Cloud Storage of DOE-Funded DAS Data for Public Accessibility

Nicole Taverna and Michael Rossol
National Renewable Energy Laboratory (NREL), Golden, CO
DAS Virtual Workshop and Tutorial
August 12, 2020
Geothermal Data Repository (GDR)

- The Department of Energy’s (DOE’s) Geothermal Technologies Office’s (GTO’s) **public data repository for data associated with geothermal research**
- Data is:
  - **Submitted** by researchers
  - **Curated** by NREL
  - **Made publicly accessible** through a variety of nodes
DAS Data Storage Caveats

• Too large to efficiently upload directly to GDR
• Difficult to find an organization willing to cover storage costs indefinitely
• Storage on hard drives:
  – Drive failure is the number one cause of data loss
  – Reduces public accessibility
Amazon Web Services (AWS) Simple Storage Service (S3)

<table>
<thead>
<tr>
<th>Object</th>
<th>Timestamp</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>20160308/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20160311/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20160312/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20160313/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20160314/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20160315/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20160321/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20160322/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20160323/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20160324/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20160325/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20160326/</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Showing 1 to 17 of 17 entries
Open Energy Data Initiative (OEDI)

- CLOUD PARTNER RELATIONSHIPS
- INNOVATIVE DATASET ACCESS
- DATA LAKE & ANALYTICS

https://data.openei.org/
How OEDI Data Lake Works

1. The incoming flow represents data—structured and unstructured—from high-value sources ranging from energy, laboratory, and resource data, analytic tools, use cases, scientific reports, and more.

2. The reservoir of water allows for a confluence of data, making it accessible for analysis in new ways.

3. Cloud vendors enable sustained access to data hosting solutions on a variety of public platforms.

4. The outflow of water is the analyzed energy data, filtered and streamlined for greater access and analysis of data, accelerating new insights and innovation.
AWS Data Lake

Datasets

Cloud Optimized Tools

Analysis

AWS S3

HSDS

AWS Glue

Amazon Athena

Tableau

Jupyter
The Highly Distributed Scalable Service (HSDS)
HDF on the Cloud

Big Idea: Map individual HDF5 objects (datasets, groups, chunks) as Object Storage Objects

Each chunk (heavy outlines) get persisted as a separate object

Parallel requests to S3 allow the HSDS service to scale to the current service demand while not introducing bottlenecks into data flow at the point of data retrieval.

Image Credit: HDF Group.
PoroTomo DAS Dataset: Cloud Storage

- Brady’s Hot Springs
- ~40 TB of horizontal array data
  - High cost of storage
- Trenched (horizontal) DAS array (DASH)
- Downhole (vertical) DAS array (DASV)
- Stored in SEG-Y and hdf5 format
  - Available via HSDS coming soon

Mystery event from PoroTomo DAS dataset (Miller et al.)
# Porotomo DAS Data on GDR

https://gdr.openei.org/submissions/980

## Porotomo Natural Laboratory Horizontal and Vertical Distributed Acoustic Sensing Data

### 11 Resources

<table>
<thead>
<tr>
<th>Resource Type</th>
<th>Description</th>
<th>View</th>
</tr>
</thead>
<tbody>
<tr>
<td>DASH Data in OEDI S3 Viewer</td>
<td>Link to Porotomo DASH data in Open Energy Data Initiative (OEDI) data viewer. Allows users to browse and download individual or groups of files.</td>
<td>View</td>
</tr>
<tr>
<td>DASH Data on AWS in SEG-Y format</td>
<td>Location of Porotomo DASH data on Amazon Web Services S3 Management Console in SEG-Y format.</td>
<td>View</td>
</tr>
<tr>
<td>DASV Data in OEDI S3 Viewer</td>
<td>Link to Porotomo DASV data in Open Energy Data Initiative (OEDI) data viewer. Allows users to browse and download individual or groups of files.</td>
<td>View</td>
</tr>
<tr>
<td>DASV Data on AWS in SEG-Y Format</td>
<td>Location of Porotomo DASV data on Amazon Web Services S3 Management Console in SEG-Y format.</td>
<td>View</td>
</tr>
<tr>
<td>Interactive Jupyter Notebook DAS Tutorial</td>
<td>Interactive Jupyter Notebook that provides a tutorial for working with DAS data, in SEGY format, using the Porotomo dataset</td>
<td>View</td>
</tr>
<tr>
<td>Porotomo Dataset Documentation</td>
<td>Documentation for the Porotomo dataset</td>
<td>View</td>
</tr>
</tbody>
</table>