National Aeronautics and Space Administration

Tackling the Four V's with NEXUS

Frank Greguska, Kevin Gill, Thomas Huang, Joseph Jacob, Nga Quach, and Brian Wilson
Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109-8099, U.S.A.

Volume

14.6 PB
Archived Earth Observation Data[1]

Velocity

16.0 TB / Day
Earth Observation Data Archived Daily[1]

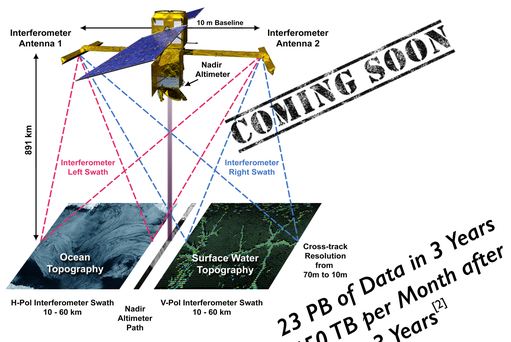
Variety

9,462
Unique Earth Observation Datasets[1]

Veracity✓

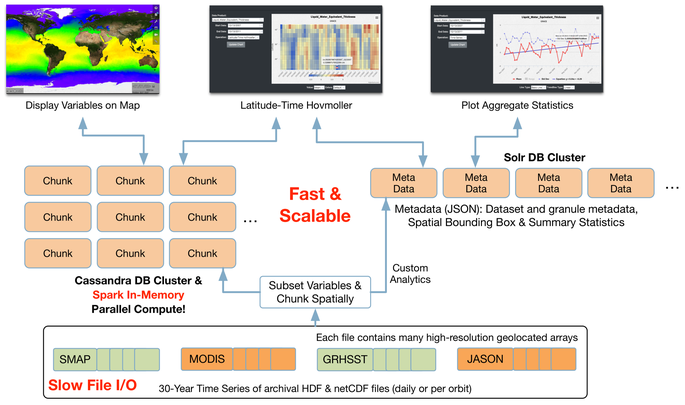
1,127 Level 1 Products
2,210 Level 2 Products
961 Level 3 Products
961 Level 4 Products[1]

Surface Water and Ocean Topography (SWOT)[2]



23 PB of Data in 3 Years
450 TB per Month after 3 Years[2]

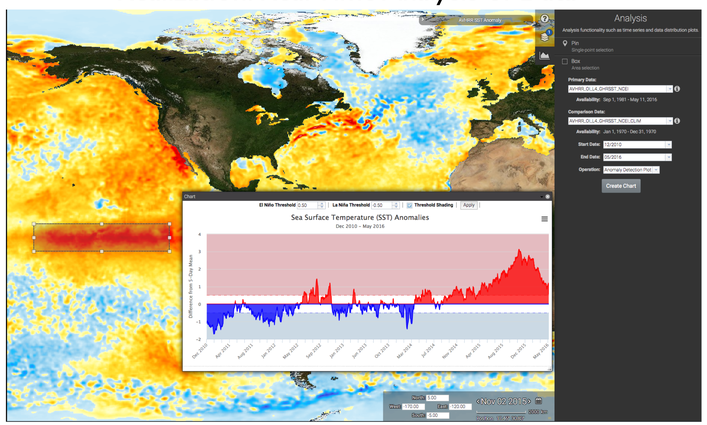
NEXUS[3]

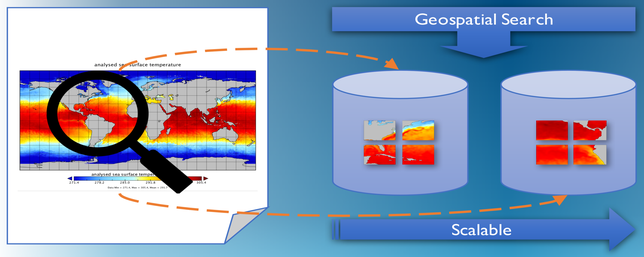


Open Source Technology Developed at JPL

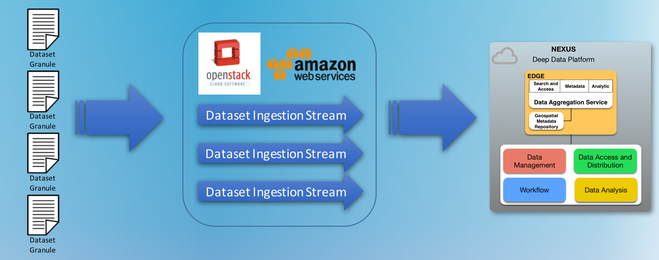
- On Premise or Cloud Ready
- Horizontally Scalable
- Geospatial Search Capabilities
- Data Products Tiled for Fast Access

Realization: Real Time Analysis of Data[4]

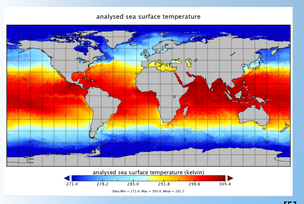




- Tiling Algorithms
  - Break datasets into smaller tiles
  - Index tiles for fast spatial lookup
- Horizontally Scalable Storage

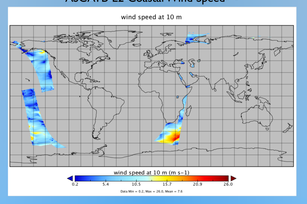


- Dataset Ingestion Streams
  - Cloud Ready
  - Scale Horizontally



AVHRR_OI-NCEI-L4-GLOB-v2.0 Sea Surface Temperature[5]

ASCATB-L2-Coastal Wind Speed[6]



- Configurable Streams
  - Multiple Tiling Algorithms
  - Support L2 Swath Data
  - Support L3/L4 Gridded Data

$$c = \text{Number of tiles desired}$$
$$d = \text{Number of dimensions}$$
$$L_d = \text{Length of dimension } d$$
$$S_d = \text{Step size for dimension } d$$
$$S_d = \left\lfloor \frac{L_d}{\frac{1}{c}} + \frac{1}{2} \right\rfloor$$

```
def filter_empty_tiles(self, tile):
    # Only supply data if there is actual values in the tile
    if tile.data.size - numpy.count_nonzero(numpy.isnan(tile.data)) > 0:
        yield tile.data
    else:
        print "Discarding data %s from %s because it is empty" % (tile.section_spec, tile.granule)
```

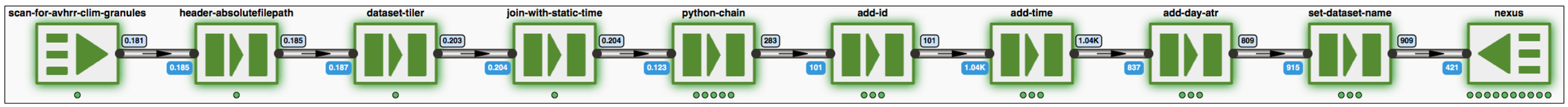
- Pluggable validation checks

```
def transform(self, tile):

    # Subtract 360 if the longitude is greater than 180
    tile.data.longitudes[longitudes > 180] -= 360

    yield tile.data
```

- Data transformation



Stream For AVHRR_OI-NCEI-L4-GLOB-v2.0 Sea Surface Temperature

© 2017 California Institute of Technology. Government sponsorship acknowledged. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsements by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

[1] https://earthdata.nasa.gov/about/system-performance/eosdis-annual-metrics-reports
[2] http://swot.jpl.nasa.gov/science/resources
[3] https://github.com/dataplumber/nexus
[4] https://oceanxtremes.jpl.nasa.gov
[5] https://podaac.jpl.nasa.gov/dataset/AVHRR_OI-NCEI-L4-GLOB-v2.0
[6] https://podaac.jpl.nasa.gov/dataset/ASCATB-L2-Coastal

Questions or Comments?

greguska@jpl.nasa.gov