



# 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](mailto:richard.baker@solers.com)

Phone: 240-790-3338

[www.solers.com](http://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)

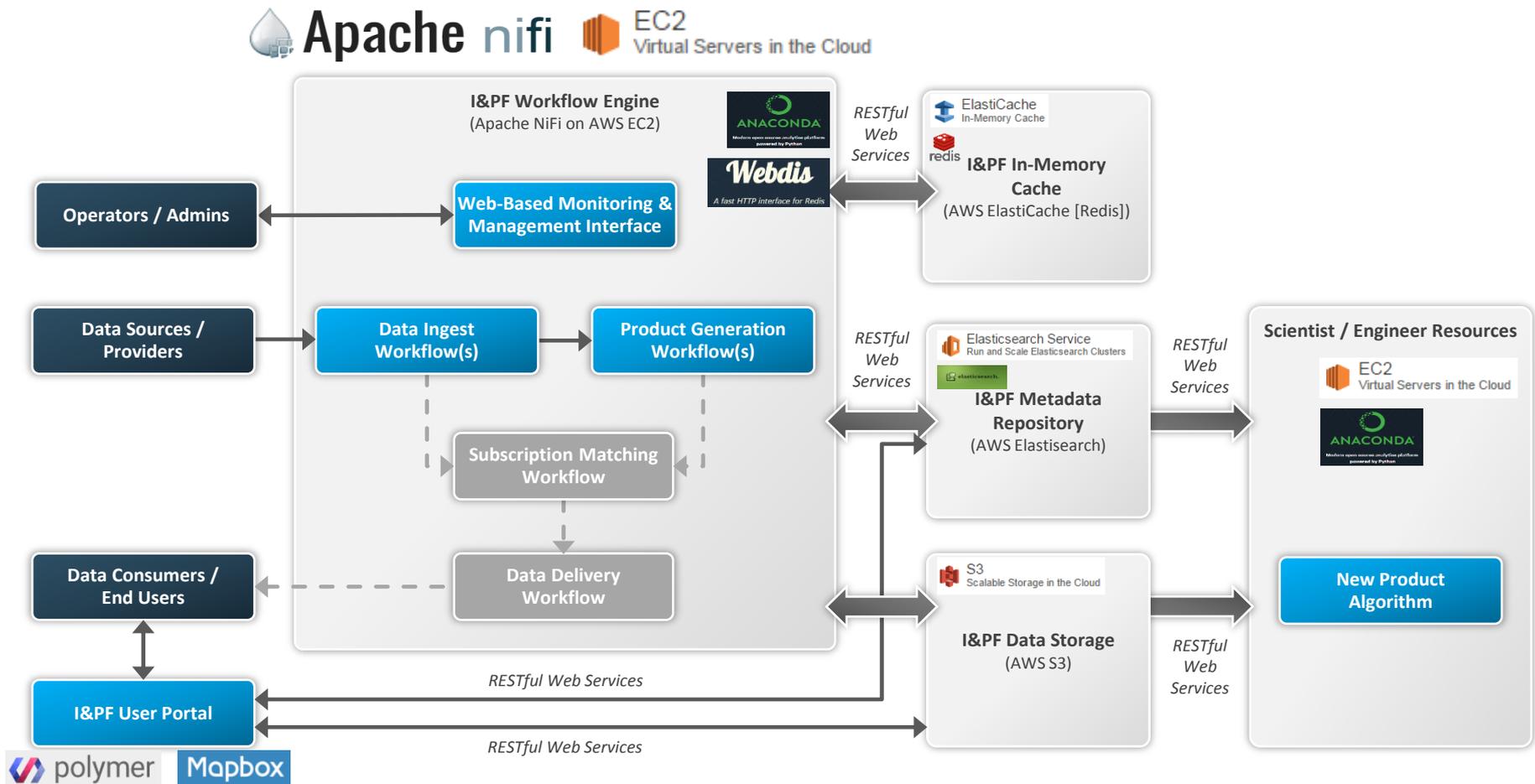
➤ Leverages readily available commercial Amazon Cloud services:



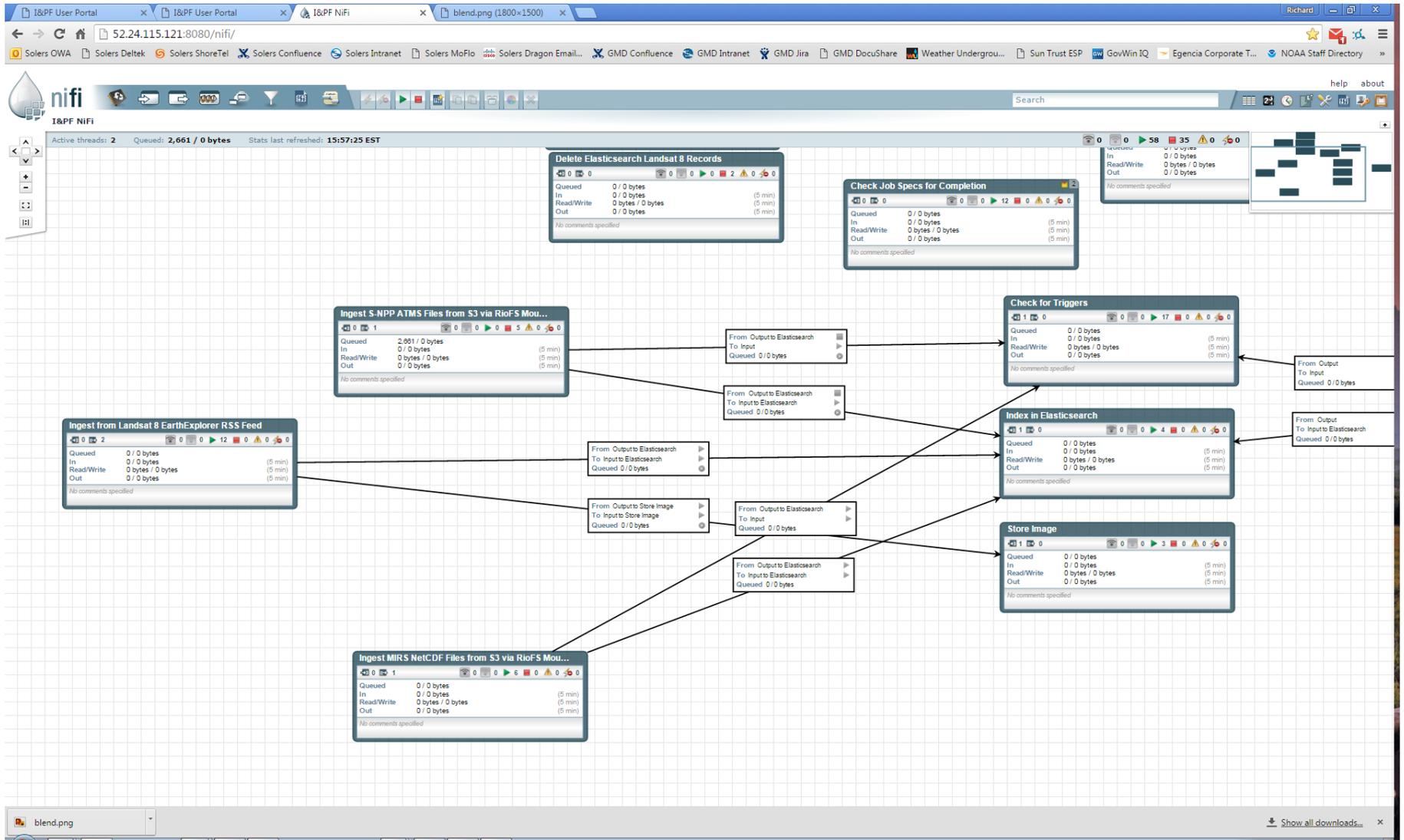
➤ Leverages readily available open source technologies:



# 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

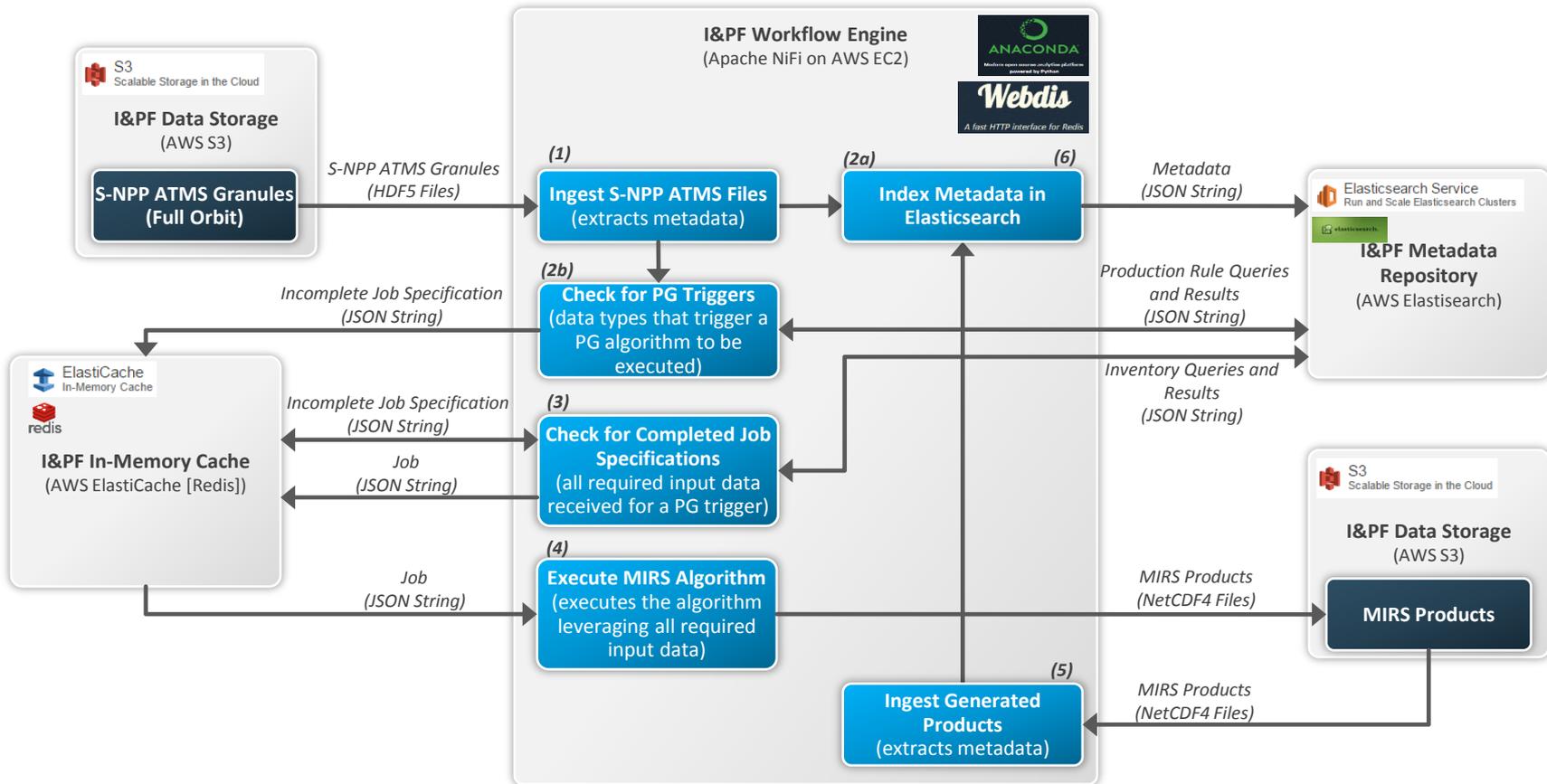
- Ingests and inventories NOAA Nexrad II Weather Radar data sets that were published on AWS S3 as part of the NOAA Big Data Project
- Makes NOAA Nexrad II Weather Radar data sets searchable and accessible

## ➤ 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

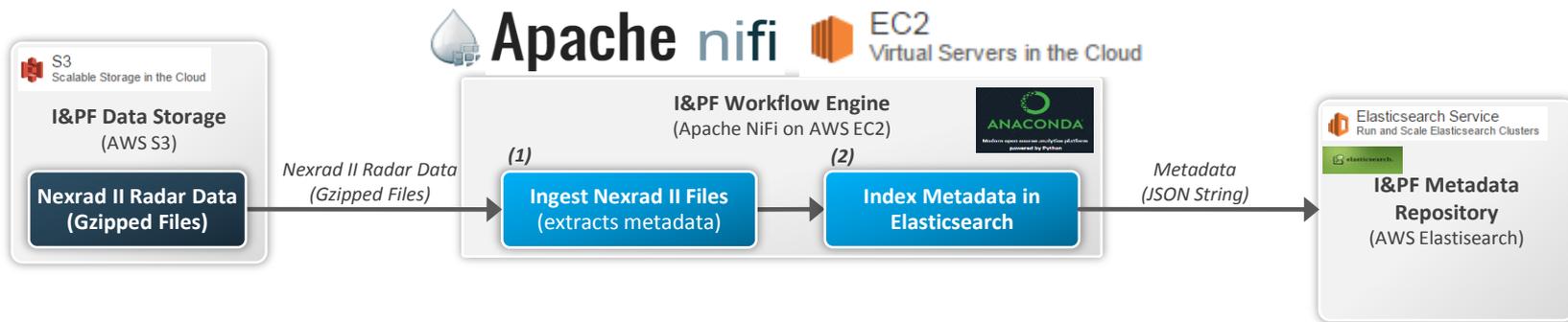
# NOAA S-NPP ATMS and MIRS

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



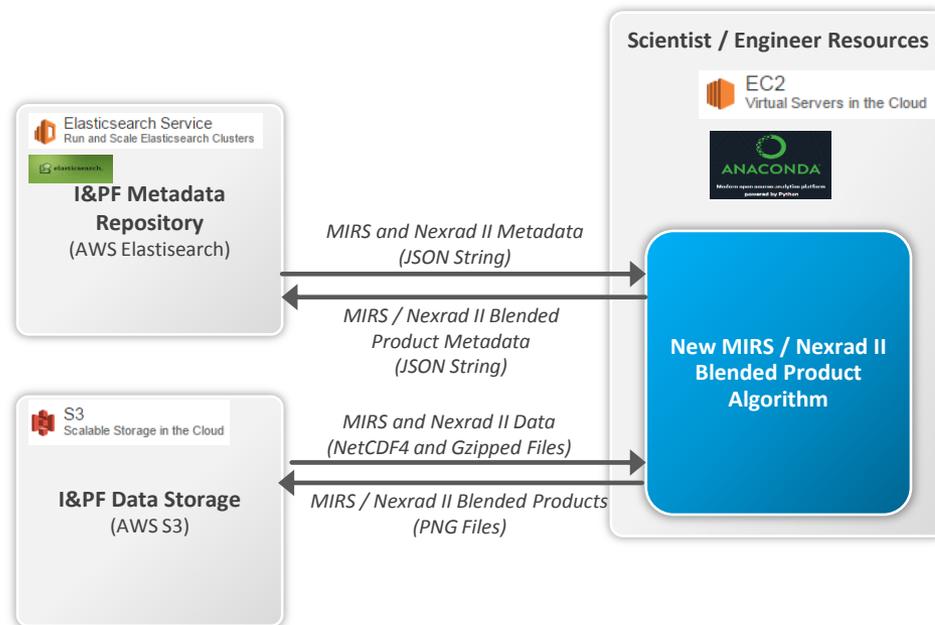
# NOAA Nexrad II Weather Radar

*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*

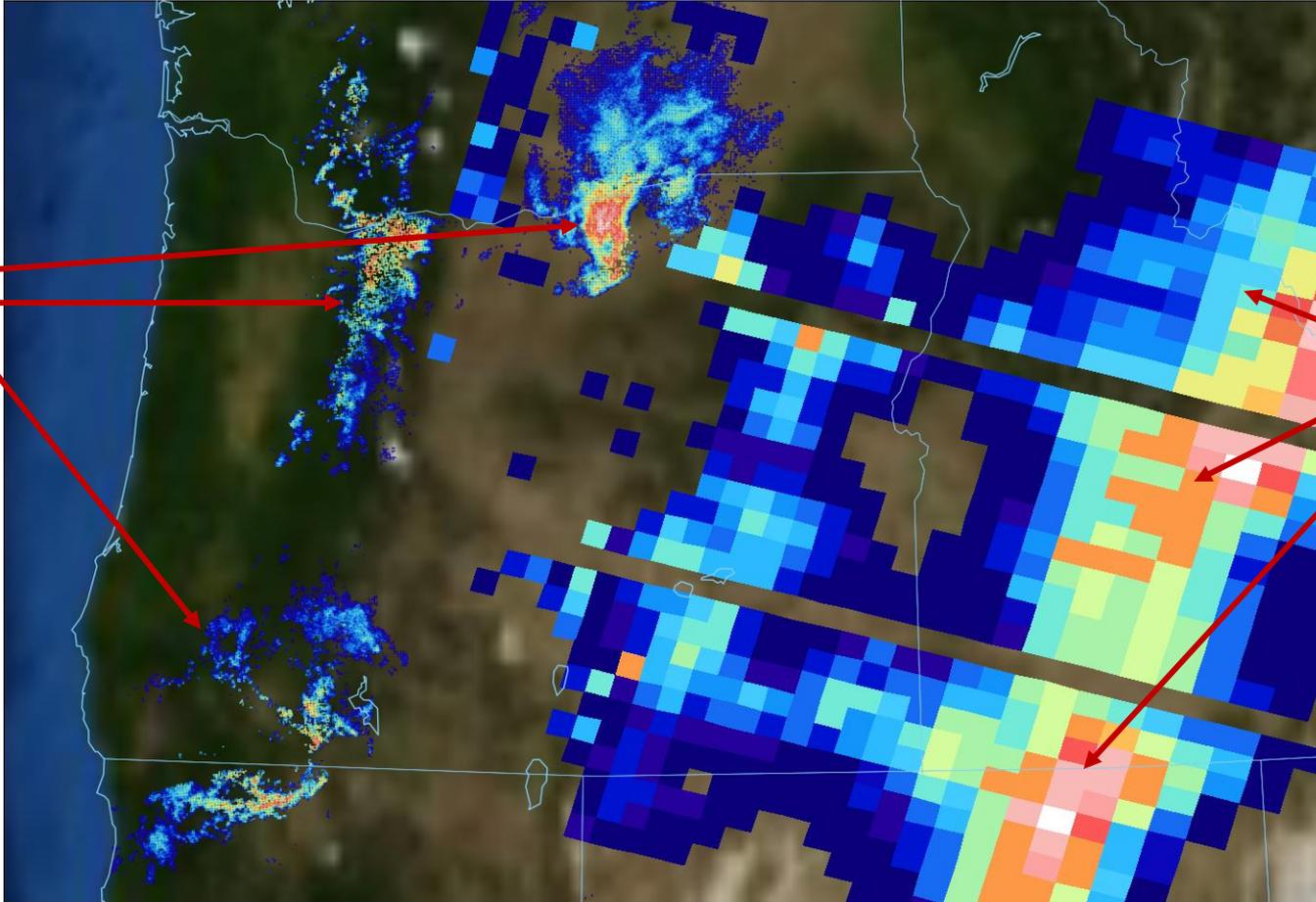


# MIRS / Nexrad II Blended Product

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

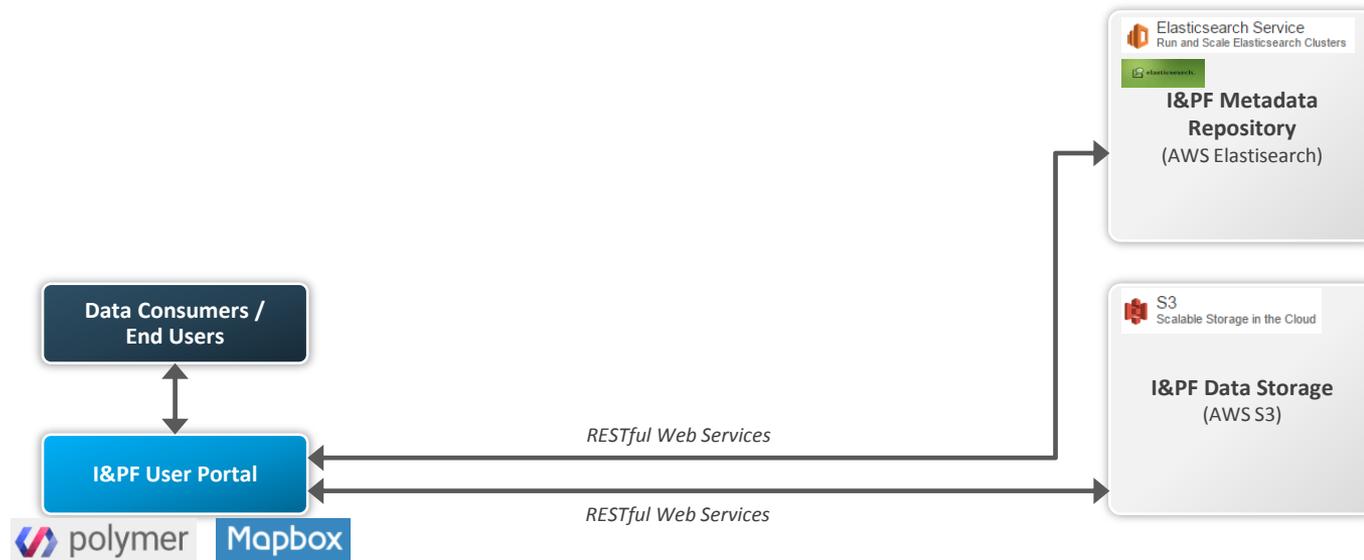
**Nexrad II  
Weather  
Radar**

**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&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

Include Map Area

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

**Search Results**

Product ID	Product Short Name	Start Time	End Time	Download Data Set
SATMS_npp_d20151202_t1954330_e1955046_b21231_c20151214162	ATMS-SDR	2015-12-02T19:54:33.018Z	2015-12-02T19:55:04.639Z	<a href="#">Download Data Set</a>
SATMS_npp_d20151202_t1954010_e1954326_b21231_c20151214162	ATMS-SDR	2015-12-02T19:54:01.018Z	2015-12-02T19:54:32.639Z	<a href="#">Download Data Set</a>
SATMS_npp_d20151202_t1953290_e1954006_b21231_c20151214162	ATMS-SDR	2015-12-02T19:53:29.018Z	2015-12-02T19:54:00.639Z	<a href="#">Download Data Set</a>
SATMS_npp_d20151202_t1952570_e1953286_b21231_c20151214162	ATMS-SDR	2015-12-02T19:52:57.018Z	2015-12-02T19:53:28.639Z	<a href="#">Download Data Set</a>
SATMS_npp_d20151202_t1952250_e1952566_b21231_c20151214162	ATMS-SDR	2015-12-02T19:52:25.018Z	2015-12-02T19:52:56.639Z	<a href="#">Download Data Set</a>
SATMS_npp_d20151202_t1951530_e1952246_b21231_c20151214162	ATMS-SDR	2015-12-02T19:51:53.018Z	2015-12-02T19:52:24.639Z	<a href="#">Download Data Set</a>

Mapbox © OpenStreetMap Improve this map

52.24.115.121/s3-ipf-data/products/ATMS-SDR/SATMS\_npp\_d20151202\_t1954330\_e1955046\_b21231\_c20151214162406256277\_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>

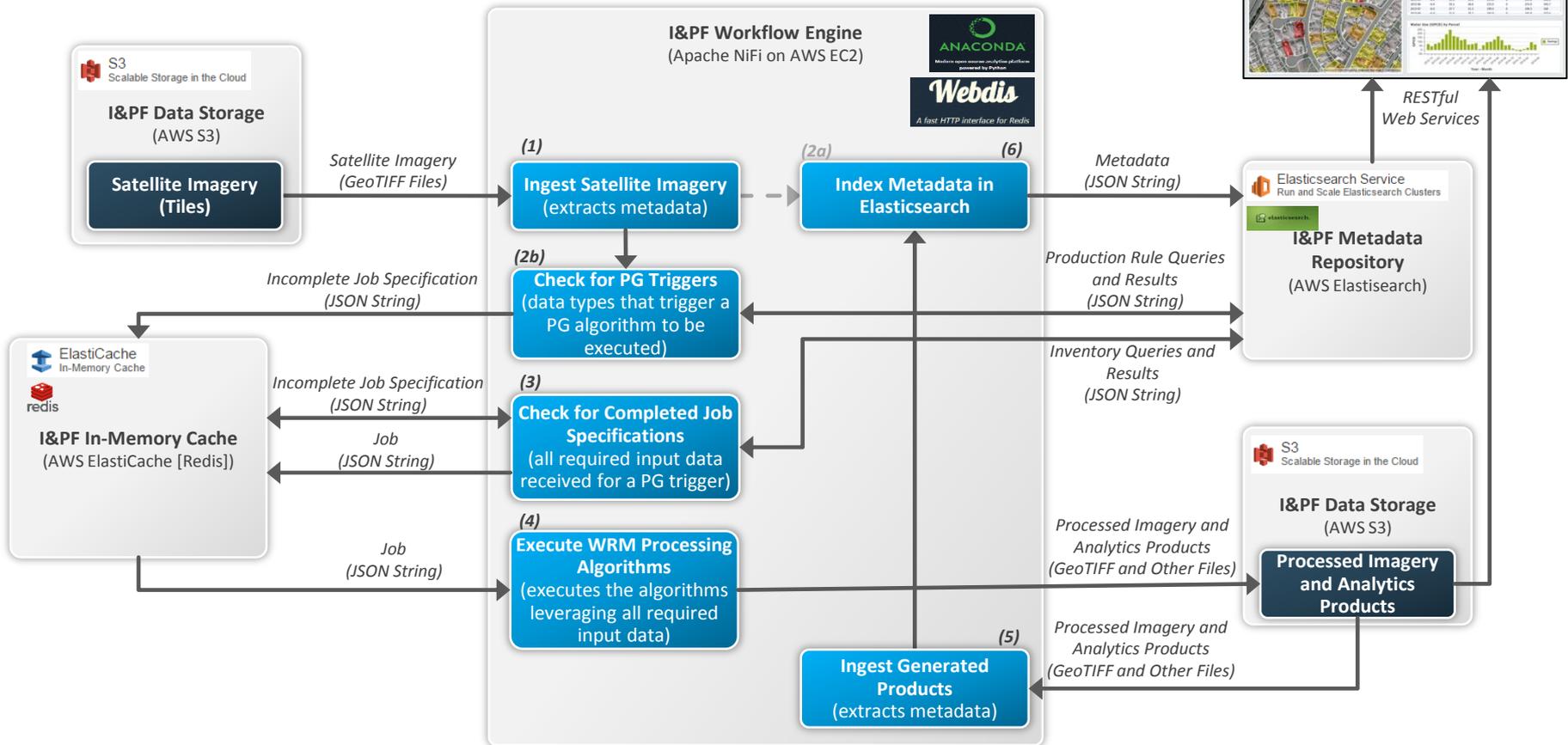
# OmniEarth WRM (Planned)

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


**Apache nifi**

**EC2**  
 Virtual Servers in the Cloud


 OmniEarth WRM Application/UI



# 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)



# 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:

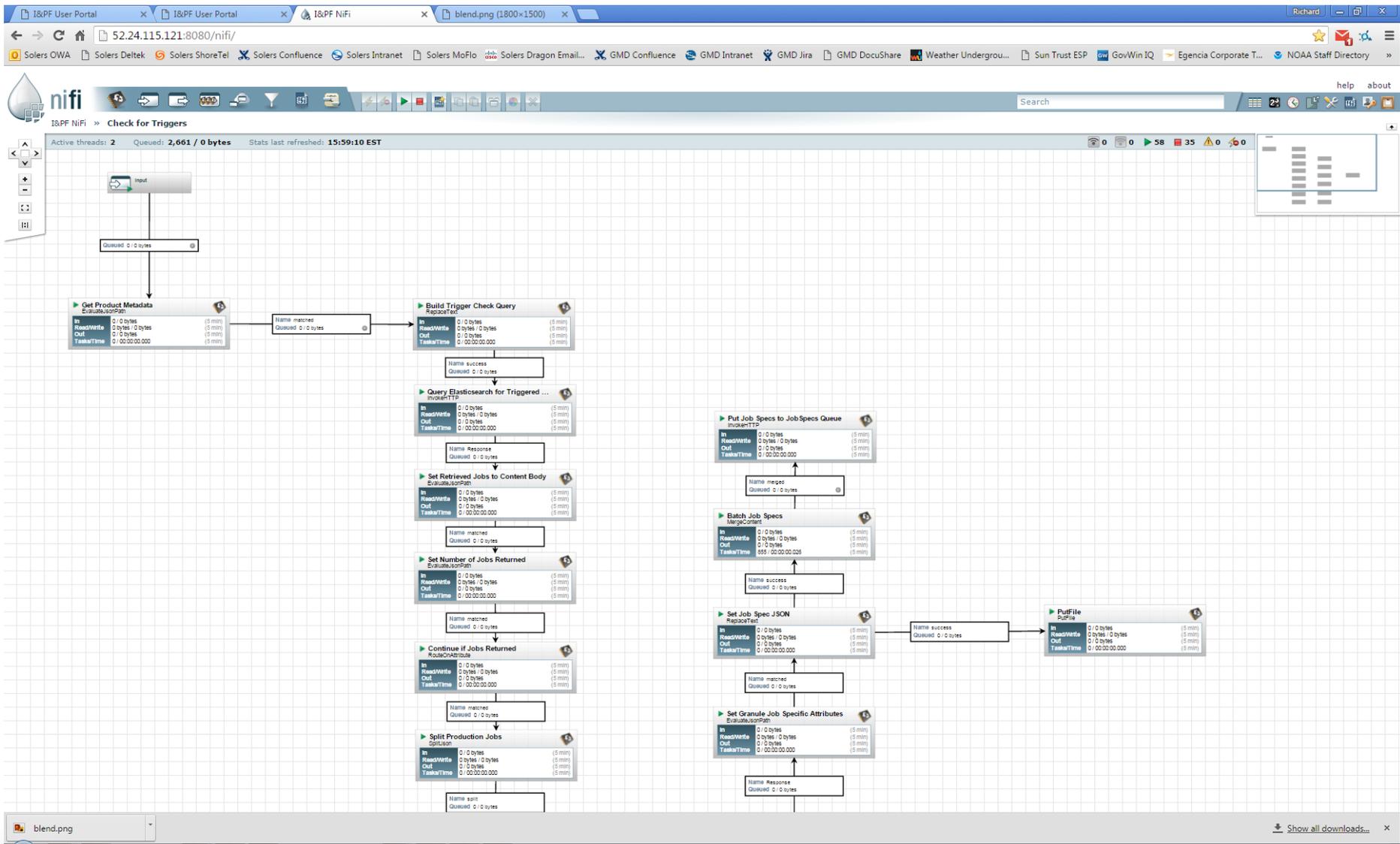
- AWS EMR (Managed Hadoop/Spark): <http://aws.amazon.com/elasticmapreduce>  
<http://hadoop.apache.org>  
<http://spark.apache.org>
- Elastic Found (Elastic's Hosted Elasticsearch on AWS): <http://www.elastic.co/found>

# 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](#)

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](#)

The map displays a satellite image of the MIRS IMG data over a geographic area covering parts of South Dakota, Nebraska, and Iowa. The data is represented by a color-coded grid overlaying the satellite imagery. The grid shows a mix of dark blue and light blue colors, indicating varying levels of data density or intensity. The map includes labels for major cities and states, such as Minneapolis, South Dakota, Nebraska, and Iowa. The map is powered by Mapbox and OpenStreetMap.

# 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 | Max Results: 10

Start Time: 00:00:00.000 | End Time: 23:59:59.999

Include Map Area

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

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

Map showing coverage areas (numbered 4-10) over the Pacific Northwest and Mountain West regions.