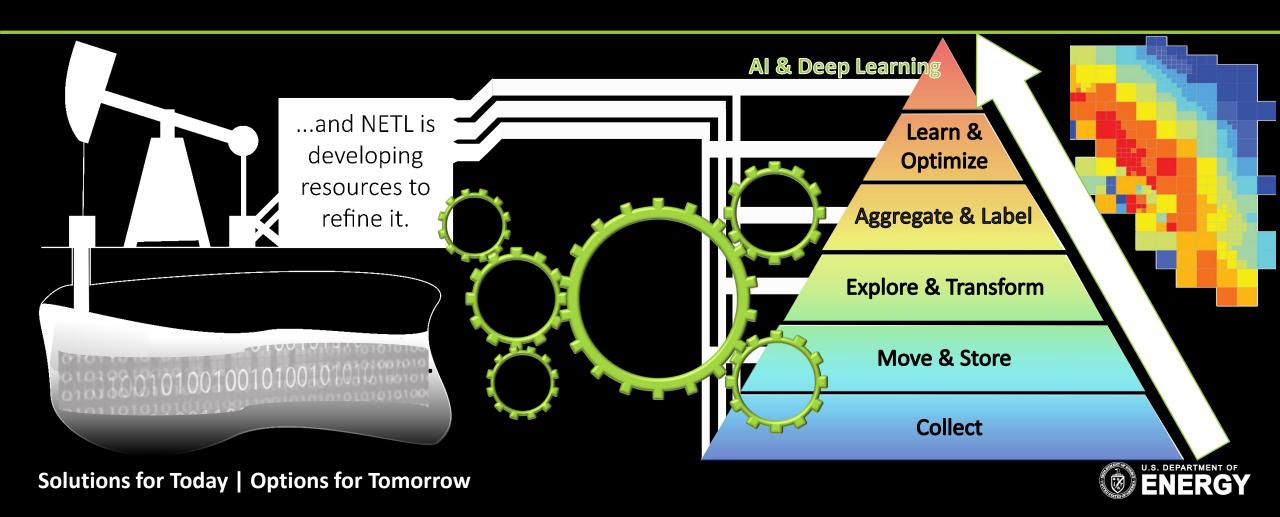
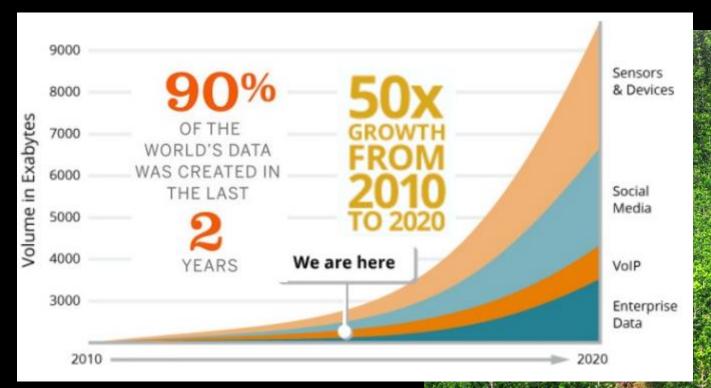
# Data is the new oil...





As access to open, authoritative data increases science driven analyses face challenges to **efficiently find**, **integrate** and **use** these resources





- Volume, variety, and velocity of data online is growing... exponentially
- How will you parse the tree from the forest?



### **Use Case: FTP Data Mining: Hadoop + EDX**

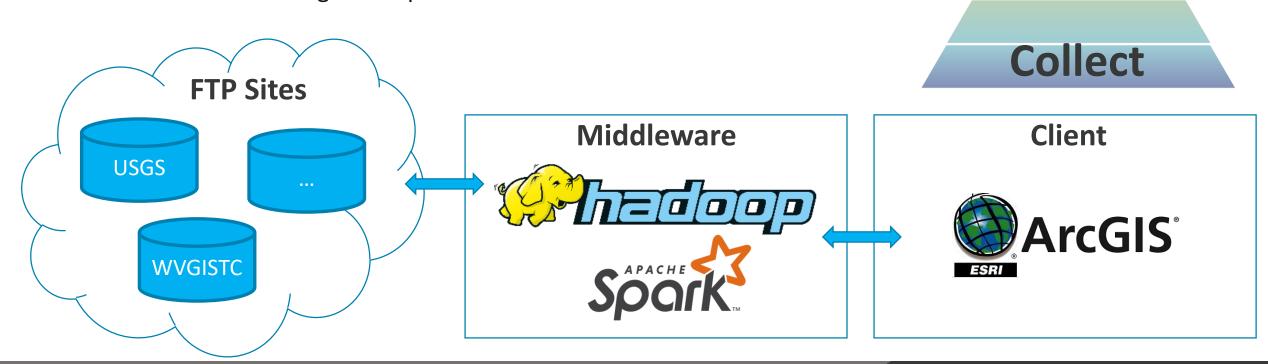


#### • Problem:

Need to search data in FTP silos (millions of files, spatial and contextual)

#### • Solution:

Index FTP silos using Hadoop



## **NETL's Big Data Discovery Ecosystem (To Date)**







#### **Data Collection:**

- FTP Recursion
- WWW Crawl

Metastore (Hive, HBase)

### Data Analysis:

- Phrase Generation
- Relevance Analysis
- Geoprocessing











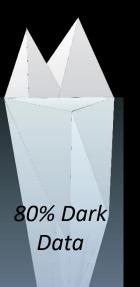
tank aboveground pre-treatment dehydration hydrogen welding fraction glycol pipeline production floating Gathering fuel Transmission shipping belowground emissions station water condensate parts liquefaction seal dioxide removal natural valve salt failure tee storage stack billion offshore flow Boosting propane void platform service diesel carbon dehydrator roof ppb ship Distribution block holdup recovery crossed dry top-down contaminant facility terminal fugitive iron intrastate line separator consequence million plant blowdown ratio refinery pump refinery pump processing hydrocarbon unit



NATIONAL ENERGY TECHNOLOGY LABORATORY

Convert Search
Terms & Phrases
into an Open O&G
Spatial Database

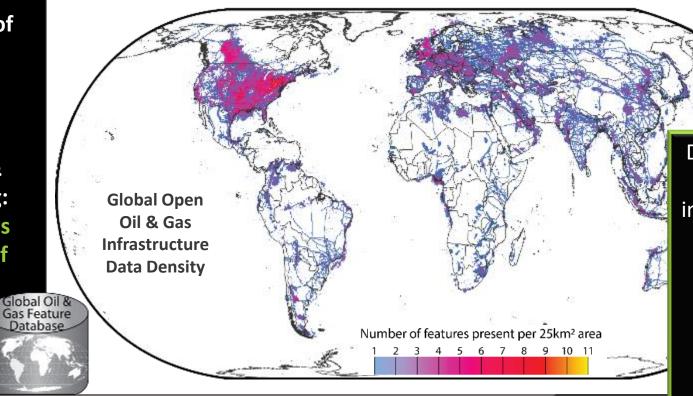
Machine Learning Automated
Approach- A tool that scans "seed"
resources and identifies relevant
keywords, then crawls the web and
parses the data for integration



4 months period of performance

Acquisition of disparate data by country, region, & continent totaling:

- ~800 datasets
- >4 Millions of features



Discovered & integrated open data sources of information related to oil & gas infrastructure across the globe







## Learn & Optimize

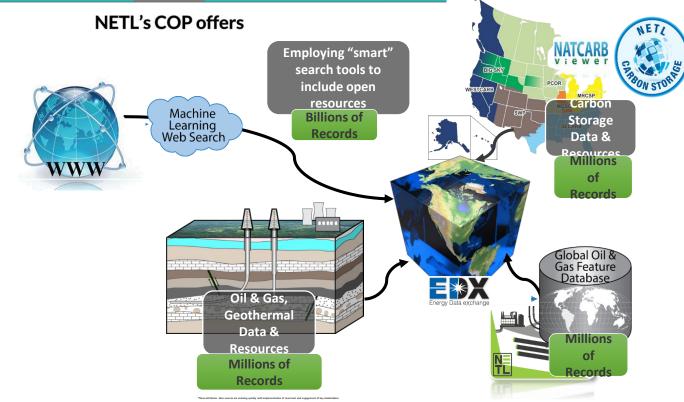




### **Future Big Data Development & Analysis**

- **EDX Cloud Services**
- **Living Database**
- **Common Operating** Platform for Data **Analytics**
- **Geocube Spatial Data** Viewer
- **Fuzzy Logic Analytics** (SIMPA)

- **AWS** Development
- Integration with decades of DOE R&D
- Federating Open Source data
- EDX & GeoCube (search and location)
- ID Data gaps in subsurface puzzle







# Data is the new oil...



