CI/CS WORKSHOP
THE COMMUNITY TOGETHER

ResearchSOC  CI CoE Pilot
Low-cost, interactive access and visualization of large scale scientific data

Steve Petruzza, Giorgio Scorzelli, Valerio Pascucci
Scalable Deployment: Exploration of 3.5TB of Weather/Climate Data in Real Time

Workflow

- Data creation
- Data Management
  - Processing
  - Analysis
  - Visualization

Distributed Resources

- 3.5 PB of data store in NASA
- Primary ViSUS server in LLNL
- Secondary ViSUS server in Utah
- Clients connect remotely
- Work without additional HPC resources
Containerized on-demand conversion and streaming service (for ESGF data)
High Performance Data Movements for Real-Time Access to Large Scale Experimental Data (Dockerized server)

• Using a desktop client (or a webviewer) Prof. Ashley Spears was able to see the data being acquired at APS from her office at UoU

• Webviewer Demo: Aluminum Foam of similar size
NEON AOP data access

- NEON has a large amount of data that is shared with the community through their data portal.
- There exist APIs to download those data in bulk (per site, per year, per data product, now also by area).
- For some data, such as sensor measurements, the portal provides an interactive navigation system.
- For others, like Airborne Observation Platforms data, there is a long list of image files...
- There is a need to present all AOP data interactively, where the users can preview, navigate, and select/access/download the data they need.

![Image of 2D wind speed and direction graph and AOP data table with site, month, and size columns.](image-url)
CiCOE data access/visualization/management efforts

- AOP data ingestion and publication
- Multiresolution streaming data access (via javascript, python, C++)
- Experimented with time series and hyperspectral data
- Experimented with mixed tile sources (Google Earth+AOP data)
- NEON endpoint for data discovery and viewer embedding
- Deployment experiments on CloudLab
Data ingestion and publication

- R scripts to download AOP dataset (byFileAOP), can we do better?
- Data processing and management:
  - extract the bounding boxes information from GeoTIFF and generate scripts to convert each dataset
  - Query NEON APIs to derive “month” value from file paths
  - Populate Utah endpoint database
  - Update the streaming server with the new datasets to make available
Streaming server/data portal

- Apache module
- Provides streaming access to data hosted locally or remotely
- Available within a Docker container or standalone installation
- Web UI to manage and ingest new datasets
Utah NEON APIs endpoint

- Handles requests to `/neonapi/products/{productCode}`
- Follows NEON APIs syntax
- Provide configuration strings to access a specific dataset
- Datasets could be relocated to other streaming servers transparently

```json
{
  "data": {
    "productCode": "DP3.30010.001",
    "siteCodes": [
      {
        "siteCode": "ABBY",
        "availableMonths": [
          "2018-07",
          "2017-06"
        ],
        "availableDataUrls": [
          "server=https://k2fdataportal.sci.utah.edu/K2Fmod_visuŠK3F&dataset=DP3.30010.001-D16-2018_ABBY_2-L3-Camera-Mosaic_V01",
          "server=https://k2fdataportal.sci.utah.edu/K2Fmod_visuŠK3F&dataset=DP3.30010.001-D16-2017_ABBY_1-L3-Camera-Mosaic-V01"
        ]
      },
      {
        "siteCode": "ARIK",
        "availableMonths": [
          "2017-05"
        ],
        "availableDataUrls": [
          "server=https://k2fdataportal.sci.utah.edu/K2Fmod_visuŠK3F&dataset=DP3.30010.001-D10-2017_ARIK_1-L3-Camera-Mosaic-V01"
        ]
      }
    ]
  }
}
```
• Component embedded as an iframe
• Datasets and time navigation
• The parent window pass the dataset settings (retrieved from the endpoint) to the iframe (GET) which provides the rest of the functionalities

**NEON integration (live)**

**Basic RGB**

**Vegetation indices (hyperspectral)**

**University of Florida deep learning tree classification (live):**

This is a link to the viewer with the trees

**Detailed view**

**Link to their main project**

**Visualizations**

**AOP Data Viewer**

This viewer allows for interactive exploration of normalized anomaly data from the Alabama Observation Platforms (AOP). Change the field site and flight for this data product using the tools below to stream different data into view. Pan and zoom to view or download higher resolution imagery. This pilot data viewer is provided through a collaboration with the U.S. Forest in the Laboratory of USDA and more updates are planned for the future.
Interoperability

• Explored solutions to integrate in the same visualization multiple “tile” sources
• Proof of concept of use AOP data and Google Earth
• New version of data format and server will allow to visualize AOP data in their geographical context
```
C:\Visus\Visus\python36\python.exe -m pip install pygobject
C:\Visus\Visus\python36\python.exe -m pip install pymap3d
```
Progressive vs Linear Computation of Time Averages for Climate Simulations

- Linear data access vs Progressive data access
- Linear average computation vs Progressive average computation
- Target result: Linear (standard approach) vs Progressive (OpenVisus)
Integrated Data Acquisition, Management and Computation for Neuroscience

(1) Data Source

(2) Preliminary Interactive Analytics

(3) Asynchronous Parallel Processing

(4) Interactive, Exploratory Assessment and Feedback
Remote Monitoring of Data Quality During Acquisition
Online Acquisition and Interactive Visualization of Terascale Microscopy
A Virtual Reality Visualization Tool for Neuron Tracing (VRNT)

Conducted a design study with neuroanatomists at the Moran Eye Center to develop a new tool for manual neuron tracing in VR

Tracing works similar to tilt brush, faster and more intuitive than desktop software
Topology Aided Neuron Tracing in a Virtual Environment

We use the Morse-Smale Complex to find a set of candidate neuron arcs
Evaluation in a neuroscience lab and outreach in a science museum
Python (Jupyter) scripting for local/remote data processing and visualization on demand

PyDataset and PyViewer to simplify data exploration

```python
element=(0.04, 0.95, 0.05, 0.73, 0.15, 0.15+0.1)
dataset=LoadDatasetPy("visus.idx")
RGB, bounds=dataset.readData(element,-6)
...
viewer=PyViewer()
viewer.addVolumeRender(RGBA, bounds)
```
Demo: Interactive Remote Analysis and Visualization of 6TB Imaging Data

- EM datasets of resolution: 130Kx130Kx340

6.4 TB of raw data

Web Server
Demo: Interactive Remote Analysis and Visualization of 6TB Imaging Data

- EM datasets of resolution: 130Kx130Kx340

6.4 TB of raw data
Deployment targets and extensions

**Live addition to NEON website**
- Identify data products
- Complete data ingestion pipeline
- Dedicated data portal
- Management of geospatial coordinate system
- Interoperability with google maps for context
- Implement testing procedures
- Basic enhancements of visualization and navigation
- Improve embedding (e.g., share)

**Continuing support and advanced features**
- Scaling of and addition of new data products
- Improve color blending of orthomosaics
- Python support (for download and scripting)
- Collaborative user interface
- Support for LIDAR data
- Support for user/NEON data upload
References

• Utah endpoint
  https://neon.visus.org/neonapi/products/{productCode}

• Endpoint and web viewer source code repository
  https://github.com/sci-visus/neon-visus

• Documentation and other use cases
  www.visus.org