

# Options for CTA computing at CSCS

Swiss CTA Days, 12<sup>th</sup> January, 2022

Pablo Fernandez, Andrii Neronov, Miguel Gila

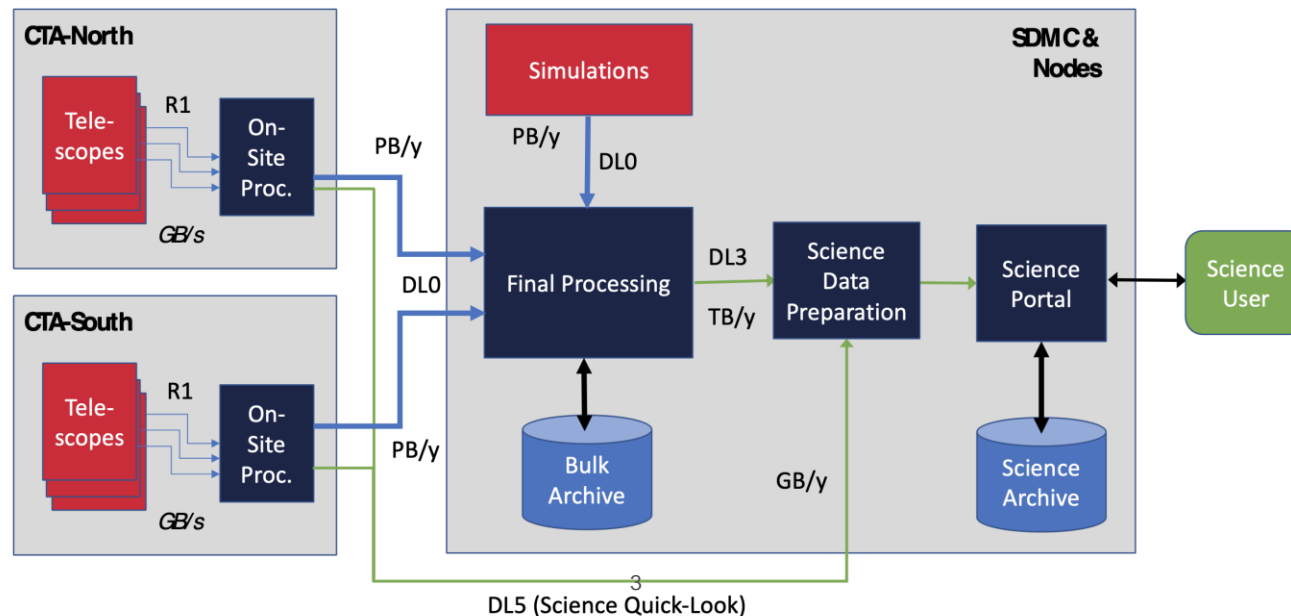
# About CSCS

- CSCS is the Swiss National Supercomputing Center, located in Lugano
  - Develops and provides the key supercomputing capabilities (knowhow and infrastructure) required to solve important problems to science and/or society.
- National and international collaborations
  - MeteoSwiss
  - CHIPP (Swiss WLCG Tier-2)
  - Other CH institutions (e.g. PSI, UZH, Empa...)
  - Member of EuroHPC (LUMI consortium), PRACE, PASC, etc.



# CTA Swiss off-site DC

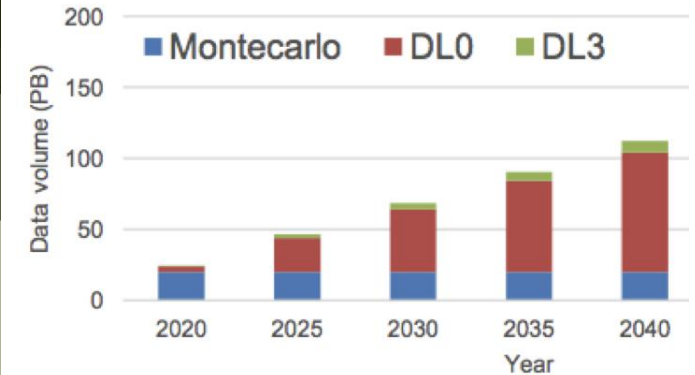
- Goal of the Swiss DC
  - Archive the raw data
  - Perform the standard data analysis: reconstruction of parameters of gamma-ray and cosmic ray “events” detected by the telescopes
  - Run Scientific analysis of the event data aimed at production of “high-level” data products ultimately used by astronomers: sky images, spectra and timelines of flux variations of astronomical sources.
- The Swiss DC will be developed through a collaboration involving EPFL and CSCS
  - With support from the CTAO teams at the University of Geneva and ETHZ.



# Off-site Data Centre implementation (OFF-SITE IC-INFRA document)

| Year                                      | 2021          | 2022          | 2023          | 2024          | 2025          |
|---|---------------|---------------|---------------|---------------|---------------|
| Cumulated disk storage TB                 | 5793          | 11682         | 17558         | 23433         | 26568         |
| Cumulated volume on tape (TB)             | 13061         | 27230         | 41298         | 55320         | 64437         |
| Total KHS06.sec                           | 1 206 948 329 | 1 221 066 374 | 1 258 106 334 | 1 318 173 186 | 1 396 887 463 |
| Total KHS06.sec per data centre           | 603 474 165   | 610 533 187   | 629 053 167   | 659 086 593   | 698 443 732   |
| <b>Equiv. nb of cores per data centre</b> | <b>1 914</b>  | <b>1 936</b>  | <b>1 995</b>  | <b>2 090</b>  | <b>2 215</b>  |

– 10 Gbps network between Off-site Data Centres  
 – 250 TB of high-performance shared storage per Data Centre  
 – Disk storage: 10% of all data + all simulated data  
 – Tape storage: two copies of all data, distributed between Off-site Data Centres  
 – Computing: one reprocessing of all data per year  
 – Database revers  
 – storage supporting POSIX, https, xrootd, GridFTP protocols



We need to be prepared for the large data volumes that are coming later on

- 2022-2025 “development” mode:
  - the first CTAO data (LST-North in particular) as well as simulated data will flow in the data center
  - Series of technical & data challenges
- After 2025 the model changes into “operation”
  - Heavy-duty cycle real-data processing
  - User support



**CSCS**

Centro Svizzero di Calcolo Scientifico  
Swiss National Supercomputing Centre

**ETH** zürich

# CSCS Infrastructure

---

## World's Most Powerful AI-Capable Supercomputer?



### CSCS, Hewlett Packard Enterprise and NVIDIA Announce World's Most...

12.04.2021

"Alps" system to advance research across climate, physics, life sciences with 7x more powerful AI capabilities than...

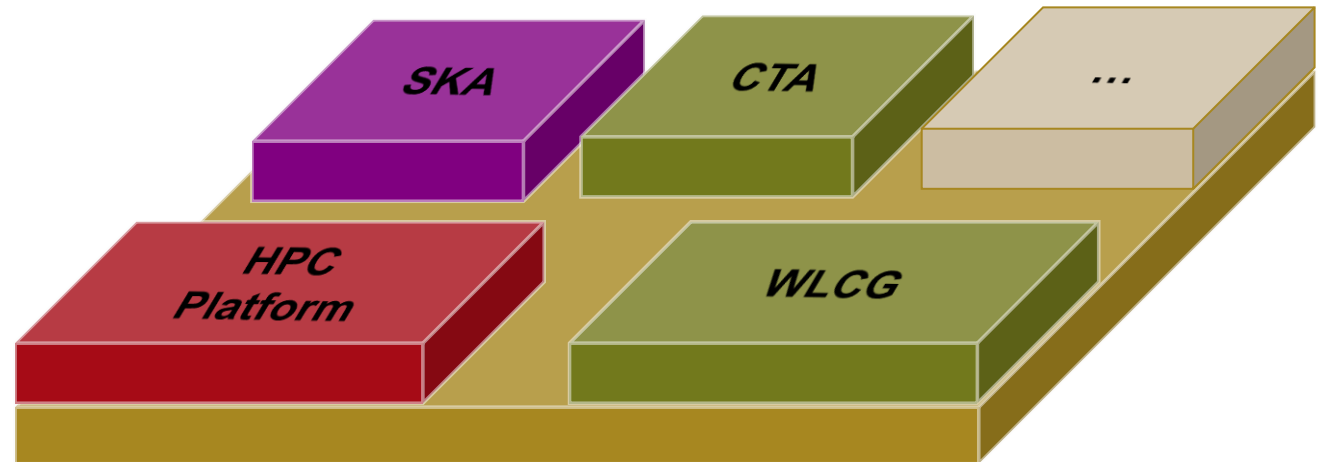
MORE

MORE SCIENCE



# Taking platforms further in Alps

- Very flexible infrastructure that allows for plenty of customization
  - CTA will look very similar to WLCG
- Customization comes at a cost
  - Similar platforms are easier to maintain





# Alps advantages

- Shared multi-purpose infrastructure
  - Economies of scale
  - No dedicated hardware, procurement is done asynchronously to the different use cases
  - Multiple architectures (e.g. CPU-only, GPU, multi-GPU, disks, flash, tapes...)
- Elasticity
  - Independent of funding cycles, can grow and shrink resources
  - Pay-per-use business model
- Green datacenter
  - PUE 1.15
- CTA, SKA and WLCG can live on the same infrastructure and share the micro-service architecture



**CSCS**

Centro Svizzero di Calcolo Scientifico  
Swiss National Supercomputing Centre

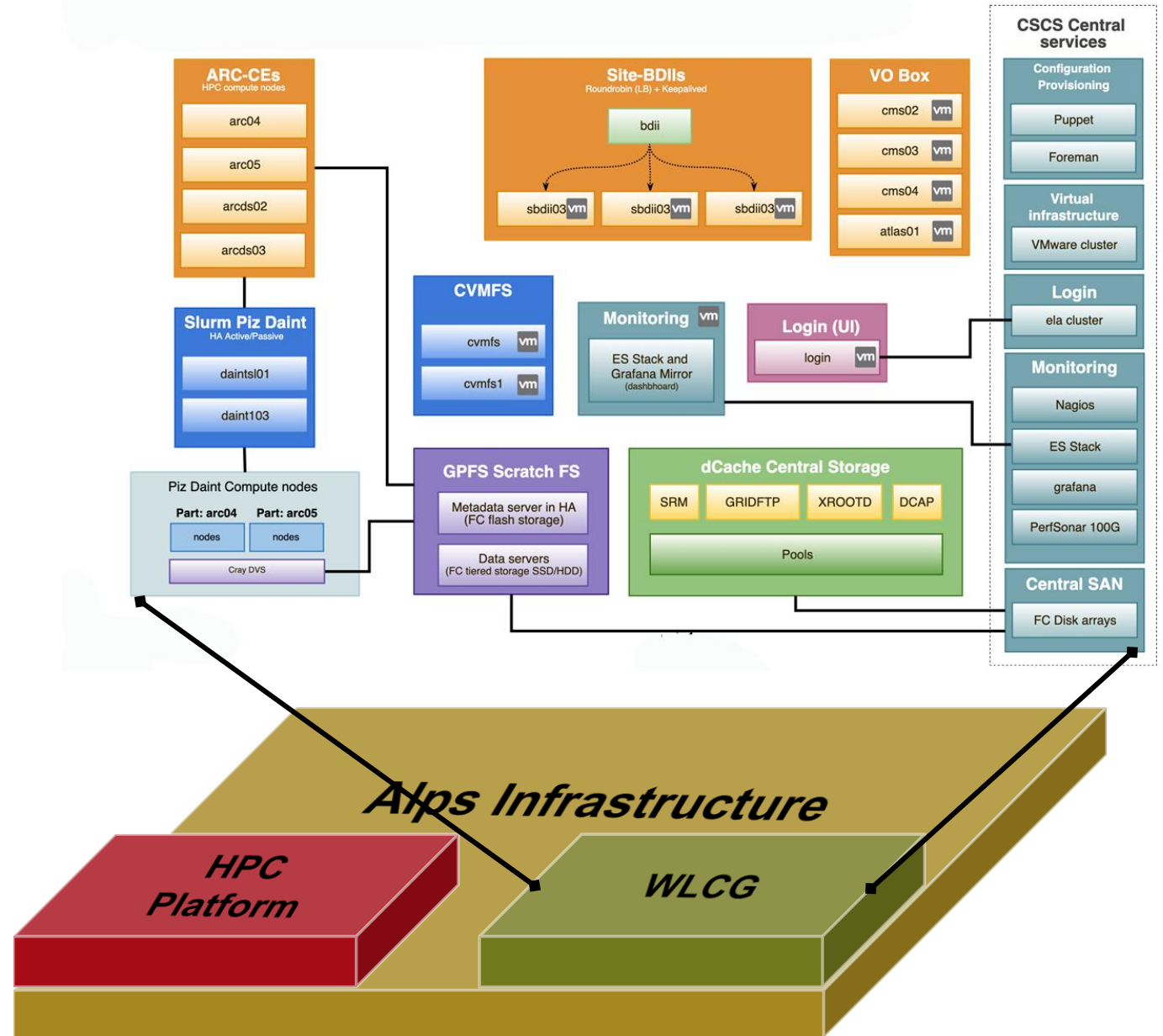
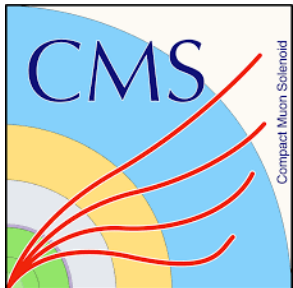
**ETH** zürich

# From WLCG to CTA

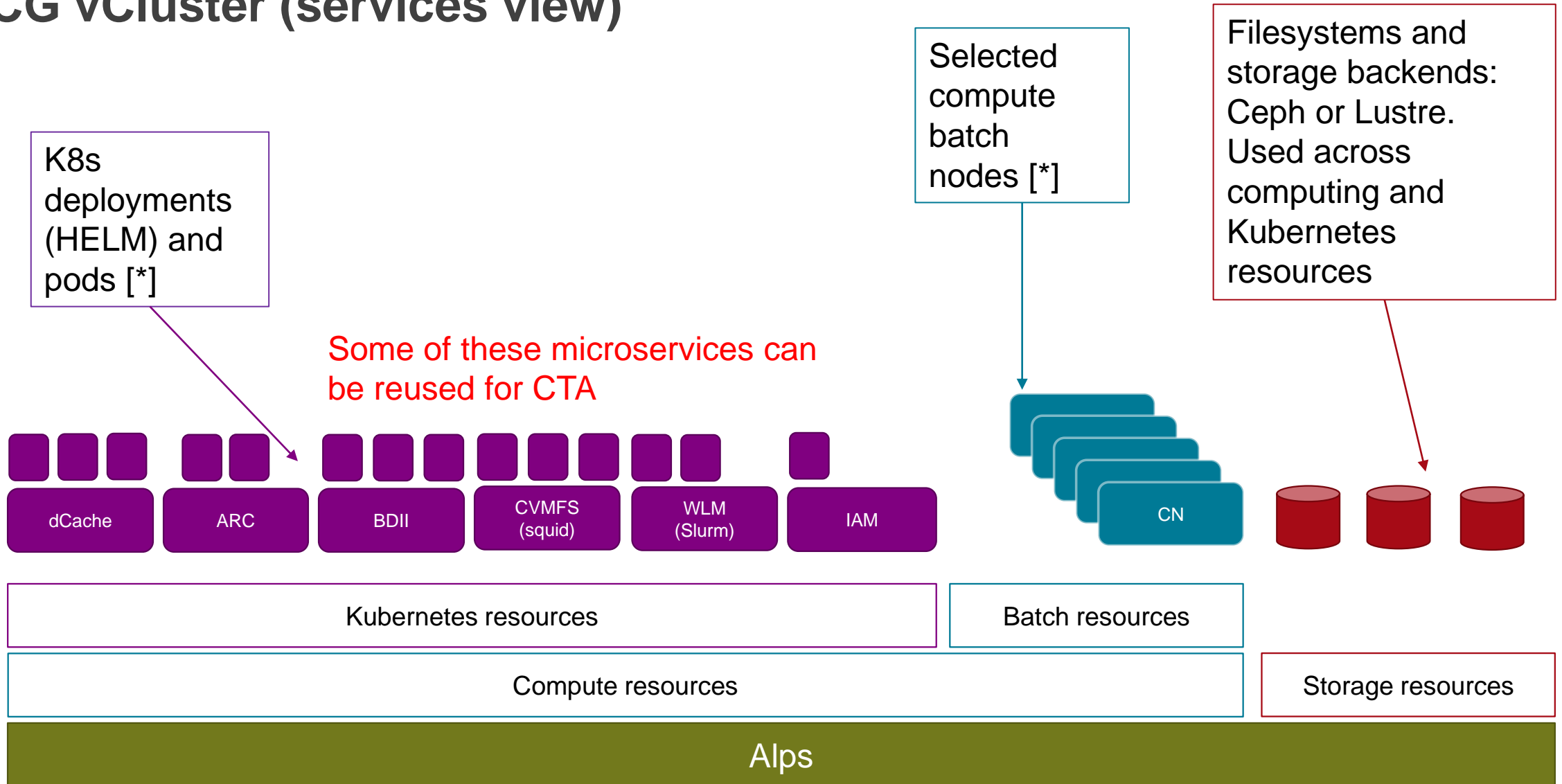
---

# WLCG Grid Middleware layer

- Lots of services to “seamlessly” connect to the WLCG federation
- Collaboration between 4 organizations in order to make things work

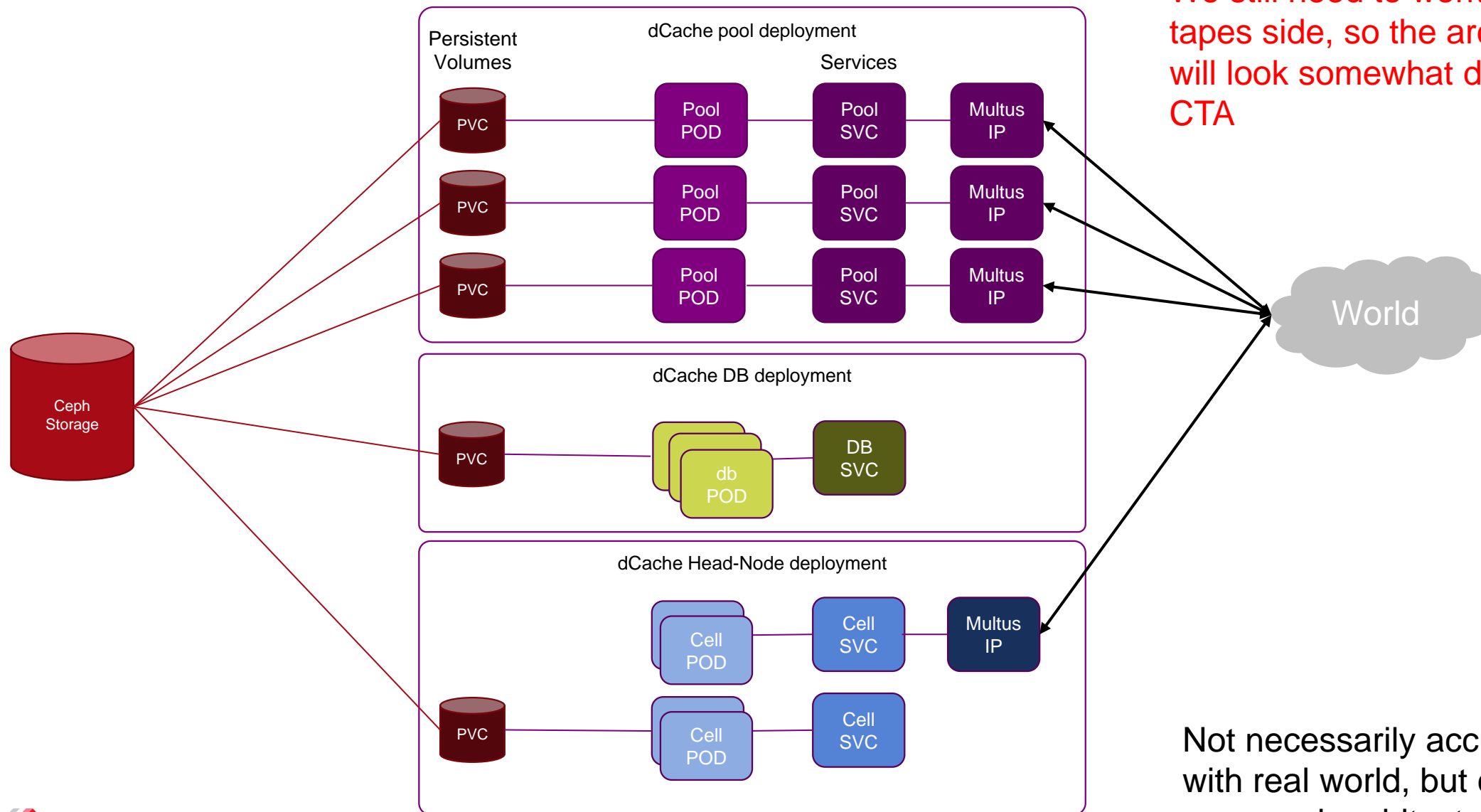


# WLCG vCluster (services view)



[\*] Could replicate multiple times depending on environment (dev/stage/prod)

# WLCG dCache (K8s architecture)



Not necessarily accurate 1:1 with real world, but conceptual proposed architecture **ETH** zürich

# Tape library

- CSCS operates a tape library (~100 PB), which can be used by CTA
  - To reduce costs
  - Can be plugged into dCache (effort planned for 2023)
  - Already provided infrastructure (including tapes) for the FACT project back in 2019-2020.
- IBM TS4500, currently with TSM
  - Evaluating other options



# What's next?

- Gather team
  - Internal CSCS is partly present but we need to hire an engineer
  - Engineer & Data Scientist at EPFL, contract is not in place yet
- Meet with architects & other sites (January 17<sup>th</sup>)
- Architect solution
  - Bulk Archive of raw observational and simulated data
    - dCache + IAM + Network? Map use-case into Alps
      - Alps needs to be ready as well (work will continue until mid'23)
    - Connecting dCache into our tape archive
  - High-performance computing (HPC) system for DPPS
  - Science archive
  - Open data access and data analysis platform for SUSS
- Operations
  - Coordination with other off-site data centers (and with SDMC)
  - Participate in technical and data challenges
  - Start processing LST1 data
  - Scale up to regular data processing for observational and simulated data
  - Eventually (by 2025 maybe) distribution of science data to users and user support

