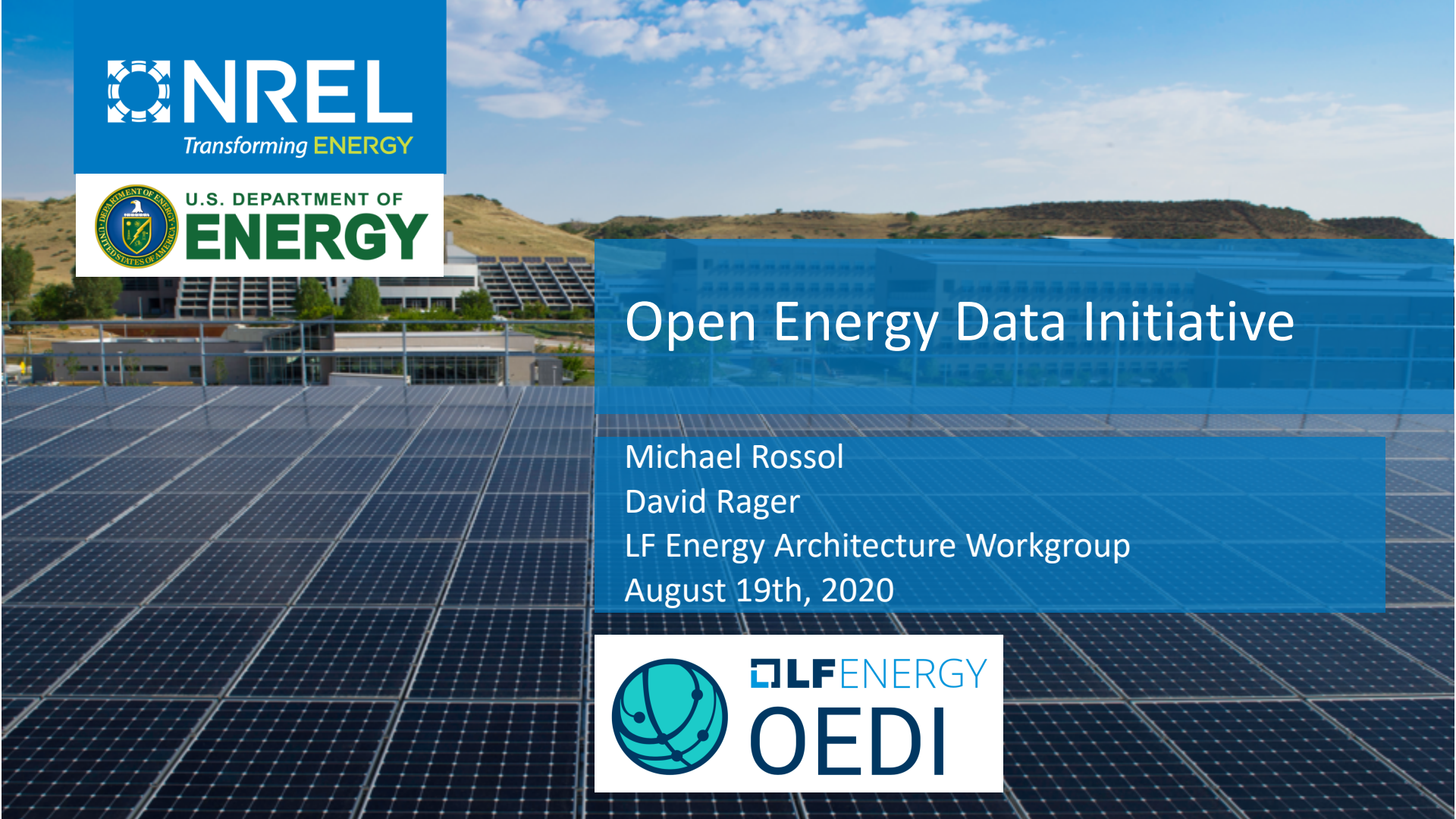


Open Energy Data Initiative

Michael Rossol
David Rager
LF Energy Architecture Workgroup
August 19th, 2020

LF ENERGY
OEDI

What is OEDI?

The Open Energy Data Initiative (OEDI) aims to improve and automate access of high-value energy data sets across the U.S. Department of Energy's (DOE's) programs, offices, and national laboratories.

Sponsored by DOE, this platform is being implemented by the National Renewable Energy Laboratory (NREL) to make data actionable and discoverable by researchers and industry to accelerate analysis and advance innovation.

Why OEDI?

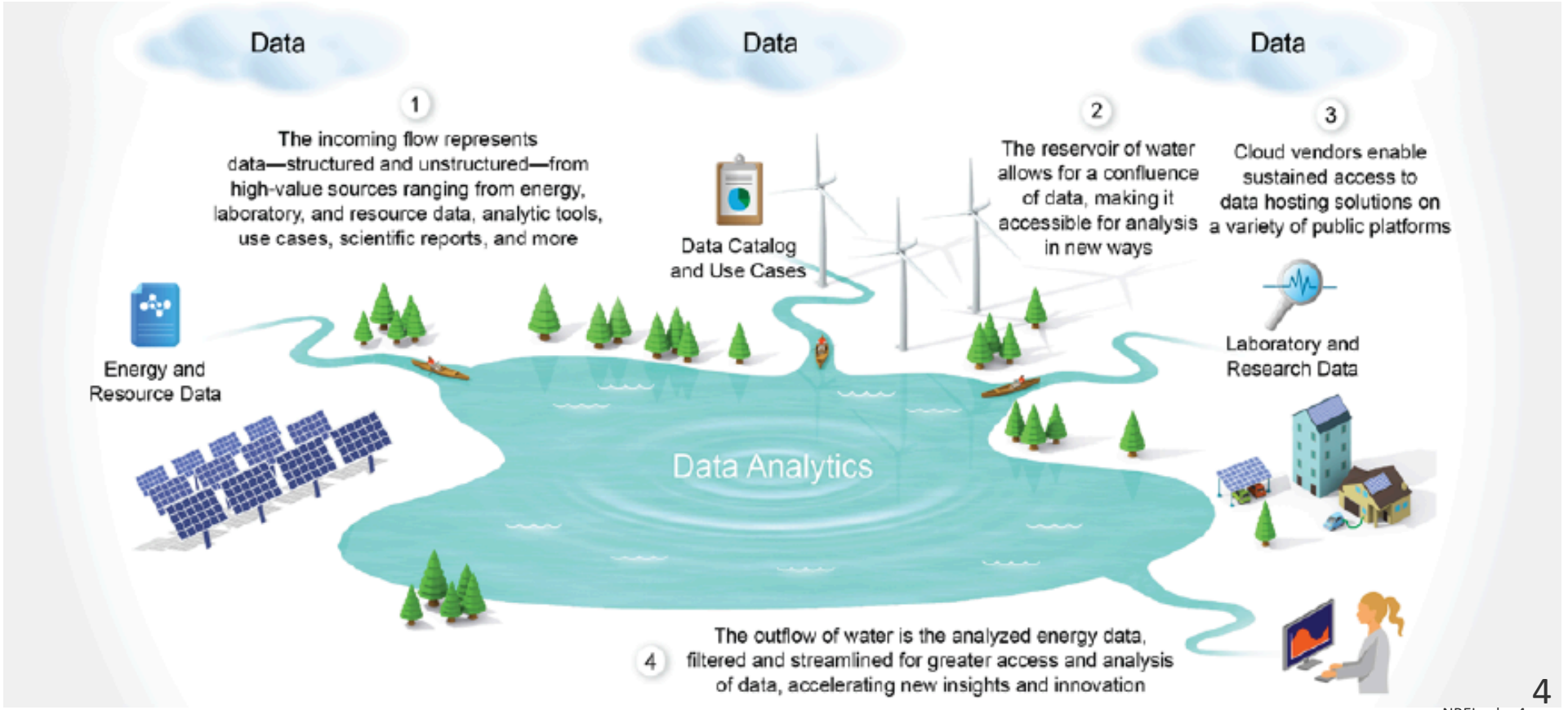
CLOUD PARTNER RELATIONSHIPS

INNOVATIVE DATASET ACCESS

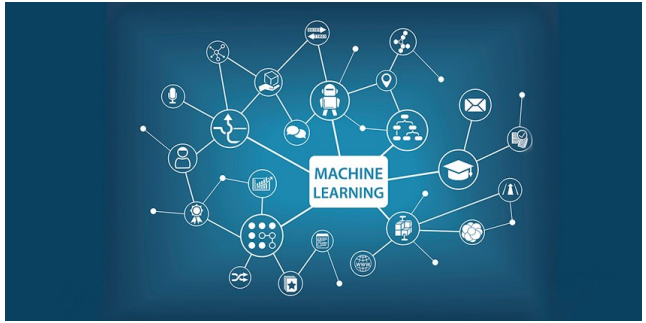
DATA LAKE & ANALYTICS

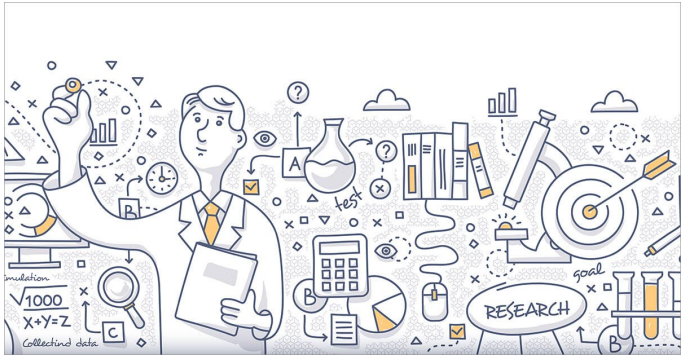
What is OEDI?

All DOE offices + 17 National Labs



What are we enabling?





OEDI Datasets

Available

- National Solar Radiation Database (NDRDB)
- Wind Toolkit Dataset
- US Historical Wave Data
- Porotomo DAS Data
- Utility Rate Database
- LBL Tracking the Sun Data
- PV Rooftop DataBase

In progress

- SMART-DS
- Cities-LEAP
- HTEM Database (High Throughput Experimental Materials)

To Be Linked

- GOES NOAA data
- US Wind Farm Database (USWFDB)
- Geothermal Data Repository (GDR)
- Marine Hydro-kinetic Data Repository
- Live Wire

Under consideration

- PV Module/Cell Performance Data
- GHCN-D Climate Dataset
- ???

What other high value datasets should be included sooner, rather than later?

Amazon Web Services (AWS) Data Lake

Datasets

Cloud Optimized Tools

Analysis

AWS S3

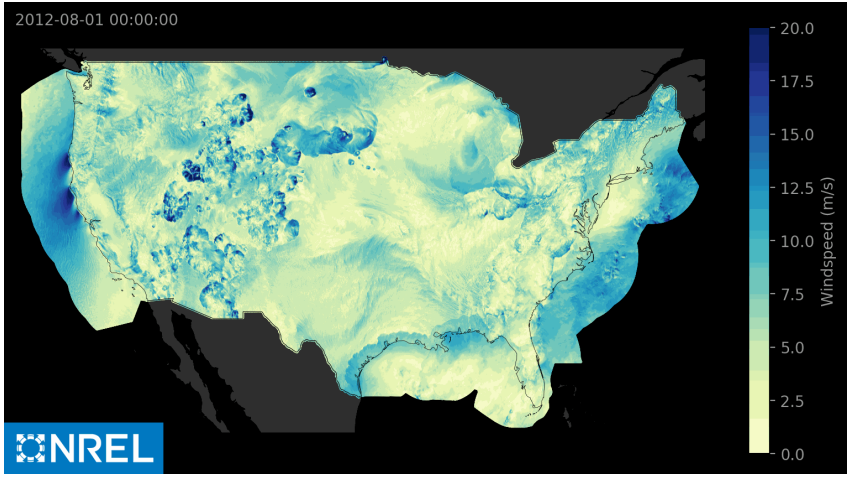
HSDS

AMAZON ATHENA

Optimizing spatio-temporal data for the cloud

National Solar Radiation Database (NSRDB)

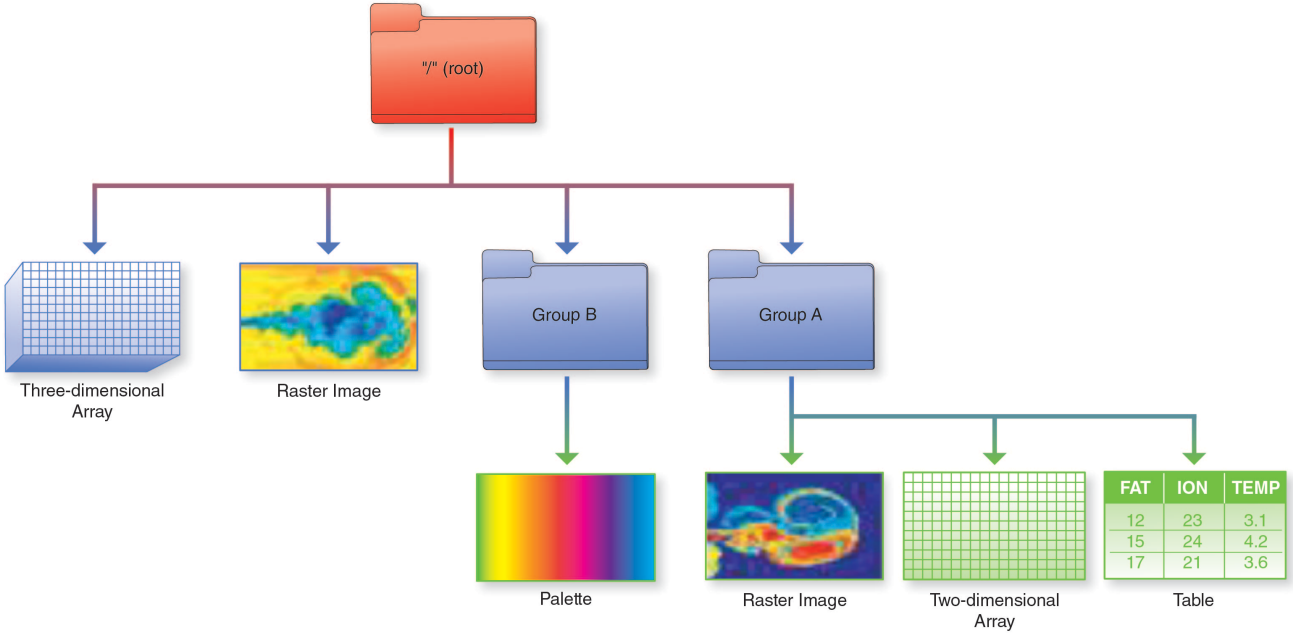
WIND Toolkit



	NSRDB*	WTK
Years	1998-2019	2007-2013
Spatial Resolution	4km x 4km	2km x 2km
Temporal Resolution	30 min	5 min
Geographic Extent	Western Hemisphere	CONUS
File Format	HDF5	HDF5
Size	36 TB	400 TB

*2018 and 2019 5min – 2km is also available

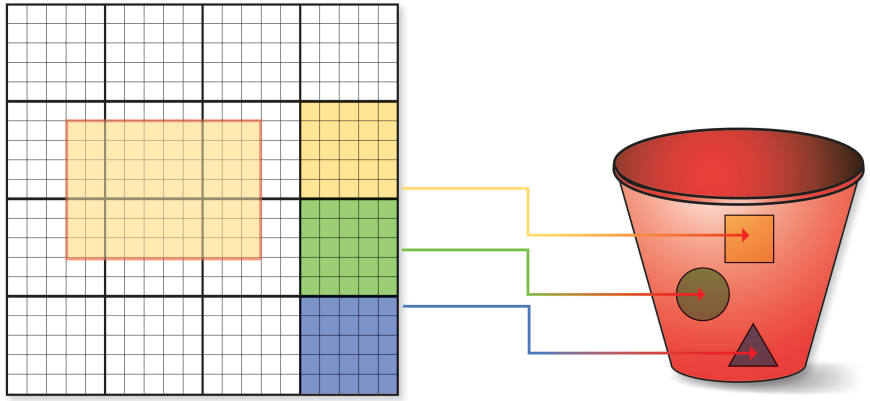
What is HDF?


"/" (root)
Three-dimensional Array
Raster Image
Group B
Group A
Palette
Raster Image
Two-dimensional Array
FAT ION TEMP
12 23 3.1
15 24 4.2
17 21 3.6
Table

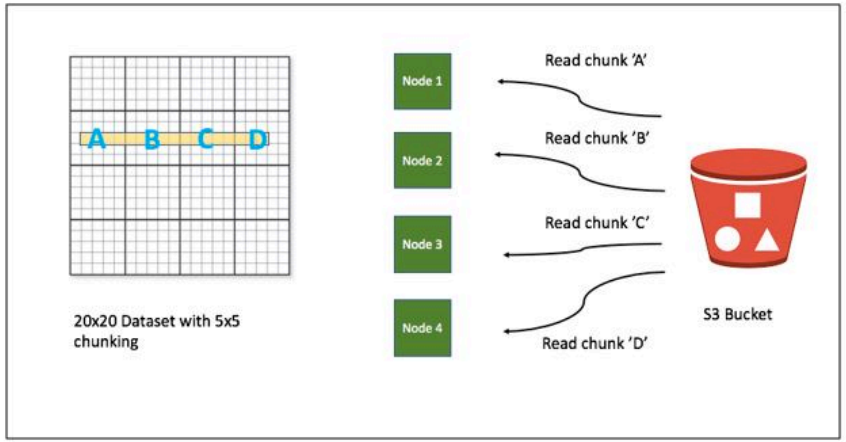
HDF for the cloud -> HSDS

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

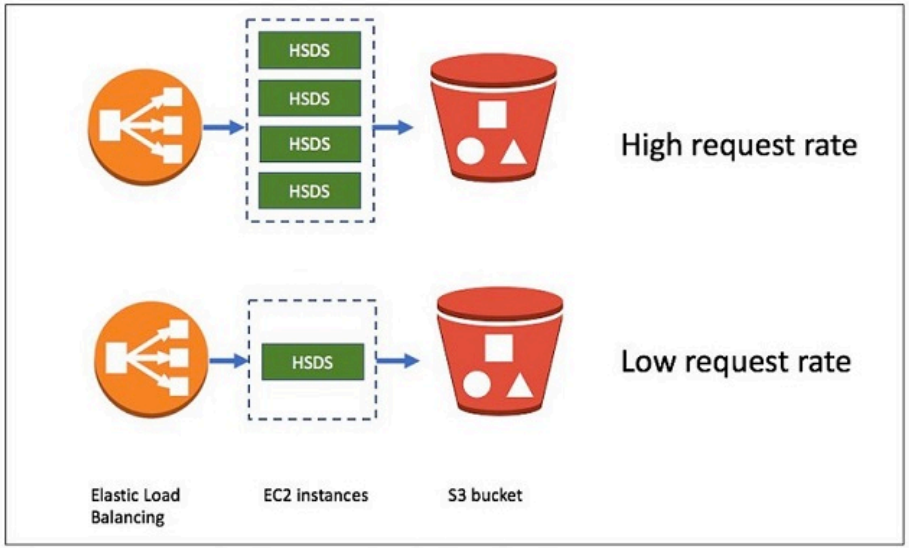
Each chunk (heavy outlines) get persisted as a separate object



Solution: Highly Scalable Data Service (HSDS) HDF + AWS



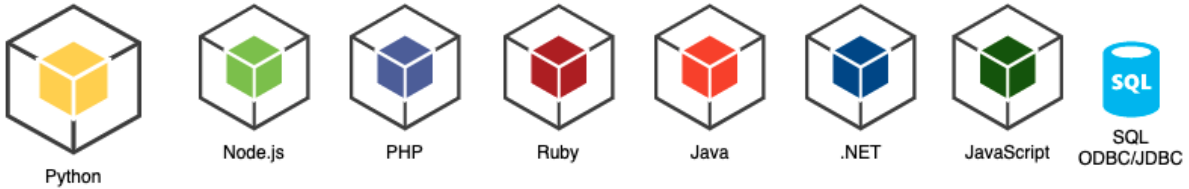
Parallel requests to S3 allow the HSDS service to scale to the current service demand while not introducing bottlenecks into data flow at the point of data retrieval.
Image Credit: HDF Group.



The HSDS service responds to the request volume by elastically scaling resources.
Image Credit: HDF Group

AWS Athena

API support for multiple languages:

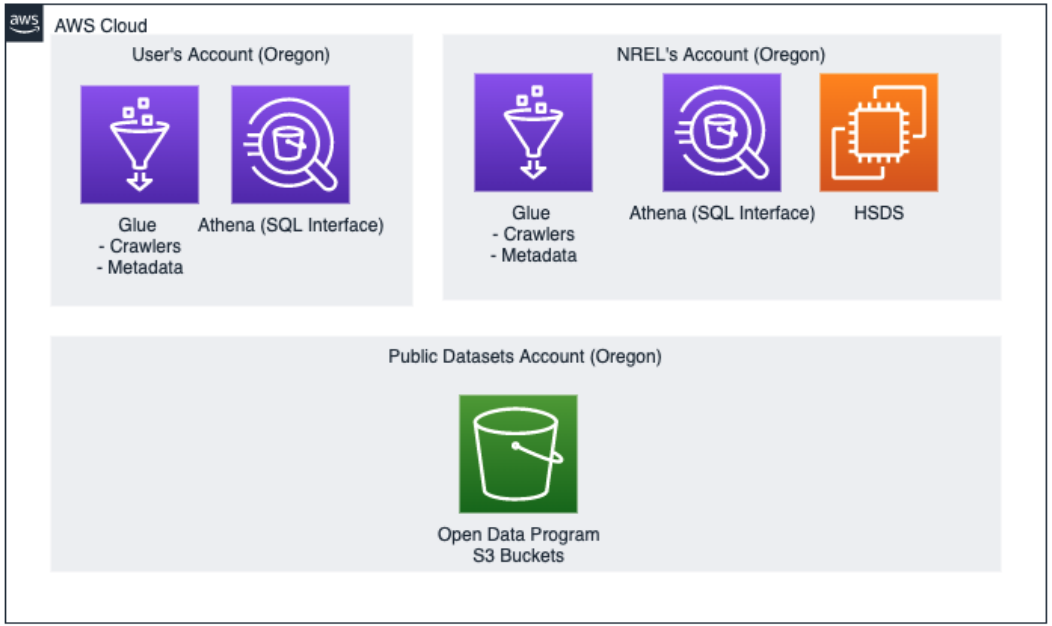


Leverages AWS Services

- Athena – cluster per query
- Glue – simplifies metadata collection and storage

OEDI helps partition the data in S3 for optimal query times

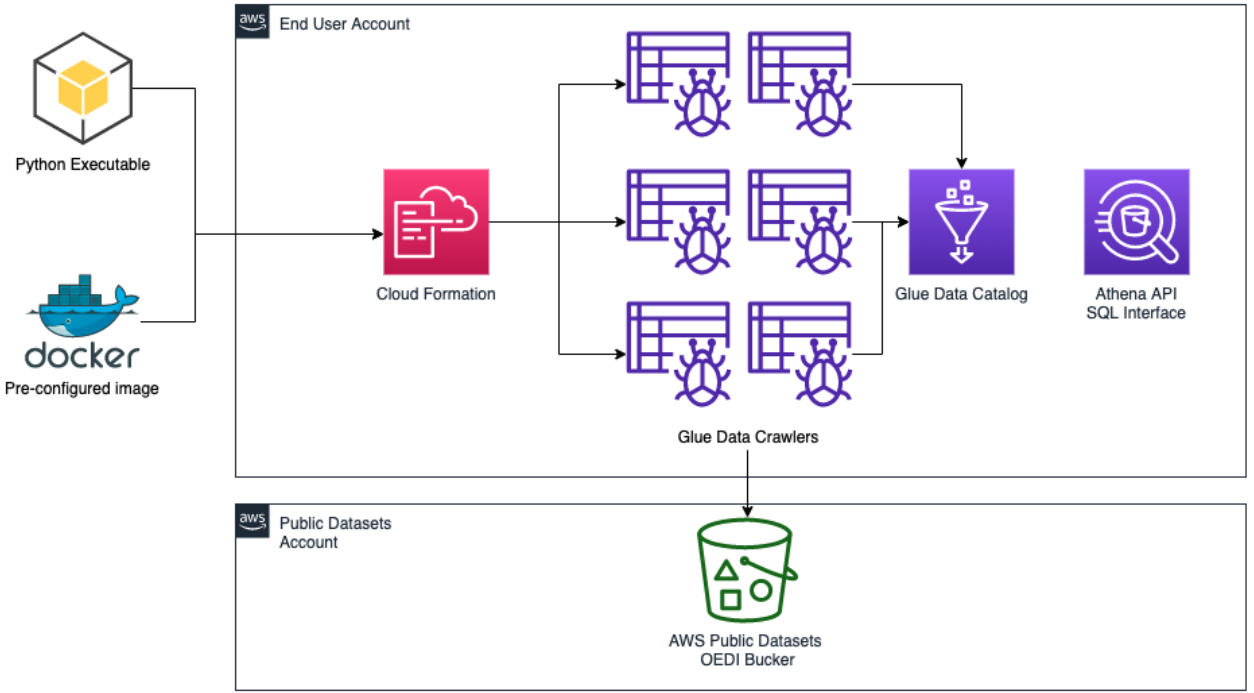
Data Formats Supported
Parquet (preferred), JSON, CSV, TSV, ORC



OEDI Distribution to Enable the AWS APIs:

Python or Docker initiate CloudFormation (infrastructure as code) and the execution of the Glue crawlers.

Once the crawlers create the metadata in Glue Athena is prepared for use with all the included data.



Glue Data Catalog

Table Metadata

Name	nrel_pv_rooftops_aspects
Description	
Database	oedi_database
Classification	parquet
Location	s3://nrel-pv-rooftops/aspects/
Connection	
Deprecated	No
Last updated	Wed Jun 10 08:41:39 GMT-600 2020
Input format	org.apache.hadoop.hive.ql.io.parquet.MapredParquetInputFormat
Output format	org.apache.hadoop.hive.ql.io.parquet.MapredParquetOutputFormat
Serde serialization lib	org.apache.hadoop.hive.ql.io.parquet.serde.ParquetHiveSerDe
Serde parameters	serialization.format **1**
	sizeKey **183832454617** objectCount **166**
	UPDATED_BY_CRAWLER **nrel-pv-rooftops-aspects**
Table properties	CrawlerSchemaSerializerVersion **1.0** recordCount **42354990**
	averageRecordSize **404** CrawlerSchemaDeserializerVersion
	compressionType **none** typeOfData **file**

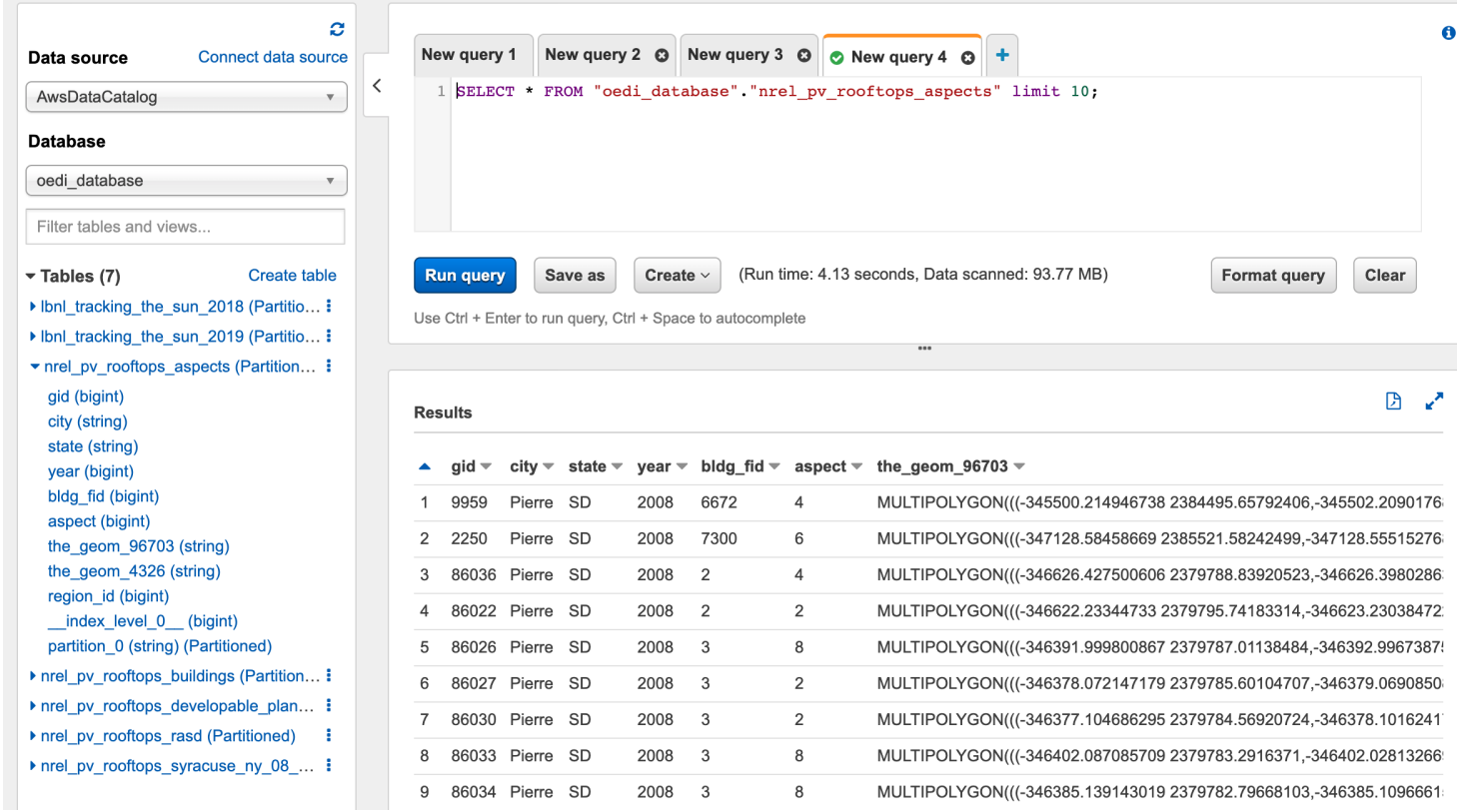
Table Schema

	Column name	Data type
1	gid	bigint
2	city	string
3	state	string
4	year	bigint
5	bldg_fid	bigint
6	aspect	bigint
7	the_geom_96703	string
8	the_geom_4326	string
9	region_id	bigint

Partitions

boise_id_07	View files
laguardiajfk_ny_07	View files
tulsa_ok_08	View files
philadelphia_pa_07	View files
dayton_oh_12	View files
worcester_ma_09	View files
huntsville_al_09	View files
providence_ri_04	View files
wichita_ks_12	View files
cleveland_oh_12	View files
greensboro_nc_09	View files
seattle_wa_11	View files

Athena Query Interface



Questions?

- michael.rossol@nrel.gov
- david.rager@nrel.gov
- data.openei.gov
- github.com/openEDI/documentation
- github.com/nrel/hsds-examples
- registry.opendata.aws