Cloud Based Ingest & Processing Framework (I&PF)

- Currently a Solers Internal Research & Development (IR&D) project

- Primary Objectives:
  - *Enable fast/easy integration* of data sources, product algorithms, and data consumers within a cloud based workflow (or “data pipeline”) framework
  - Provide *easy to use web-based user interfaces* for search and access (for end users), as well as workflow monitoring and management (for system operatorsadmins)
  - Provide *RESTful web services* for other developers, scientists, etc. to *discover and access* the ingested/processed *data and metadata*, for use in other research / engineering initiatives (e.g., developing a new product algorithm)

- Leverages readily available commercial Amazon Cloud services:
  - EC2: Virtual Servers in the Cloud
  - Elasticsearch Service: Run and Scale Elasticsearch Clusters
  - ElastiCache: In-Memory Cache
  - S3: Scalable Storage in the Cloud

- Leverages readily available open source technologies:
  - Apache nifi
  - polymer
  - Mapbox
  - Webdis: A fast HTTP interface for Redis
  - Anaconda: Modern open source analytics platform powered by Python
Cloud Based I&PF Architecture

I&PF Workflow Engine
(Apache NiFi on AWS EC2)

Web-Based Monitoring & Management Interface

Data Ingest Workflow(s)

Product Generation Workflow(s)

Subscription Matching Workflow

Data Delivery Workflow

I&PF Metadata Repository
(AWS Elastisearch)

I&PF Data Storage
(AWS S3)

I&PF In-Memory Cache
(AWS ElastiCache [Redis])

RESTful Web Services

RESTful Web Services

RESTful Web Services

RESTful Web Services

RESTful Web Services

RESTful Web Services

Operators / Admins

Data Sources / Providers

Data Consumers / End Users

I&PF User Portal

Scientist / Engineer Resources

New Product Algorithm

Data Delivery Workflow

Data Ingest Workflow(s)

Product Generation Workflow(s)

Subscription Matching Workflow
I&PF Apache NiFi Workflow Engine
I&PF Initial Demonstration Use Cases

- **NOAA S-NPP ATMS and MIRS**
  - Ingests and inventories Suomi National Polar Partnership (S-NPP) Advanced Technology Microwave Sounder (ATMS) granules
  - Generates Microwave Integrated Retrieval System (MIRS) products from the ATMS granules
  - Makes ATMS granules and MIRS products searchable and accessible

- **NOAA Nexrad II Weather Radar**
  - Ingests and inventories NOAA Nexrad II Weather Radar data sets that were published on AWS S3 as part of the NOAA Big Data Project
  - Makes NOAA Nexrad II Weather Radar data sets searchable and accessible

- **MIRS / Nexrad II Blended Product**
  - Leverages the available MIRS products and NOAA Nexrad II Weather Radar data sets to produce a new blended product that combines the MIRS snow/water data with the Nexrad II radar data over mountainous regions
NOAA S-NPP ATMS and MIRS

Use Case: I&PF leverages automated workflows to ingest S-NPP ATMS granules, and generate MIRS products from them.
NOAA Nexrad II Weather Radar

Use Case: I&PF leverages automated workflows to ingest Nexrad II Weather Radar data
New MIRS / Nexrad II Blended Product Algorithm

Use Case: Scientist/Engineer leverages the ingested/processed MIRS products and Nexrad II Weather Radar data from the I&PF to develop a new MIRS / Nexrad II Blended Product algorithm.
MIRS / Nexrad II Blended Product

MIRS (SWE: 1.25-5.03 mm) and NexRad (REF: 20.0-51.6243 dBZ): Western US: Dec 2, 2015

Nexrad II Weather Radar

MIRS Snow/Water
I&PF User Portal

Provides a web-based user interface for data consumers / end users to discover, access, and visualize the data and metadata that has been ingested and processed by the I&PF
I&P User Portal: S-NPP ATMS Discovery and Access

![Search Results](image-url)

1. **Product**: SATMS_app_d20151202_11554330_e1955406_b21231_c201512141600
   - **ID**: SATMS_app_d20151202_11554330_e1955406_b21231_c201512141600
   - **Product Short Name**: ATMS-SDR
   - **Start Time**: 2015-12-02T19:34:23.018Z
   - **End Time**: 2015-12-02T19:34:46.692Z
   - **Download Data Set**

2. **Product**: SATMS_app_d20151202_11554010_e19554326_b21231_c201512141600
   - **ID**: SATMS_app_d20151202_11554010_e19554326_b21231_c201512141600
   - **Product Short Name**: ATMS-SDR
   - **Start Time**: 2015-12-02T19:34:01.018Z
   - **End Time**: 2015-12-02T19:34:23.692Z
   - **Download Data Set**

3. **Product**: SATMS_app_d20151202_11553220_e1955406_b21231_c201512141600
   - **ID**: SATMS_app_d20151202_11553220_e1955406_b21231_c201512141600
   - **Product Short Name**: ATMS-SDR
   - **Start Time**: 2015-12-02T19:34:00.692Z
   - **End Time**: 2015-12-02T19:34:23.692Z
   - **Download Data Set**

4. **Product**: SATMS_app_d20151202_11552570_e19553286_b21231_c201512141600
   - **ID**: SATMS_app_d20151202_11552570_e19553286_b21231_c201512141600
   - **Product Short Name**: ATMS-SDR
   - **Start Time**: 2015-12-02T19:35:29.018Z
   - **End Time**: 2015-12-02T19:34:46.692Z
   - **Download Data Set**

5. **Product**: SATMS_app_d20151202_11552250_e19552566_b21231_c201512141600
   - **ID**: SATMS_app_d20151202_11552250_e19552566_b21231_c201512141600
   - **Product Short Name**: ATMS-SDR
   - **Start Time**: 2015-12-02T19:35:25.018Z
   - **End Time**: 2015-12-02T19:34:46.692Z
   - **Download Data Set**

6. **Product**: SATMS_app_d20151202_11951530_e19552446_b21231_c201512141600
   - **ID**: SATMS_app_d20151202_11951530_e19552446_b21231_c201512141600
   - **Product Short Name**: ATMS-SDR
   - **Start Time**: 2015-12-02T19:35:18.018Z
   - **End Time**: 2015-12-02T19:35:46.692Z
   - **Download Data Set**

**Include Map Area**
- NW: 53.44437822485651, -126.62792966875
- NE: 53.44437822485651, -71.09557031249999
- SW: 29.879755334062977, -71.09357601249999
- SE: 29.879755334062977, -71.09557031249999

[Mapbox](image-url): OpenStreetMap improve this map
OmniEarth Water Resource Management (WRM)

OmniEarth utilizes large satellite imagery sets combined with advanced machine learning algorithms to classify land cover for purposes of determining outdoor water budgets at the parcel level. These budgets aid water agencies in drought-ridden communities in the US to best target water over-users.

This Use Case includes:
- Ingesting satellite imagery required by OmniEarth’s WRM processing chain
- Creating workflows to automate their WRM processing chain that produces processed imagery and analytics products, which are utilized by their user-facing WRM Application / User Interface (UI)
- Leveraging RESTful web services to automate the provisioning of the processed imagery and analytics products to their user-facing WRM Application/UI

WRM Information: [http://water.omniearth.net](http://water.omniearth.net)
OmniEarth WRM (Planned)

Use Case: I&PF leverages automated workflows to ingest satellite imagery and execute the OmniEarth WRM processing chain

Apache NiFi

I&PF Workflow Engine (Apache NiFi on AWS EC2)

1. Ingest Satellite Imagery (extracts metadata)
2a. Index Metadata in Elasticsearch
2b. Check for PG Triggers (data types that trigger a PG algorithm to be executed)
3. Check for Completed Job Specifications (all required input data received for a PG trigger)
4. Execute WRM Processing Algorithms (executes the algorithms leveraging all required input data)
5. Ingest Generated Products (extracts metadata)
6. Metadata (JSON String)

I&PF Data Storage (AWS S3)

Satellite Imagery (GeoTIFF Files)

Elasticsearch Service Plan and Scale Elasticsearch Clusters

I&PF Metadata Repository (AWS Elastisearch)

RESTful Web Services

Processed Imagery and Analytics Products

Processed Imagery and Analytics Products (GeoTIFF and Other Files)

I&PF In-Memory Cache (AWS ElastiCache [Redis])

Elasticache In-Memory Cache

Incomplete Job Specification (JSON String)

Job (JSON String)

Job (JSON String)

Incomplete Job Specification (JSON String)

Production Rule Queries and Results (JSON String)

Inventory Queries and Results (JSON String)

Ingested Satellite Imagery (GeoTIFF Files)

I&PF Metadata Repository

Job (JSON String)

Processed Imagery and Analytics Products (GeoTIFF and Other Files)

Processed Imagery and Analytics Products

Ingested Generated Products (extracts metadata)

Copyright © 2016 Solers, Inc.
I&PF Future Considerations

Resource Management, Job Scheduling, and “Big Data” Analytics
- Provide load distribution and auto-scaling for concurrent data processing / algorithm execution tasks
- Provide a “Big Data” analytics platform that can leverage the ingested/processed data and metadata from the I&PF for large-scale analytics tasks
- Currently evaluating Hadoop YARN and Apache Spark via AWS Elastic MapReduce (EMR)

Elastic’s Found as a Replacement for AWS Elasticsearch
- AWS Elasticsearch is Amazon’s hosted Elasticsearch service (currently only supports Elasticsearch v1.5.2, and has no involvement from Elastic)
- Found is Elastic’s own hosted Elasticsearch service on AWS (latest version and features)
Ingest and processing framework for commercial small satellite startup companies

- Enable them to quickly get their satellite data ingested, processed, and available to users via a scalable cloud-based workflow or “data pipeline” framework, without requiring on-premise infrastructure

Development, integration, and test environment for Government (and commercial) satellite ground systems

- Perform calibration and validation of new product algorithms that leverage multiple satellite (and other) data sets within a scalable cloud-based framework, prior to integrating them into operations, without requiring on-premise infrastructure
I&PF Technologies

Current Technologies:

- Webdis (RESTful HTTP Interface for Redis): [http://webd.is](http://webd.is)
- Google Polymer (Web Framework): [http://polymer-project.org](http://polymer-project.org)

Future Technology Considerations:

- Elastic Found (Elastic’s Hosted Elasticsearch on AWS): [http://www.elastic.co/found](http://www.elastic.co/found)
BACKUP
I&P User Portal: MIRS Discovery and Access

Search Results

Product ID: AV0kLm_L02sDA0qfr1U
Product Short Name: MIRS IMG
Start Time: 2015-12-07T08:27:42Z
End Time: 2015-12-07T08:28:15Z

Download Data Set

Include Map Area

NW: 46.004592355774482, -104.83154296875
SW: 40.254370602072649, -104.83154296875
NE: 46.004592355774482, -90.494453125
SE: 40.254370602072649, -90.494453125

Mapbox © OpenStreetMap Improve This map
I&PF User Portal: Nexrad II Discovery and Access