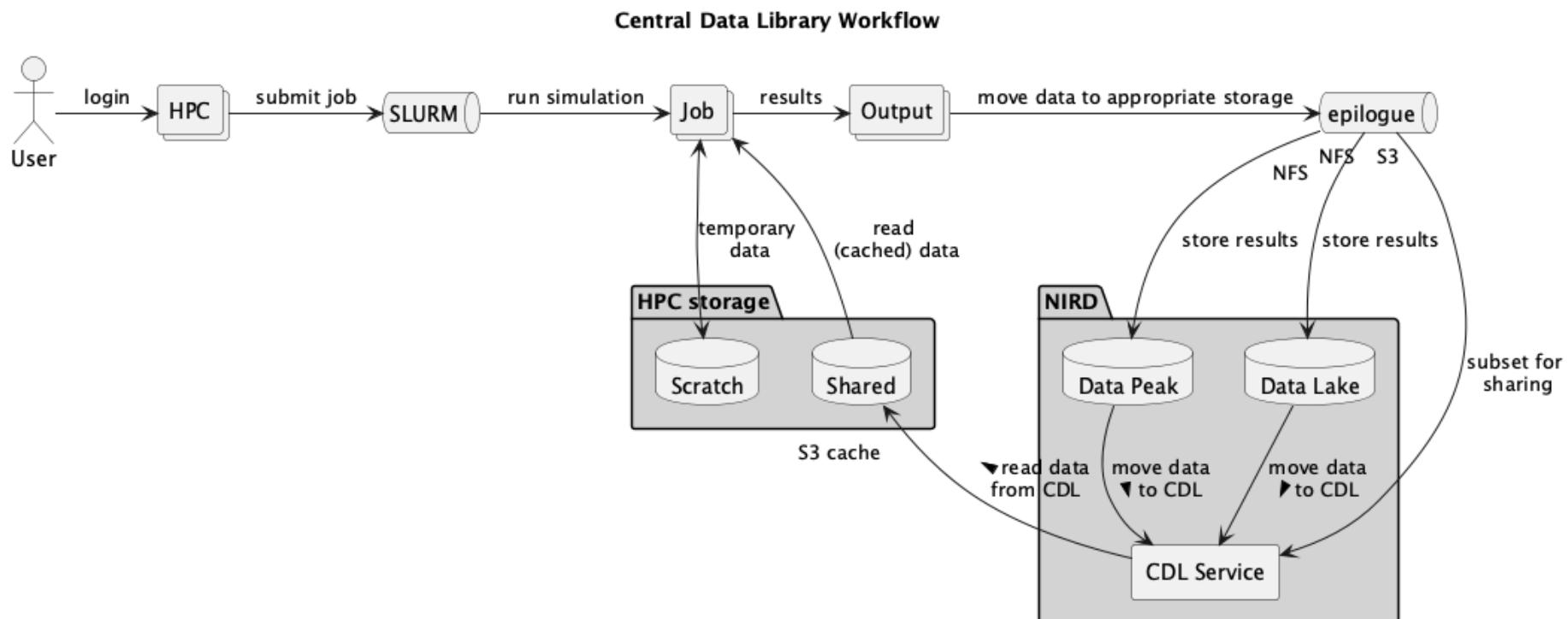


What is the CDL?

- What is the Central Data Library?
 - Long-term storage of non-persistent input datasets enriched with metadata
 - Sharing input data and data libraries with multiple projects
 - Share AI-models and AI-training data with multiple projects and collaborators
 - Access to data via the S3 protocol and S3 API

CDL workflow



Current status

- Launched in 2025
- Allocation on NIRD (DL)
- In the process of collecting / migrating datasets from HPC
(/cluster/shared) to NIRD
- CDL datasets visible as read-only S3 buckets
- S3 account and bucket management not yet coupled with MAS

Planned improvements

- Enabled data caching on HPC
- Streamlined application and allocation process
 - Management of S3 accounts, roles and buckets from MAS
 - Support for S3 system accounts
- Introduction of research objects
- Support for large-language models
 - Likely XeT storage
- Addition of a data management platform
 - Likely iRODS

Benefit for the users

- Central storage for shared datasets
- Free of charge*
- Easy collaboration
- Improved data governance and data management

** see CDL policy and user contribution model*

References

- <https://www.sigma2.no/service/nird-central-data-library>
- <https://www.sigma2.no/central-data-library-policy>
- https://documentation.sigma2.no/files_storage/nird/cdl.html