

Data Centers and the Future of CLASS breakout session notes

About 45-50 attendees in the room.

Online Participants: Anna Milan, Bhavana Rakesh, Dave Jones, Pamela Taylor, Peter Grimm and 1 other

Nancy Ritchie:

- Gave introduction to the session
- History of CLASS:
 - In the beginning started development in NESDIS/OSD

David Kending (CLASS Systems Eng) - Lessons Learned from Suomi NPP

- CLASS now has data rate of 4.5 TB/day w/NPP
- Focusing on using these lessons learned on GOES-R
- NESDIS IDPS is the largest source of data going into CLASS today:
 - 119 data products
 - GRAVITE next largest (1.3 TB/day)
- Data going out of CLASS
 - 3 interfaces
 - SDS
 - GRAVITE
 - NESDIS STAR
 - 16 active Data Center Subscriptions
- CLASS ingest between 10/28/2011 and 3/31/2012 - >11 million files
 - Challenge to maintain these files in a database
- QUESTION: How is metadata at collection and dataset level maintained for these files
 - Discovery level and granular metadata stored in database.
- CLASS has requirement to have 4x throughput vs. ingest rate.
 - QUESTION: Where did this requirement come from?

Lessons Learned:

- Challenge 1 -- Scope creep: changing data formats, new data sources to archive
 - Lesson Learned -- So...Need to ensure that requirements are tightly managed and that an accurate baseline is developed
- Challenge 2 -- data specific development work
 - Lesson Learned --
- Challenge 3 --
 - Lesson Learned --

Program-level NPP Challenges & Lessons Learned:

- 'Scope' creep - baseline all documents, publish change control (tighter management)
- Data development work outside of scope of CLASS, including development of new sub-setting tool - need program & data centers to identify specific requirements for data & tools. Better coordination with CLASS for requirements gathering. Apply standard tools.
- Late decision for 24x7 support - need advanced notice (9 months) to hire staff
- Stakeholders not all involved early - established a user forum with CLASS involvement
- Not everyone had access to the same documentation so testing and development were often happening in different environments and with different requirements - communication/distribution of key documents across team is essential
- No defined test schedule early, no validation methods & lack of communication with external test teams - Need detailed integration and test planning early in the project life cycle. Test planning should include approach, methodology, verification methods, artifacts, expected results and detailed test cases.

Data Centers and the Future of CLASS breakout session notes

- Translation and tracing of requirements across organizations (e.g. from NASA, to NOAA, to CLASS) was poor.

QUESTION: how is the tracing of requirements documented now?

- System interfaces (between testing and development) and communication is key.

- Test data no current with documentation - Sample test data must be provided early and match current documentation version.

- CLASS needs test data 6 months prior to needing operational functionality because CLASS has a six month development schedule. Not getting test data in a timely manner is an indication that the project is off track or has issues.

Need better process for tracking defects during external testing - need external test tracking tool.

Testing in operations was beneficial but test environment should also mimic operations - Performance and Load (PaL) testing needs to be enhanced

Definition of system freeze not fully understood - can patches be done during testing period? - Need clear expectations and direction (and understanding) of what is wrapped up in this.

- Are system patches allowed?

Internal Lessons Learned:

- CLASS organization has changed. Now project focused. GOES-R team will follow that campaign throughout its lifecycle.

QUESTION: How scalable is this? At peak there were 30 people working on NPP team.

- Now have faster turnaround on minor system changes (don't need to go through full CCB)

Dean Carter (CLASS Operations Manager) - CLASS Present and Future

- Reviewed current CLASS configuration

- 14 CLASS Data Families being ingested currently - 4 more planned

- CLASS adheres to OAIS Reference Model

- Archived modeled data sits in NOMADS, not CLASS

- Exploring the cloud as a complement for the CLASS system; pilot expected in FY13

Ken Casey - A Data Center perspective on the vision for CLASS.

- Overview of the three NOAA National Data Centers

- Data Centers = in this context, they are climatologists, oceanographers, geophysical, meteorologists

- provide scientific expertise to users of data - provide services, lead communities

- not little "dc" data center. More science stewardship not just IT.

Brennan Hay - A tale of 3 storage requirements:

- Science Operations Requirements

- High performance

- Direct attached, SAN or Networked

- Format (blocks and files)

- Low latency

- Appropriately sized for data sets

- Structured and unstructured data

- Administrative System Storage Requirements (40-60 NOAA systems)

- General storage pool

- May include PII

- Low performance for users

- High performance for servers

- Extremely duplicative

- Format (blocks and files)

2012 NOAA EDM Conference
Tuesday, May 15-17

Data Centers and the Future of CLASS breakout session notes

- No NOAA enterprise high performance storage solution
- CLASS and HPCS - archive storage within NOAA - why aren't CLASS & HPCS sharing resources?
- Need service oriented capabilities?

<end session notes>