

Data Services

ALCF

Aditya Tanikanti
Venkatram Vishwanath
Sam Foreman
Benoit Cote

Data Services

Data Management Systems

- Conducted a comprehensive evaluation of prominent Data Management Systems, including Dataverse, Invenio, Ckan, and the Globus data portal, utilizing the CELS VM.
- This assessment aims to automate the curation and visualization of data sourced from the Eagle file system. Our objective is to seamlessly integrate these enhancements into ACDC.

Database as a service

- Initial benchmarking for renowned databases such as Postgres, Mongo, and Neo4j on a single ALCF node.
- To aid users, we've provided an intuitive setup process accompanied by detailed instructions. These leverage singularity/apptainer containers, enabling users to effortlessly deploy popular databases on ALCF login/compute nodes. You can find more details [here](#).
- Our vision is to offer persistent databases for users, allocated on a segment of ALCF computes and synced with ALCF file systems.

MLOps

- We have undertaken the initial setup and performance assessment of leading MLOps utilities, encompassing tools like MLflow, MLDE, WandB, and Neptune, among others.
- Goal is to establish a portal for users, granting them access to cutting-edge MLOps features such as hyperparameter tracking and model serving.

DOI Minting Service

- We've made notable progress by collaborating with OSTI to furnish DOIs via their API.
- Our primary objective is the seamless integration of this feature within the existing ACDC infrastructure.

Data Services

Container Registry and Services

- We have established a container registry leveraging Github Packages, which generously offers free registry services for both Enterprise Github users and individual contributors.
- To simplify user integration and scalability on platforms like Polaris and Sunspot, we've presented guidelines on utilizing base images embedded with Pytorch, Tensorflow, and mpich.
- For further details, please visit:
[Argonne LCF Container Registry](#)
[Argonne LCF Packages](#)

Inference as a service

- Our ongoing efforts are channeled towards executing canned applications through platforms like Globus Compute, FuncX, and Parsl.
- Our immediate focus is centered around the provision of LLMs as a service on both Sunspot and Polaris.

IBM GSS PAIRS Climate data portal

- We are currently in active collaboration with IBM to incorporate their PAIRS application on Polaris. This application stands out due to its effortless querying capabilities for climate raster data located on ALCF file systems.
- Interestingly, this is the powerhouse software behind the IBM Weather App. Our ultimate vision is to empower users with the ability to seamlessly query the vast reservoir of approximately 4PB+ climate data housed on the Eagle File System.

We aim to elevate the quality of our offerings by forging partnerships with CPS, through in-depth case studies on innovative service infrastructures and tools, as well as improved data management strategies and services.