2017-09-14 Update

Sprint 32

Sprint 32 starts next week. Notes are available on the Wiki:

https://opensource.ncsa.illinois.edu/confluence/display/NDS/2017-09-12+Sprint+planning

The focus for this sprint includes:

- NBI Pilot
- NCSA Industry demo(s)
    - NBI?
    - NCSA Genomics
    - TERRA stitching case
- Website and documentation updates

NBI Pilot

We have a preliminary version of the NBI MongoDB running in Labs Workbench with a sample Jupyter environment. This includes a complete set of the NBI records (17 million) and a notebook based on an example provided by our contact. Craig will meet with Robin next week (9/21) to discuss.

I've created a Github repository with basic instructions and additional information:
https://github.com/craig-willis/nbi-pilot

The NBI pilot is beneficial in several ways:

- Connects NDS to Bridge Health group through Midwest Big Data Hub
- Represents another example of NDS hosting "active" data (e.g., in ready-to-analyze framework)
- Provides immediate access to a resource and example analysis for Bridge Health community

NCSA Industry

We will be presenting the Labs Workbench platform at the NCSA Industry conference in October. The main focus is demonstrating Workbench integration with larger analytical platforms, including Hadoop/Spark and HPC (with a priority on Spark).

Ben has installed a small Spark cluster for integration testing. We are considering using the NBI data (above) to demonstrate additional capabilities beyond MongoDB. We are also in contact with the NCSA Genomics group and CyberGIS for potential access to existing Spark clusters with known science cases. In a meeting with the CyberGIS group on 9/14, we may have access to a Twitter dataset with some geospatial analysis that could be of interest.

We will also go forward with the creation of a Jupyter notebook to demonstrate the TERRA-REF full-field image stitching process. The basic use case is to create a full-field representation (and eventually mosaic) of crop data from high-resolution sensor data that may be applicable to other domains.

Website and documentation updates

We will be working to refine the description of the NDS Labs service to prioritize the "sandboxing" capability near research data, reducing focus on the "playground" for research data management tools. We will also correct some problems with existing pilot descriptions, specifically MDF (broken links, inaccurate information, etc).

Einstein Toolkit Tutorial instance

The Einstein Toolkit group submitted a pilot request for us to support an instance of the Labs Workbench service for their ongoing tutorial and workshop support. This will be a dedicated instance hosted at NCSA (Nebula, NDSLabs project) that will enable users to run through a Jupyter-based tutorial used during the previous Einstein Toolkit School. We expect to expand this instance to support workshops in the future. The Einstein Toolkit will likely be a good demonstration case for SC17.

SC17 planning

The focus of the SC17 demonstration will be Workbench to HPC. At this point, we're focused on the TERRA-REF image stitching and Einstein Toolkit cases, but are open to other suggestions. We plan to demonstrate the following:

1. New user launches tutorial or exploratory interactive environment in Workbench
2. The user runs some small-scale analysis via interactive environment in Workbench
3. The user runs the same analysis at scale on an HPC system, via the Workbench system.

The goal is to demonstrate how Workbench supports "scaling up" the research process from a single interface.

QMCDB

Kandace received a request for information about the QMC-DB database, apparently hosted by NDS Labs. This was an early pilot project. We've determined that the VMs exist but are shutdown in OpenStack and have reached out to Ray Plante for information about accessing the systems. The PI is interested in resuming the pilot, if possible. This would be similar to the NBI pilot (above).

Features

In addition to the above demonstrations and instances, we are prioritizing the following features:

- Oauth/OIDC support (single sign-on)
- LDAP support for authorization
- Commercial cloud support
- Expanding storage support (currently tied to Gluster)
- Improved security

bioCADDIE

The bioCADDIE final report was delivered on 9/5. We're waiting for the final invoices to post, but everything must be complete by 9/31.