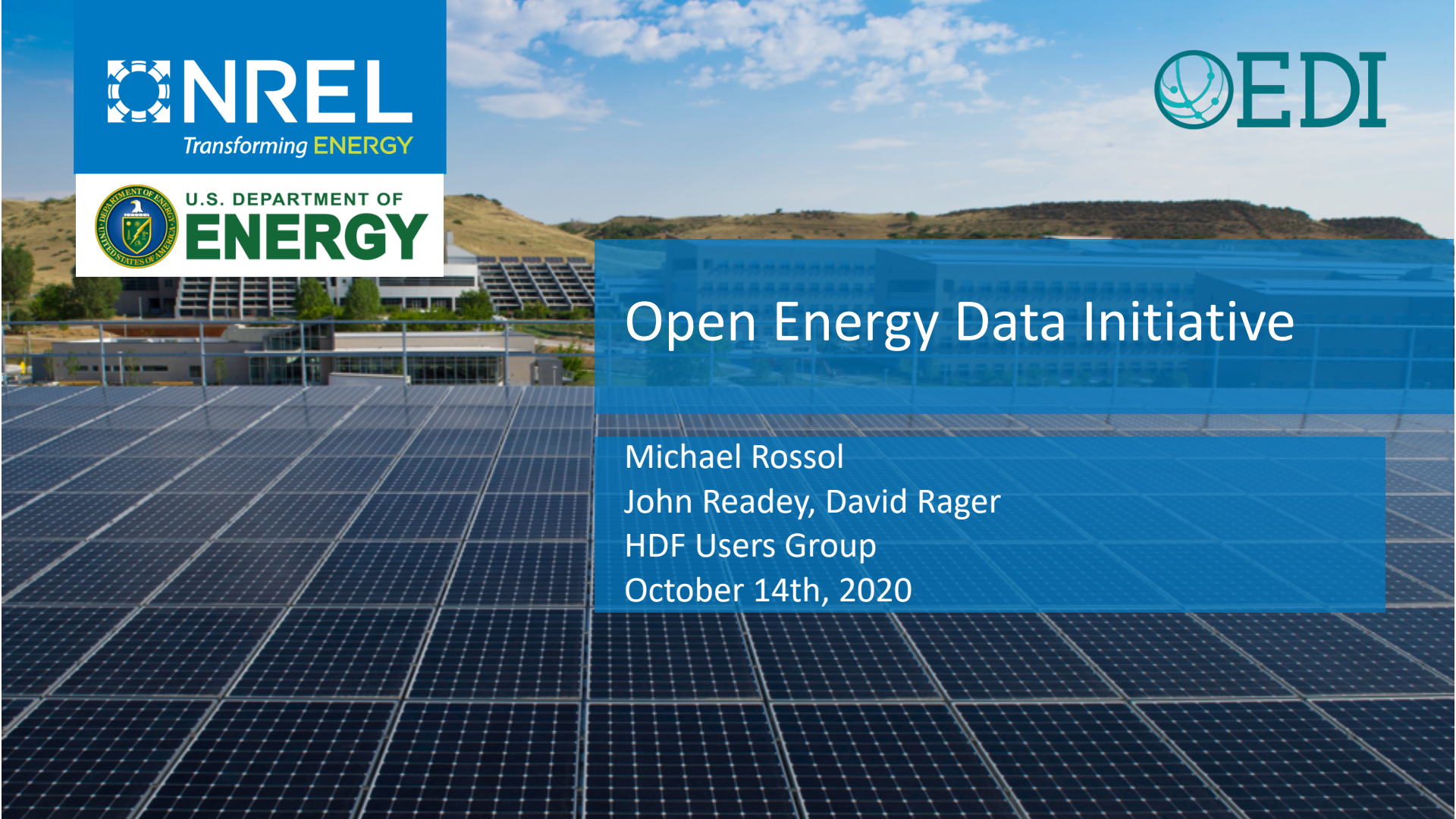


Open Energy Data Initiative

Michael Rossol
John Readey, David Rager
HDF Users Group
October 14th, 2020

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?

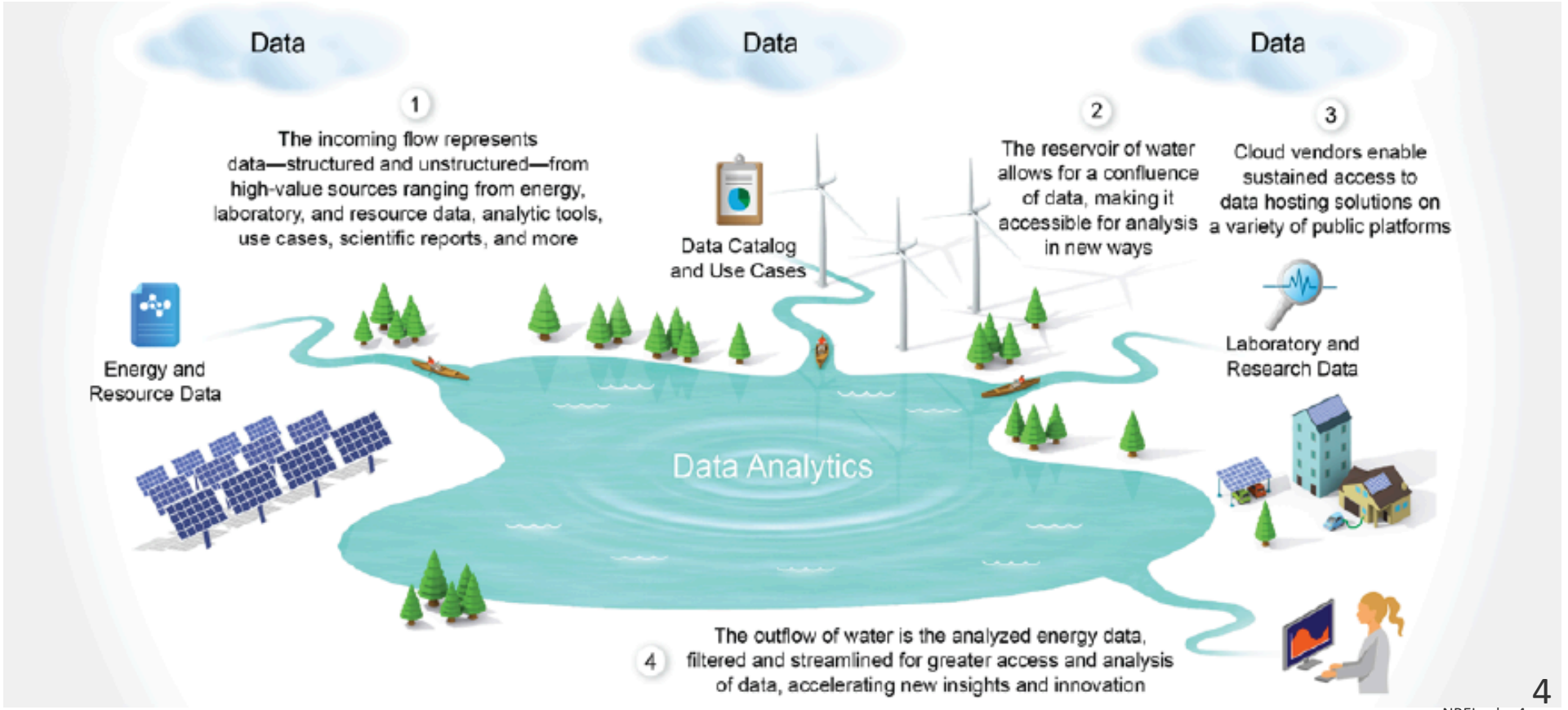
CLOUD PARTNER RELATIONSHIPS

INNOVATIVE DATASET ACCESS

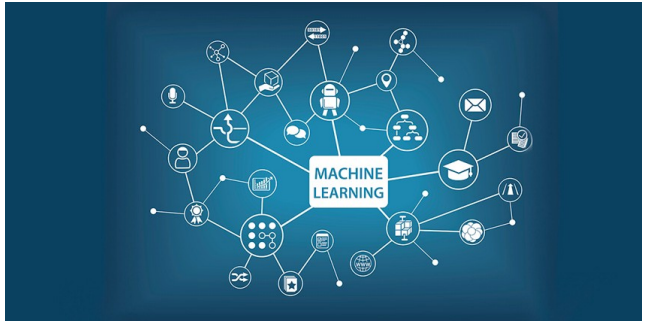
DATA LAKE & ANALYTICS

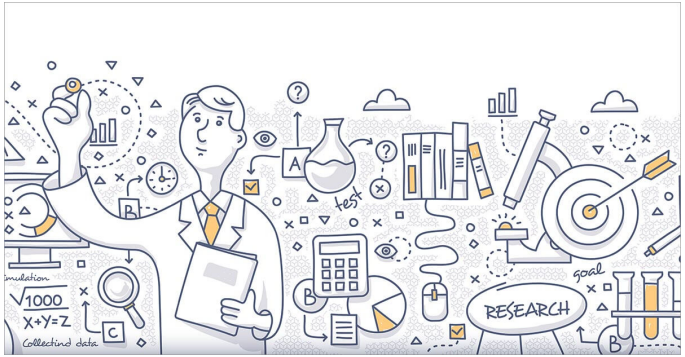
What is OEDI?

All DOE offices + 17 National Labs

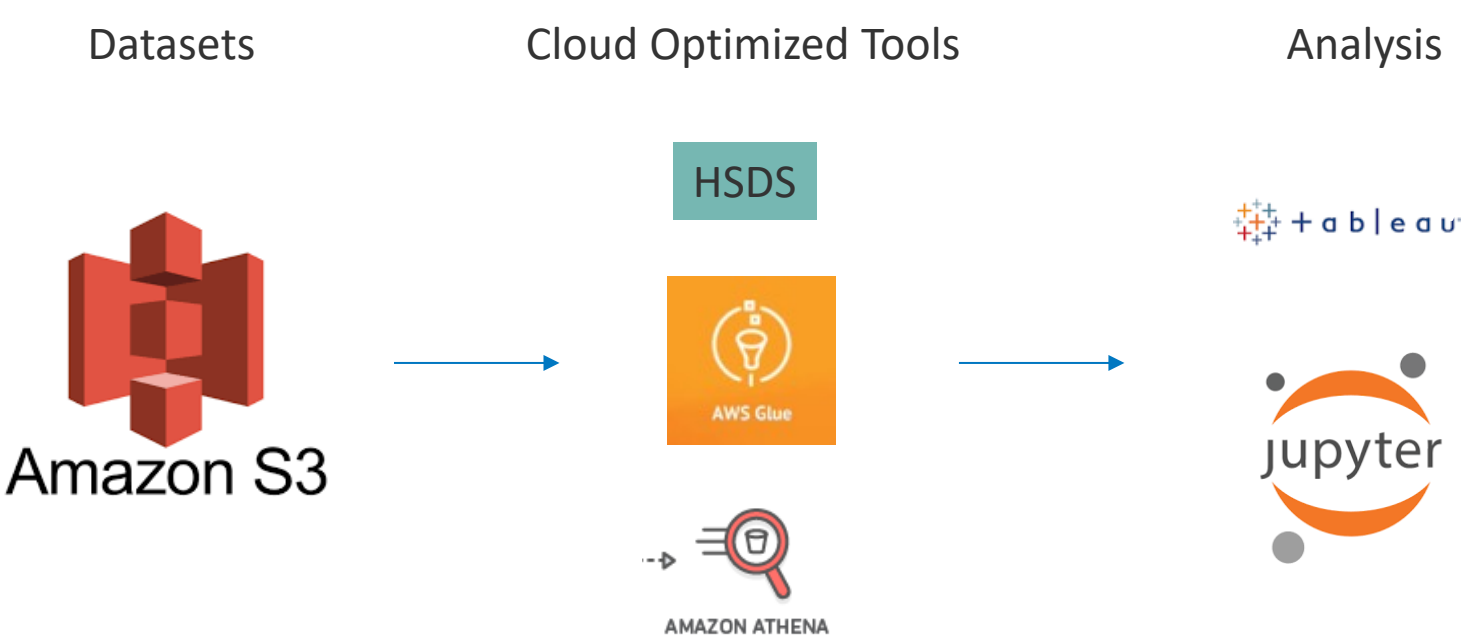


What are we enabling?





Amazon Web Services (AWS) Data Lake



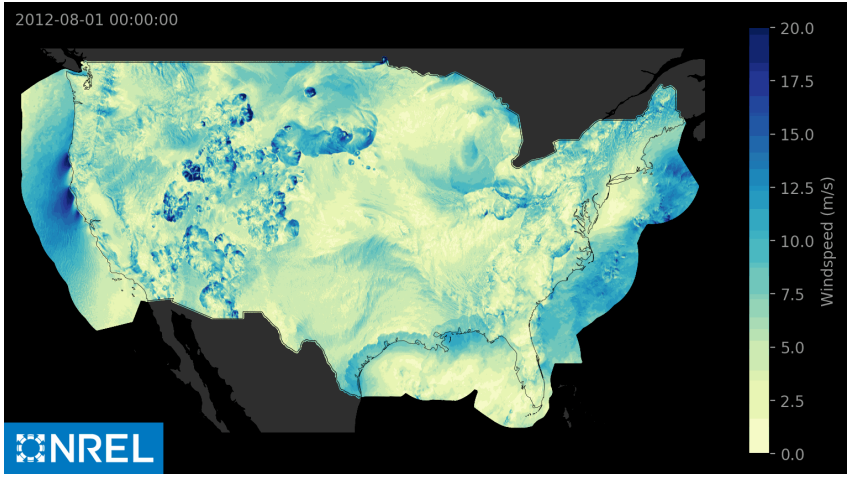
Disseminating spatio-temporal data from the cloud

National Solar Radiation Database (NSRDB)

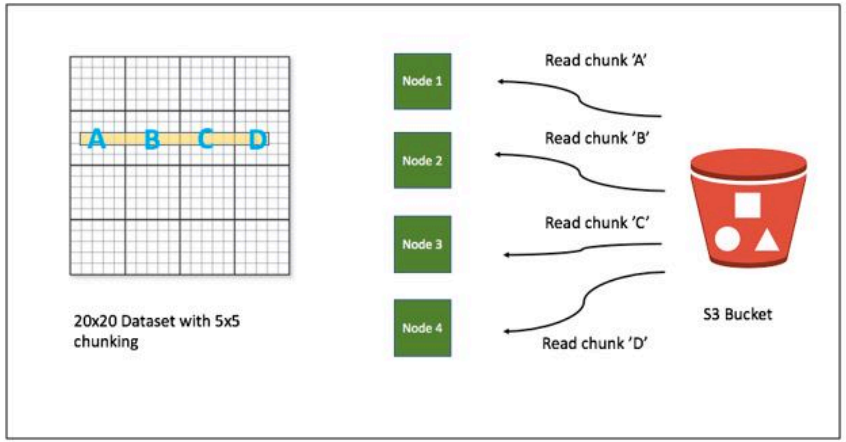
	NSRDB*	WTK
Years	1998-2019	2007-2014
Spatial Resolution	4km x 4km	2km x 2km
Temporal Resolution	30 min	5 min
Geographic Extent	Western Hemisphere	North America
File Format	HDF5	HDF5
Size	46 TB	424 TB

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

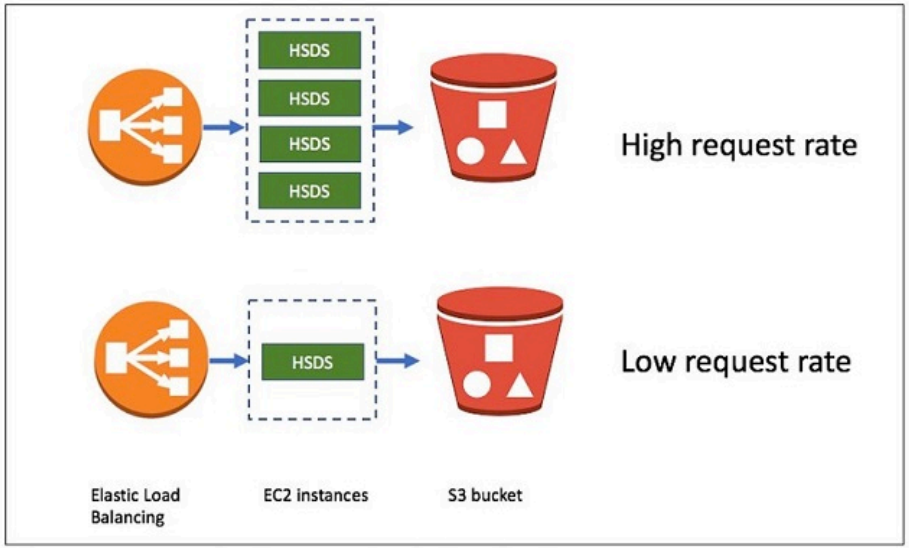
WIND Toolkit



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.

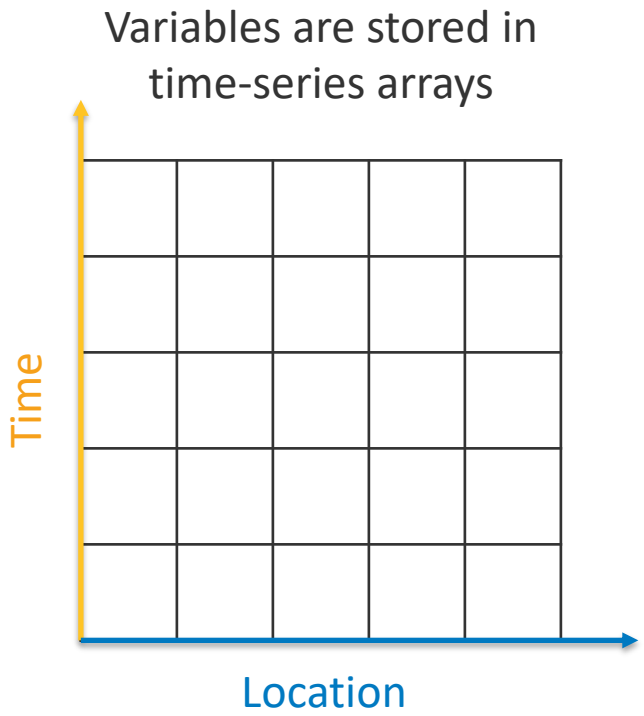


The HSDS service responds to the request volume by elastically scaling resources.
Image Credit: HDF Group

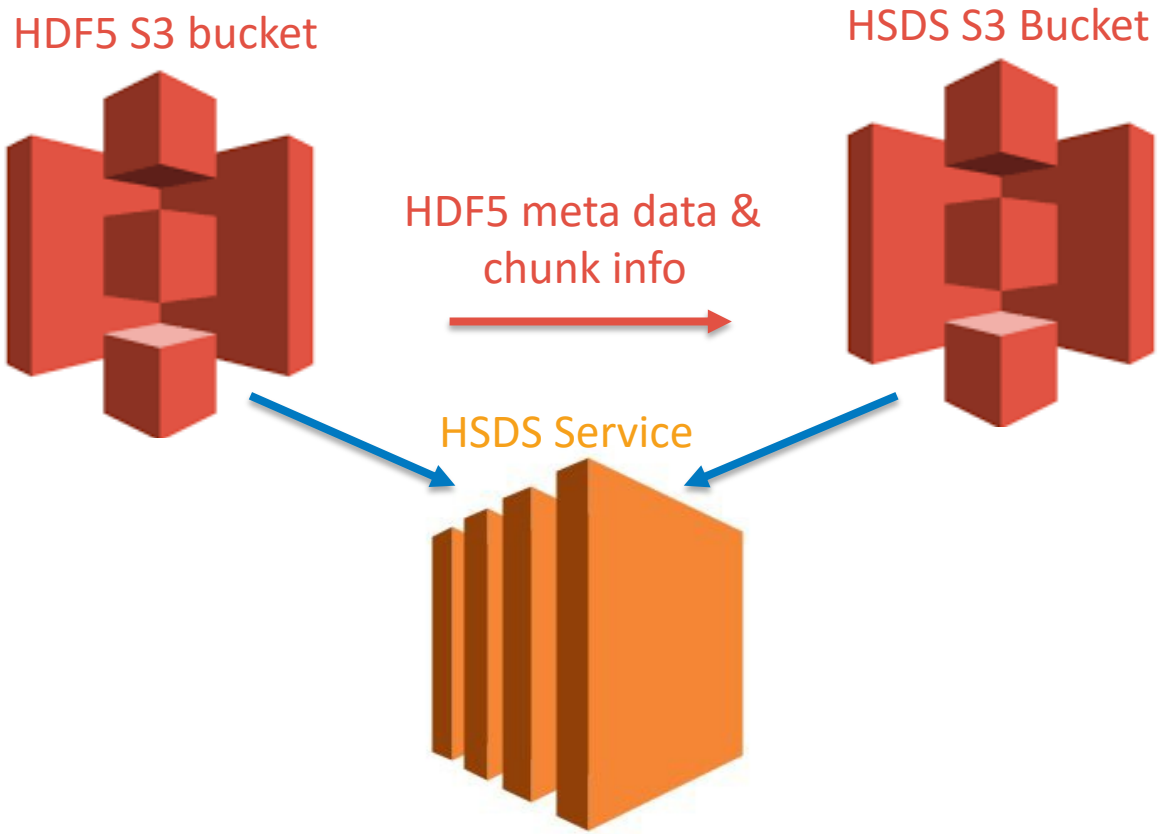
Leveraging HDF5 file format

wtk_conus_2010.h5
coordinates Dataset {2488136, 2}
meta Table {2488136}
precipitationrate_0m Dataset {8760, 2488136}
pressure_100m Dataset {8760, 2488136}
relativehumidity_2m Dataset {8760, 2488136}
temperature_100m Dataset {8760, 2488136}
time_index Dataset {8760}
winddirection_100m Dataset {8760, 2488136}
windspeed_100m Dataset {8760, 2488136}

Time-series datasets are self documented with:
- units, description, scale factor attributes



Leveraging HSDS



NRELs HSDS resources
In partnership with the AWS open data program

HDF5 buckets:
- s3://oedi-data-lake/umcm
- s3://nrel-pds-nsrdb
- s3://nrel-pds-porotomo/H5
- s3://nrel-pds-wtk
- s3://wpto-pds-us-wave

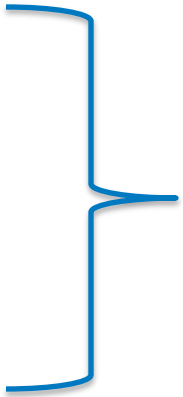
HSDS bucket:
- s3://nrel-pds-hsds

HSDS Service:
- Endpoint: https://developer.nrel.gov/api/hsds
- API Key: https://developer.nrel.gov/signup/

Unique advantage of HSDS

5min WTK data for CONUS (17 TB per year)
HDF5 files by hub-height:

- wtk_conus_2010_0m.h5
- wtk_conus_2010_2m.h5
- wtk_conus_2010_10m.h5
- wtk_conus_2010_40m.h5
- wtk_conus_2010_60m.h5
- …
- wtk_conus_2010_200m.h5



Single HSDS ‘file’ sourced for multiple .h5 files

/nrel/wtk/conus-5min/wtk_conus_2010.h5

Unique challenges of HSDS

- Auto-scaling:
  - to handle stochastic service usage patterns
  - to leverage HSDS' parallel framework for large request
- Loading large datasets:
  - Easiest to read meta-data from a local file
  - This is not always the most convenient
- Fixing data bugs:
  - Easy when data is loaded into HSDS
  - Cumbersome when HSDS is reading from .h5 files on S3

Questions?

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