A HIGH-PERFORMANCE SCIENCE-DMZ SYSTEM FOR LARGE-SCALE DATA ORGANIZATION AND DISTRIBUTION

Argonne NATIONAL LABORATORY

PETREL DATA SERVICE

SYLVESTER JOOSTEN
sjoosten@anl.gov

EIC Software Consortium
May 13, 2020

THE SCIENCE-DMZ MODEL

A Scalable Network Design Pattern for Optimizing Science Data Transfers

More info: https://fasterdata.es.net/science-dmz/

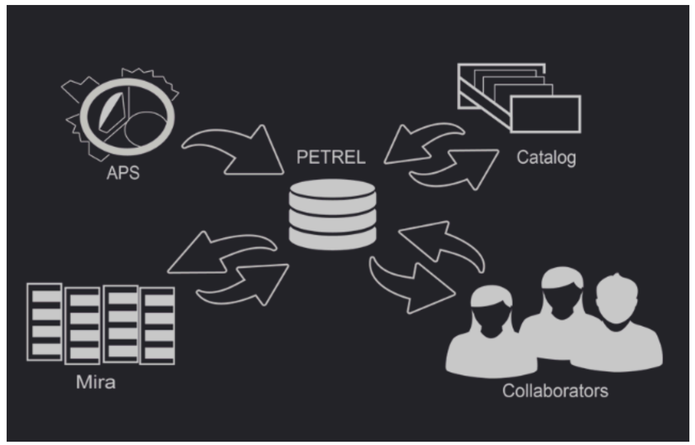
- Designed by ESnet engineers
- Distinct from general-purpose business or “enterprise” systems”
- Network architecture explicitly designed for high-performance applications
- Dedicated systems for data transfer
- Security policies and enforcement mechanisms tailored for high-performance science environments

PETREL

Data management and sharing system developed by ALCF and Globus

PETREL

More info:
https://dl.acm.org/doi/10.1145/3332186.3332241
https://www.globus.org/globus-user-story-petrel-argonne (old)
https://petrel.alcf.anl.gov



- Implements science-DMZ model
- Enables storage and sharing of large-scale datasets between collaborators in different institutions without local accounts
- Login through Globus using own laboratory or university's federated login
- In third incarnation, growing rapidly towards 17PB of usable space
- Seamless integration with OSG and other HPC assets, laboratory farms and local machines.

PETREL FOR EICUG YR EFFORT

- Secured starting allocation of 100TB (to be increased as needed)
- Permissions are setup so all files are read-accessible with a Globus account (which everyone already has!)
- Write-permission can be obtained (on request) by being added to the EIC Globus group.
- Fine-grained write permissions are possible, right now not needed (?)
  - *Maybe consider mirroring YR subgroups in directory structure?*
- Could start out by copying over existing simulations from BNL
- Can provide synergy with BNL Box approach

PROPOSED ADDITION TO EICUG STORAGE PAGE

Worldwide access through Petrel/Globus

Petrel is a Globus-enabled data service for researchers that provides a simple and intuitive interface for self-managed project-based data sharing. Our Petrel allocation is 100TB (more can be added if needed). The existing pre-grenerated Monte-Carlo data will be made available soon. You can find more info on Petrel here.

Because of the Globus back-end it is easy to move, share and discover data via a single interface, regardless if you are working with on a HPC facility, computer farm or your local machine. All major laboratories and supercomputing facilities, as well as most universities support Globus. This allows you to use your existing laboratory (ANL, BNL, JLab, LBNL, etc.) or even university credentials to access the files. To access the storage space, log in to globus.org with your existing laboratory or university credentials and access the `petrel#eic` endpoint. It's as easy as that!

If you encounter any issues, or if you want write-access to the storage space, you can contact Sylvester or Markus.

TLDR: login to globus.org and access the `petrel#eic` endpoint