

2012
Belle2, BES3, CTA adopt DIRAC

2021
Python3 full support

- 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 infrastructure projects

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
- DIRAC-Rucio joint workshops to elaborate common strategies towards better integration of the projects