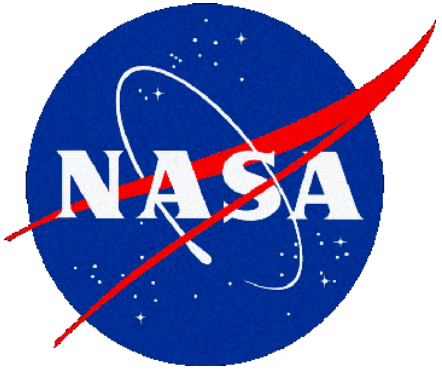


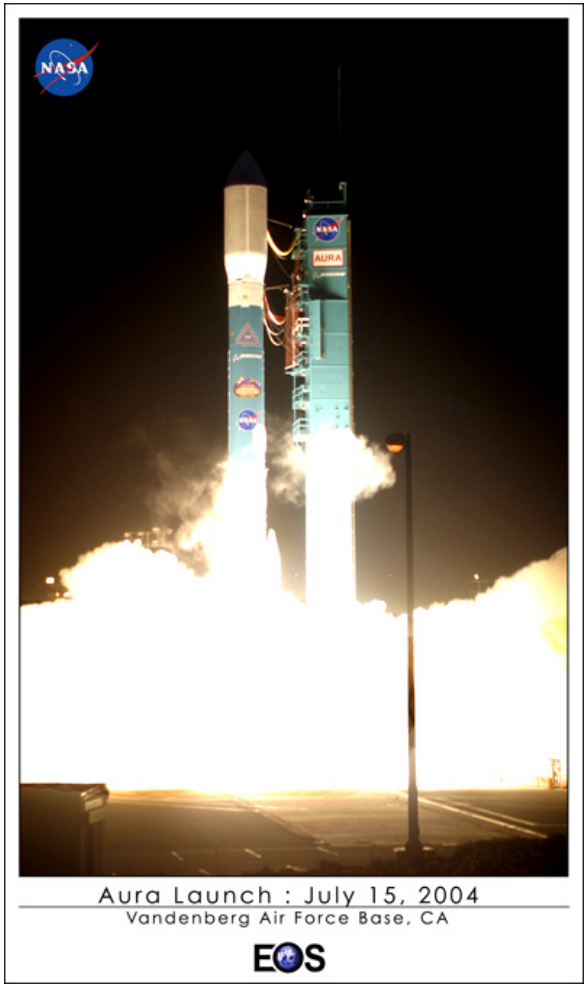
Testing the Archivas Cluster (ArC) for Ozone Monitoring Instrument (OMI) Scientific Data Storage

Curt Tilmes

NASA Goddard Space Flight Center

- Introduction
- OMI Data Processing System
- Archivas
- Testing
- Planned Architecture

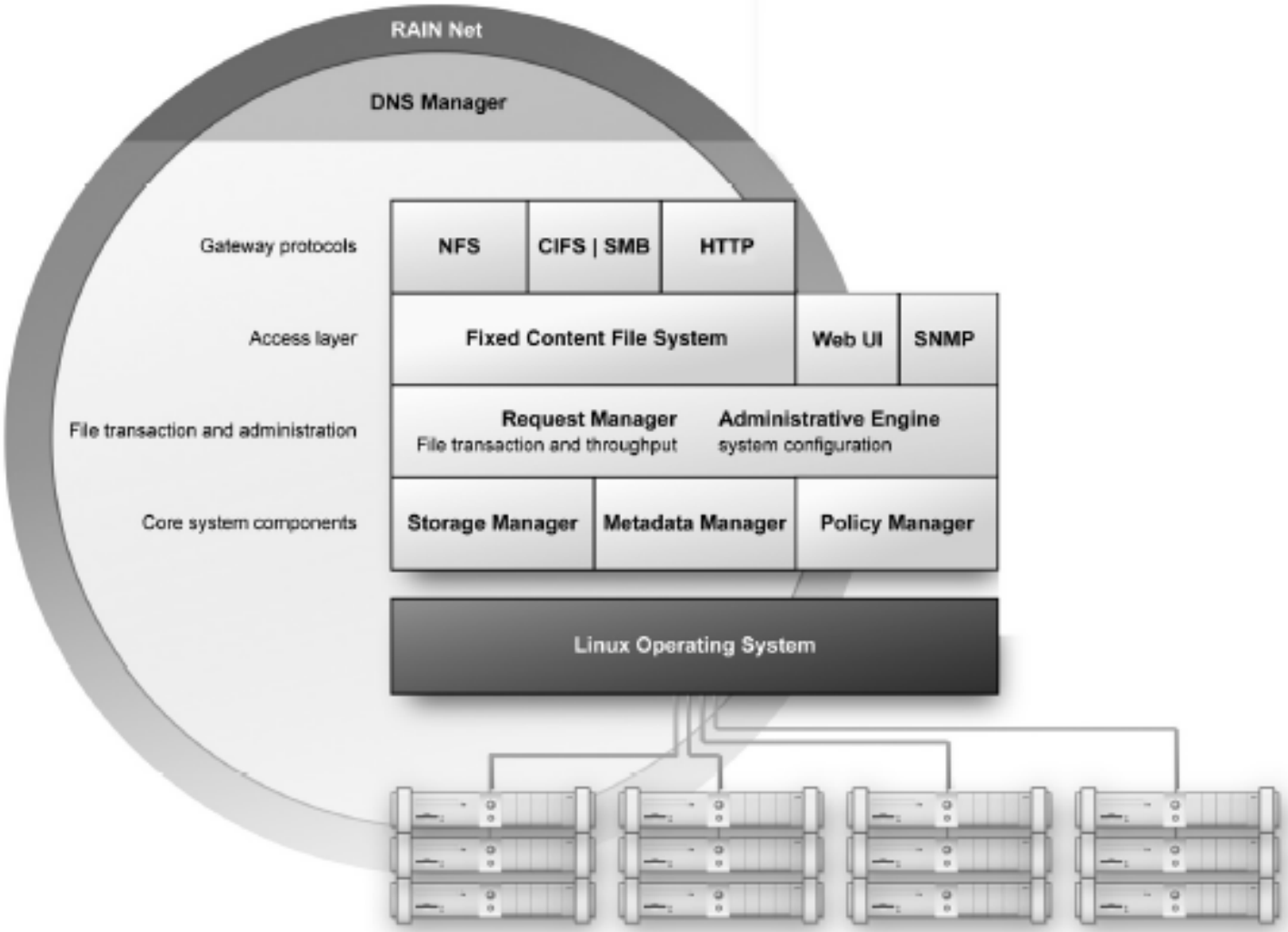
- NASA Goddard Space Flight Center
- Ozone Monitoring Instrument (OMI) on EOS Aura spacecraft
- Netherlands Agency for Aerospace Programs (NIVR) in collaboration with the Finnish Meteorological Institute (FMI) and the Royal Netherlands Meteorological Institute (KNMI) sponsored OMI construction.



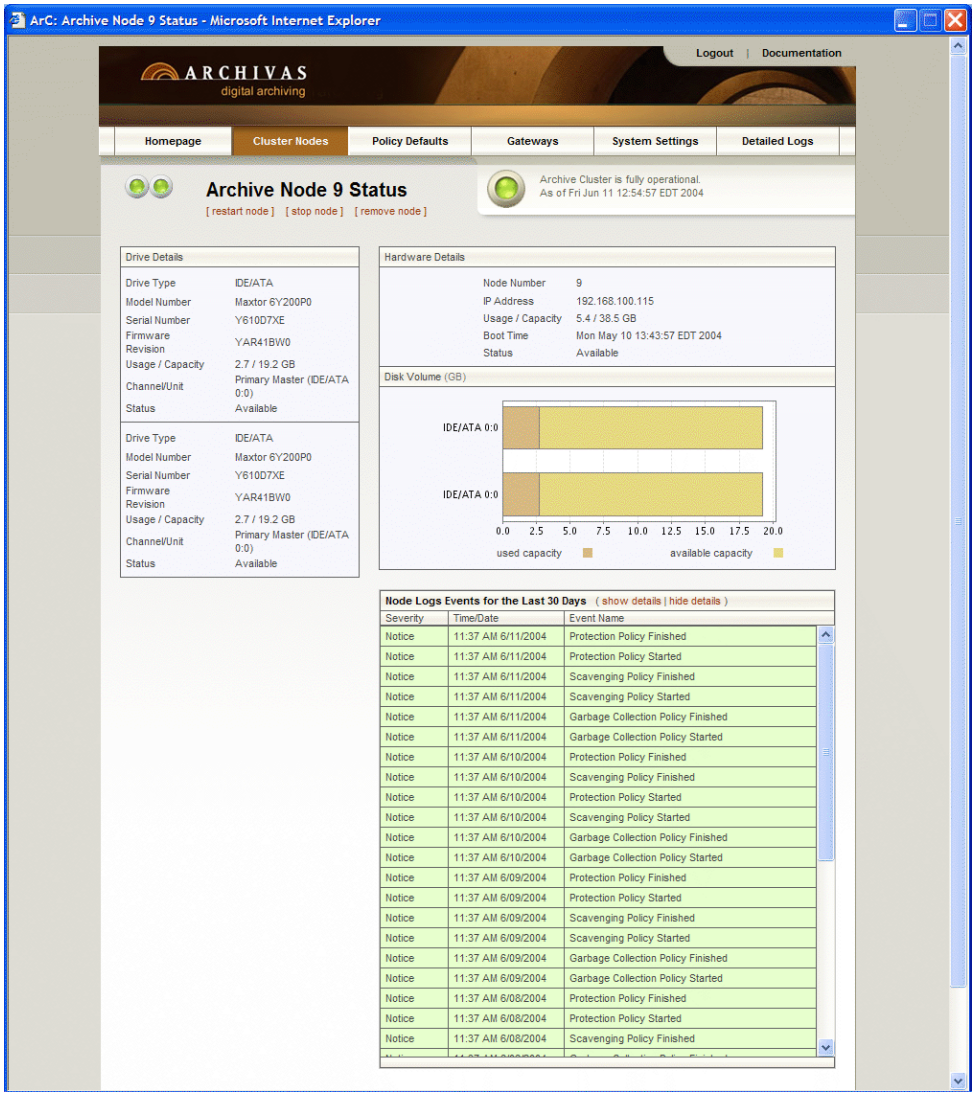
- Approximately 100GB data/day
  - Level 0 – Raw instrument data
  - Level 1 – Geolocated/Calibrated
  - Level 2 – Geophysical parameters
  - Level 3 – Gridded in Space or Time
- Level 0 data received from Goddard Earth Science Distributed Active Archive Center (GES-DAAC)
- Run Level 1 algorithm to geolocate/calibrate
- Run a series for Level 2 algorithms to detect geophysical parameters and produce products for total column ozone, aerosols, NO2, SO2, formaldehyde, ozone profile, etc.
- Plan to produce Level 3, gridded, daily products later this year
- All data sent to the GES-DAAC for long term Archive and Distribution to the public

- Archivas, Inc. – see booth in exhibit hall
- Archivas Cluster (ArC) is a software product for data storage using a cluster of commodity, low cost Intel Linux hardware with local storage
- ArC provides a scalable global filesystem (a single namespace across the entire storage cluster) with high availability, high reliability with easy administration
- Development funded in part through a NASA Small Business Innovative Research (SBIR) grant
- The OMI project became an internal NASA partner to ensure the product will ultimately meet NASA needs
- Archivas presents an alternative to existing data servers

Archivas Internal Architecture



Archivas Cluster – Administrative Console



- SBIR Phase I : IBM cluster of 6 nodes with 5TB of data, using 100BaseT
- Focused mainly on exercising functionality rather than performance.
- Used the HTTP gateway to push data from the OMI (first simulated data, then real data)
- Several new releases fixed bugs and expanded functionality
- Filled the 5TB several times
- Stored over 12 million files
- Retrieved/verified over 2 million files
- Unplanned disk failure occurred, and was handled gracefully by the system

- SBIR Phase II : Dell cluster with 36TB of data, using Gigabit Ethernet
- The OMI production system will push its data products to ArC using HTTP
- We will allow scientists direct, read-only NFS access to the Arc filesystem, allowing them to easily run ad-hoc analyses across the entire OMI data set on a cluster of client workstations
- Storage cluster easily expanded by adding more nodes

Planned Architecture