



# Utilizing Big Data Technologies to Mine Weather Information for Forensic Analytics

**Mr. Chad Hutchins, Naval Research Laboratory, Monterey CA.**

**Talk Given By: Dr. John Sample, Naval Research Laboratory, Stennis Space Center, MS**

Oceans in Action Workshop. August 17-18, 2016

# Problem Set

In order to understand the future behaviors of targets of interest, the Intelligence Community (IC) gathers all available information that could have a potential correlation to the target, including ocean and atmospheric (METOC) information. Authoritative METOC information on what occurred in the past is called *historical METOC* and the process of correlating events to historical METOC is called *forensic METOC*.

Current methods of finding and querying historical METOC information on the TOP SECRET/SCI (TS/SCI) network is very labor intensive and requires moving information from the UNCLASSIFIED network to the TS/SCI network.

To address fast querying and data storage on the TS/SCI network NRL developed **Ellipsis**, an environmental intelligence system for forensic analytics.

# Ellipsis Services

Ellipsis uses the Cloudera Enterprise distribution of Hadoop (CDH 5.7) to provide quick discovery and access to historical METOC information via JSON web services.

Services are built for machine to machine communication to allow a user to use their own preferred viewer.

## Services Include:

- Point
- Data within Point Radius
- Rectangle
- Station Specific

## All services with the following time selection:

- Single Point in Time
- Broken Time Range
- Continuous Time Range
- Data within Radius

Example query:

Return all stations within 10Km of  
36.59, -121.89 for 2010, June, 15<sup>th</sup>, all  
hours.

```
/ellipsis-ws/rest/metoc/getMetocInRadius?lat=36.59&lon=-121.89&time=&years=2010&months=06&days=15&hours=&fields=&metocVars=*&radius=10&stationId=&bbox=&
```

Query to reply: 4.88 seconds

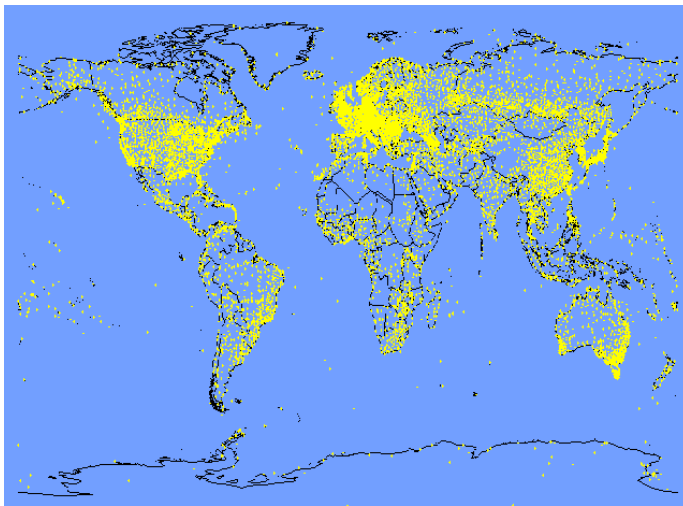
Snippet of Response:

```
{
  "ICOADS": {
    "data": []
  },
  "ISD": {
    "data": [
      {
        "wxstation": "994019",
        "hr": "3",
        "mo": "6",
        "LON": "-121.88",
        "slp": "1015.6",
        "dy": "15",
        "elev": "2",
        "visivar": "None",
        "visivarqual": "1",
        "skyceildet": "9",
        "atm": "12.2",
        "LAT": "36.6",
        "distance": "1.27",
        "UID": "9q924qz_994019_99999",
        "yr": "2010",
        "wban": "99999",
        "datetime": "201006150300"
      },
      {
        "wxstation": "994019",
        "hr": "1",
        "mo": "6",
        "LON": "-121.88",
        "slp": "1016.2",
        "dy": "15",
        "elev": "2",
        "visivar": "None",
        "visivarqual": "1",
        "skyceildet": "9",
        "atm": "13.0",
        "LAT": "36.6",
        "distance": "1.27",
        "UID": "9q924qz_994019_99999",
        "yr": "2010",
        "wban": "99999",
        "datetime": "201006150100"
      },
      {
        "wxstation": "994019",
        "hr": "5",
        "mo": "6",
        "LON": "-121.88",
        "slp": "1015.9",
        "dy": "15",
        "elev": "2",
        "visivar": "None",
        "visivarqual": "1",
        "skyceildet": "9",
        "atm": "11.4",
        "LAT": "36.6",
        "distance": "1.27",
        "UID": "9q924qz_994019_99999",
        "yr": "2010",
        "wban": "99999",
        "datetime": "201006150500"
      },
      {
        "wxstation": "994019",
        "hr": "18",
        "mo": "6",
        "LON": "-121.88",
        "slp": "1015.4",
        "dy": "15",
        "elev": "2",
        "visivar": "None",
        "visivarqual": "1",
        "skyceildet": "9",
        "atm": "11.3",
        "LAT": "36.6",
        "distance": "1.27",
        "UID": "9q924qz_994019_99999",
        "yr": "2010",
        "wban": "99999",
        "datetime": "201006151800"
      },
      {
        "wxstation": "994019",
        "hr": "22",
        "mo": "6",
        "LON": "-121.88",
        "slp": "1015.3",
        "dy": "15",
        "elev": "2",
        "visivar": "None",
        "visivarqual": "1",
        "skyceildet": "9",
        "atm": "12.3",
        "LAT": "36.6",
        "distance": "1.27",
        "UID": "9q924qz_994019_99999",
        "yr": "2010",
        "wban": "99999",
        "datetime": "201006152200"
      },
      {
        "wxstation": "994019",
        "hr": "13",
        "mo": "6",
        "LON": "-121.88",
        "slp": "1014.5",
        "dy": "15",
        "elev": "2",
        "visivar": "None",
        "visivarqual": "1",
        "skyceildet": "9",
        "atm": "10.8",
        "LAT": "36.6",
        "distance": "1.27",
        "UID": "9q924qz_994019_99999",
        "yr": "2010",
        "wban": "99999",
        "datetime": "201006151300"
      },
      {
        "wxstation": "994019",
        "hr": "6",
        "mo": "6",
        "LON": "-121.88",
        "slp": "1016.0",
        "dy": "15",
        "elev": "2",
        "visivar": "None",
        "visivarqual": "1",
        "skyceildet": "9",
        "atm": "11.2",
        "LAT": "36.6",
        "distance": "1.27",
        "UID": "9q924qz_994019_99999",
        "yr": "2010",
        "wban": "99999",
        "datetime": "201006150600"
      },
      {
        "wxstation": "994019",
        "hr": "10",
        "mo": "6",
        "LON": "-121.88",
        "slp": "1015.1",
        "dy": "15",
        "elev": "2",
        "visivar": "None",
        "visivarqual": "1",
        "skyceildet": "9",
        "atm": "10.9",
        "LAT": "36.6",
        "distance": "1.27",
        "UID": "9q924qz_994019_99999",
        "yr": "2010",
        "wban": "99999",
        "datetime": "201006151000"
      }
    ]
  }
}
```

# Ellipsis Datasets

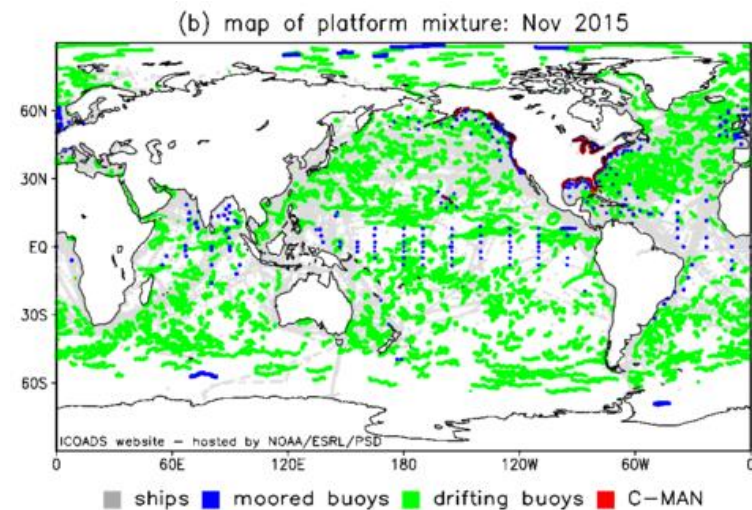
## Integrated Surface Database (NOAA/NCEI)

- Global hourly and synoptic observations
- 1901-2015
  - Initial dataset
  - Will be regularly updated



## International Comprehensive Ocean-Atmosphere Dataset (NOAA/NCEI)

- 10/1662-10/2008
  - Initial dataset
  - Will be regularly updated



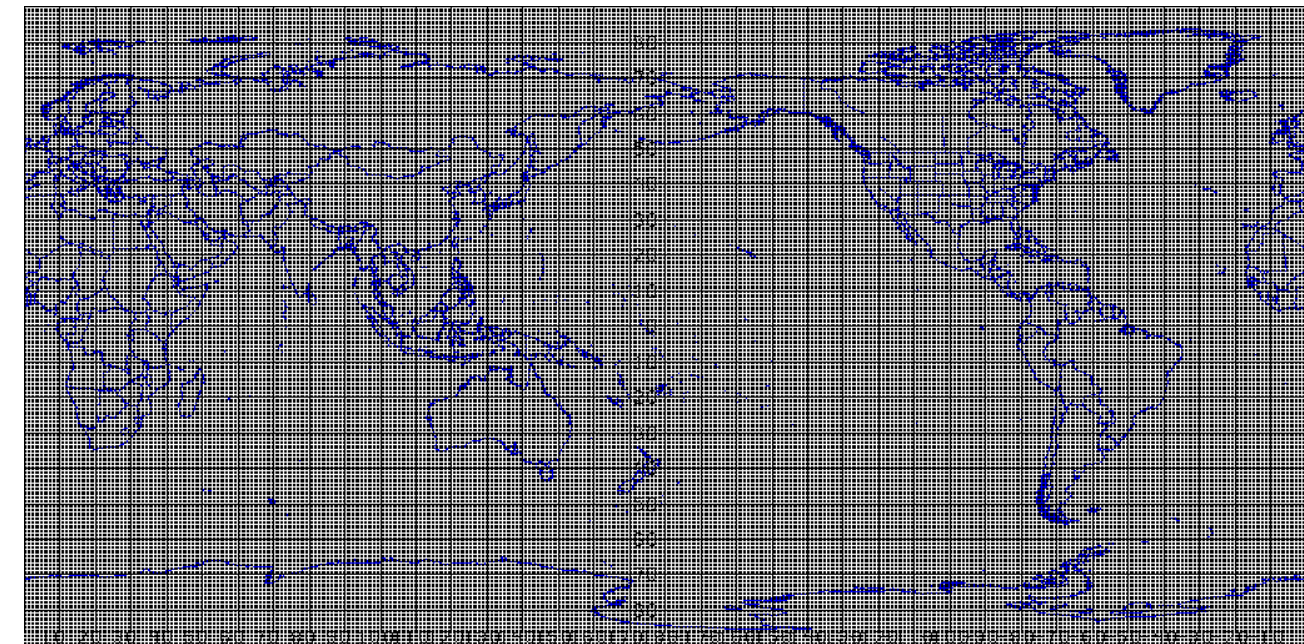
# Ellipsis Datasets

To provide context to non-meteorologists and oceanographers, the NOAA WAVEWATCH III Climate Forecast System Reanalysis (CFSR) is used to provide

- Mean
- 90<sup>th</sup> Percentile (Extreme Max)
- 10<sup>th</sup> Percentile (Extreme Min, not used for wind data)
- Standard Deviation

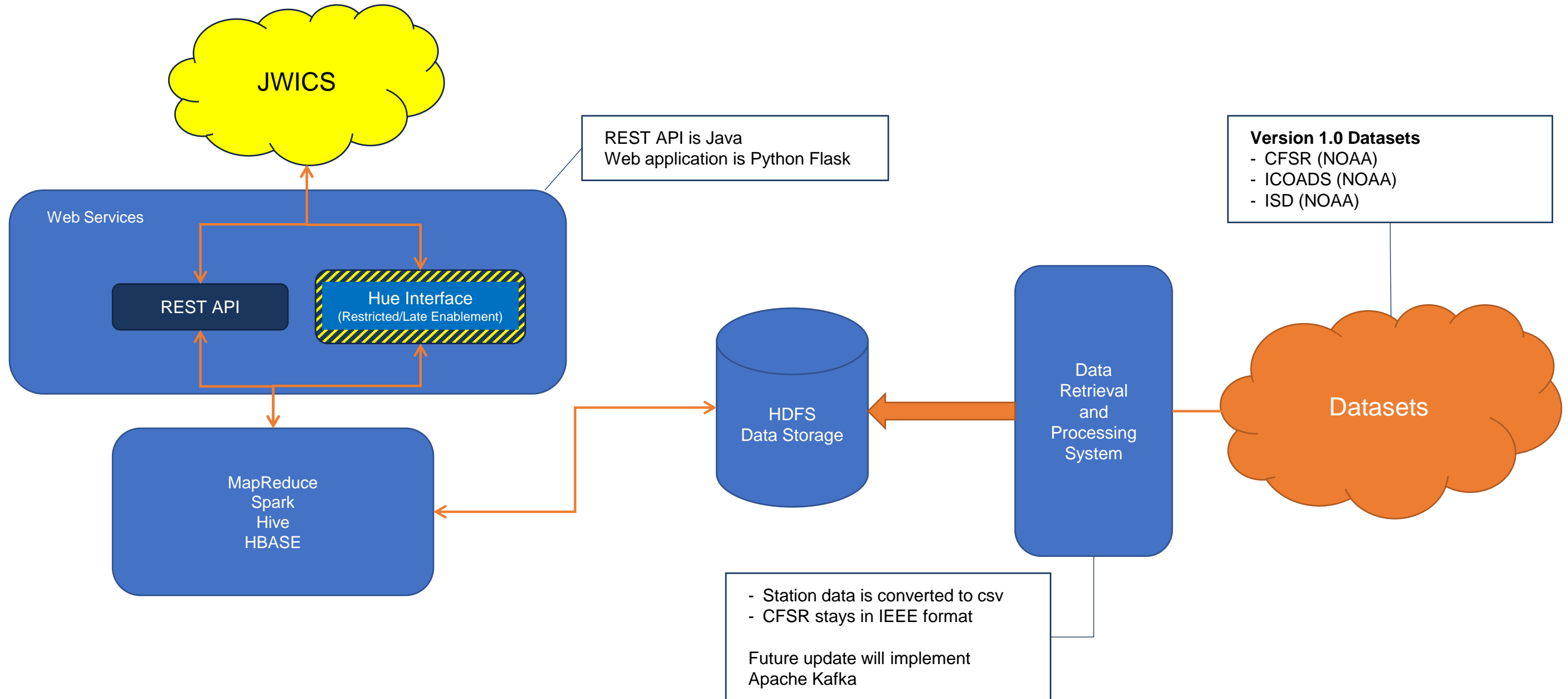
For the following variables

- 2m Air Temperature
- 2m Wind Speed and Direction
- Significant Wave Height



CFSR Grid Resolution (1/2  
Degree)

# System Diagram



# Ellipsis Development Hardware

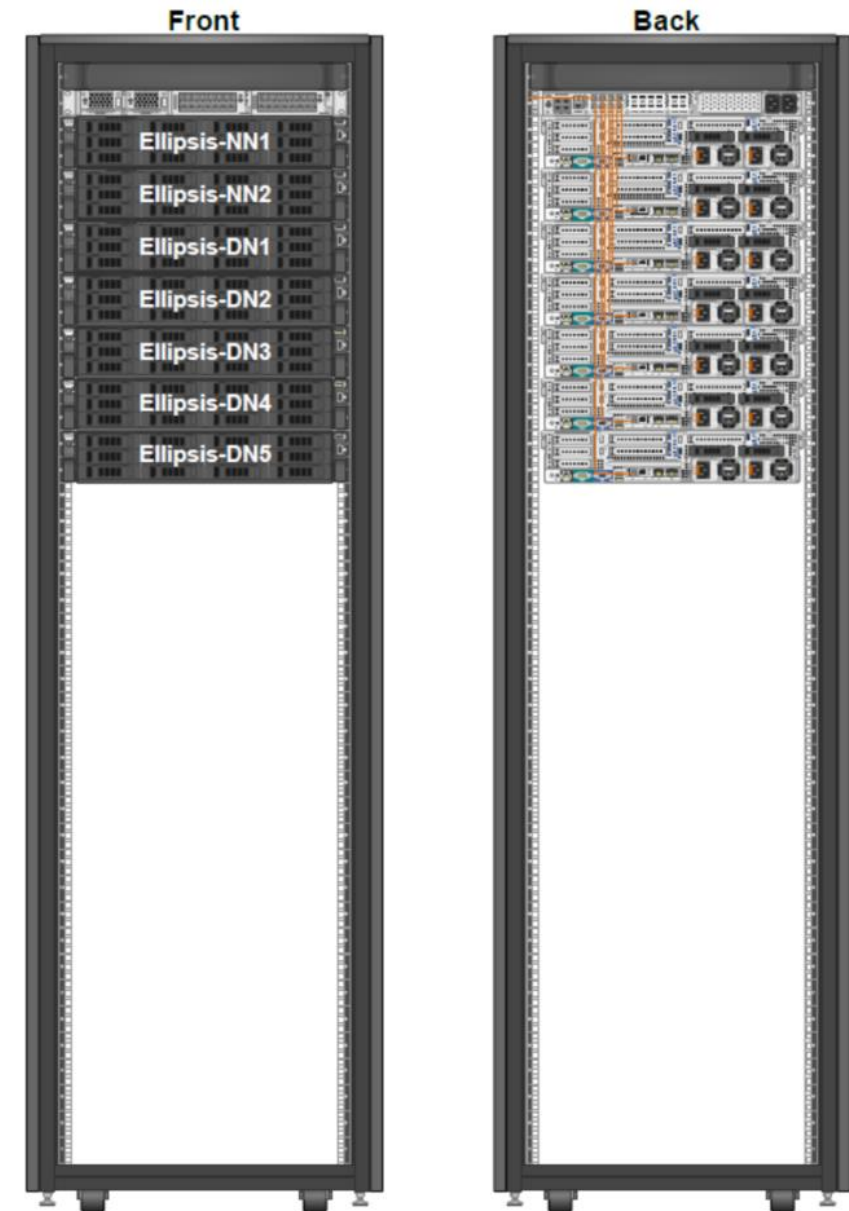
Hardware consist of 7, 2U Dell PowerEdge R730xd Dual Xenon 3.4GHz with 128GB of memory.

<b>CPU:</b>	6 Cores x 2 Sockets (24 w HT)
<b>Memory:</b>	128GB @ 2133MHz
<b>Boot SSD:</b>	120GB x2
<b>HDFS HDD:</b>	2TB x 12

Interconnect is Cisco Nexus 5010 with 10GB fiber connections.

107T HDFS disk space.

System is underpowered on memory and cannot support Cloudera Impala.





# Future Features

- Support streaming datasets using Apache Kafka.
  - Continuous observation stream.
- Allow spatial-temporal forensics. Given conditions at one location at a specific time, search for similar conditions elsewhere in space and time.
  - Allows analysts to assess the spatial and/or temporal uniqueness over the conditions at the location of interest.
    - Was this location special from a METOC sense?
    - Was this time special from a METOC sense?
- Anticipatory analytics.
  - Alerting based on when subscribed conditions at an area are expected based on environmental forecasts.
    - Useful for when IC events are believed to have been tied to METOC conditions.