HST Data Processing

Operations and Engineering Division
Data Systems Branch
Mark Kyprianou
Highlights - Infrastructure

• Deployed database clustering technologies to DADS/OPUS systems
  – Internal and Public databases
  – Improve uptime, reliability and performance

• Updated/added new web servers
  – 3 independent environments:
    • Dev, Test and OPS
    • Allows for isolation in the software life cycle
Highlights - Instruments

• Science data processing enhancements
  – Astrodrizzle (ACS and WFC3)
  – CALACS w/CTE
  – COS reprocessing/cataloging enhancements
  – CALCOS with no IRAF
  – STIS flight software problem with time tag data
  – STIS Reprocessing to create valid CCI files for post-SM4 data
Retrieval Times

Linux Request Retrieval Times

- Network issues due to security scan, fixed in subsequent builds.
- Hardware failure on weekend of 6/25.
- ACS CTE activated May 16, 2012.
Future Development
Future Development

• Infrastructure
  – Red Hat OS upgrade for DADS/OPUS servers
  – Purchased NAS storage (EMC Isilon)
    • Very scalable for Big Data
    • Blades support
      – High IOPS
      – CPU processing
      – Storage
Future Development

• Instrument Data Processing enhancements
  – Keywords
  – Algorithmic enhancements

• Reprocessing planned for ACS, WFC3 and COS

• Iraf Free: WFC3 and STIS
  – In test
HST Upgrade Projects
HST Upgrade Projects

• Background Automated Reprocessing (BAR)
  – Integrated reprocessing
  – Online cache

• Workflow Manager (WFM)
  – Technology refresh for work flow processing

• Calibration Reference Data System (CRDS)

• Driven by analysis of current systems targeting JWST

• HST Mission Office elected to adopt these enhancements
Background Automated Reprocessing (BAR)

- Handle reprocessing activities in an integrated fashion
  - Use prioritization scheme to control processing sequence
- Provide online access to files
- Enhance tiered file storage sub-system
  - Store data to provide direct access
  - Rework file handling to be mission independent
Work Flow Manager (WFM)

• Upgrade workflow processing
  – OPUS technology refresh is over due
  – Limited functionality for expansion (cloud computing, additional compute resources)

• Condor: Open source, high-throughput computing software framework has been selected
  – Allows to access compute resources on different levels/rules
Overview: BAR/WFM/CRDS
Current Status

• BAR
  – SDPQ - complete
  – File Storage to online cache – in progress
  – Working on distribution updates

• WFM
  – Condor
    • Enterprise level, Open source workflow package
    • Open Workflow Language (OWL)
      – Blackboard, template for data centric processing
  – Shoveler - complete
  – Prototype CALACS pipeline

• Continued development thru FY13
Questions/Comments