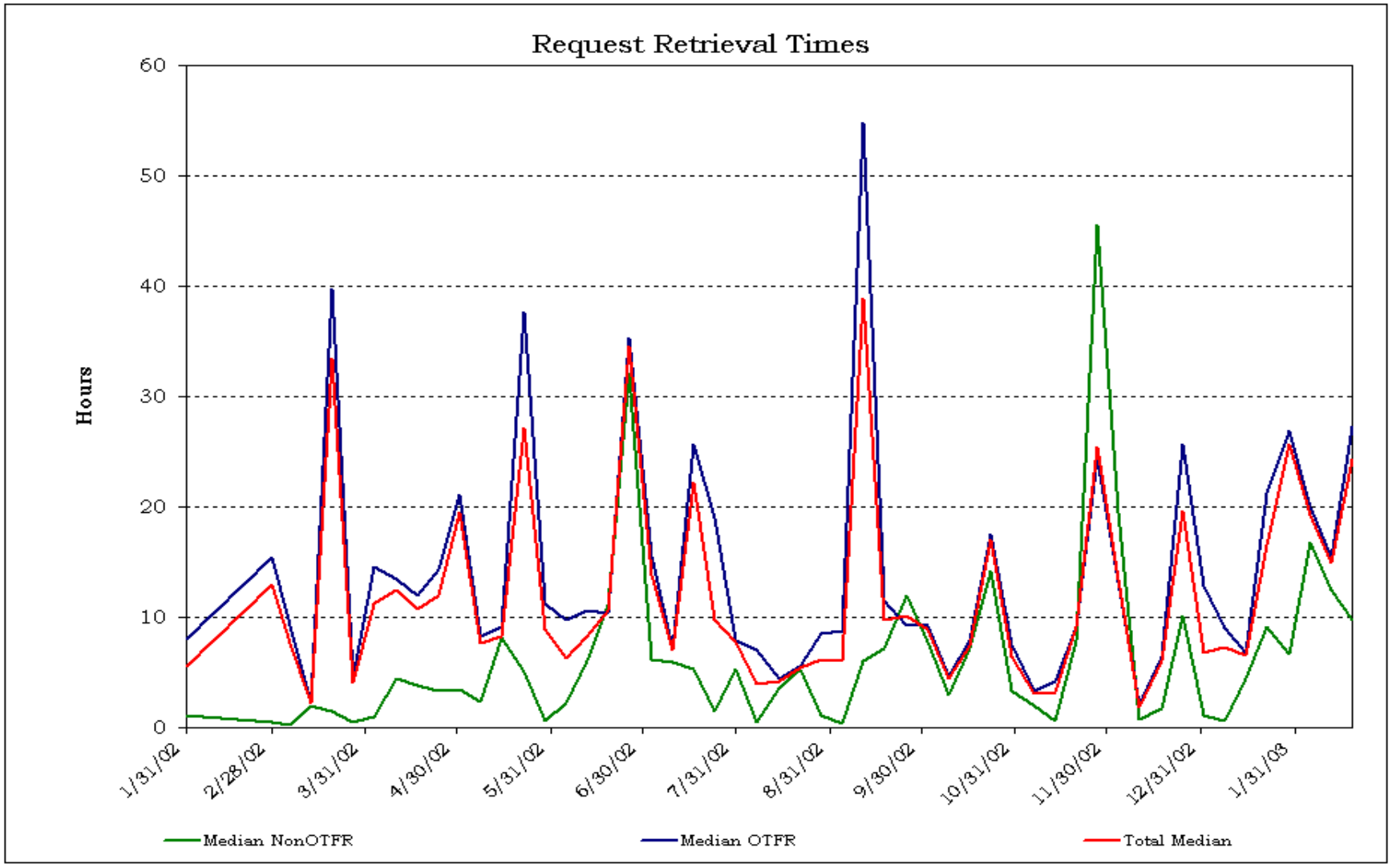
Future Directions in HST Data Processing

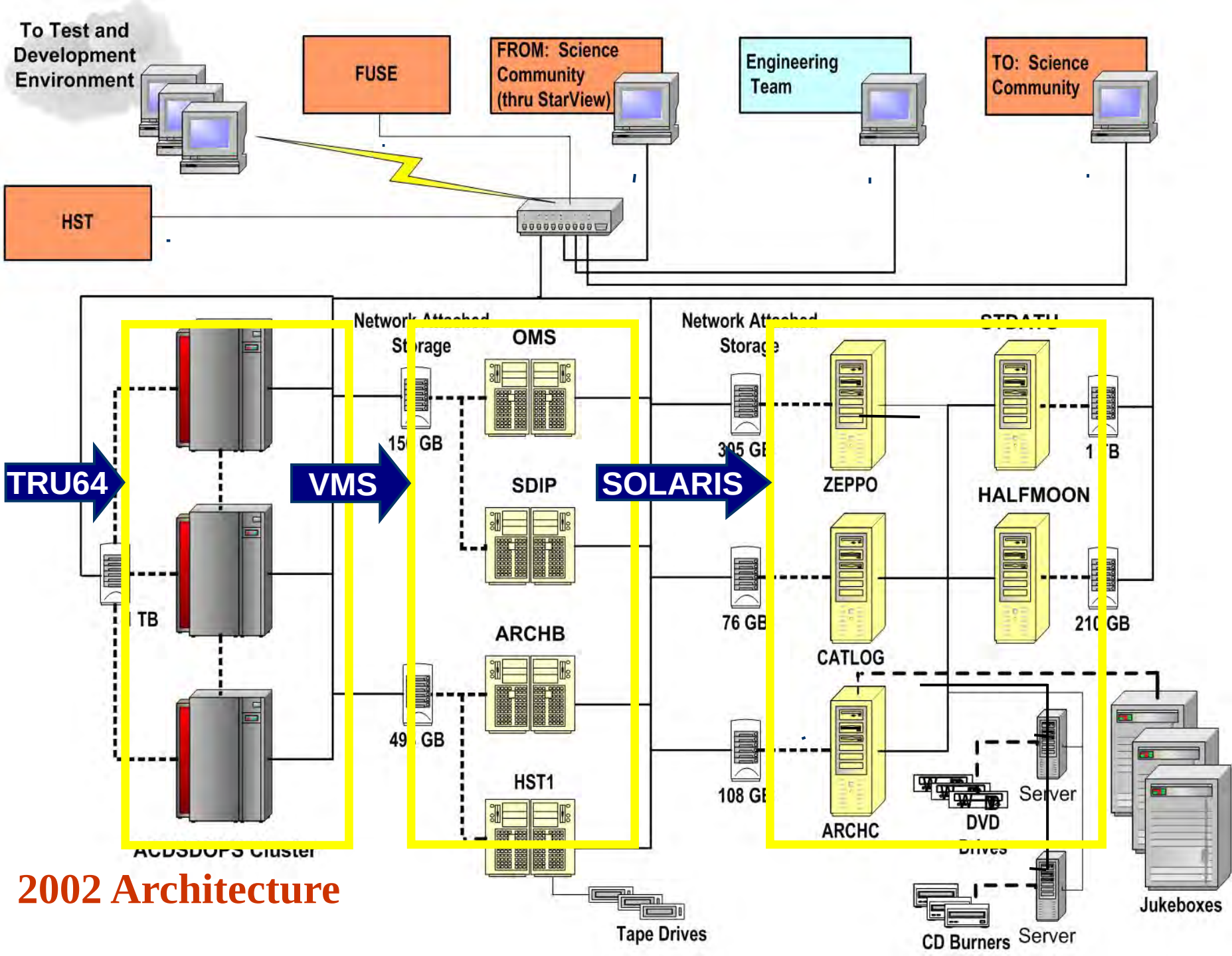
19 November 2004

History (2002-2003)

- Several parts of the system exhibited unacceptable performance
  - Processing of data from HST to the Archive
  - Response time to user requests for data from Archive
- Several specific causes
  - NFS mount problems
  - Disk corruption in OPUS
  - Jukebox problems
  - Other specific hardware problems
- Symptomatic of more general problems with the data processing systems

History (2002-2003)





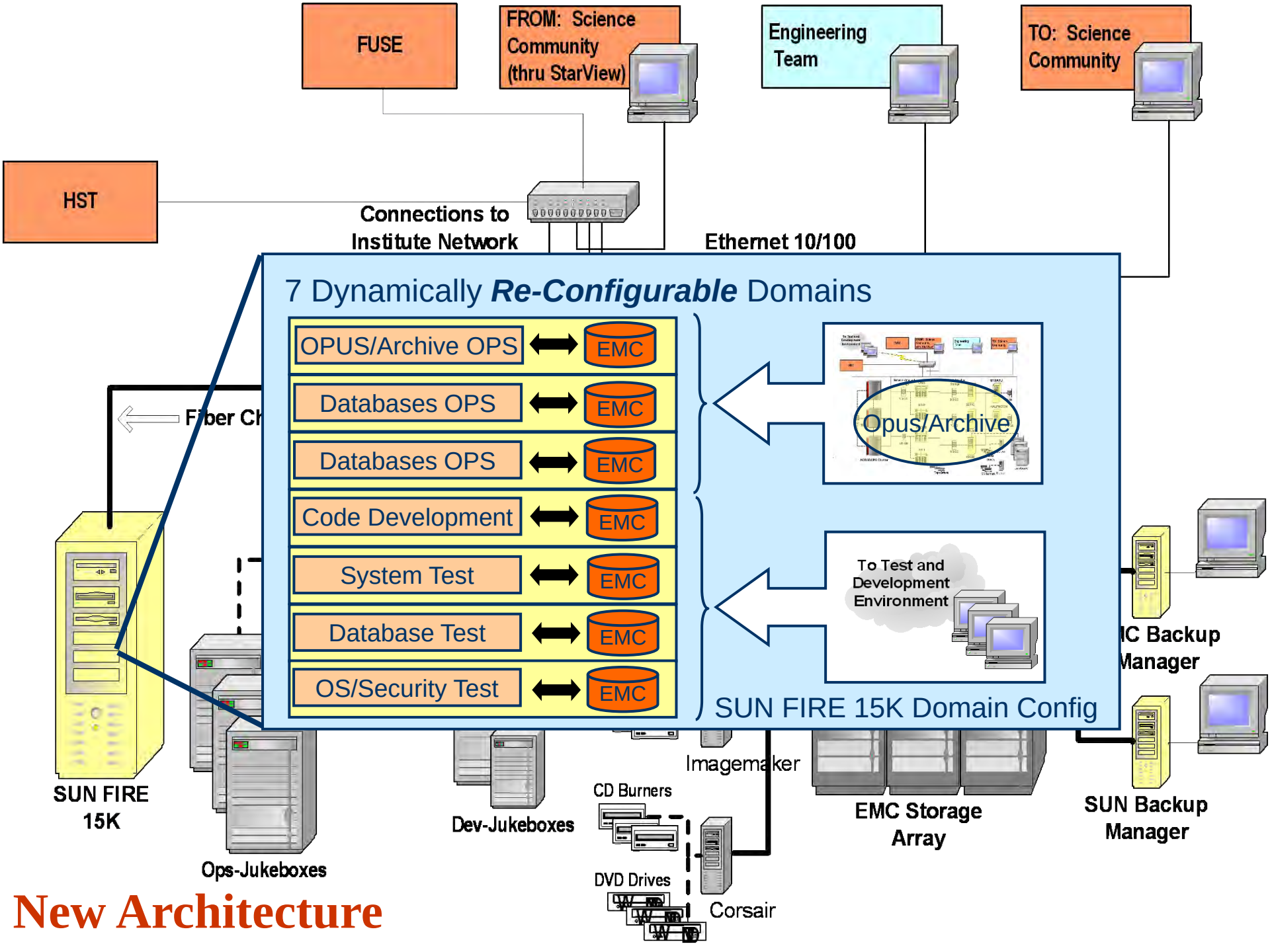
2002 Architecture

Goals for re-implementation (2003-2004)

- Reduce total cost of data processing activities
- Simplify hardware and network architecture
  - Reduce Operating Systems from 3 to 1
  - Consolidate many boxes into two highly reliable boxes
  - Flexible allocation of computing resources
- Provide common development, test, and operational environments
  - Development and test systems could not replicate load of operational systems
- Improve ability to capture performance data, metrics, etc.

Goals for re-implementation (2003-2004)

- Continue planned pipeline evolution
  - DADS Distribution redesign provides more flexibility to users and operators
  - Storing copy of Raw Data on EMC will dramatically reduce load and reliance on Jukeboxes
  - Ingest redesign provides opportunity to finally end the arbitrary boundary between OPUS and DADS

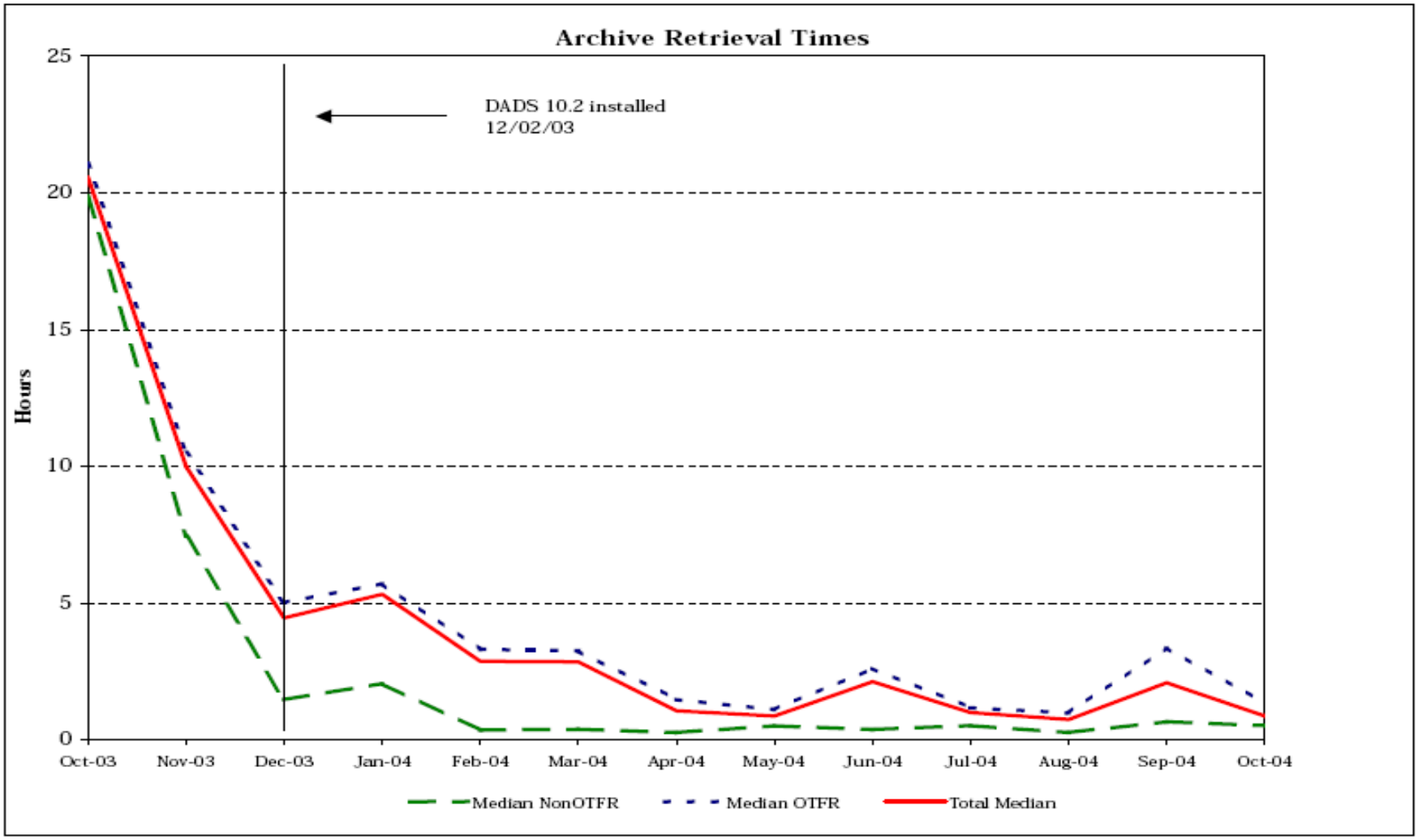


New Architecture

Re-implementation completed in 2004

- New Archive Ingest system installed June 2
  - Removes artificial boundary between OPUS pipeline and Archive
  - Removes VMS systems from input side of the Archive
- Upgraded to 64 bit compiler on June 21
- Tuned the EMC disk system over the summer
- Multidrizzle installed in ACS pipeline Sept. 22
  - Extra compute load is well within system capacity

Current Retrieval Performance



Near-term Plans

- Clean up remnants of old architecture
  - Totally decouple Jukeboxes from retrieval/ingest
  - Determine future of safestore
    - Media
    - Jukeboxes (?)
- Science value-added improvements
  - Multidrizzle
  - Accuracy of coordinates
  - Calibration upgrades (e.g. STIS close-out)

Multidrizzle

- Dither Working Group is leading development of Multidrizzle (A. Koekemoer et. al.)
- Recent steps:
  - Multidrizzle implemented in Pipeline for ACS associations
  - Same version released last week in STSDAS
- Next steps – more Instruments:
  - STSDAS version can handle STIS imaging, WFPC2 data
  - Extend association definitions for ACS, WFPC2 to cover “pos targs”, and other data groups
  - Implement WFPC2 and STIS imaging in pipeline (mostly requires additional scientific verification and validation)
  - Use for NICMOS and STIS spectroscopy will require some additional software development

Multidrizzle

- Next steps – Tweakshift
  - Currently:
    - Pipeline Multidrizzle uses headers for offsets between images
    - STSDAS version can accept measured offsets
  - Prototype Tweakshift (in STSDAS) to measure offsets from images
    - Via object identification
    - Via cross-correlations
    - Via wavelets (TBD)
  - After evaluation via STSDAS release, move to pipeline processing
    - Eliminates small pointing errors between images and orbits
    - Provides opportunity to improve absolute position reference

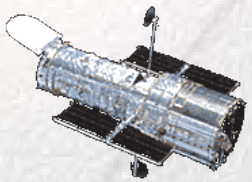
Accuracy of Coordinates

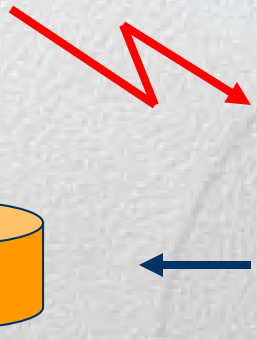
- Accuracy of WCS coordinates is ~ 1-2 arc-seconds
  - Original Guide Star Catalog limits accuracy
  - Slow Drift of FGSs relative to SIs has not been accurately calibrated
- Imaging data
  - Tying GSCII to Hipparchos frame provides opportunity to improve accuracy
  - Use objects found by Tweakshift to directly determine WCS coordinates – not affected by drift of FGSs relative to SIs
  - Evaluations have been successful
- Small FOV data sets (e.g. STIS)
  - Retroactively calibrate drift of apertures
    - Imaging data can provide FGS-to-FGS drifts
- Goal is accuracies of ~0.1-0.2 arc-seconds

Longer-Term Plans

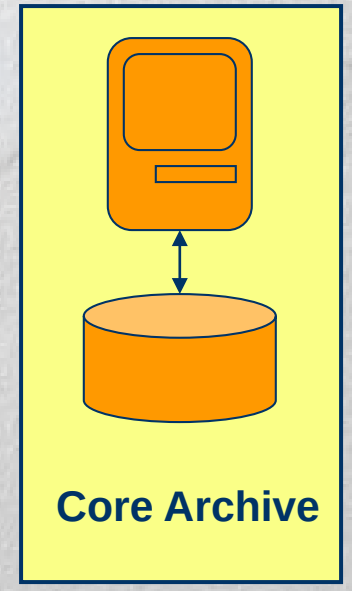
- January, 2004 decision re SM4 led us to look for ways to augment the scientific value of the HST data processing and archive services
  - Stimulate an even broader re-use of the Hubble data
  - Support a more aggressive archival research program, in the event there is a major gap between HST and JWST
- Timed well with respect STScI planning for other missions we support (Kepler, JWST, NVO, etc.)
- We will be developing a proposal for enhancing the Hubble data services
  - Minimal disruption to the core architecture we have just finished
  - Add an outer “layer” that builds on the core
    - Higher level science products
    - Faster response time
    - NVO compatible

Hubble Data Archive
Current Situation



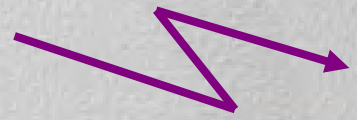


Safe Store



Core Archive

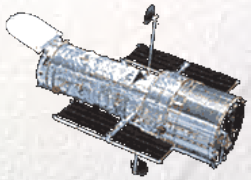
- Process data from HST
- Re-process requested data for user
- All data is available

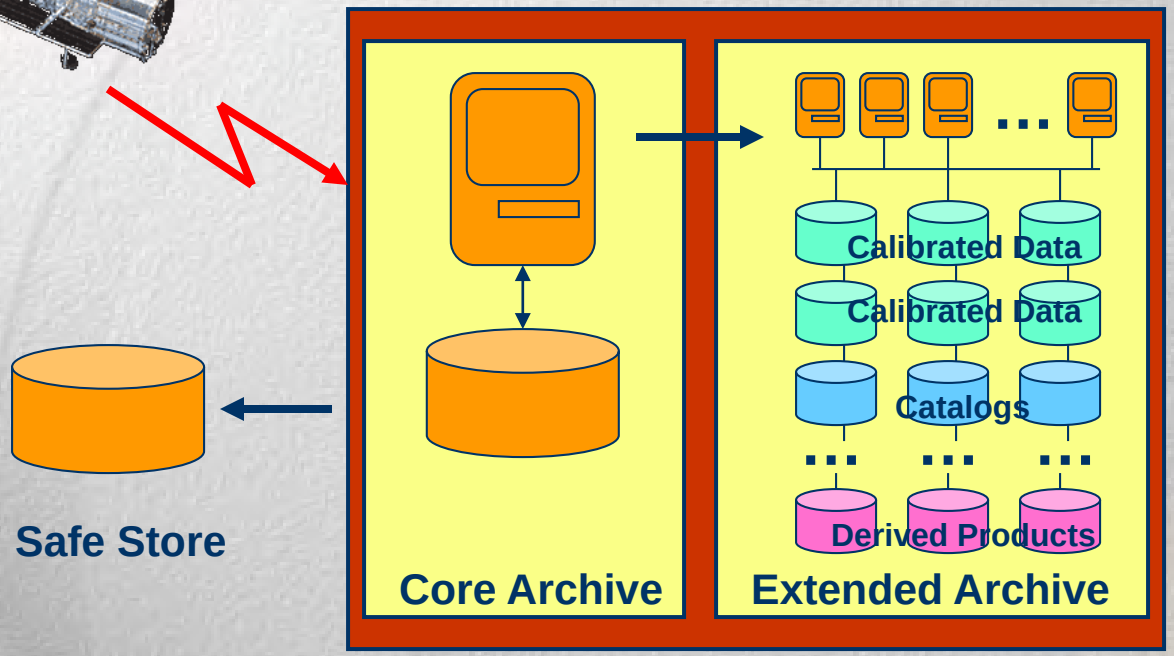


Archive User

- Retrieval Times: Hours

Hubble Data Archive
Potential Enhancements





Archive User

- Cutouts
- URL-based access
- Retrieval Times: Seconds

- Process data from HST
- Re-process data for extended archive
- All data is available

- Most calibrated data online
- Automatic updates from core
- Improved WCS coordinates
- Object catalogs
- Space/time “footprint” service
- Other high-level products…

Next steps

- Formulate specific plans and schedules for near-term (~ next year) upgrades
  - Multidrizzle
  - Coordinate accuracies
- Document our concept for enhanced HST data services for HSTP and HQ
- We welcome STUC input to both processes