Status of CEPC Distributed Computing

Xiaomei Zhang
On behalf of CEPC distributed computing group
Nov 10, 2021
The CEPC International Workshop

Contents

- **Current status**
  - **Sites and resources**
  - **Workload management and data management**
- **R&D activities**
  - **Rucio**
  - **XCache**
  - **IAM**
- **Summary**

Reminder

- CEPC data volume estimation at data-taking
  - PB scale in Higgs/W factory
  - EB scale in Z factory
- Distributed computing technology will be used to organize resources in both R&D and data-taking
  - Benefit as much as possible from WLCG middleware and experience
- The CEPC distributed computing prototype has been built up based on DIRAC
- The system was proved to work well for CEPC R&D detector simulation
- CEPC users can access the resource in the system using grid certificate with client installed

Current status

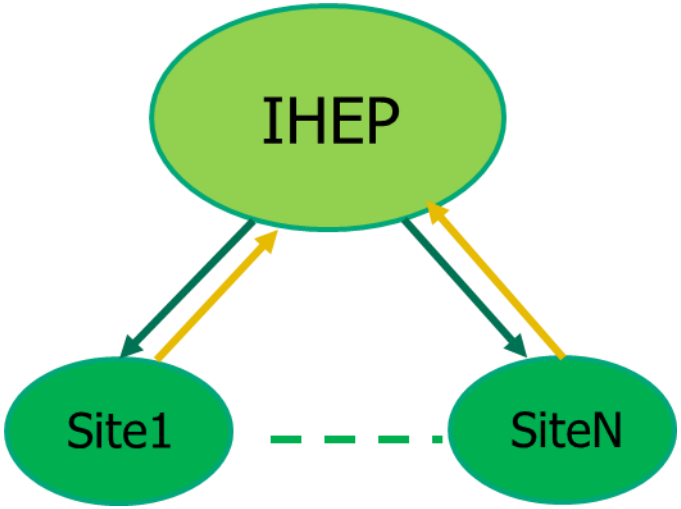
Sites and resources

- Six sites from UK and other China universities
  - ~3000 CPU cores, ~3PB disk
  - Shared with other experiments
- Plan to add 500 dedicated cores for IHEP site before end of this year
- The system has been proved to be able to work well with various resources including Grid, Cluster, Cloud, Commercial Cloud, etc

Site Name	CPU Cores
Grid.IHEP.cn	500
CLOUD.IHEPCLOUD.cn	100
GRID.QMUL.uk	1600
CLUSTER.IPAS.tw	500
CLUSTER.SJTU.cn	100
GRID.LANCASTER.uk	300
Total (Active)	**~3000**

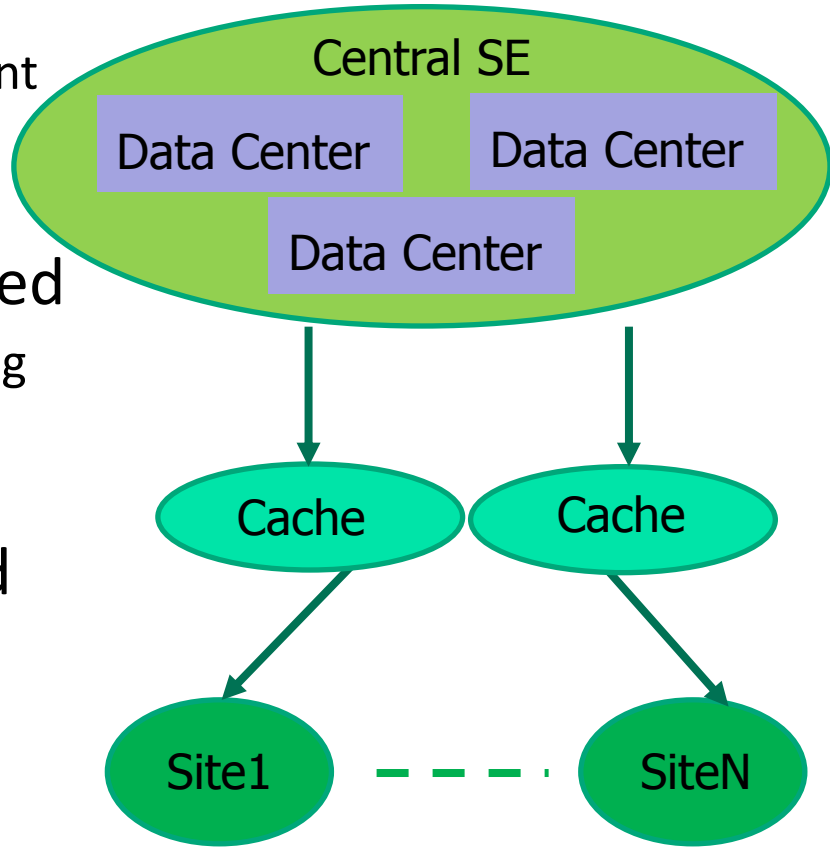
Current computing model

- IHEP as central site
  - Event generation, MC production and analysis
  - Hold central storage for all experiment data
- Remote sites
  - MC production, no requirements of storage
- Data flow
  - IHEP -> Sites, jobs access input files from IHEP
  - Sites -> IHEP, output MC data directly transferred back to IHEP from jobs



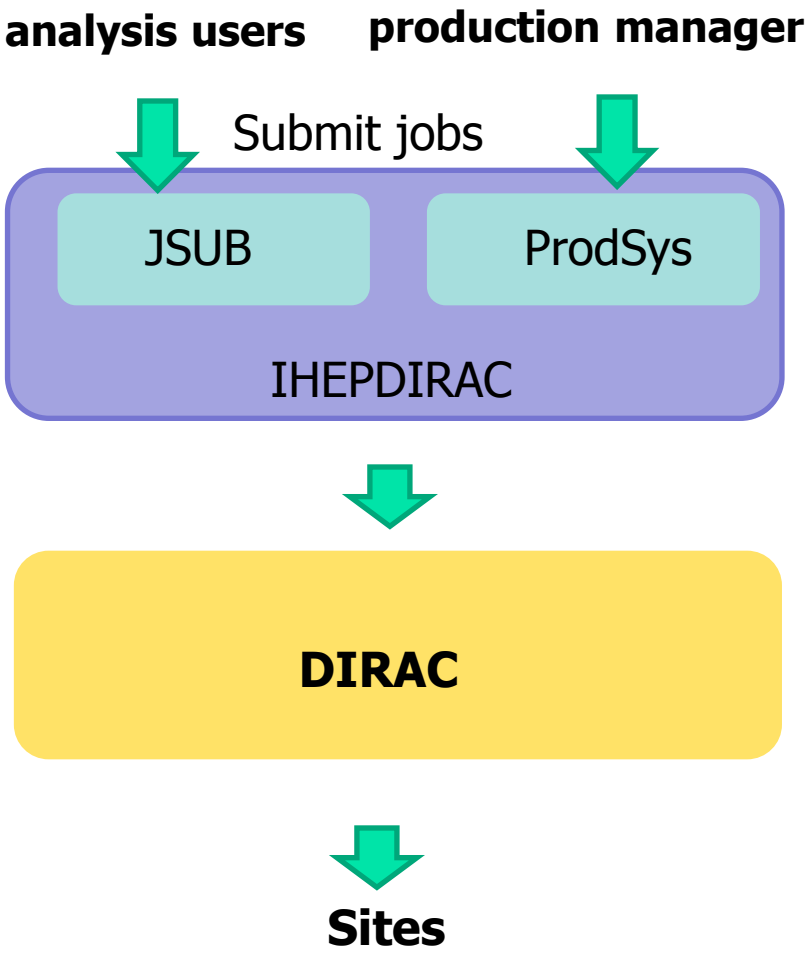
Future computing model

- For future scaling up, simple data model is not enough
  - Single-point central storage could be a weak point
  - Direct remote data access will cause high failure rate when WAN traffic is heavy
- "Data Lake" data model can be considered
  - Robust central storage with federation among big data centers
  - More efficient data access with cache layer
- Advanced data management system and data access policy are needed
  - Rucio and XCache could be a choice



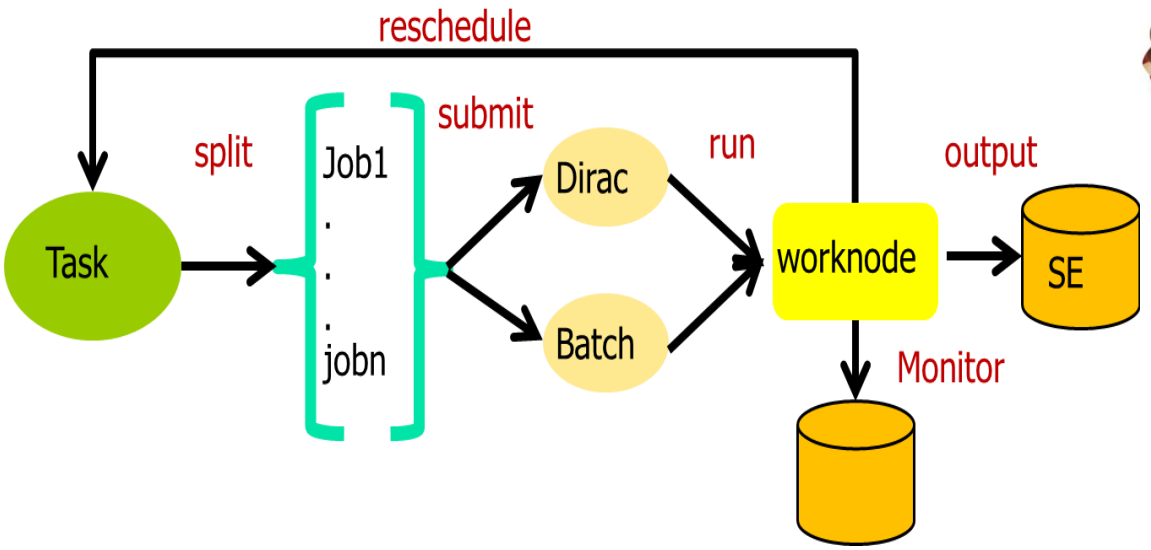
Workload management

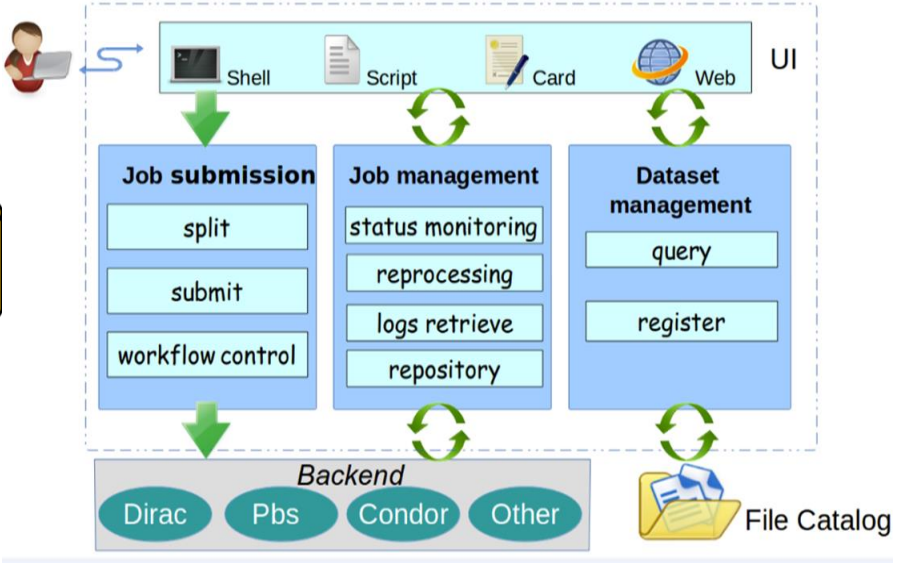
- Manage job submission and work flow
- DIRAC
  - Provide a middle layer between jobs and resources to hide complexity from users
- IHEPDIRAC
  - Job submission and management, more experiment-specific
  - JSUB
    - Massive job submission frontend for **analysis users**
  - ProdSys
    - Submit and manage production tasks for **production groups**



Analysis job submission with JSUB

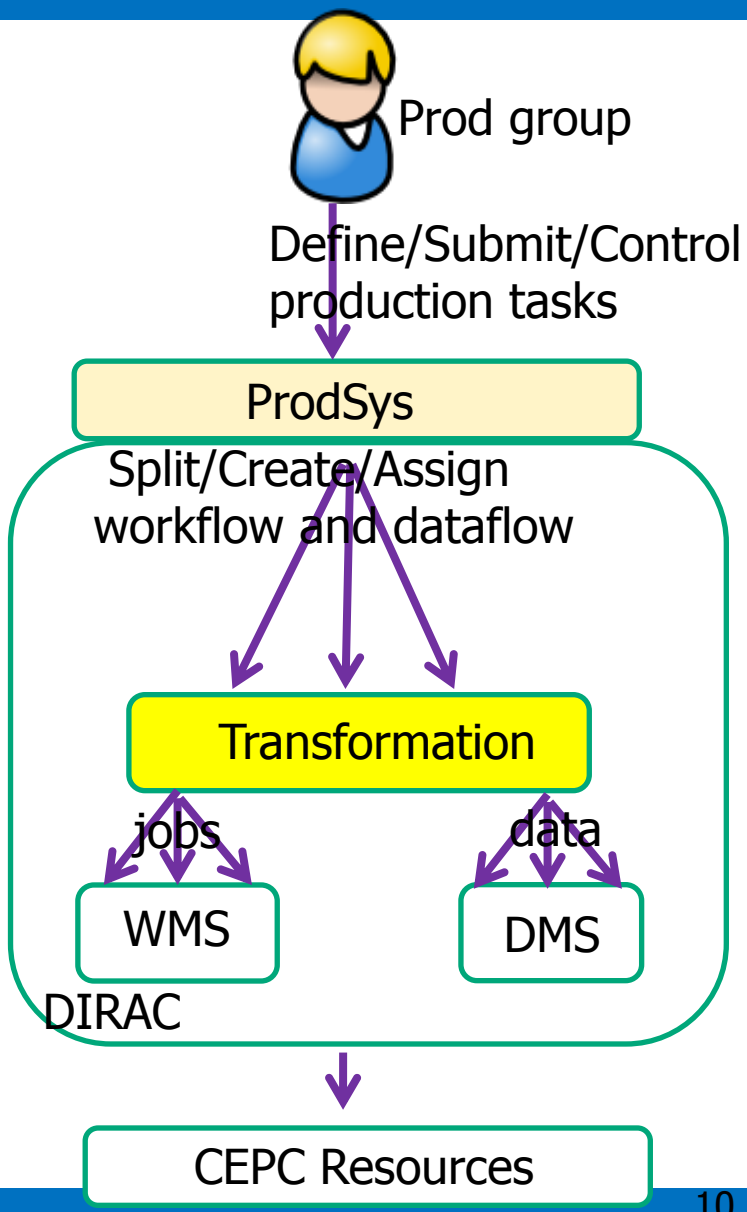
- Help manage life cycle of an analysis task in an automatic way
  - split->submit->run->status monitor->output retrieval -> reschedule
- Main infrastructure is ready, extensible
- Supports to the CEPCSW based on Gaudi is added
  - New extensions for Gaudi framework is being designed
  - Discussions with the SW framework group are needed





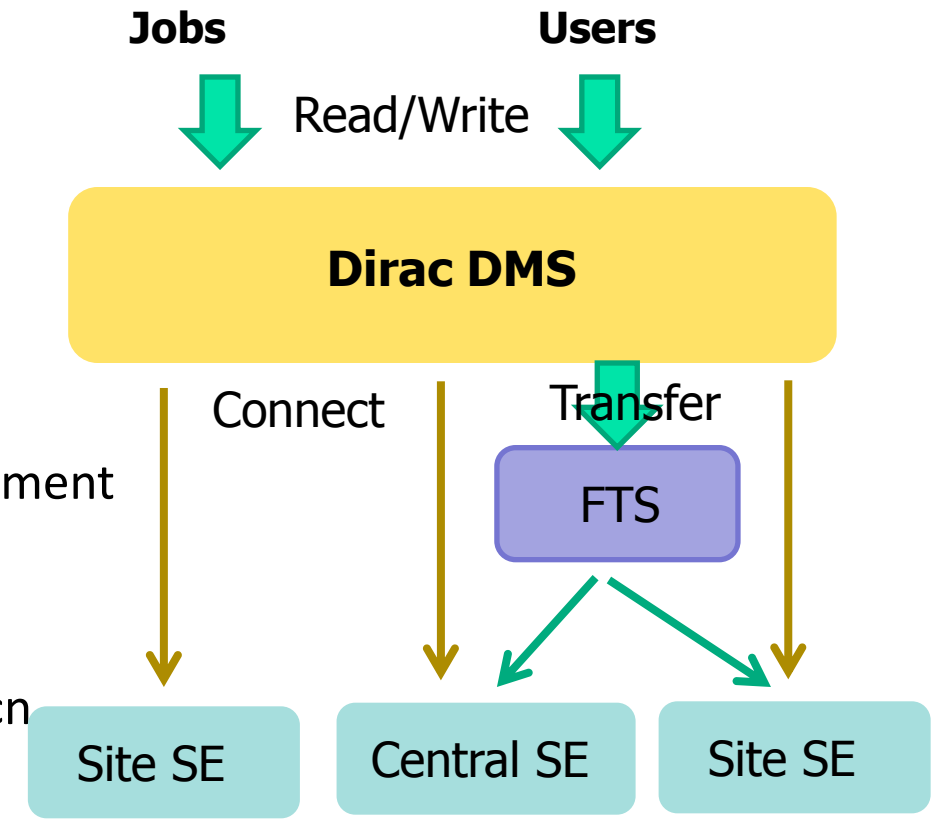
Production task management with ProdSys

- Besides submission of production tasks, ProdSys can
  - Seamlessly work with data management system
  - Deal with massive production workflow and dataflow automatically
- Core infrastructure has been set up
- Supports to the CEPCSW is to be added
  - Interface to support the CEPCSW is being designed
  - Discussions with the SW framework group needed

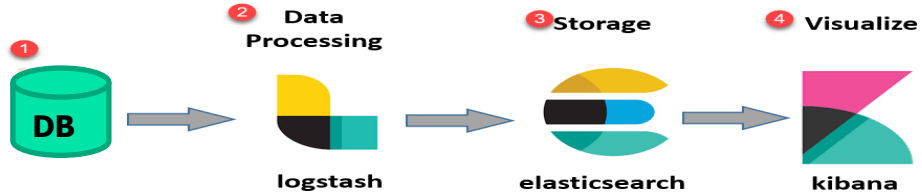


Data management

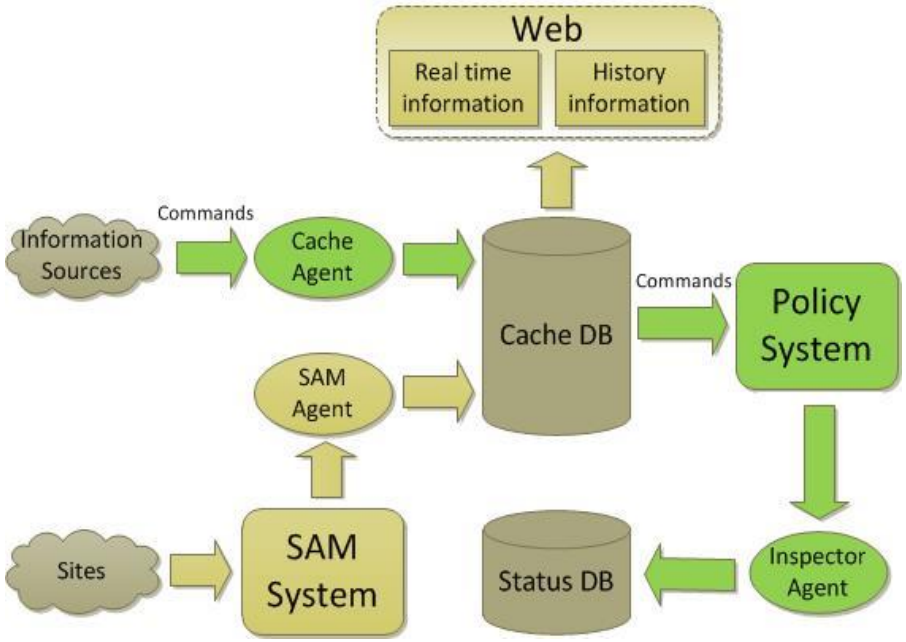
- Manage data placement and data flow globally, and provide interface for jobs and users to access data
- Three layer: Data Management, File Transfer, Storage
- DIRAC Data Management System
  - File Catalogue: global view of data
  - Meta Catalogue: dataset management
  - Rucio DM is in study for possible replacement
- FTS (File Transfer System)
  - Manage low-level file movements
  - fts3 server in IHEP: https://fts3.ihep.ac.cn
- Storage Element (SE)
  - Use StoRM, lustre as its backend
  - Support SRM and gridftp protocol access

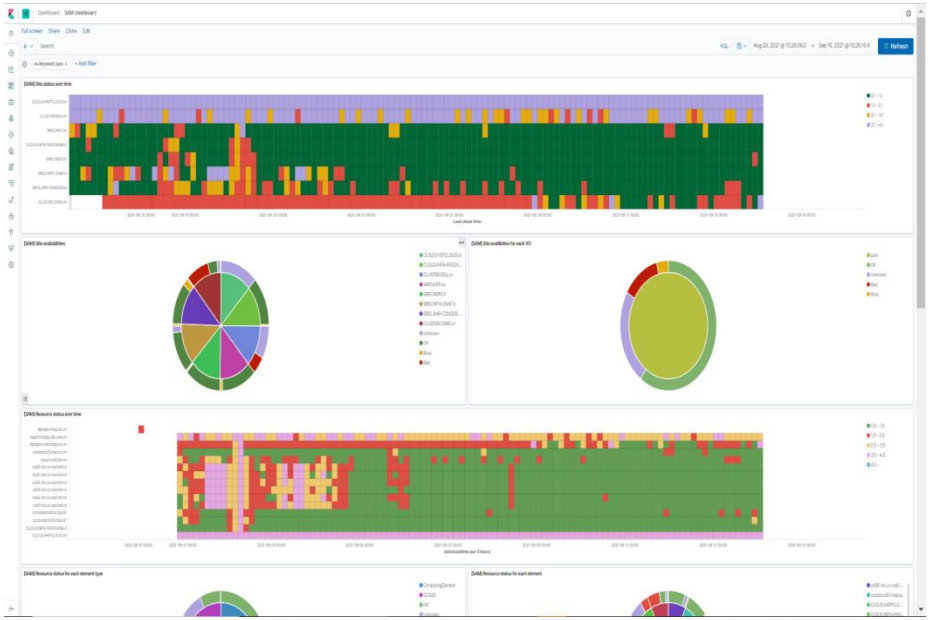


Site monitoring



- Regular site and service status check to get high availability and reliability
- Site monitoring system has been implemented in two ways
  - Active: send out standard CEPC jobs and check results regularly
  - Passive: collect user job status
- Monitoring dashboard has just set up to give a view of sites status using Logstash + ES + Kibana

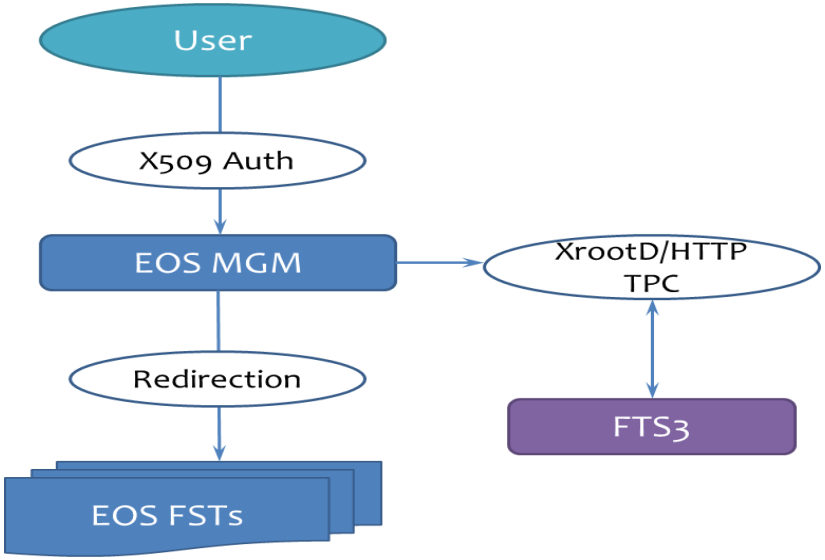




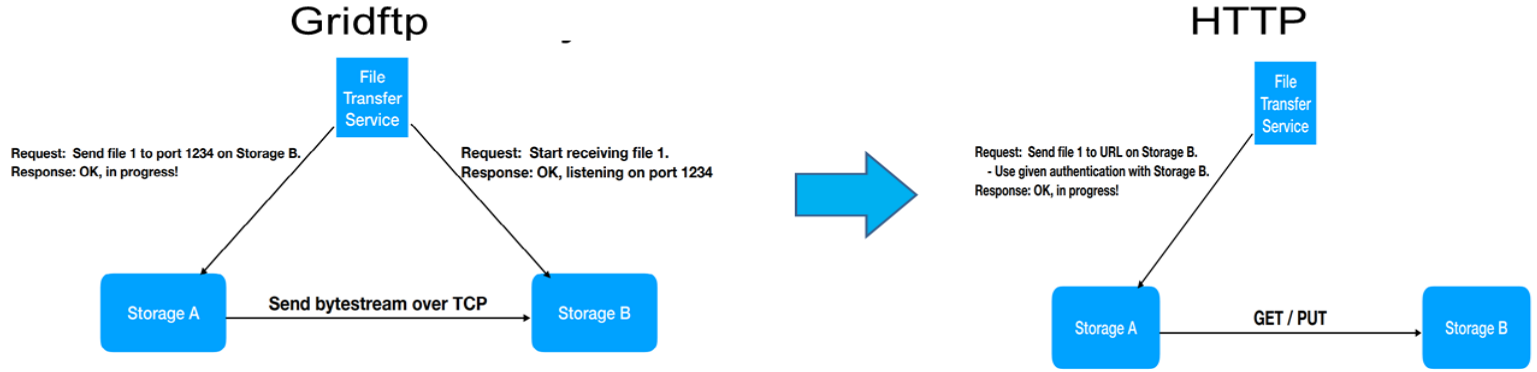
R&D activities

EOS SE and Third Party Copy (TPC)

- EOS will become main storage in IHEP
- Webdav (HTTP) are taking place of Gridftp as the TPC baseline protocol in WLCG
- Study on EOS SE with HTTP TPC support has been carried out
  - Testbed was set up and registered into DIRAC/Rucio
  - HTTP TPC is being tuned and tested

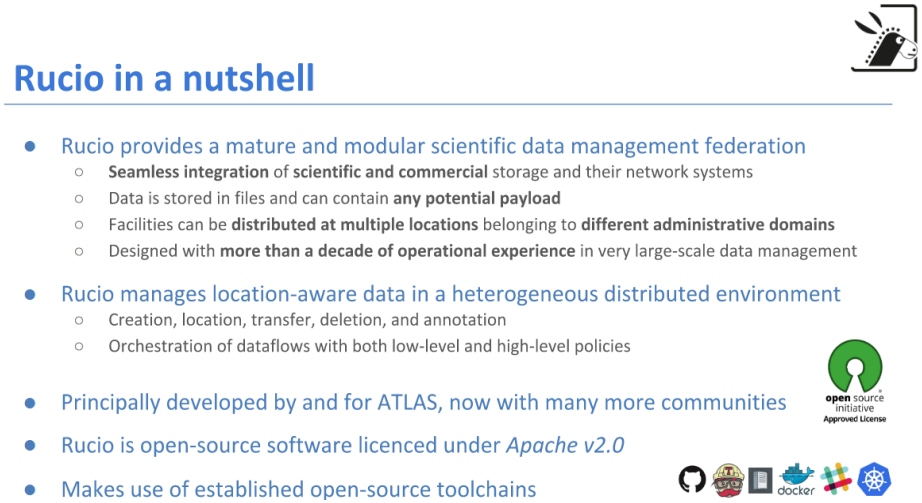


The client initiates the third party copy by **issuing a COPY request to either the source or destination endpoint**



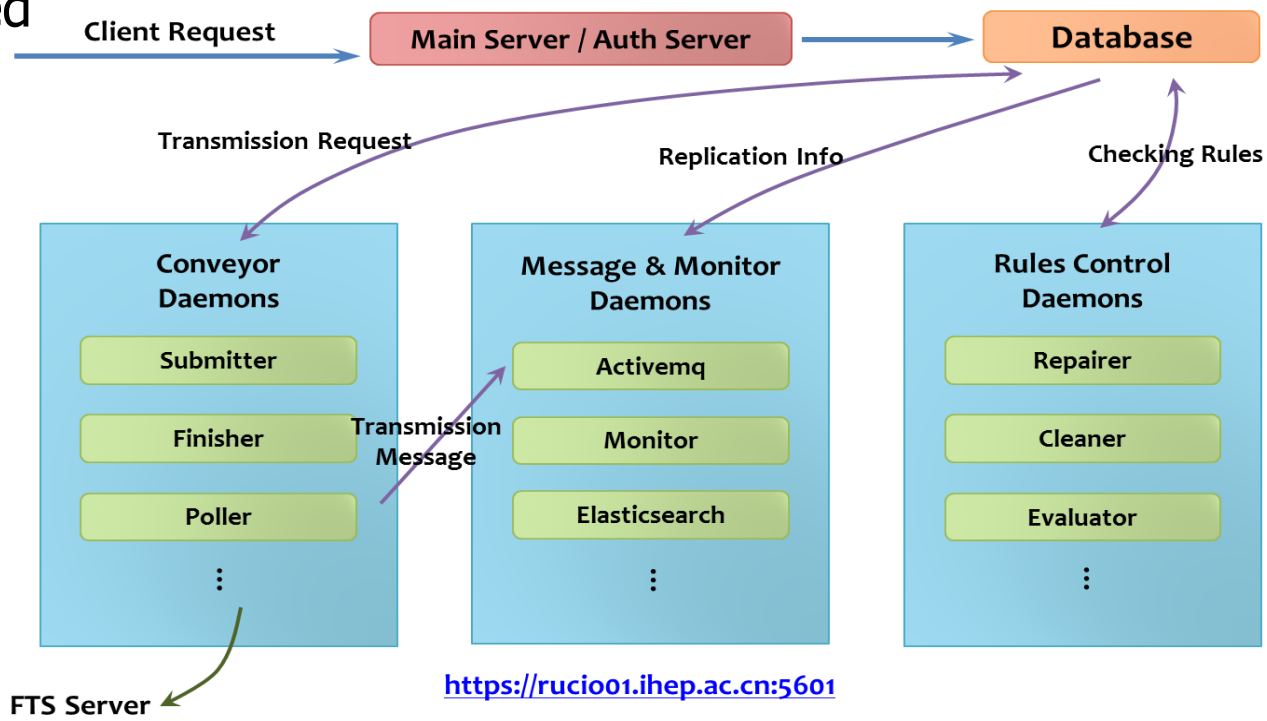
Rucio

- Rucio is an data management system which can provide the functionalities needed to manage SEs, data and data flow globally
- Rucio is developing into a common standard for scientific data management, and widely evaluated and used in production by many experiments and WLCG middleware
- In 2019 CEPC International Workshop, Martin Barisits from Rucio was invited to give a report on “Rucio - Scientific Data Management”



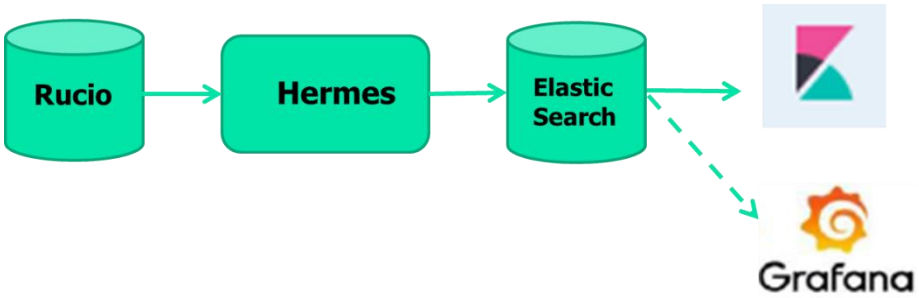
Rucio testbed set-up

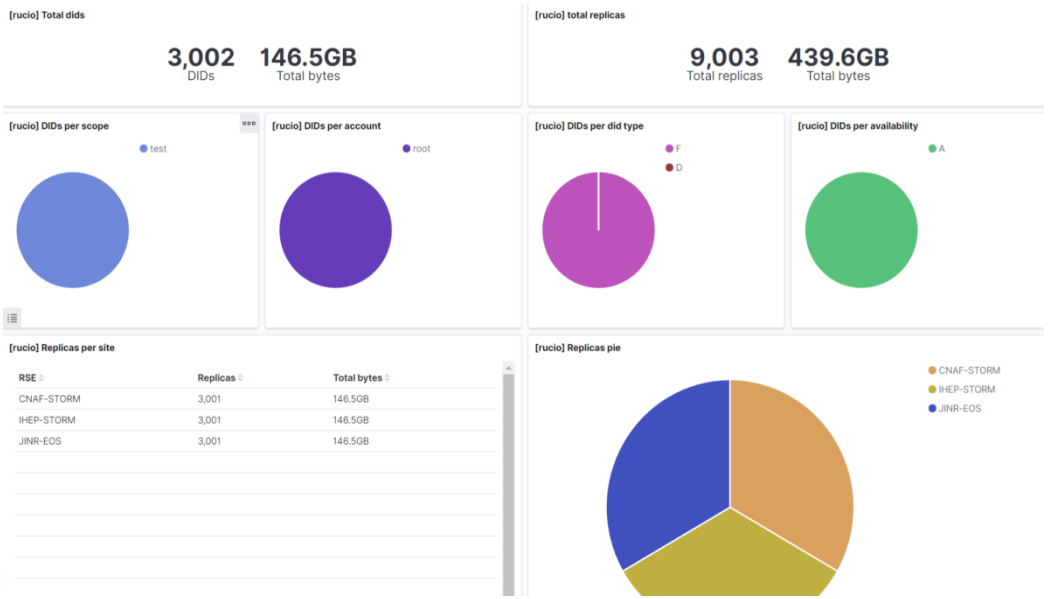
- Rucio testbed has been deployed with container in IHEP, including
  - Main server and Authority server
  - Database: PostgreSQL
  - Run independent daemons for transmission and monitoring
- Grid authentication allowed:
  - X509 certification + VOMS proxy+ User/Password
- Storage Elements registered
  - StORM, EOS, dCache
- Protocols supported:
  - SRM, gridftp, xrootd
- Upload/download
  - gfal2
- Transfer with FTS3
  - Connect to IHEP FTS server



Rucio monitoring set-up

- Monitor has been enable by connecting to ElasticSearch + Kibana
  - Hermes daemon in Rucio collect info of data and transfers
  - Info is sent to ES and shown in Kibana, or later in Grafana
- Rucio data and transfer status can be shown in Kibana





Integration of DIRAC and Rucio

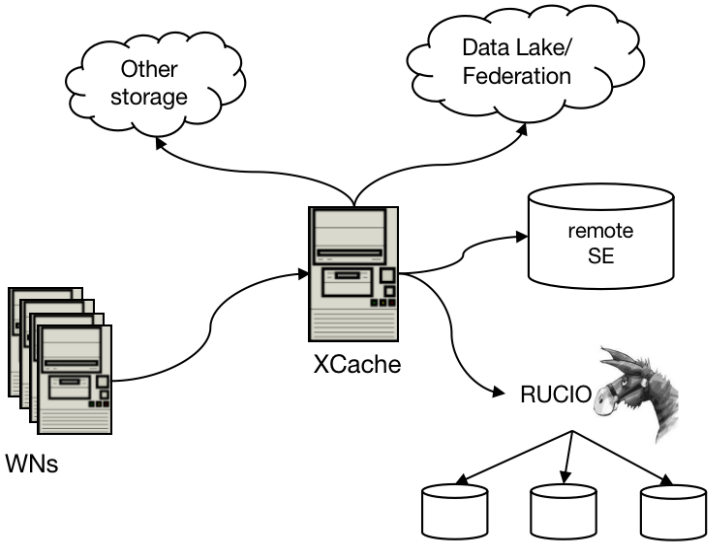
- To use DIRAC as WMS, Rucio as DMS, integration of DIRAC and Rucio is a issue to be considered
- BelleII has taken first step on the integration
- In 2020 CEPC International Workshop, Cedric Serfon from BelleII gave a talk on “Belle II data management with Rucio”
  - Integration is not straightforward, need much efforts

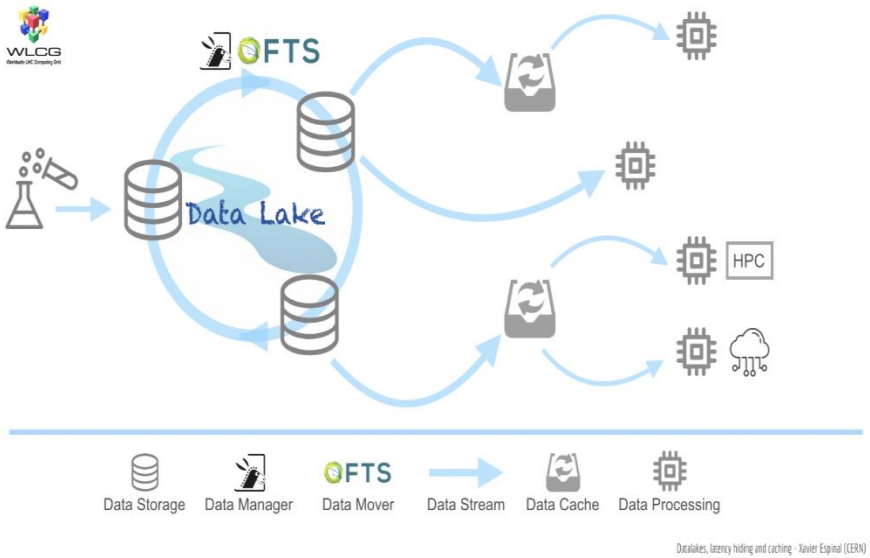
Steps on migration to Rucio

- Set up Rucio testbed (Done)
- Customize Rucio for CEPC experiments (Doing)
- With DIRAC highly extensible infrastructure, implement Rucio as a File Catalogue plugin -- “Rucio File Catalogue” (RFC) in DIRAC
  - Rucio client is wrapped as a RFC service in DIRAC
- Develop daemons to synchronize SE registers between Ruico and DIRAC
  - Implemented as DIRAC agents
- Test with RFC functions through DIRAC interface
- Test with the ProdSys system to have a unified control of workflow and dataflow from DIRAC side

XCache

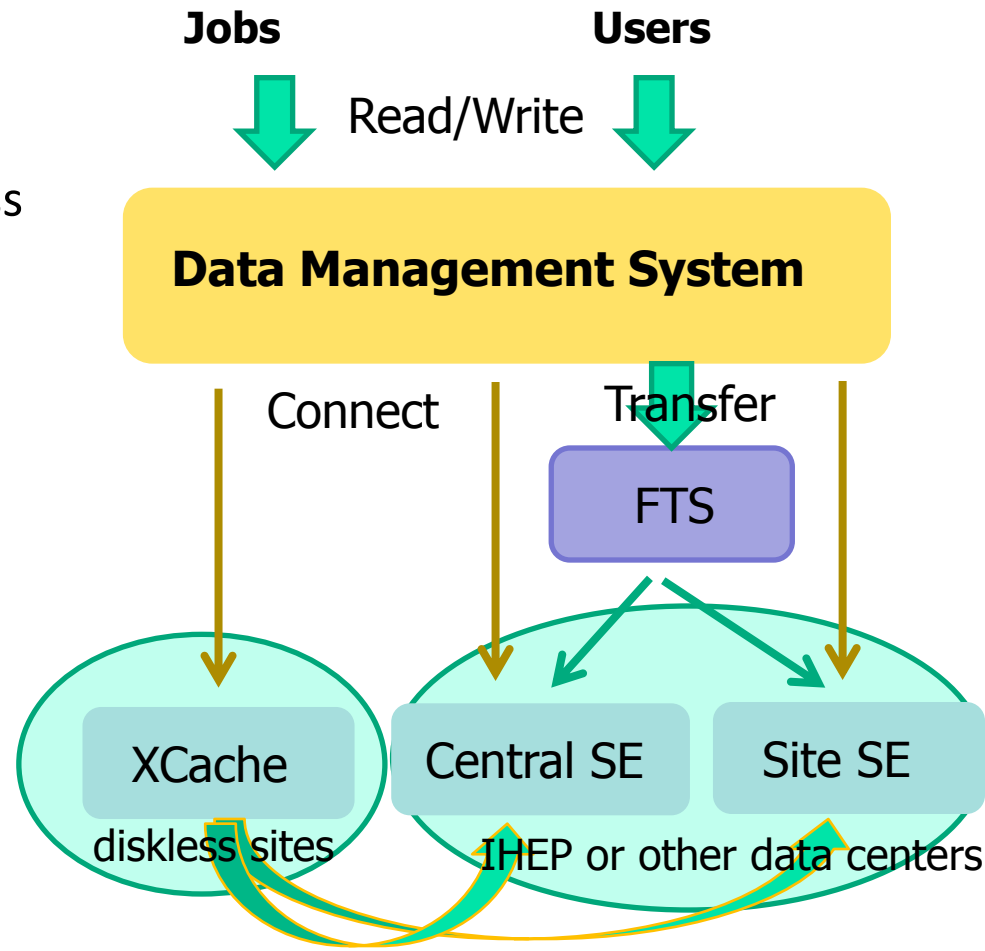
- XRootD Proxy Cache system supporting ROOT and HTTP protocol
- XCache can achieve highly efficient remote data access with high cache hit
- With plugin architecture, XCache is clusterable, easy to scale up
- Ruico and XCache play an important role in future “Data Lake or federation” model





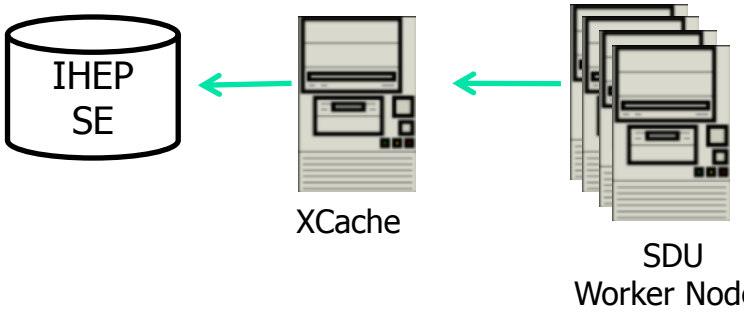
XCache for diskless sites

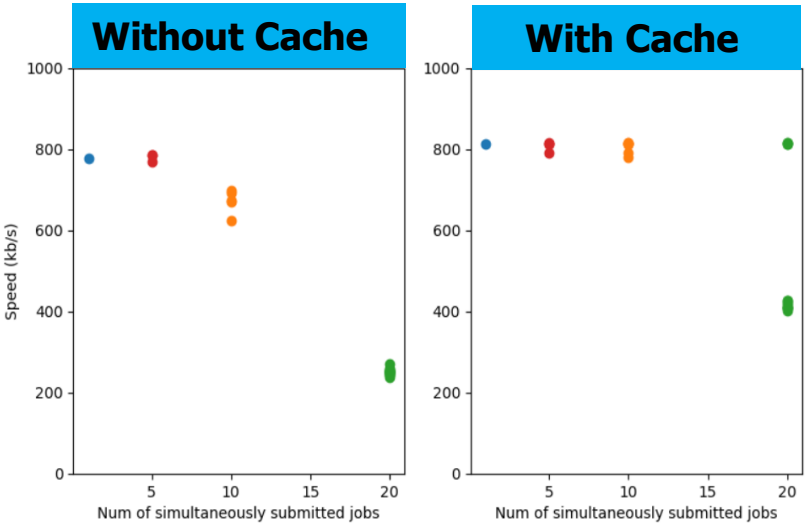
- XCache could be very helpful for small sites without enough manpower for a decent SE
  - XCache can act as volatile SE for diskless sites
- Jobs in diskless sites can transparently use it to do cache-ware remote access
  - Register XCache into DIRAC or Rucio
  - Set data source to be SEs from IHEP or other data centers
- This way can help reduce burden of central SE and network traffic of IHEP



XCache set-up and tests

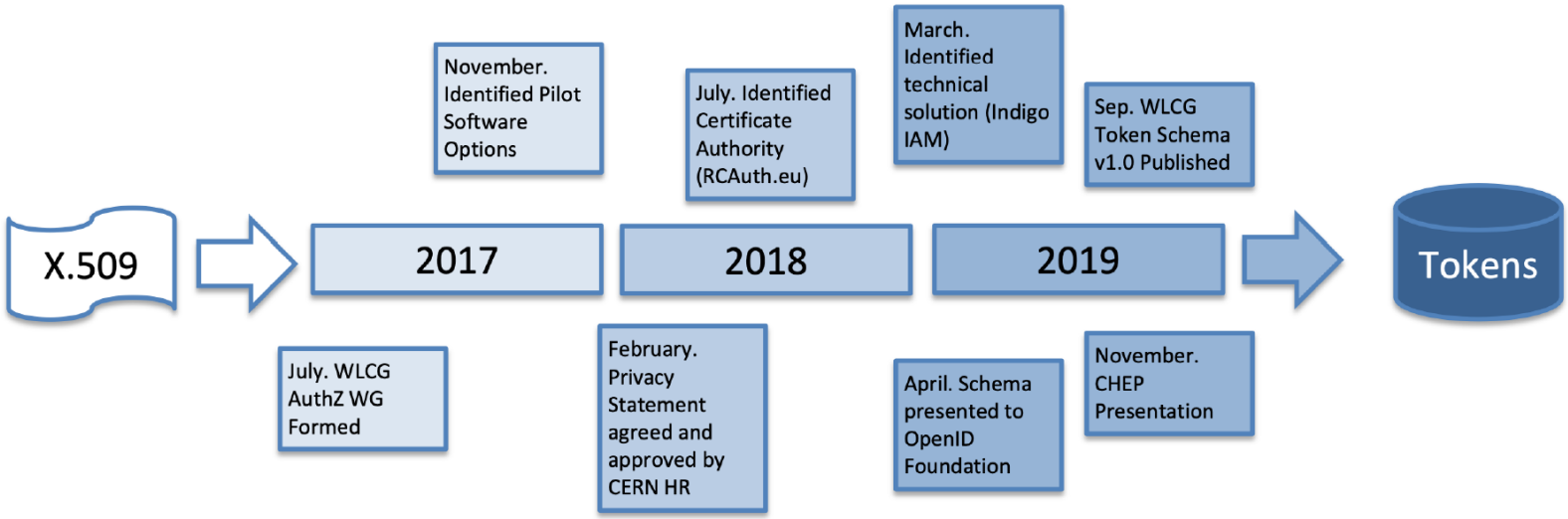
- XCache server was set up at SDU
  - Use X509 based certificate authentication
  - XRootD and https protocol are supported
- XCache was used through JSUB
  - XCache server registered in DIRAC
  - JSUB checks with DIRAC and use XCache when downloading input data
- With the above setup, jobs running in SDU can automatically read data through XCache which can use IHEP SE as data source
- Tests with JUNO jobs at SDU has shown the advantage of XCache
  - Failure rate reduced and the speed improved





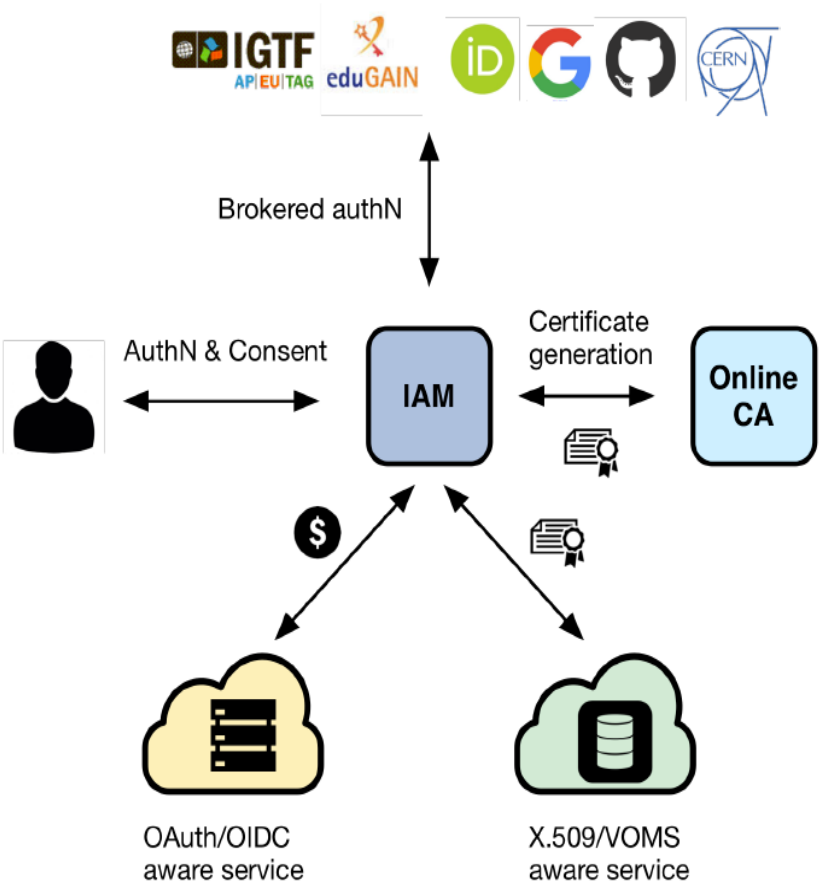
New AAI (Authentication and Authorization Infrastructure)

- WLCG considered moving away from x509 certificates towards token based authentication systems
  - Current X509-based AAI has many weak points: inflexible, not convenient….
  - Token-based authentication system is a widely-used, mature industry solution



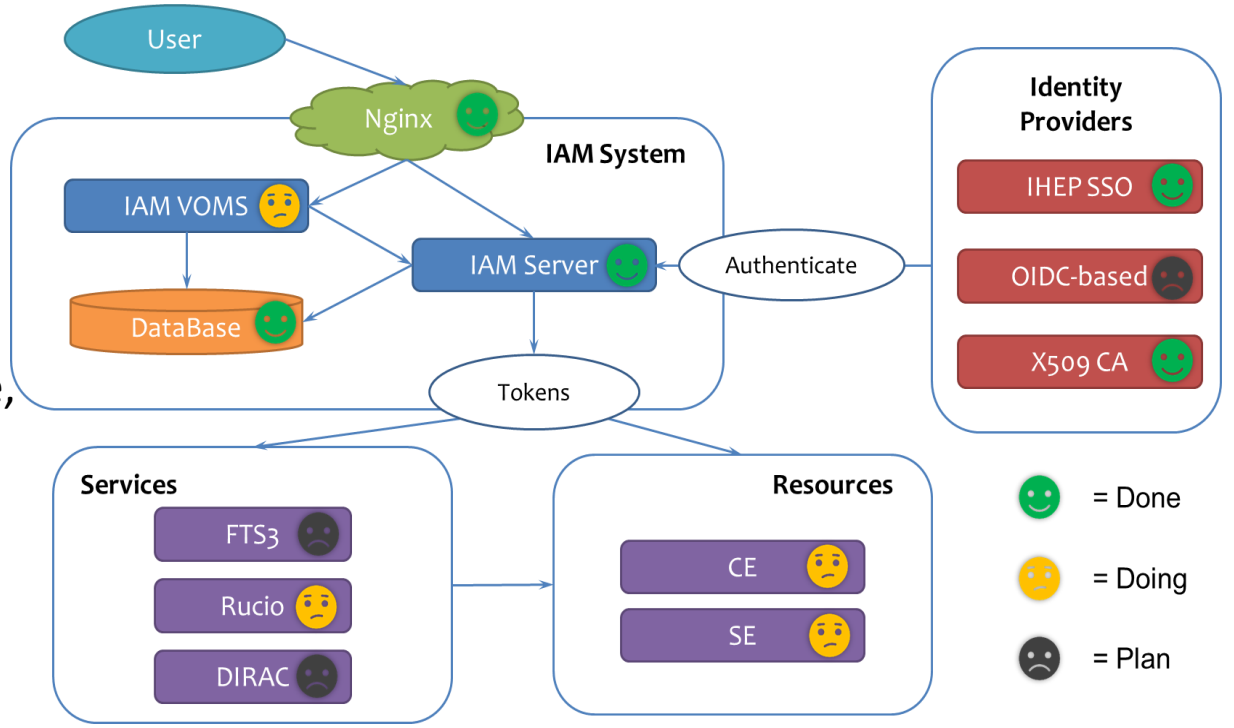
INDIGO Identity and Access Management (IAM) Service

- IAM is selected by the WLCG Management Board to be the core of the future token-based WLCG AAI
  - Support multiple authentication mechanisms
  - Interact with onlineCA to get user CA
  - Integrate existing VOMS aware service to produce VO-aware proxy
  - expose identity information, attributes and capabilities to services via JWT tokens and standard OAuth & OpenID Connect protocols
  - support Web and non-Web access, delegation and token renewal
- These functionalities are important in transition from X509 to token AAI



Status of migration to token-based AAI

- IAM service testbed has just been set up
- Need to do:
  - Connect with Identity Providers
    - SAML Home institution IdP (eg. IHEP SSO)
    - OpenID Connect (eg. Google, Microsoft)
    - X.509 certificates
  - Integrate VOMS
  - Integrate Online CA
  - Grid services support tokens
  - Resources support tokens



Summary

- The CEPC distributed computing prototype is ready for detector R&D simulation
- Submission interfaces are adding supports to the new CEPC software framework
- Studies have been carried out to closely follow the trends of WLCG evolution
  - Advanced data management and access technology in the data lake model
  - Token-based AAI with the IAM service
  - HTTP and XRootD TPC

Thank you!

Why Rucio?

- Widely evaluated and used in production by many experiments
  - Atlas, CMS, LIGO, BelleII, SKA……
- Compared with DIRAC DMS, the advantages beyond basic functionalities
  - ➔ Modern technologies, intensively developed
  - ➔ Automated and efficient data management capabilities
    - Expressive policy engines with rules for automated data flows
    - Automated corruption identification and recovery
    - Data popularity based replication ……
  - Actively support WLCG evolutions：TPC, token-based authentication, XCache…