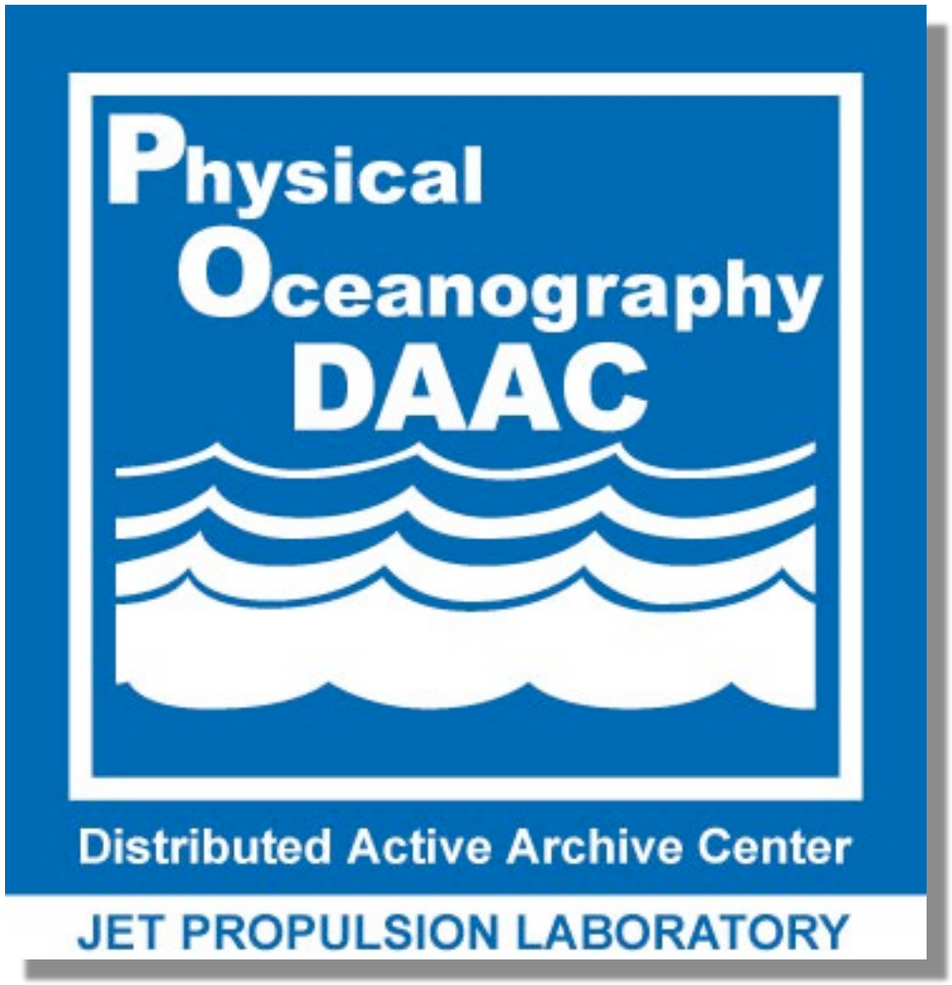

Physical
Oceanography
DAAC
Distributed Active Archive Center
JET PROPULSION LABORATORY

Data Management and Archive System 4

Federated Architecture

Thomas Huang

Core System Team Members:
Christian Alarcon, Qui Chau, Nga Chung, Mike Gangl

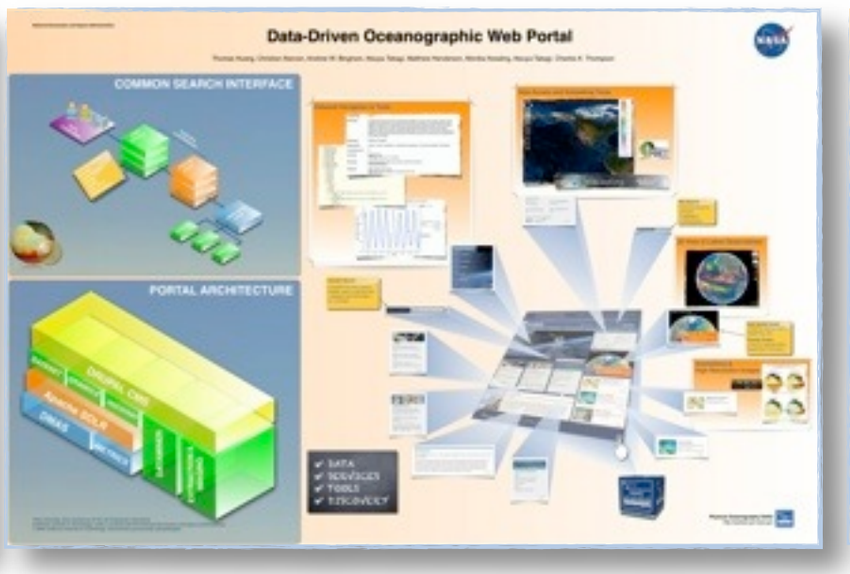
March 12, 2012

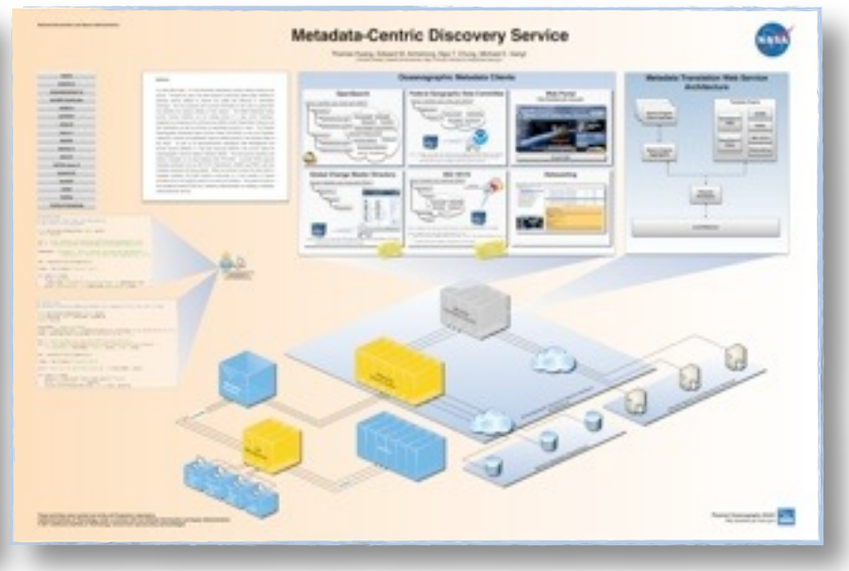
Breakthrough Data System Built With Web Services

2009 Introduces First Distributed Parallel Ingestion System DMAS

2010 Introduces Core Web Architecture for PO.DAAC Web Portal – Fast Faceted Search and Drupal CMS

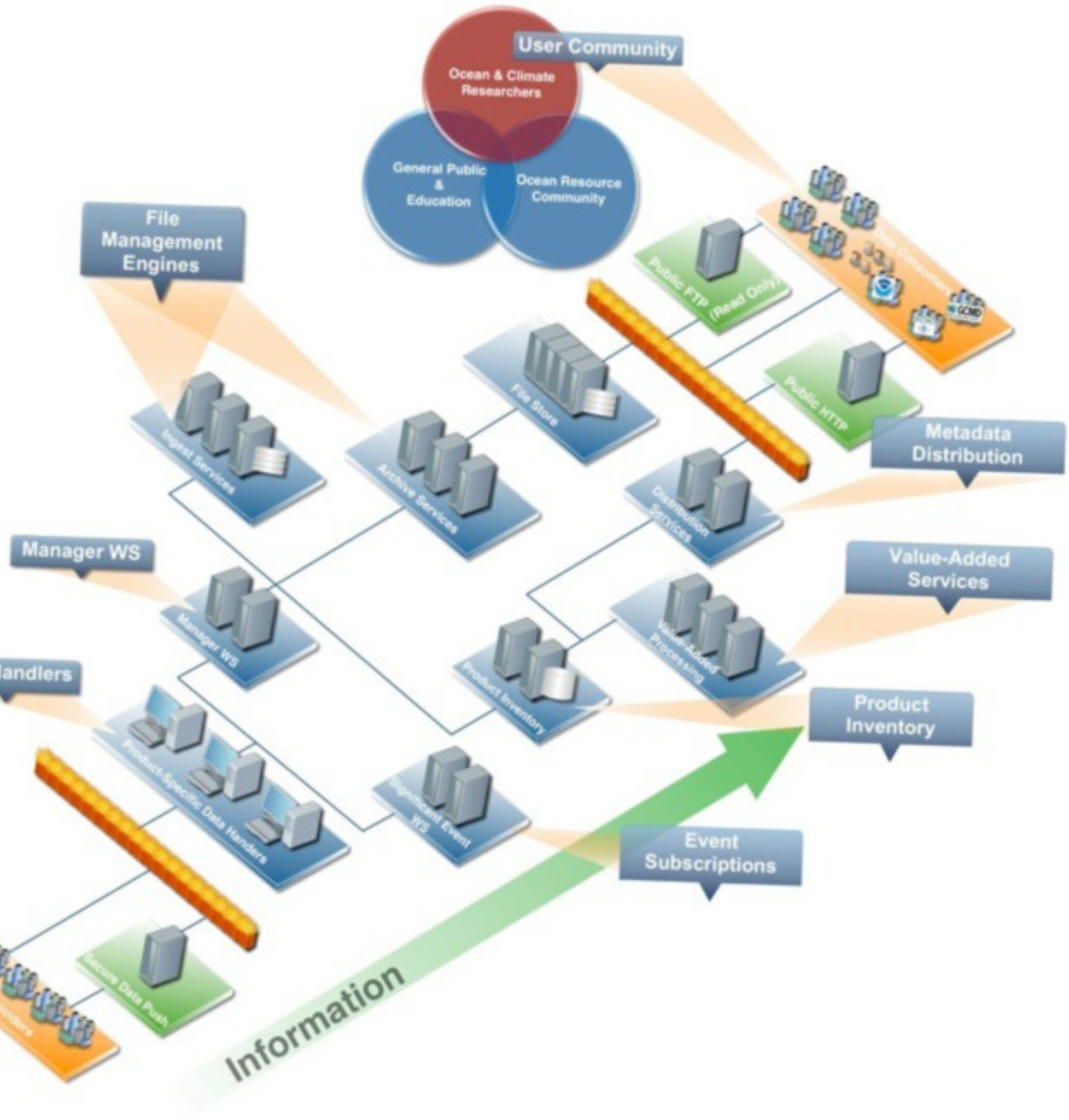
2011 Introduces Oceanographic Common Search Interface (OCSI) – OpenSearch, GCMD, FGDC, ISO, Datacasting





dmas - thuang - 03122012

DMAS - A Distributed Parallel Ingestion System



dmas - thuang - 03122C

Stats 3/2011-2/2012

1,092,542 Granules Ingested (1 Granule consists of 1..5 Files)

3000 Average Granules/Day

7000 Peak Granules/Day

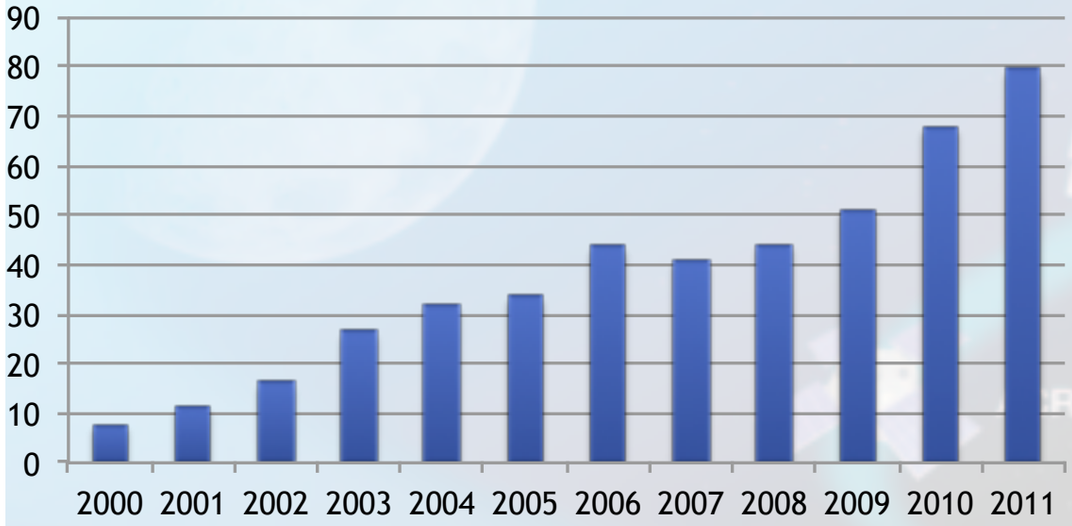
14TB In Total Volumes

152 Datasets Added

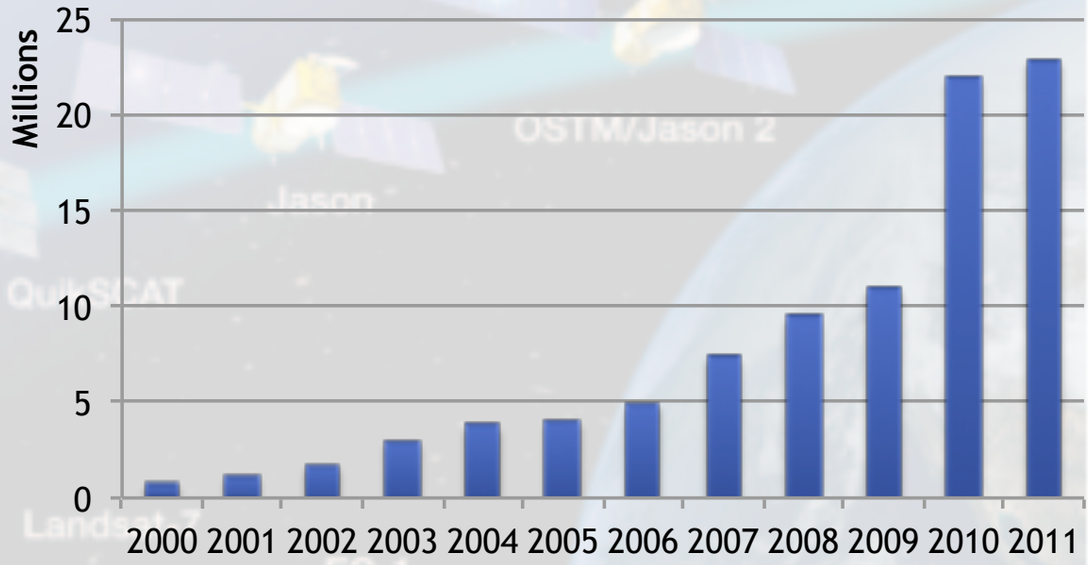
dmas - thuang - 03122012

Archive Metrics

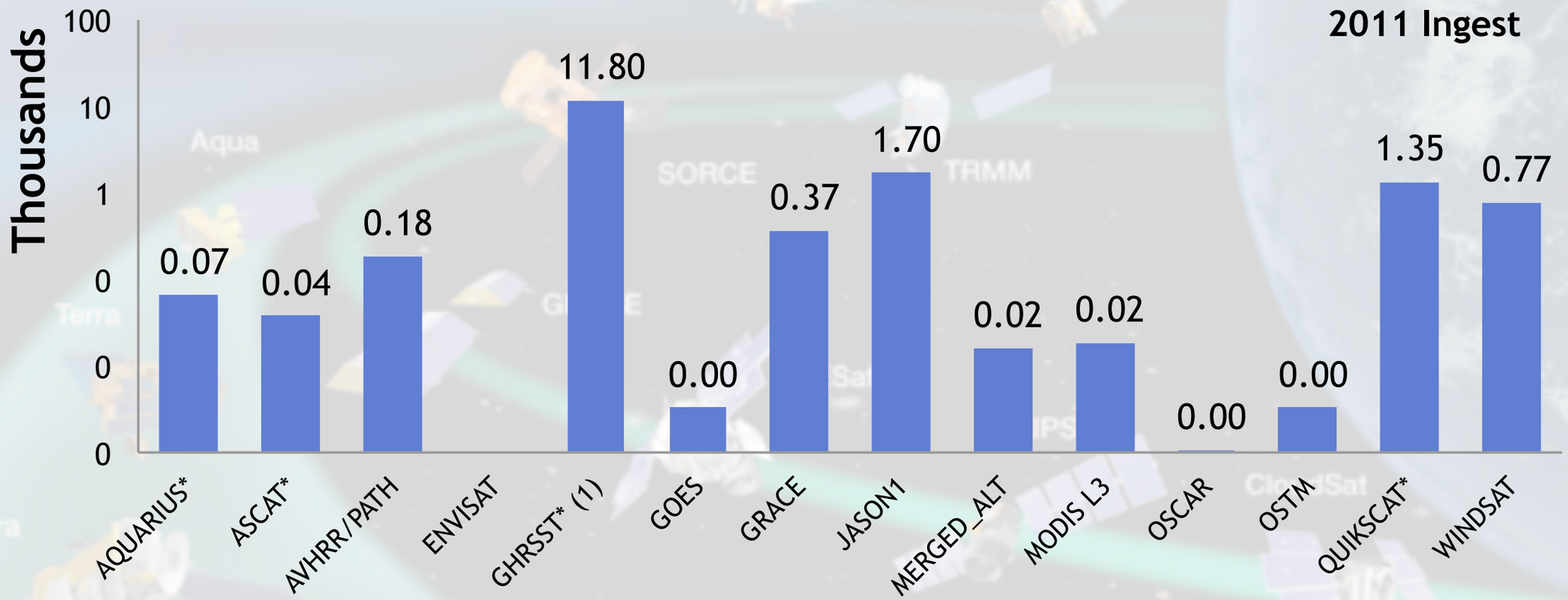
Archive (Compressed) TB



Archive Files



2011 Ingest



dmas - thuang - 03122012

Metadata-Centric Architecture

Metadata-Centric Discovery Service

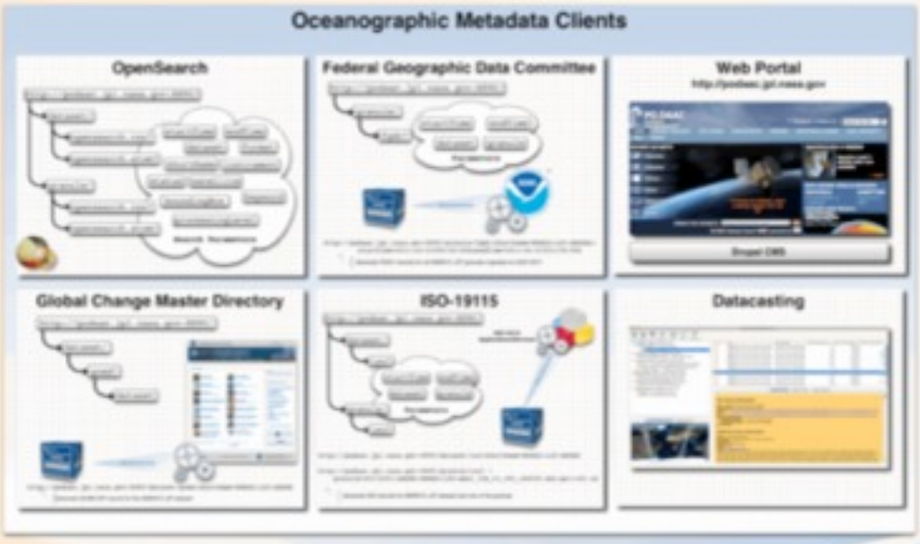
Thomas Huang, Edward M. Armstrong, Nga T. Chung, Michael E. Gangl

AQUA
AQUARIUS/SAC-D
Aura
CloudSat
GRACE
Jason-1
MODIS
NSCAT
QuikSCAT
SeaSAT
SSMI
TERRA
TOPEX-POSEIDON

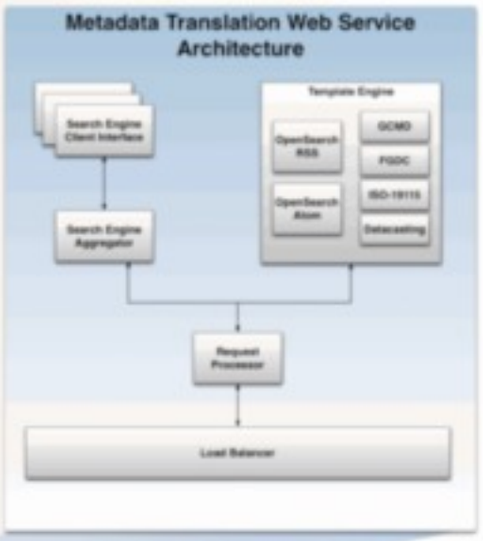
Abstract

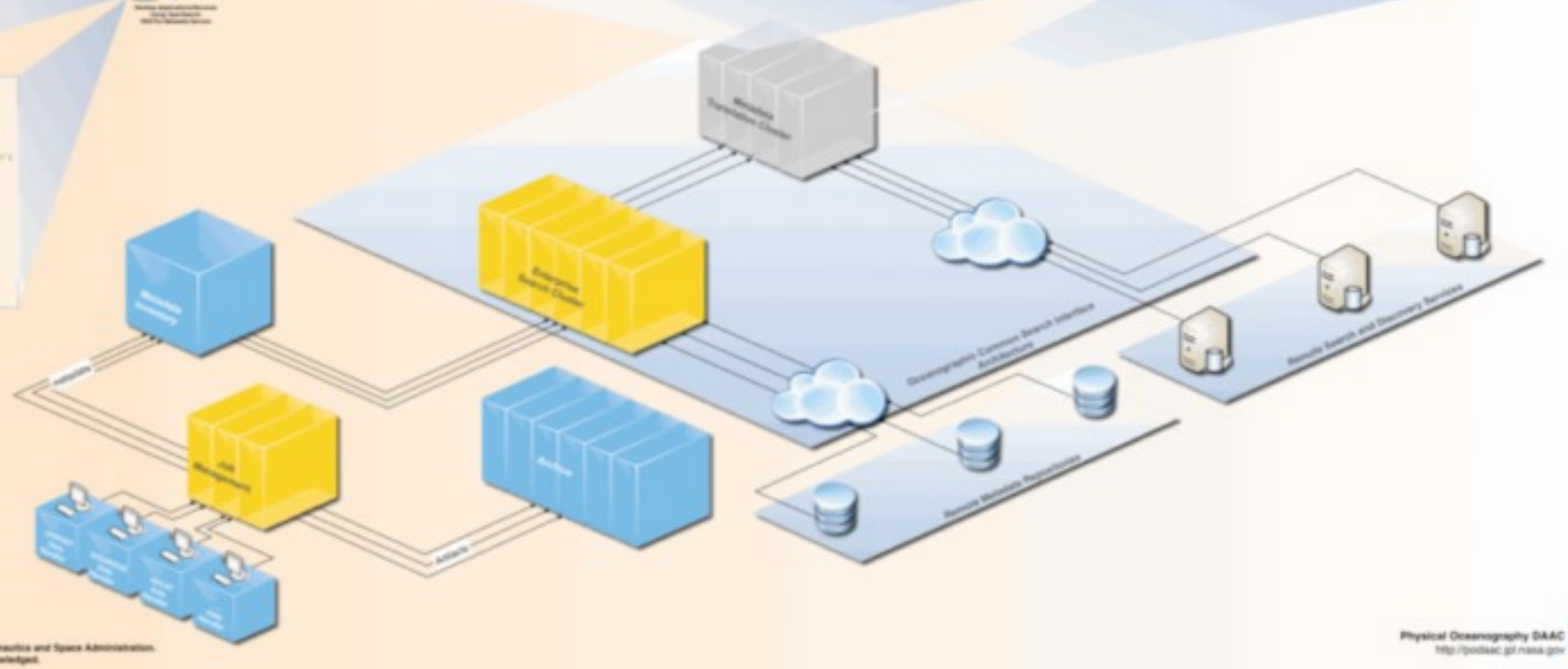
It is data about data. It is the information describing a picture without looking at the picture. Through the years, the Earth Science community seeks better methods to describe science artifacts to improve the quality and efficiency in information exchange. One the purposes are to provide information to the users to guide them into identifies the science artifacts of their interest. The NASA Distributed Active Archive Centers (DAACs) are the building blocks of a data centric federation, designed for processing and archiving from NASA's Earth Observation missions and their distribution as well as provision of specialized services to users. The Physical Oceanography Distributed Active Archive Center (PO.DAAC), at the Jet Propulsion Laboratory, archives and distributes science artifacts pertain to the physical state of the ocean. As part of its high-performance operational Data Management and Archive System (DMAS) is a fast data discovery RESTful web service called the Oceanographic Common Search Interface (OCSI). The web service searches and delivers metadata on all data holdings within PO.DAAC. Currently OCSI supports metadata standards such as ISO-19115, OpenSearch, GCMD and FGDC with new metadata standards still being added. While we continue to seek the silver bullet in metadata standard, the Earth Science community is in fact consists of various standards due to the specific needs of its users and systems. This poster focuses on the architecture behind OCSI as a reference implementation on building a metadata-centric discovery service.

Oceanographic Metadata Clients



Metadata Translation Web Service Architecture





These activities were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Physical Oceanography DAAC
http://podaac.jpl.nasa.gov

dmas - thuang - 03122012

“Capturing Good Data is like Brewing A Good Cup of Coffee”

-- Thomas Huang

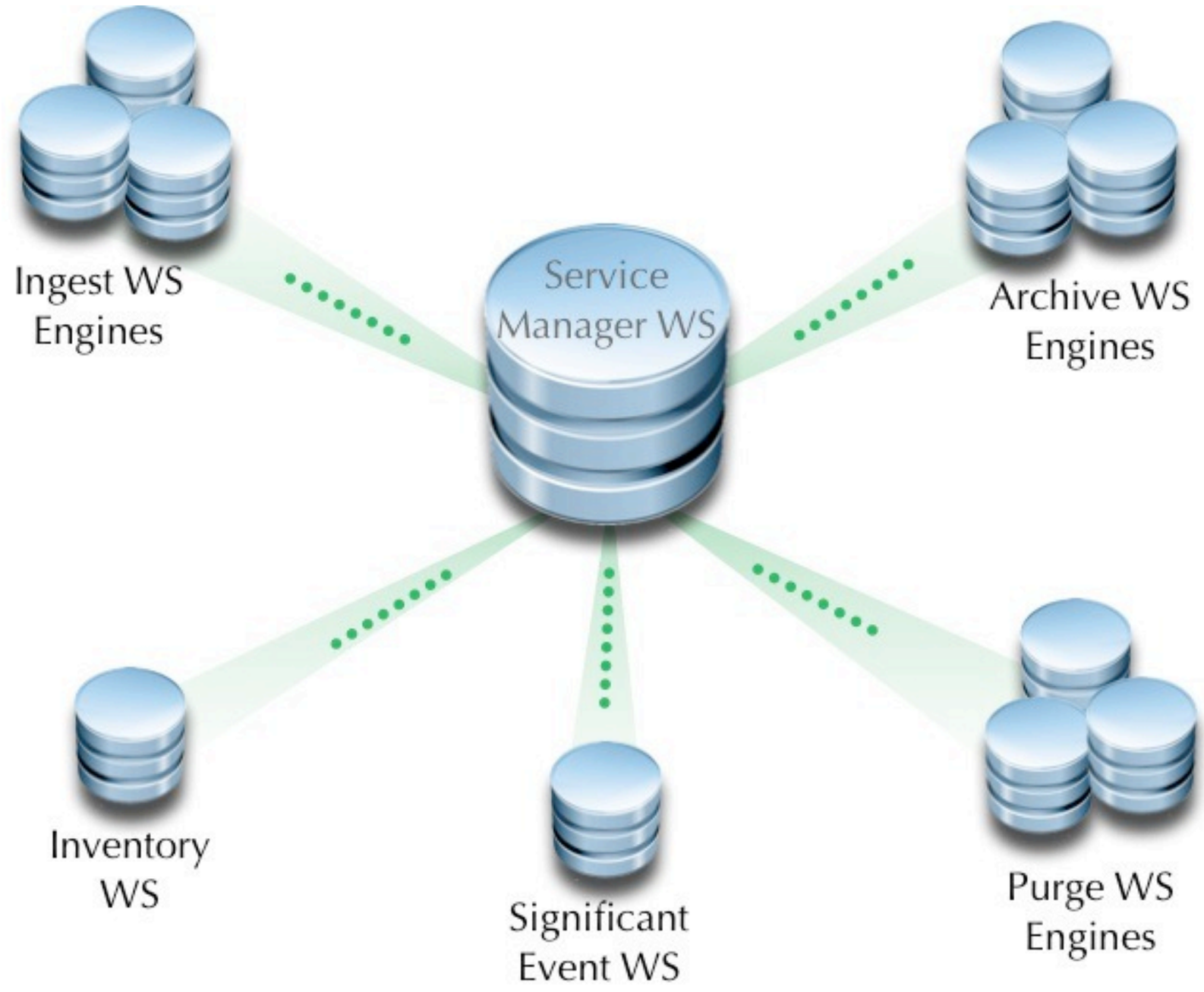
1

2

3

dmas - thuang - 03122012

DMAS' "Scale-Out" Ingestion Architecture



Dynamic Deployment of Ingest/Archive/Purge Engines

2012

dmas - thuang - 03122012

DMAS Federation

Over Capacity Avoidance

dmas - thuang - 03122012

"Scale-Out" Job Management Architecture

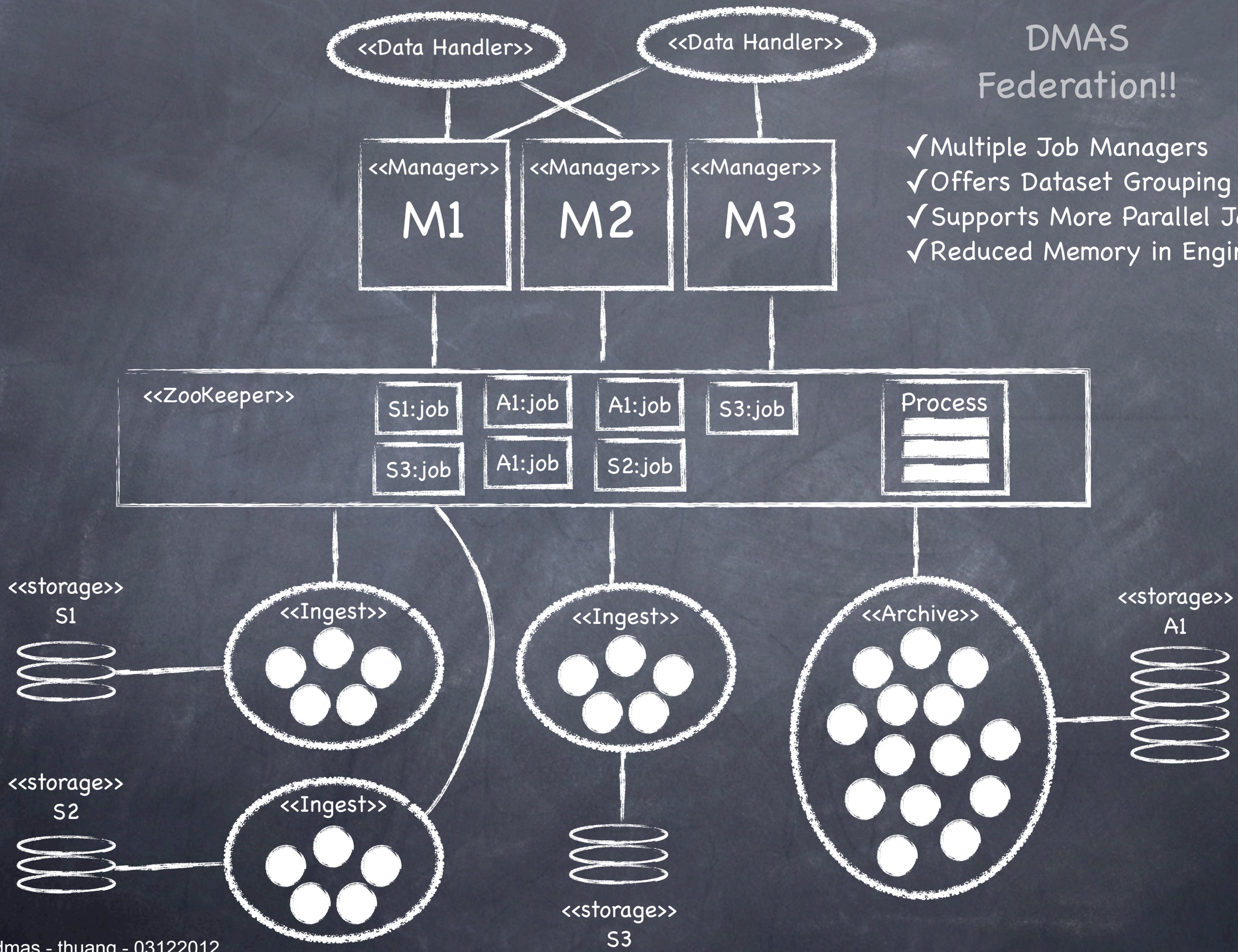
Simplify Job Assignment

No More Purge Engines

New Operator Console

DMAS Federation!!

- ✓Multiple Job Managers
- ✓Offers Dataset Grouping
- ✓Supports More Parallel Jobs
- ✓Reduced Memory in Engines



dmas - thuang - 03122012

Highly Reliable Distributed Coordination

Maintains Configuration Information

Communicate With Infinite Number of Worker Nodes

Provide Distributed Synchronization

Provide Group Services

Deploy on a Single Machine or Small Cluster

dmas - thuang - 03122012

DMAS & ZooKeeper Interactions

Engines

- ✓Register with ZK
- ✓Check Pause/Stop Signals
- ✓Fetch Ingest/Archive Job
- ✓Update Job Status

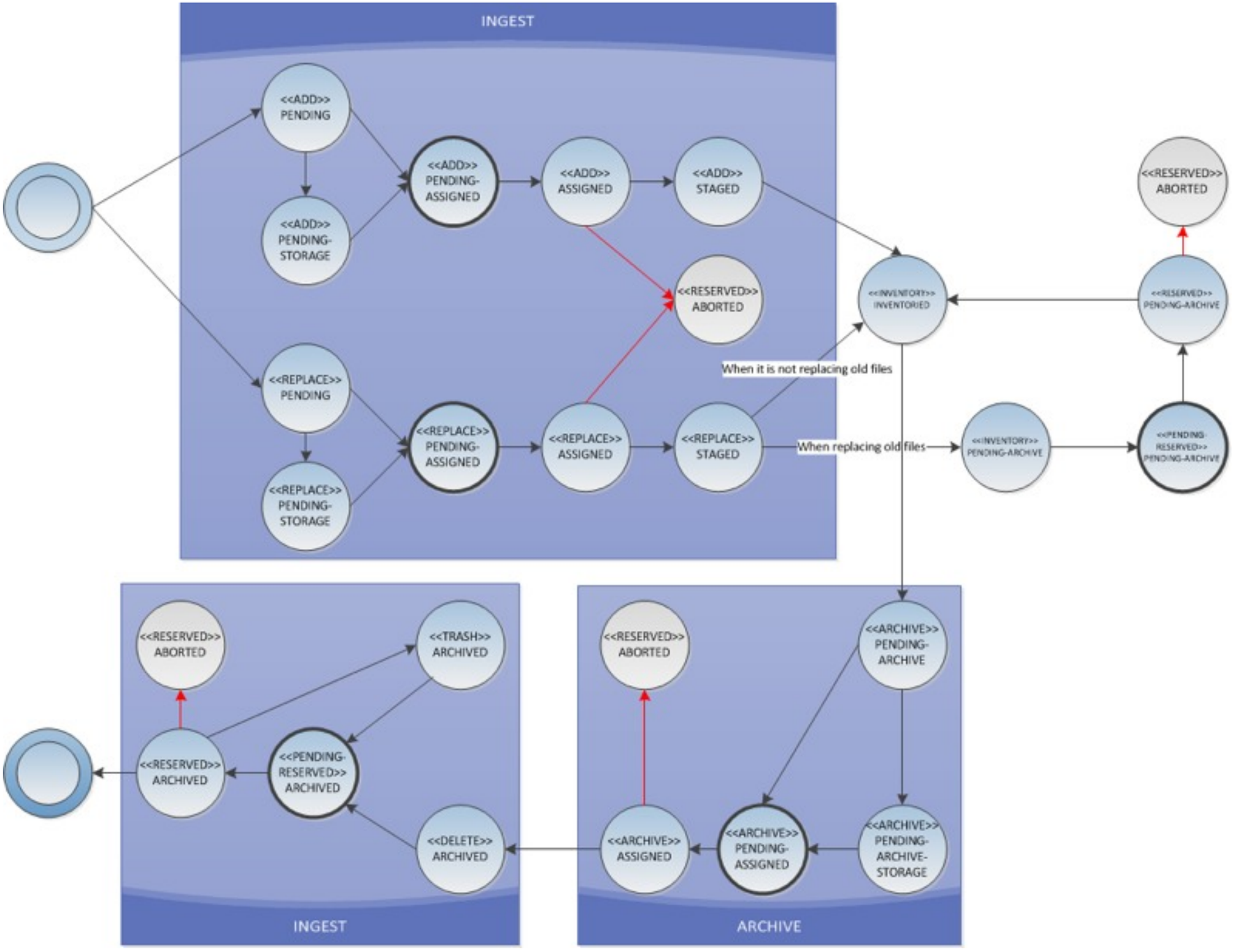
Manager

- ✓Post Ingest/Archive Jobs
- ✓Register for Notification
- ✓Upon Notified, Check Job
- ✓Remove Process Node

Operator

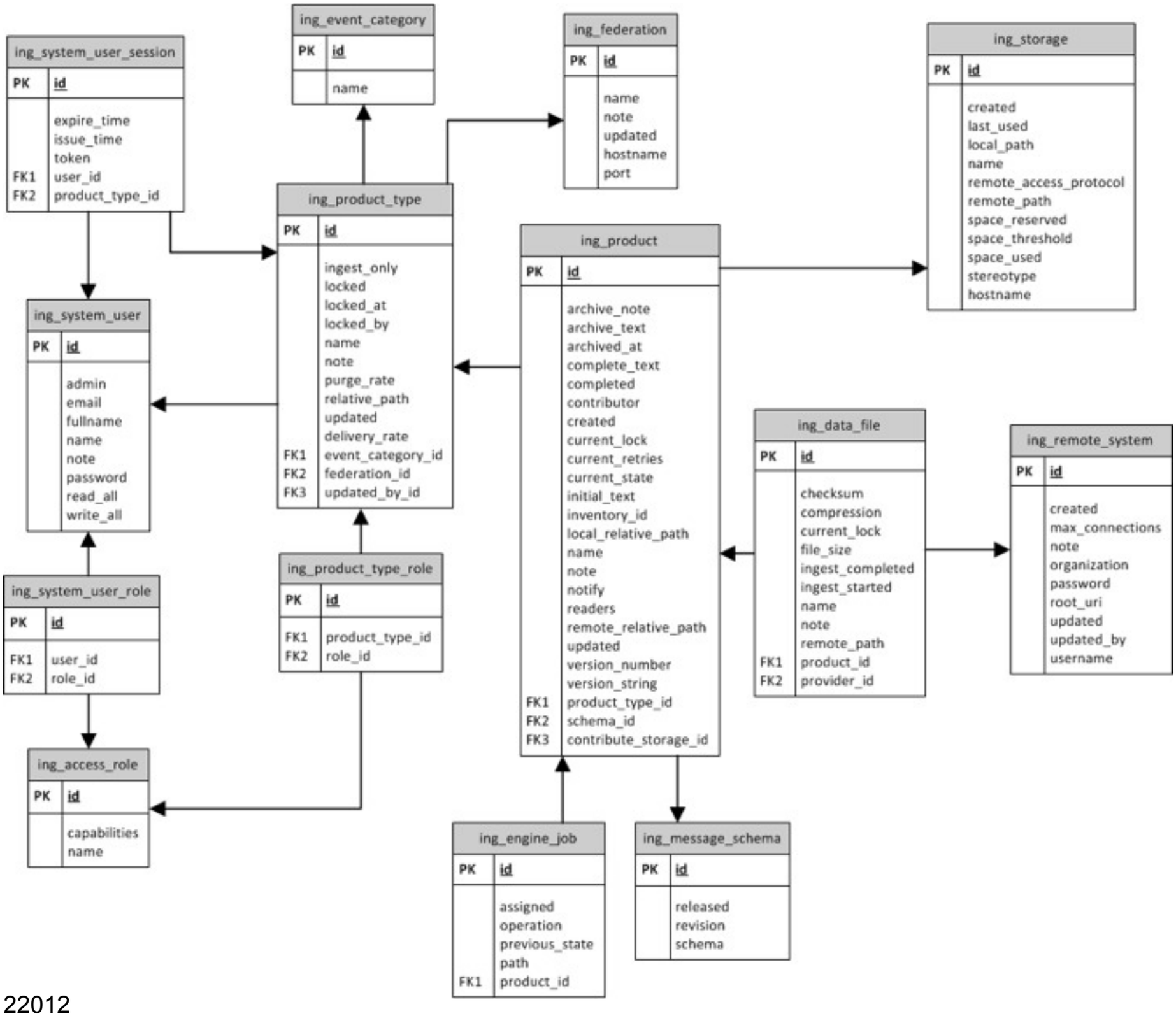
- ✓List/Check Engines
- ✓Pause/Resume Engines
- ✓Stop (Delete) Engines

Manager's Job State Machine



dmas - thuang - 03122012

Manager Data Schema



dmas - thuang - 03122012

"More Visibility and Control, Please"

OPS

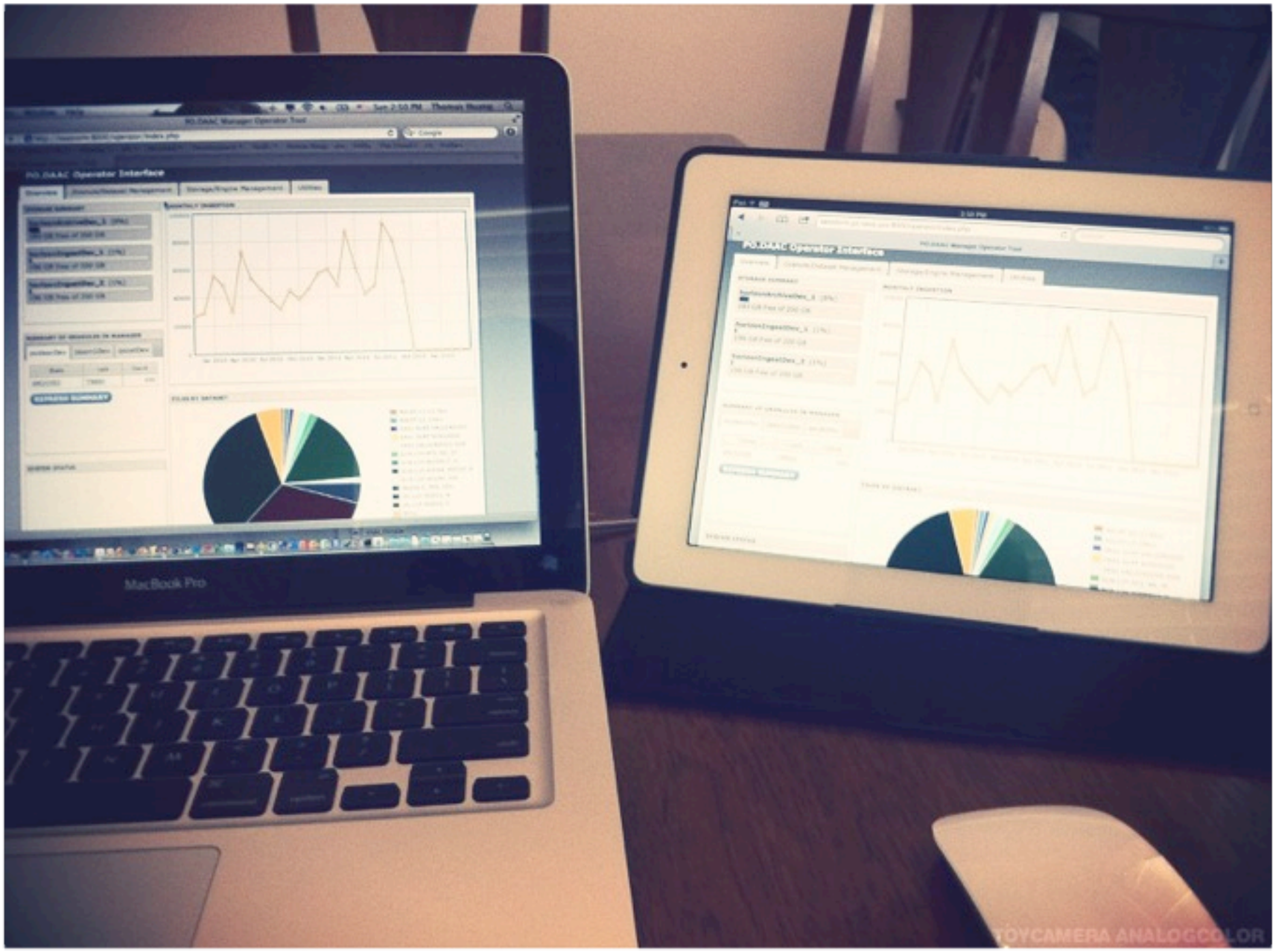
"We Are Working On It"

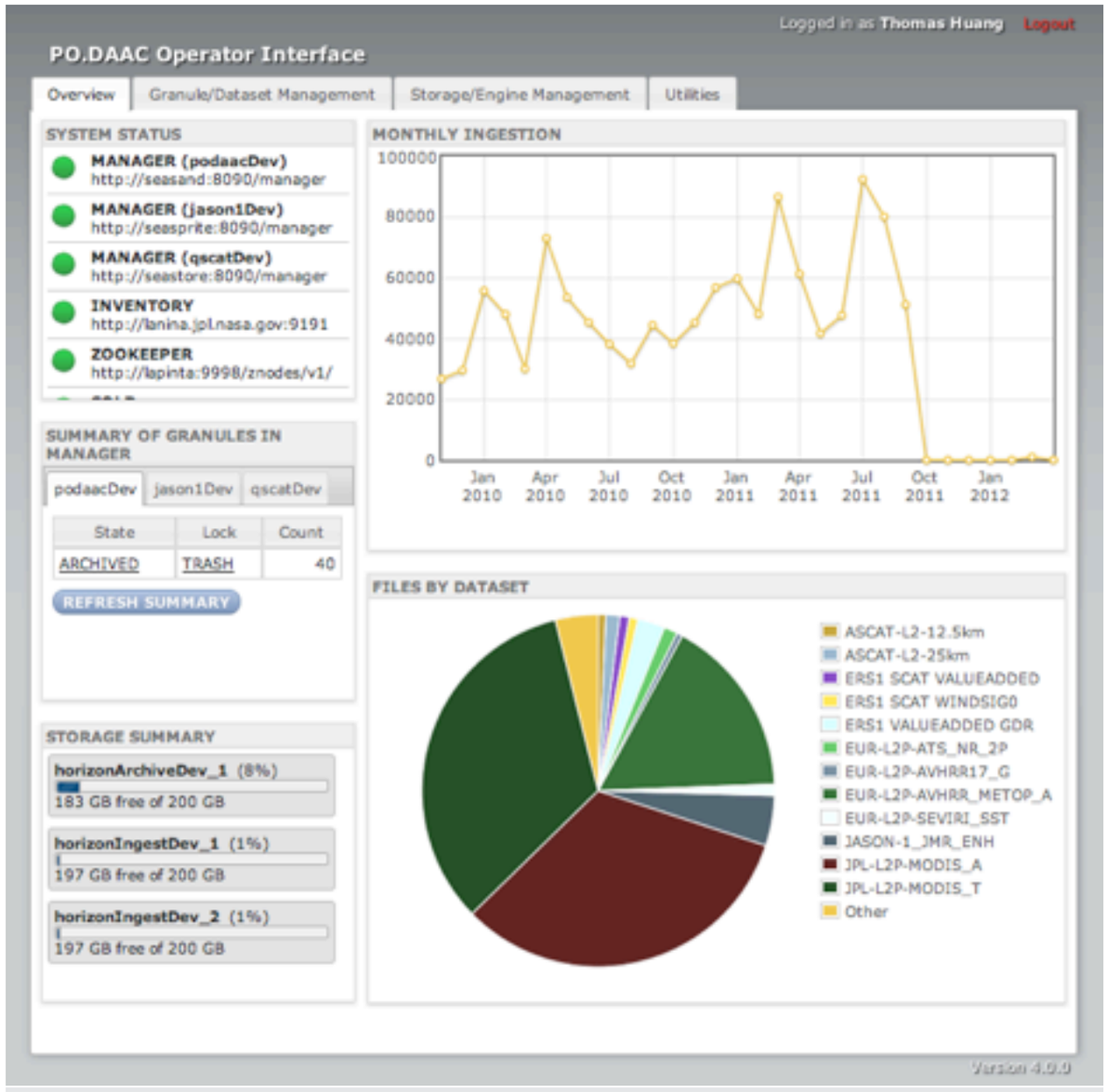
DEV

"Check This Out"

DEV

Working Late Last Night...





Meet The New Operator Console

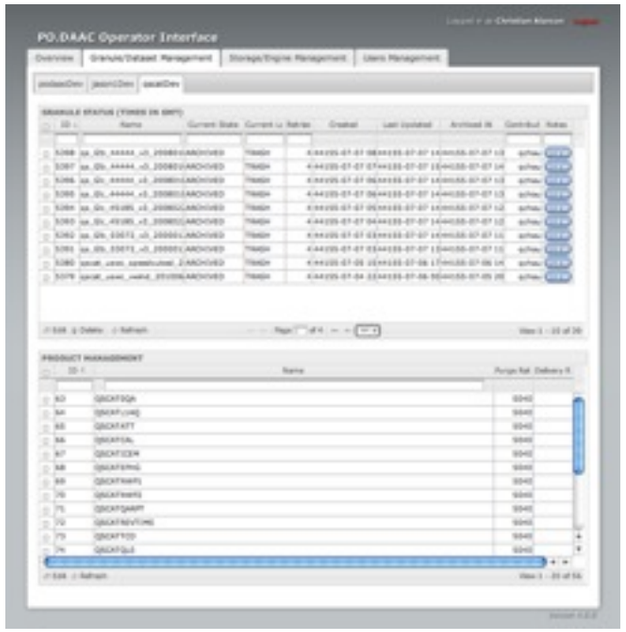
dmas - thuang - 03122012

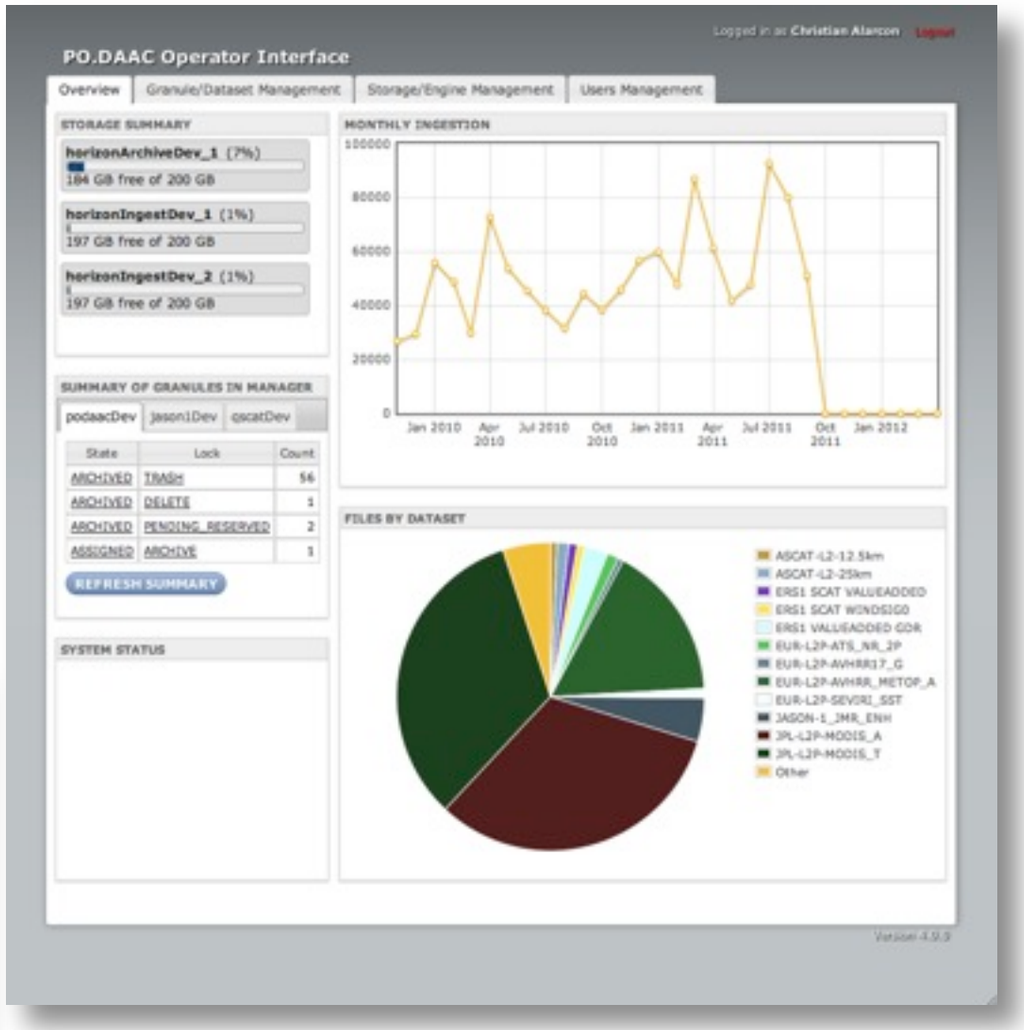
dmas - thuang - 03122012

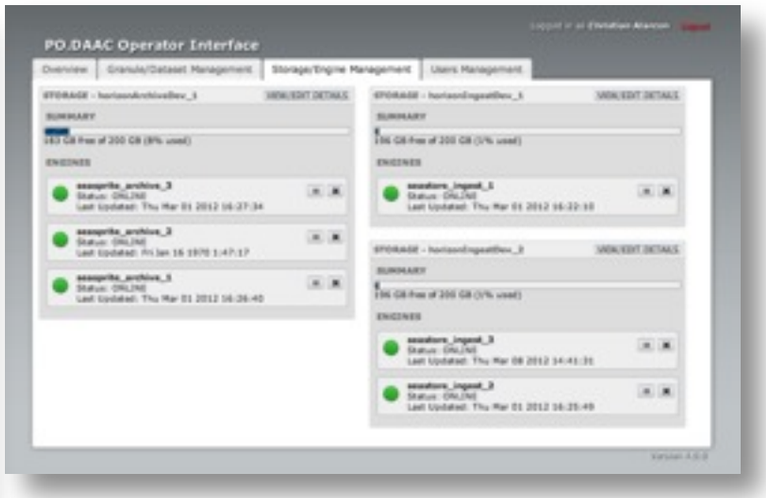
National Aeronautics and Space Administration

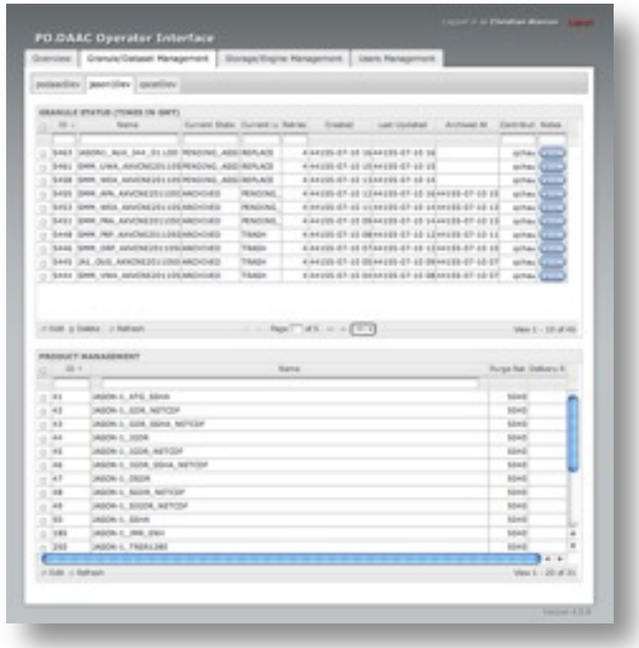
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

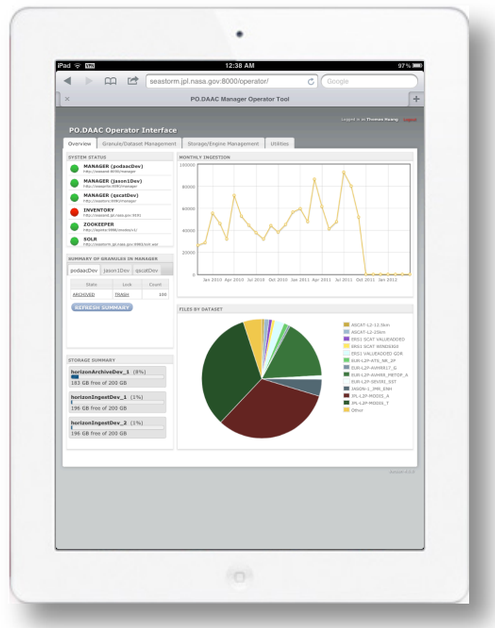
The New Operator Console



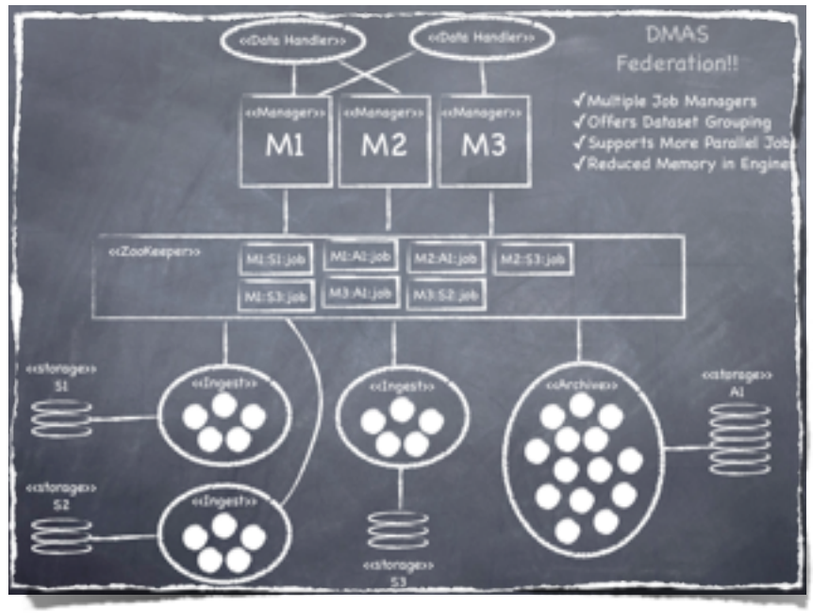








dmas - thuang - 03122012



DMAS Federation

dmas - thuang - 03122012