



Next Generation Cloud Based Ingest & Processing Framework (I&PF) for Environmental Data

GSAW 2016

Rich Baker

Chief Architect

Solers, Inc.

Email: richard.baker@solers.com

Phone: 240-790-3338

www.solers.com

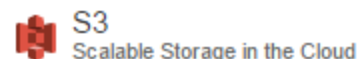
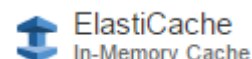
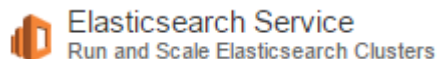
Cloud Based Ingest & Processing Framework (I&PF)

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

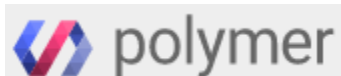
➤ Primary Objectives:

- **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 operators/admins)
- 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)

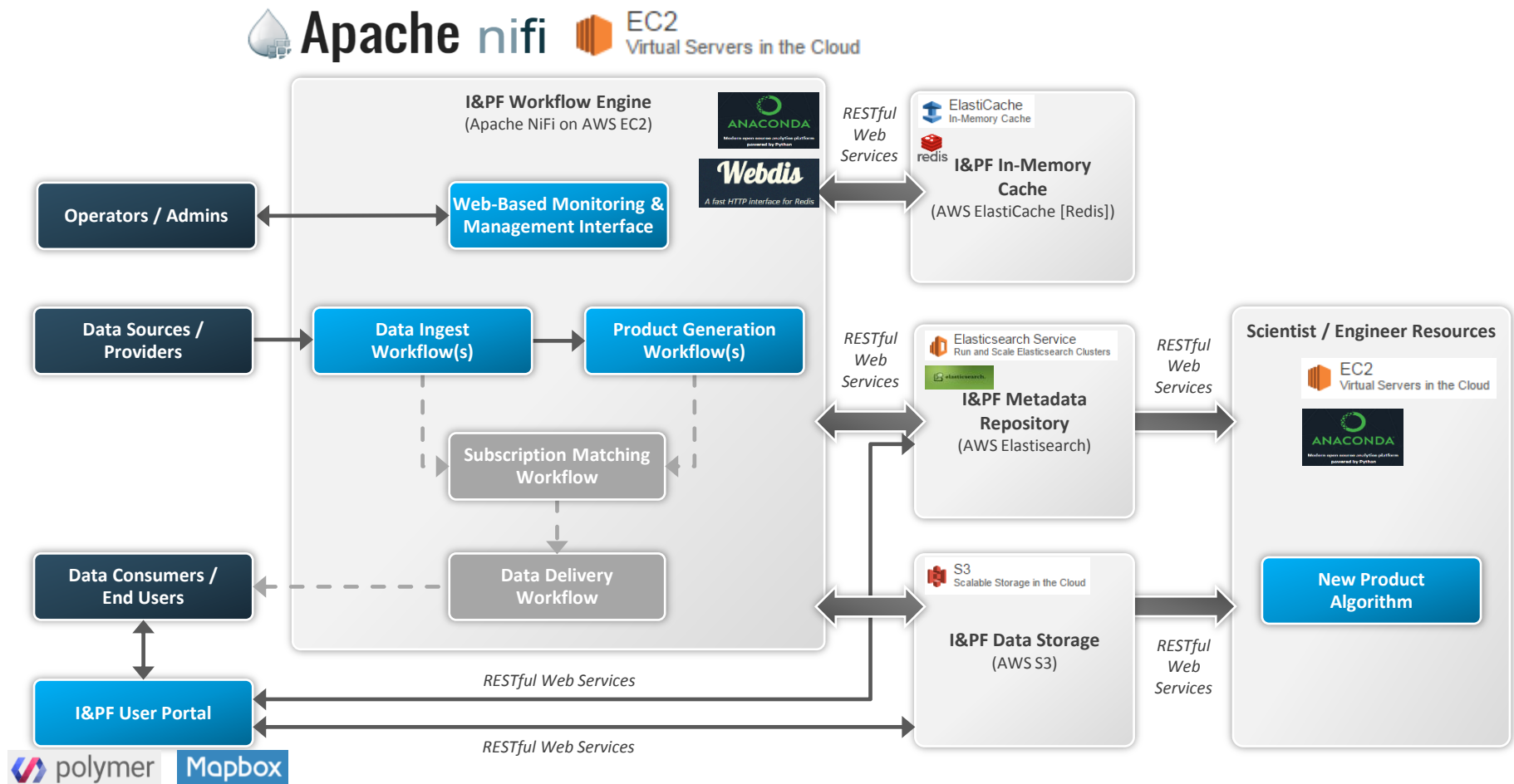
➤ Leverages readily available commercial Amazon Cloud services:



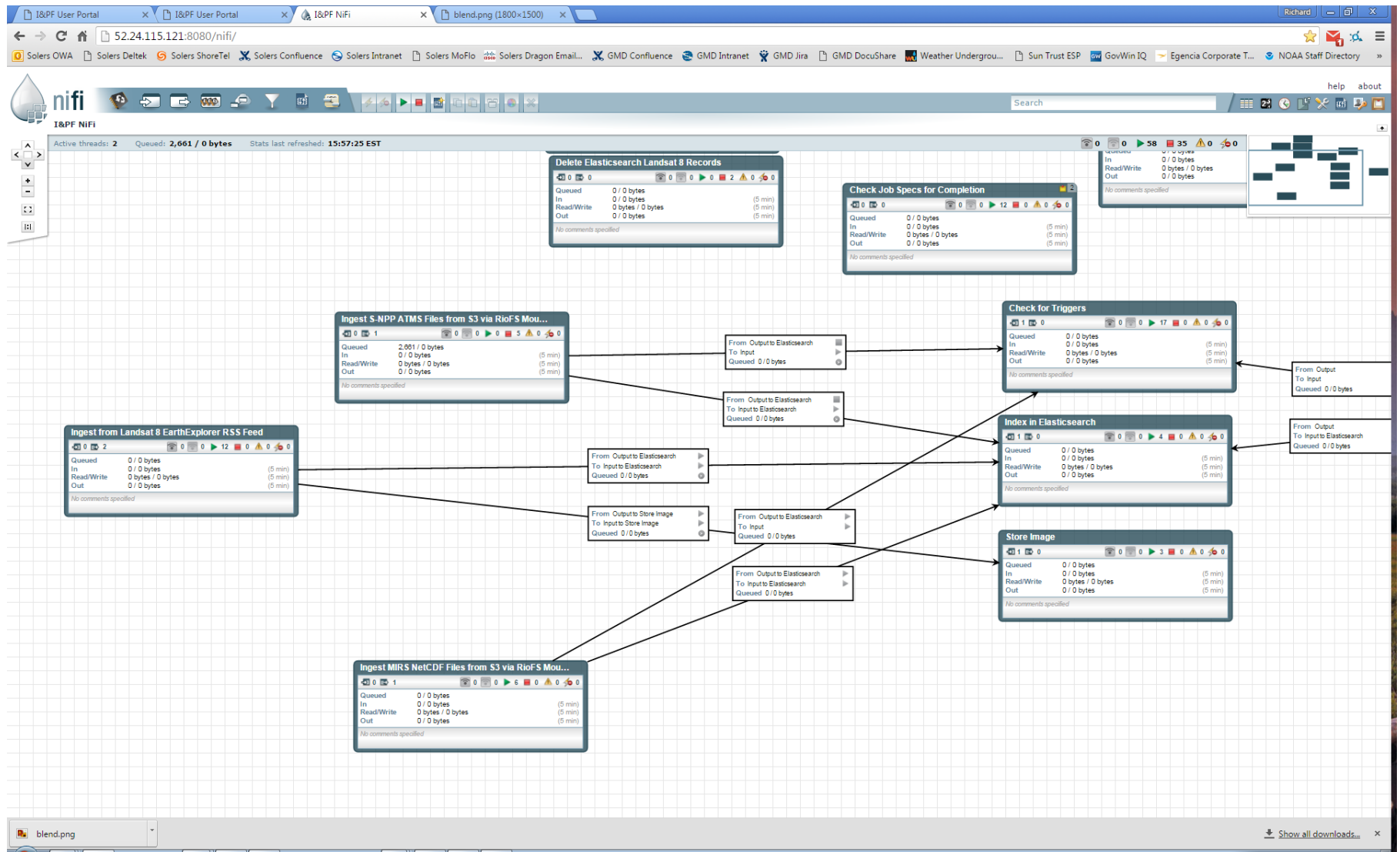
➤ Leverages readily available open source technologies:



Cloud Based I&PF Architecture



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

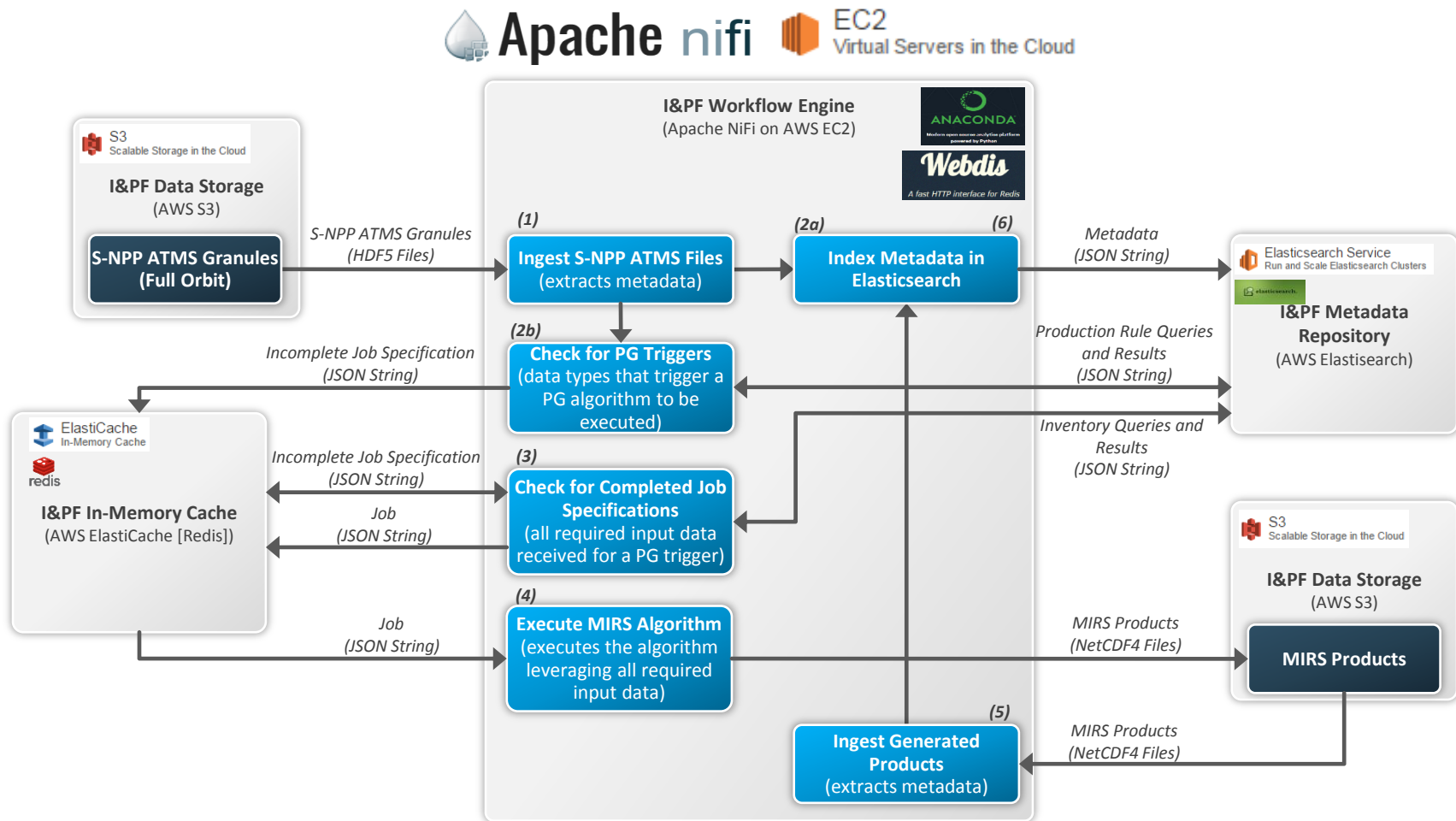
- Ingests and inventories NOAA Nexrad II Weather Radar data sets that were published on AWS S3 as part of the NOAA Big Data Project
- Makes NOAA Nexrad II Weather Radar data sets searchable and accessible

➤ 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

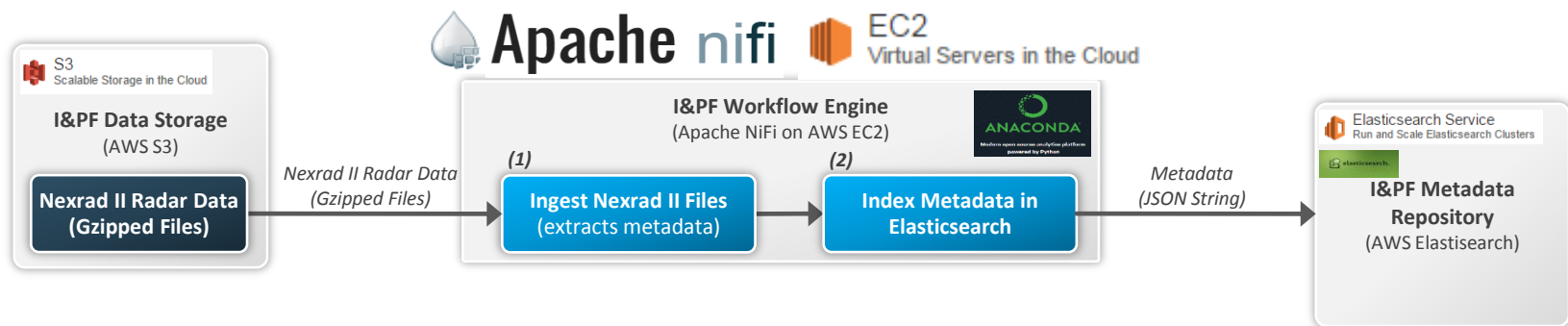
NOAA S-NPP ATMS and MIRS

Use Case: I&PF leverages automated workflows to ingest S-NPP ATMS granules, and generate MIRS products from them



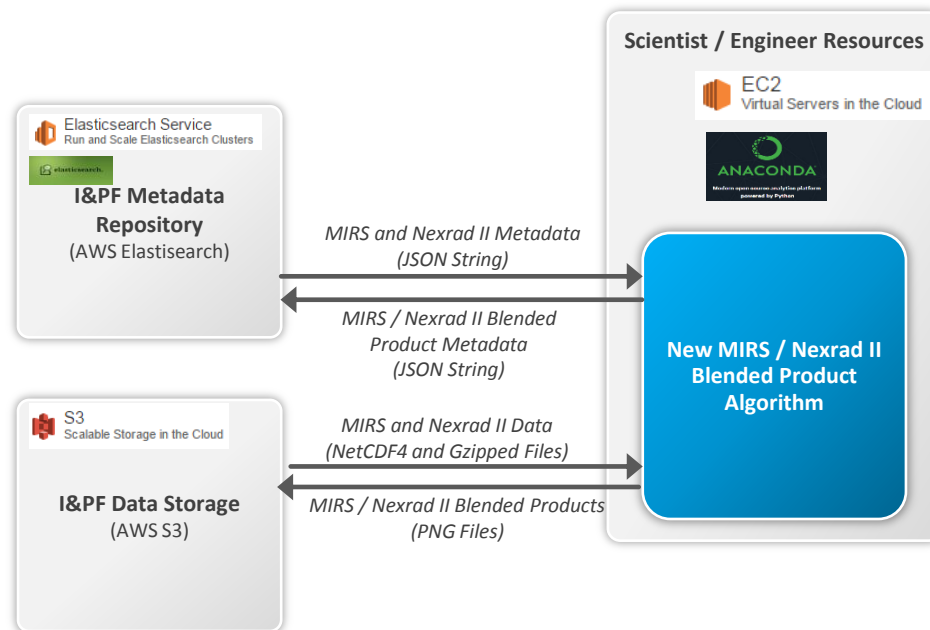
NOAA Nexrad II Weather Radar

Use Case: I&PF leverages automated workflows to ingest Nexrad II Weather Radar data

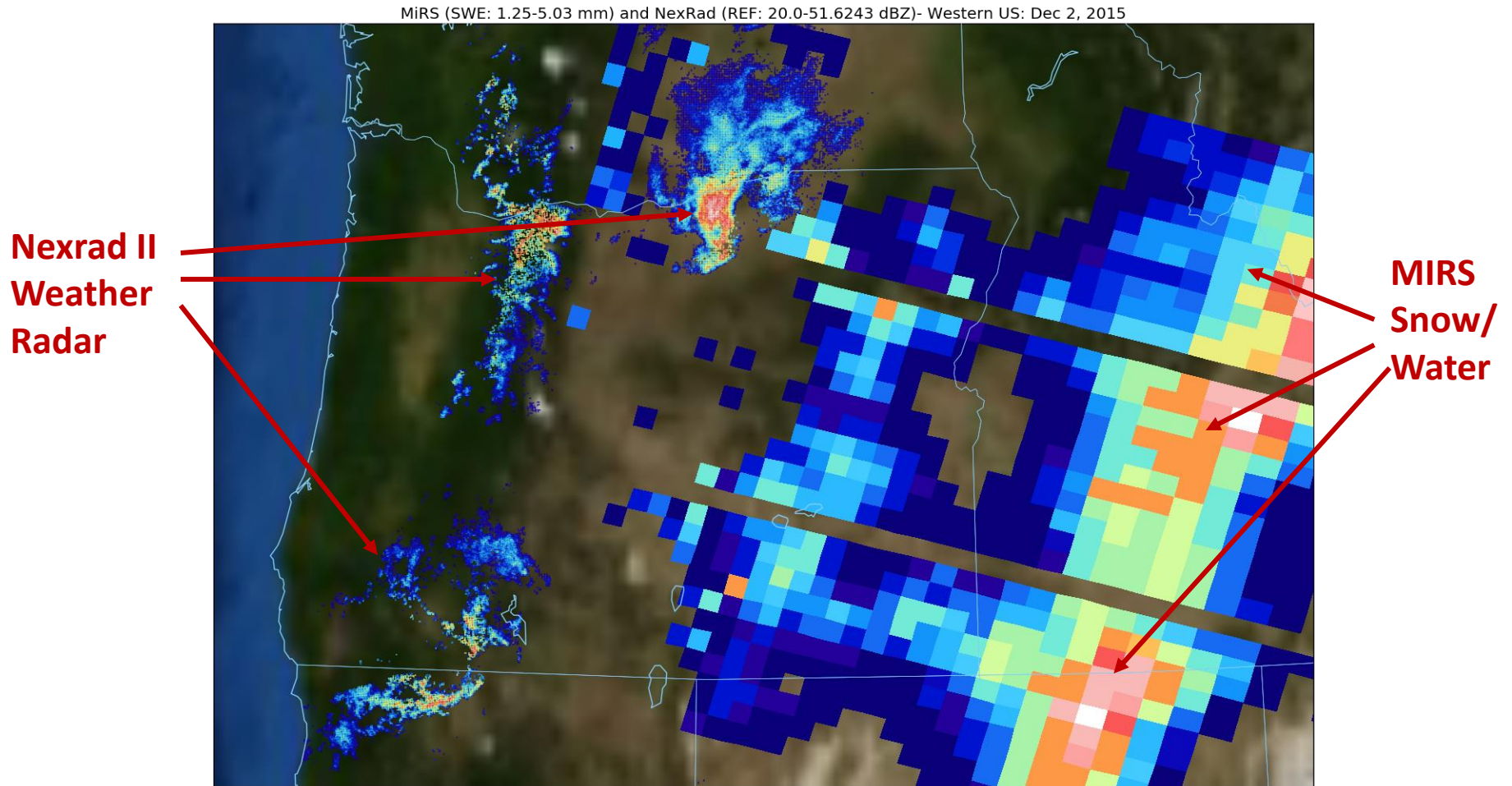


New MIRS / Nexrad II Blended Product Algorithm

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

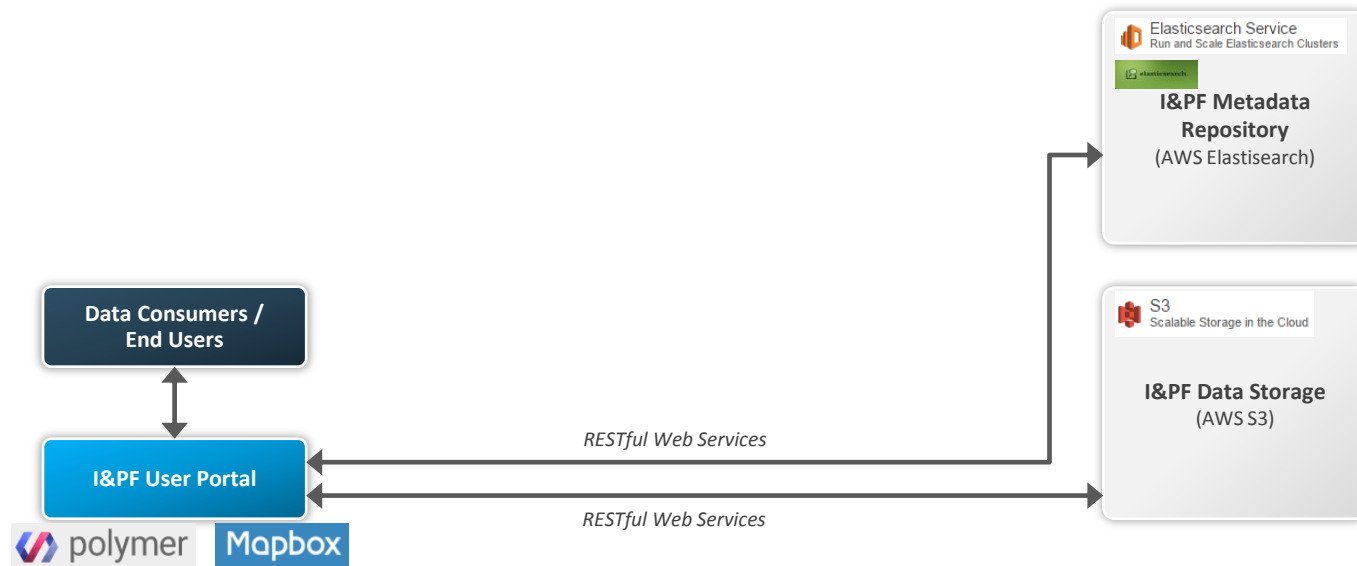


MIRS / Nexrad II Blended Product



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&PF User Portal: S-NPP ATMS Discovery and Access

SOLERS I&PF User Portal Welcome: Rich Baker Preferences Logout Search Subscriptions

Source Type: Satellite Source: NOAA S-NPP Instrument: ATMS Product Short Name: ATMS-SDR Start Date: mm/dd/yyyy End Date: mm/dd/yyyy Max Results: 10

Search Results

☒ Include Map Area

NW: 53.64463782485651, -128.6279296875 NE: 53.64463782485651, -71.05957031249999
SW: 29.878755346037977, -128.6279296875 SE: 29.878755346037977, -71.05957031249999

1 Product ID: SATMS_npp_d20151202_t1954330_e1955046_b21231_c20151214162
Product Short Name: ATMS-SDR
Start Time: 2015-12-02T19:54:33.018Z
End Time: 2015-12-02T19:55:04.639Z
[Download Data Set](#)

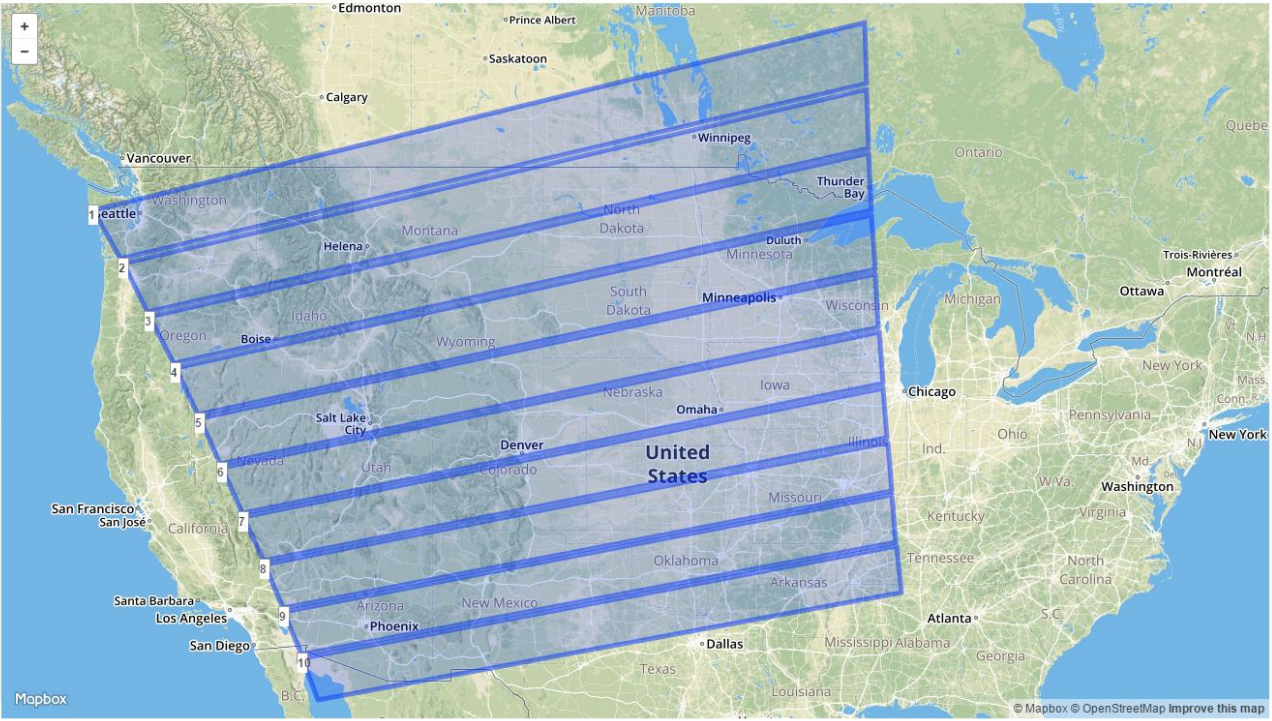
2 Product ID: SATMS_npp_d20151202_t1954010_e1954326_b21231_c20151214162
Product Short Name: ATMS-SDR
Start Time: 2015-12-02T19:54:01.018Z
End Time: 2015-12-02T19:54:32.639Z
[Download Data Set](#)

3 Product ID: SATMS_npp_d20151202_t1953290_e1954006_b21231_c20151214162
Product Short Name: ATMS-SDR
Start Time: 2015-12-02T19:53:29.018Z
End Time: 2015-12-02T19:54:00.639Z
[Download Data Set](#)

4 Product ID: SATMS_npp_d20151202_t1952570_e1953286_b21231_c20151214162
Product Short Name: ATMS-SDR
Start Time: 2015-12-02T19:52:57.018Z
End Time: 2015-12-02T19:53:28.639Z
[Download Data Set](#)

5 Product ID: SATMS_npp_d20151202_t1952250_e1952566_b21231_c20151214162
Product Short Name: ATMS-SDR
Start Time: 2015-12-02T19:52:25.018Z
End Time: 2015-12-02T19:52:56.639Z
[Download Data Set](#)

6 Product ID: SATMS_npp_d20151202_t1951530_e1952246_b21231_c20151214162
Product Short Name: ATMS-SDR
Start Time: 2015-12-02T19:51:53.018Z
End Time: 2015-12-02T19:52:24.639Z
[Download Data Set](#)



52.24.115.121/s3-ipf-data/products/ATMS-SDR/SATMS_npp_d20151202_t1954330_e1955046_b21231_c20151214162_256277_noaa_ops.h5

I&PF Ongoing Use Case

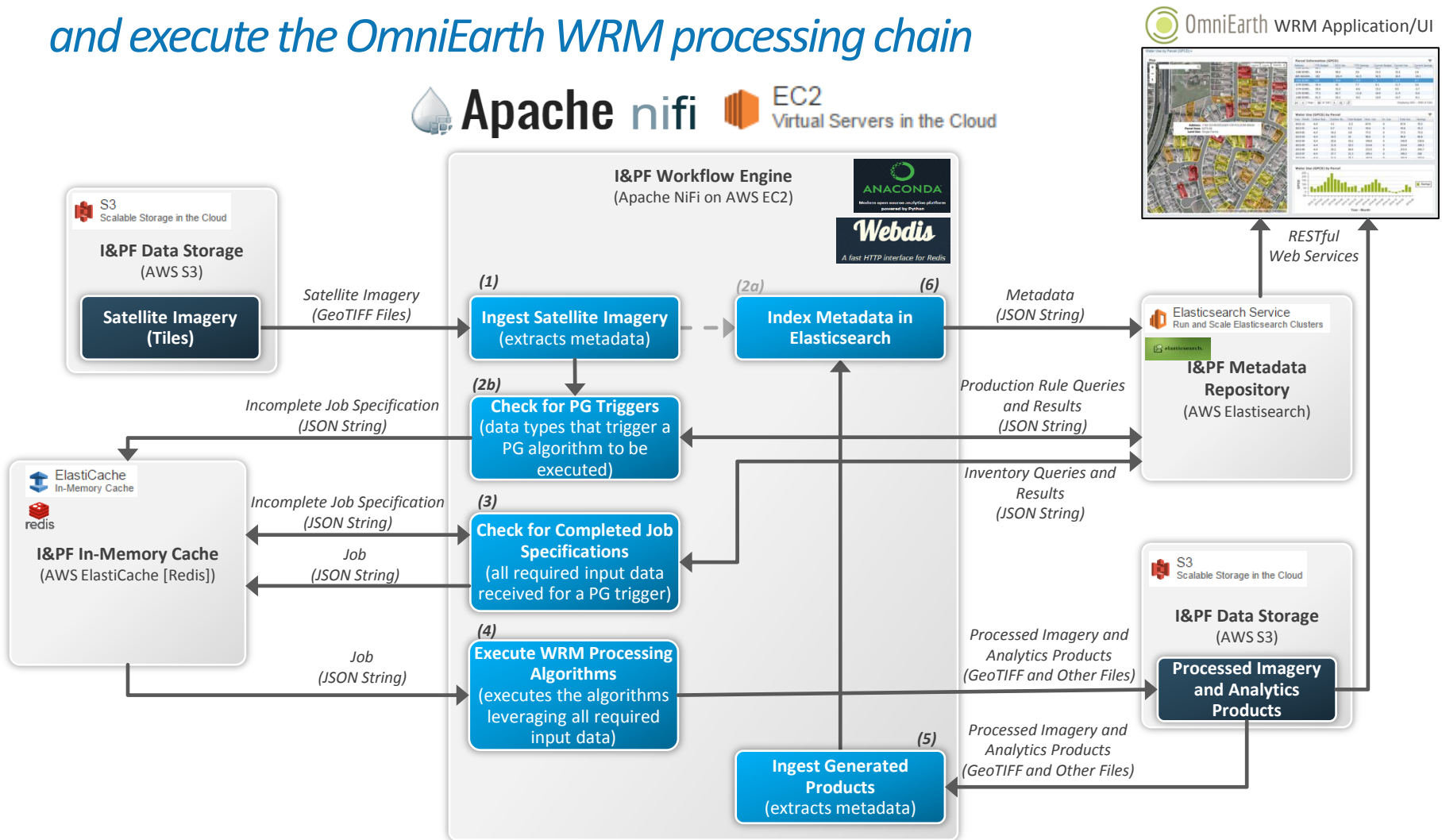
➤ OmniEarth Water Resource Management (WRM)

- 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

 OmniEarth WRM Information: <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



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)



EMR
Managed Hadoop Framework



hadoopYARN



➤ 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)



elastic



I&PF Potential Utility

➤ **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:

- Amazon Web Services (Public Cloud Services): <http://aws.amazon.com>
 - AWS EC2 (Virtualized Computing): <http://aws.amazon.com/ec2>
 - AWS Elasticsearch (Metadata Repository): <http://aws.amazon.com/elasticsearch-service>
<http://www.elastic.co/products/elasticsearch>
 - AWS ElastiCache (In-Memory Redis Cache): <http://aws.amazon.com/elasticache>
<http://redis.io>
 - AWS S3 (Data Storage): <http://aws.amazon.com/s3>
- Apache NiFi (Workflow Engine): <http://nifi.apache.org>
- Continuum Anaconda (Python Framework): <http://www.continuum.io>
- Webdis (RESTful HTTP Interface for Redis): <http://webd.is>
- Google Polymer (Web Framework): <http://polymer-project.org>
- Mapbox (Web Mapping Toolkit): <http://www.mapbox.com>

➤ Future Technology Considerations:

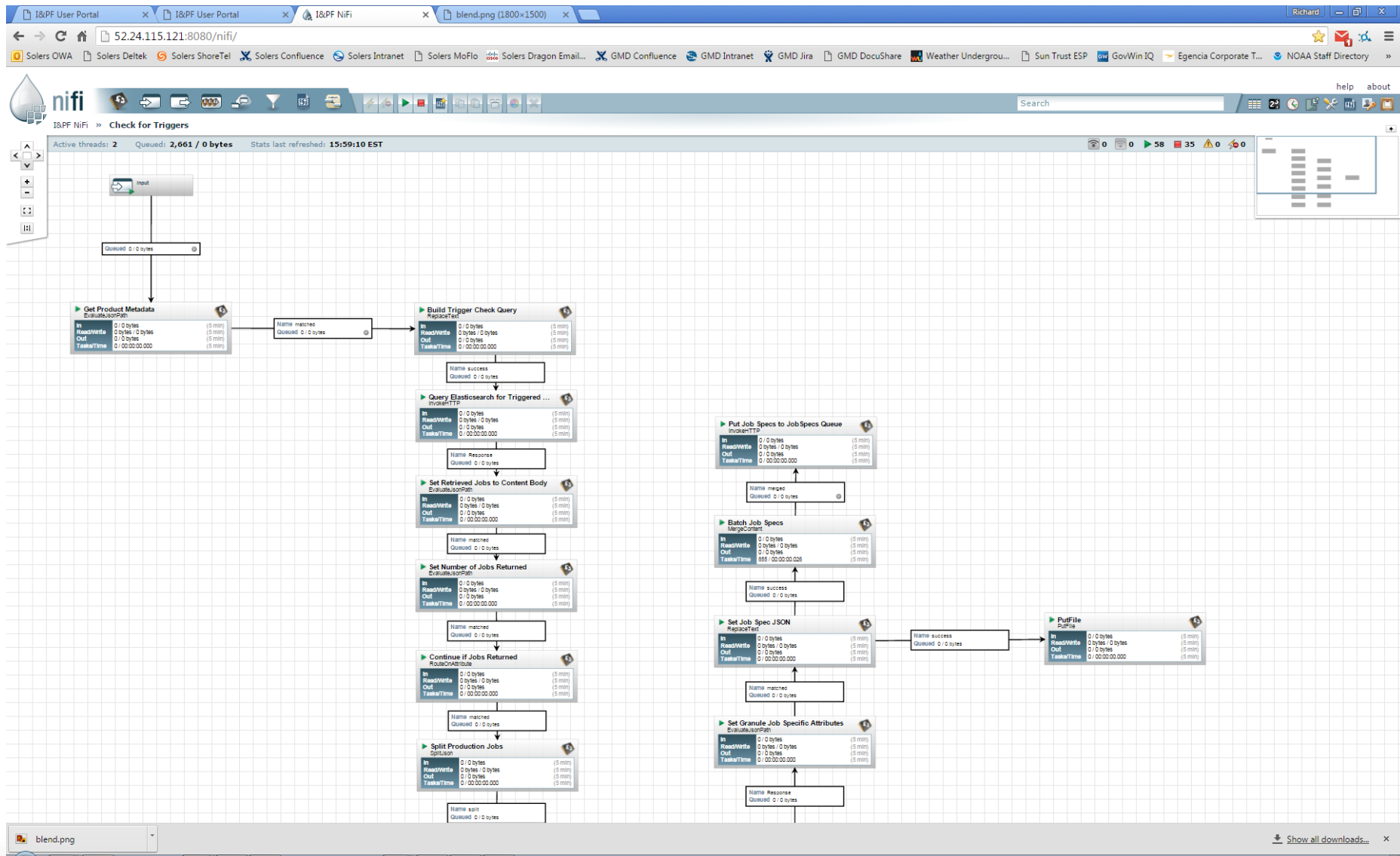
- AWS EMR (Managed Hadoop/Spark): <http://aws.amazon.com/elasticmapreduce>
<http://hadoop.apache.org>
<http://spark.apache.org>
- Elastic Found (Elastic's Hosted Elasticsearch on AWS): <http://www.elastic.co/found>

Questions



BACKUP

I&PF Apache NiFi Workflow Definition



I&PF User Portal: MIRS Discovery and Access

I&PF User Portal | Welcome: Rich Baker | [Preferences](#) | [Logout](#) | [Search](#) | [Subscriptions](#)

52.24.115.121/ipf-portal/

Solers OWA | Solers Deltek | Solers ShoreTel | Solers Confluence | Solers Intranet | Solers MoFlo | Solers Dragon Email... | GMD Confluence | GMD Intranet | GMD Jira | GMD DocuShare | Weather Underground... | Sun Trust ESP | GovWin IQ | Egencia Corporate T... | NOAA Staff Directory

Source Type: Satellite | Source: NOAA S-NPP | Instrument: Any | Product Short Name: MIRS IMG | Start Date: mm / dd / yyyy | End Date: mm / dd / yyyy | Max Results: 10

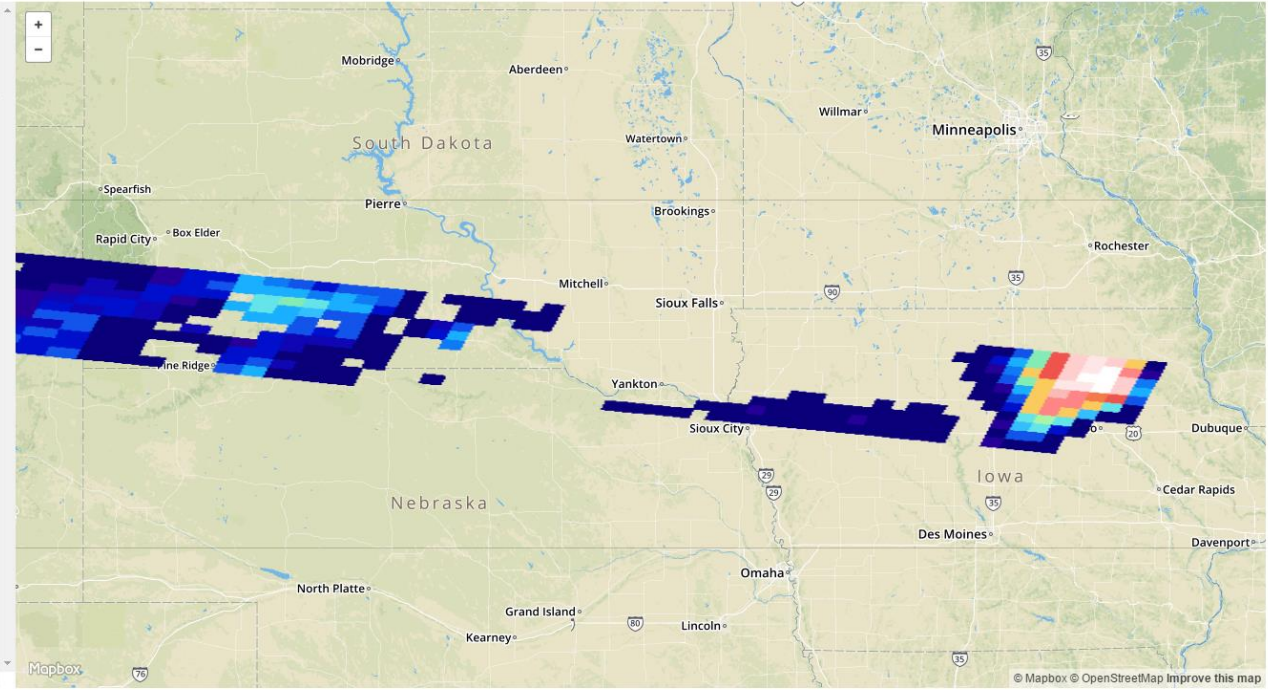
Start Time: 08:27:42.000 | End Time: 08:28:13.999

Search Results

☒ Include Map Area

NW: 46.00459325574482, -104.83154296875 | NE: 46.00459325574482, -90.439453125
SW: 40.25437660372649, -104.83154296875 | SE: 40.25437660372649, -90.439453125

1. Product ID: AVIXmLa_G2sdA0qfrs1U
Product Short Name: MIRS IMG
Start Time: 2015-12-02T08:27:42Z
End Time: 2015-12-02T08:28:13Z
[Download Data Set](#)



I&PF User Portal: Nexrad II Discovery and Access

SOLERS I&PF User Portal Welcome: Rich Baker Preferences Logout Search Subscriptions

Source Type: Radar Source: NOAA Nexrad II Instrument: Any Product Short Name: Nexrad Start Date: mm/dd/yyyy End Date: mm/dd/yyyy Start Time: 00:00:00.000 End Time: 23:59:59.999 Max Results: 10

☒ Include Map Area

NW: 48.73445537176822, -128.47412109375 NE: 48.73445537176822, -99.68994140625
SW: 36.87962060502676, -128.47412109375 SE: 36.87962060502676, -99.68994140625

Search Results

1	Product ID: KCBX20151202_235954_V06.gz Product Short Name: Nexrad Start Time: 2015-12-02T23:59:54Z End Time: 2015-12-03T00:09:24.538Z Download Data Set
2	Product ID: KRTX20151202_235730_V06.gz Product Short Name: Nexrad Start Time: 2015-12-02T23:57:30Z End Time: 2015-12-03T00:07:04.270Z Download Data Set
3	Product ID: KLRX20151202_235711_V06.gz Product Short Name: Nexrad Start Time: 2015-12-02T23:57:11Z End Time: 2015-12-03T00:06:42.591Z Download Data Set
4	Product ID: KRGX20151202_235533_V06.gz Product Short Name: Nexrad Start Time: 2015-12-02T23:55:33Z End Time: 2015-12-03T00:05:16.664Z Download Data Set
5	Product ID: KPDT20151202_235536_V06.gz Product Short Name: Nexrad Start Time: 2015-12-02T23:55:36Z End Time: 2015-12-03T00:04:47.973Z Download Data Set
6	Product ID: KMAX20151202_235717_V06.gz Product Short Name: Nexrad Start Time: 2015-12-02T23:57:17Z End Time: 2015-12-03T00:02:04.197Z Download Data Set
7	Product ID: KCBX20151202_235014_V06.gz Product Short Name: Nexrad Start Time: 2015-12-02T23:50:14Z End Time: 2015-12-02T23:59:44.187Z Download Data Set
8	Product ID: KRTX20151202_234748_V06.gz