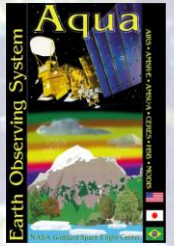
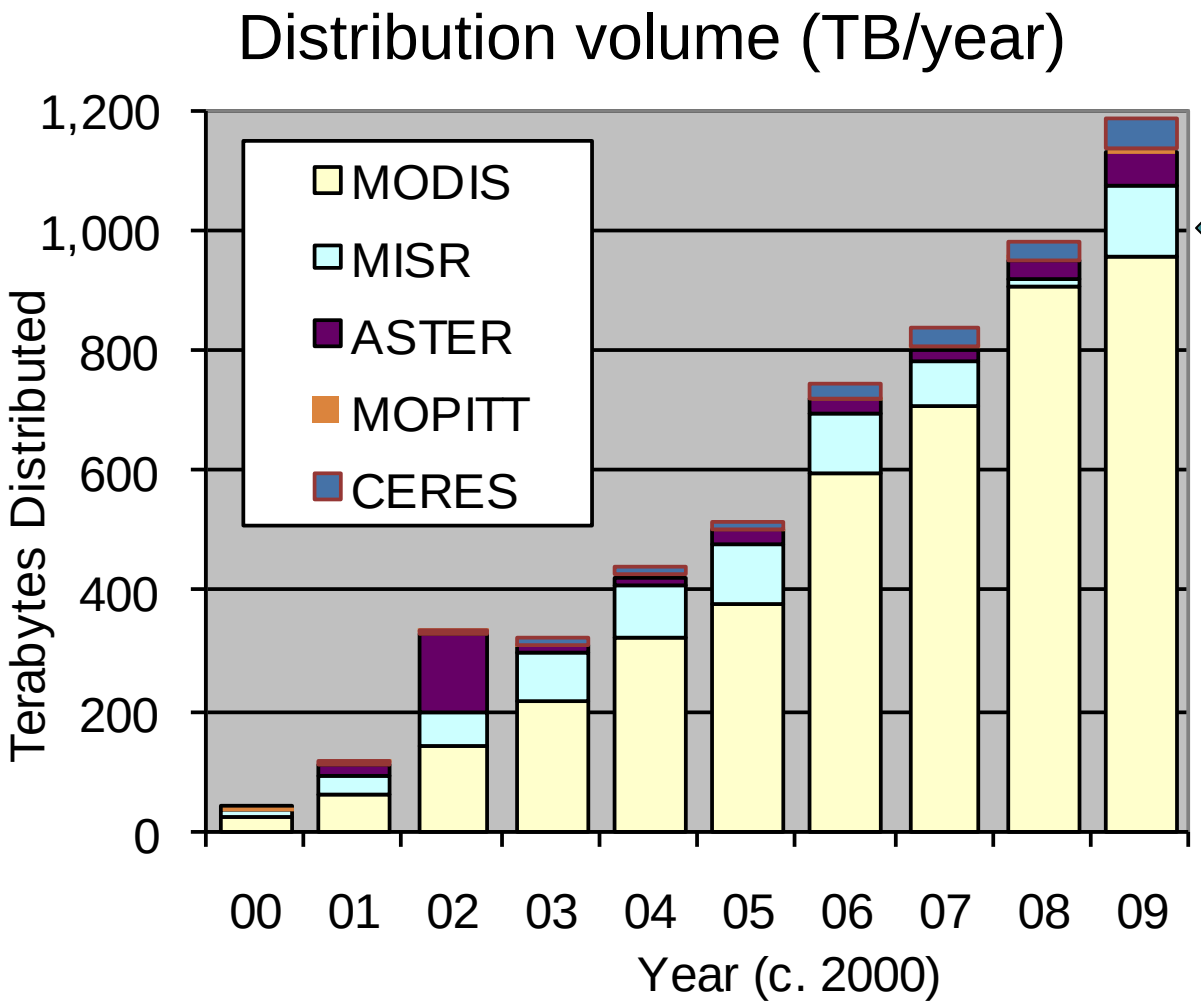
Managing MODIS Satellite Data for Global Studies

Edward Masuoka

NASA Goddard Space Flight Center, Greenbelt, MD, USA



Terra data volume distributed



1 PB

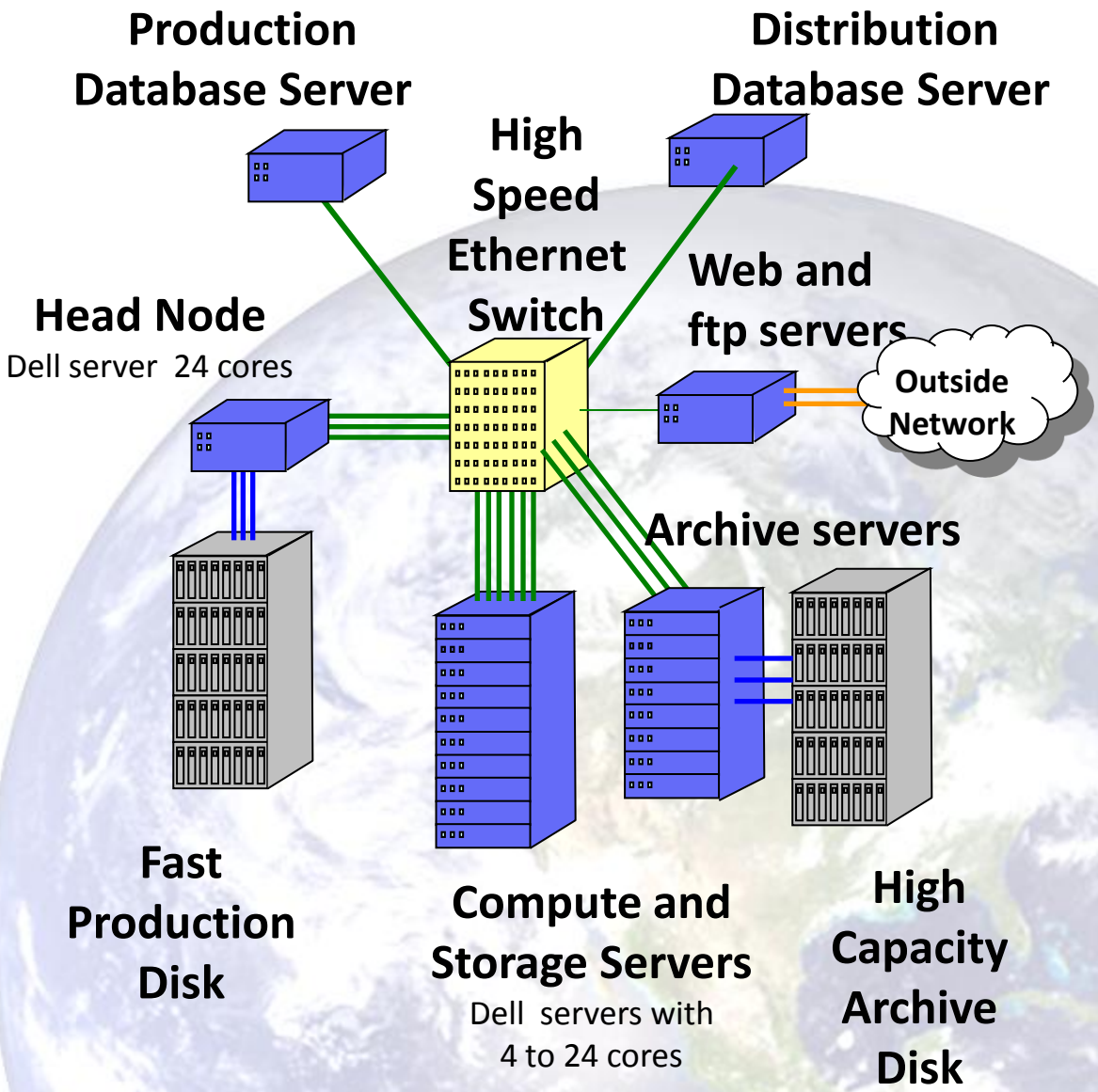
Increased an average of 21% per year over the last 7 years

Lessons from MODIS products

- Size standard products to facilitate processing, distribution and ease of use
  - 5 minutes of swath data (Level 1 and Level 2)
  - 10 degree x 10 degree tiles (Level 3)
  - Majority of products are less than 500MB/file
- Specialized products to meet emerging needs
  - Products in Climate Modeling Grid for climate modelers
  - Gap filled products for models
  - Near Real Time products for applications users (1.25 hours after acquisition)
    - GIS ready products and imagery and application specific formats
- Self documenting format (HDF) with appropriate metadata (easier to do upfront)
  - NetCDF used by Atmosphere community and modelers
  - Forward/backward compatibility- Standard products are in HDF 4 with HDF 5 converter
- Always have a fully reprocessed collection available for users
  - Reprocess the full mission in background
  - Give users time to evaluate the reprocessed products for their applications

MODAPS and LAADS

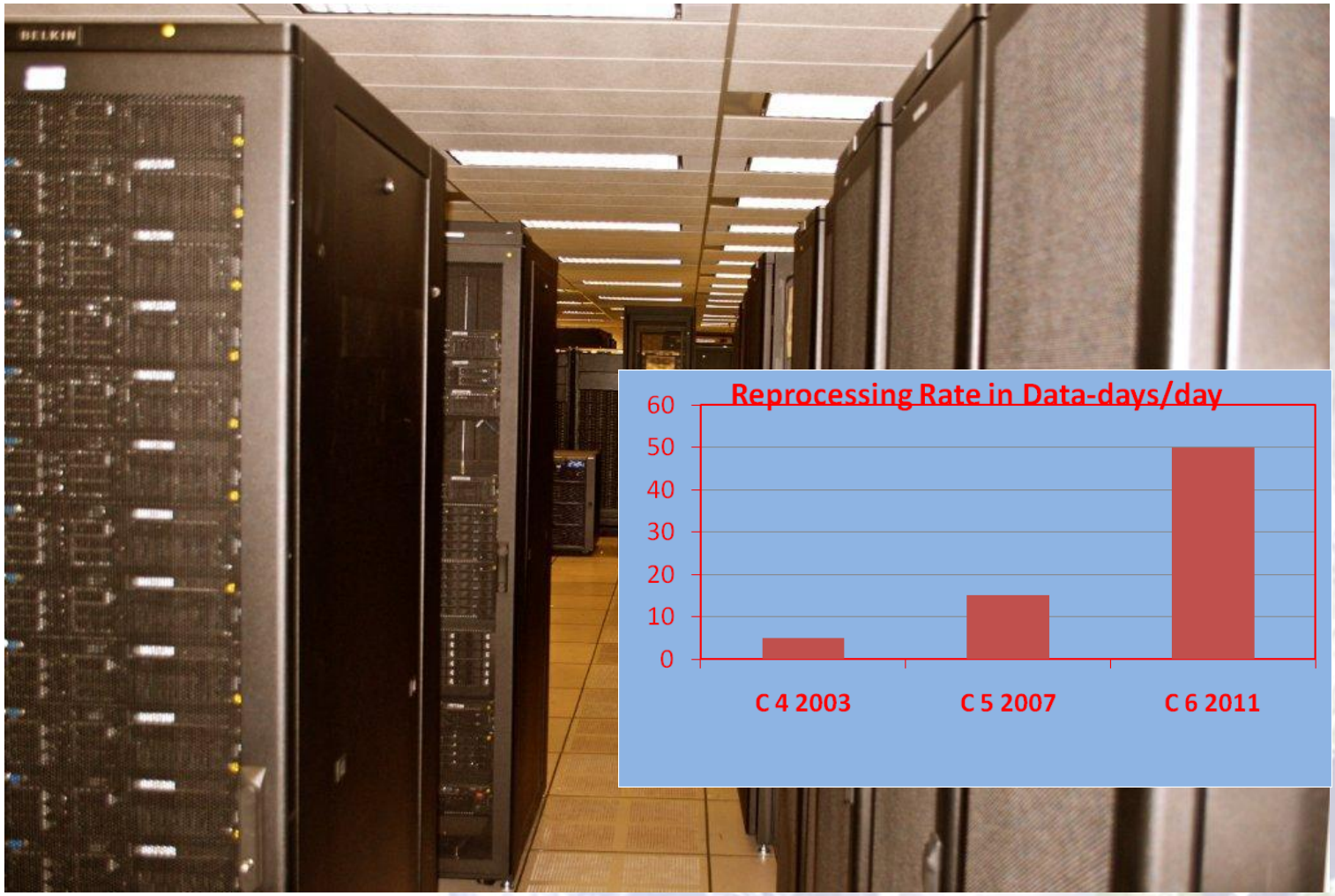
- Over 1,000 servers and 5PB of storage for MODIS, OMI and other projects.
- Middle-tier servers and disk storage
- Open-source software
- Uniform system configurations
- Automated updates daily from Depot
- Automated H/W problem reporting
- Moderate systems under Flight Projects Security Plan



MODIS Adaptive Processing System

- Scalable system for MODIS, AVHRR, VIIRS and Landsat processing
  - Designed so that processing resources can be easily moved where needed
- Built with commodity hardware
  - Easier to scale, cost savings and easier technology refresh
- Built with open source components
  - Linux, Apache, Perl, Postgres, Subversion, FUSE
- All data products are online
  - Facilitates reprocessing of Level 2 and Level 3 products
- Designed to run with limited staff
  - “Lights out” processing outside of normal business hours
  - Months of processing can be queued up and execute w/o human intervention
  - Alerts emailed to system administrators when hardware components generate warnings or fail
  - Easy to use tools for monitoring the system (Ganglia) and investigating failed jobs
- Rapid updating or provisioning of servers with science processing software, the operating system and applications (Depot and SATE)

Compute servers for global product generation

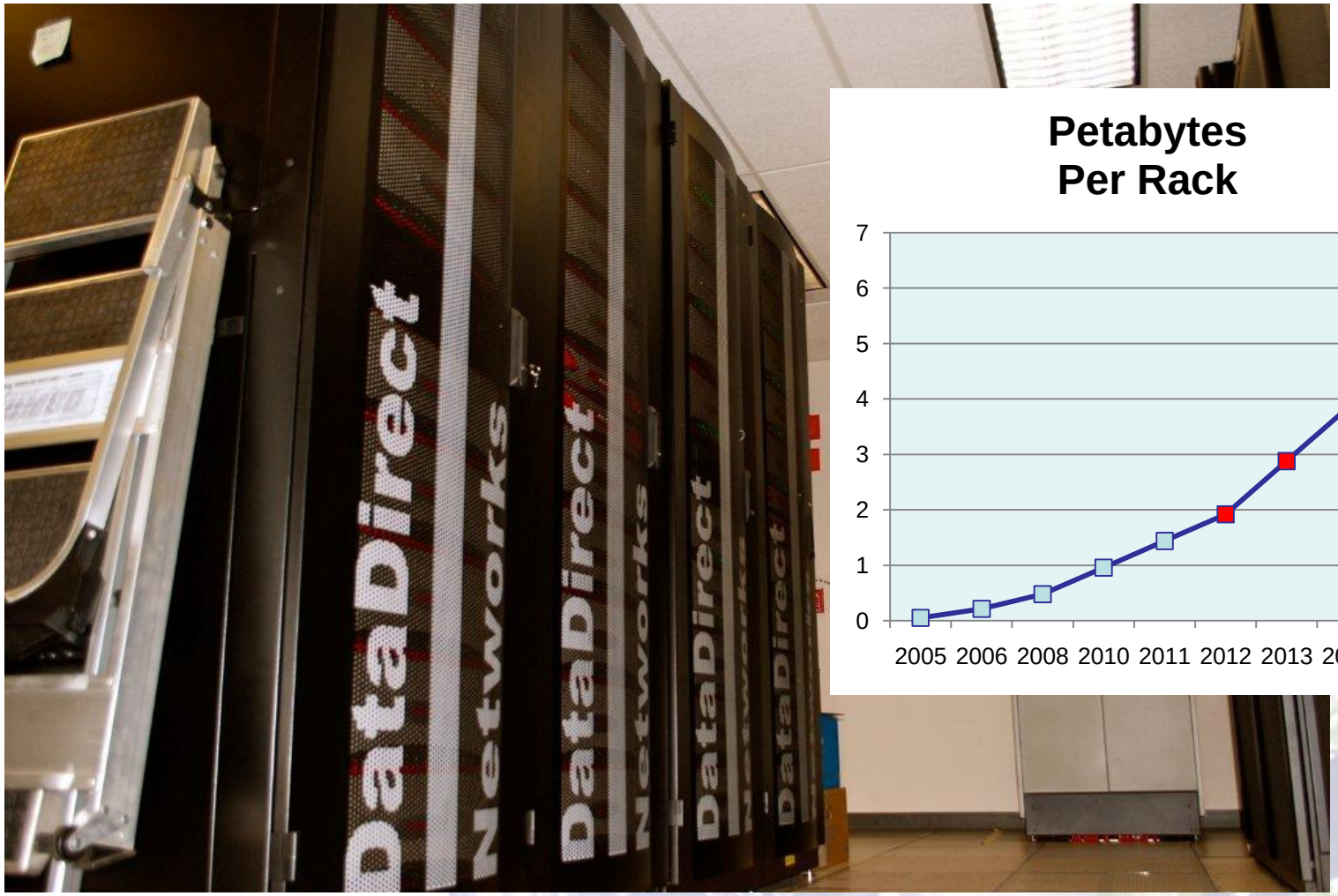


Level 1 and Atmosphere Archive and Distribution System

- Web-based search and order at http://ladsweb.nascom.nasa.gov/
- Online archive of MODIS Level 1, Atmosphere and Land products can also be accessed using ftp though a directory structure organized as:

  /allData/Reprocessing Collection #/Data Product/Year/Day
- Built upon the MODAPS framework to support post processing of products including the following operations: subset, sub-sample, mosaic, mask, parameter selection, geographic reprojection and format conversion)
- The archive also includes on-demand products which are produced when ordered by the end-user
- Web services enable machine to machine access for functions on web site
- Separating data products into archive sets to accommodate storing the large number of test results in the online archive as well as multiple reprocessing campaigns
- Other interfaces to the online archive include: an iRODS server providing access to the MODIS atmosphere products for users of the NASA Center for Climate Simulation (NCCS) and a server for EPA scientists that works in concert with visualization and analysis software running on their desktop systems to subset, sub-sample and combine MODIS atmosphere products.

Processing and Archive Storage



Concluding Thoughts

- Most innovation is not in the basic data system for processing, archiving and distribution
  - Refine and build upon mature data system software for basic functions
- Minimize movement of large data volumes in generating products
- Launch slips are usually measured in years but a small system(s) is needed for algorithm refinement, integration/testing and demonstration/evaluation of readiness are needed in the years leading up to launch
- Impact of Moore's Law on computation and storage has made many things possible and requirements grow with increased capacity/performance
- Leveraging tools, procedures and staff experience for land product quality assessment from MODIS and VIIRS allowed us to meet schedule and cost targets
- Achieved significant savings (50% or more) purchasing servers in volume of 100+ but even greater savings in manpower by adding new missions and applications within our data center (OMI, OMPS, LEDAPS, LTDR, MODIS4NACP)
- Time of science experts is the most valuable resource for making high quality products but hardest to fund over the long haul. This impacts our ability to address algorithm improvements and instrument issues.
- Immortality of key personnel is not a viable alternative to succession planning