I. INTRODUCTION

The Moderate Imaging Spectroradiometer (MODIS) is a 36-band instrument (.46nm to 14.nm) with spatial resolutions from 250 meters to 1 kilometer. The first MODIS instrument was launched on the EOS Terra spacecraft in December of 1999. The second is scheduled to launch on the EOS Aqua spacecraft in late 2001. Each MODIS instrument will produce 70GB of raw data per day from which 390GB of calibrated and earth-located radiance products (Level 1 products) and 450GB of higher-level science products will be archived and distributed to the public. During the EOS Terra and Aqua missions MODIS data production will ramp-up from today’s 710GB/day to 7,790GB/day in 2003.

MODIS science products are produced by the MODIS Adaptive Processing System (MODAPS) from calibrated radiance and earth-locations produced at the Goddard Distributed Active Archive Center (DAAC). The products are shipped to the MODIS Science Team for quality assurance and to DAACs for distribution to the public as shown in Figure 1.

On February 24, 2000 the first MODIS image of the Earth was acquired and processed. The emphasis was on getting all MODIS science products released to the public as soon as a satisfactory level of quality was achieved. After 13 months and over 250 changes to the science software, all MODIS products are being distributed to the public. The at-launch MODIS processing system, MODAPS V1, was used to produce and distribute 250GB/day of MODIS products through February 2001. Lessons learned from developing and operating the V1 system were addressed in the design of a follow-on processing system, MODAPS V2.

II. MODAPS V1: SYSTEM ARCHITECTURE AND LESSONS LEARNED FROM PROCESSING

The MODAPS V1 system produced archived and distributed MODIS products on an 80 processor Silicon Graphics Origin 2000 system with 40GB of memory, a single 10TB filesystem and 36TB of near-line storage in 3 tape libraries as shown in Figure 2.

Much of the early MODAPS development effort was focused on getting all MODIS PGEs running correctly in the production system and completing the functionality required to support data orders, production monitoring and debugging. Using a large computer with a single filesystem for processing, archiving and distribution simplified software development [1,2] as did adapting the job scheduler developed for the SeaWiFS data processing system[3] to

FIG. 2 MODAPS V1 architecture

Much of the early MODAPS development effort was focused on getting all MODIS PGEs running correctly in the production system and completing the functionality required to support data orders, production monitoring and debugging. Using a large computer with a single filesystem for processing, archiving and distribution simplified software development [1,2] as did adapting the job scheduler developed for the SeaWiFS data processing system[3] to
schedule MODIS PGEs. Loader tasks were run by the MODAPS operators to stage the required input files for production and to submit a processing recipe instance of related PGEs into the system for the time period specified as arguments to the loader. The SeaWiFS scheduler ran PGEs in separate production directories and copied the output products to archive directories for subsequent distribution to users. Using a single filesystem eliminated the I/O required to move files out of the production directories and into archive directories. When required inputs weren’t available to run a recipe in V1 for a given time period (5 minute scene) or spatial 100x100km tile) the recipe would leave gaps in production corresponding to the missing inputs. In V1 when there were missing products, it was difficult to fill in just the missing scenes or tiles since recipes would run for a minimum time period of an hour or so. Rather than create custom jobs by hand for each missing scenes or tiles, it was often simplest to rerun the entire period and let the system overwrite the previously produced scenes or tiles.

Several performance bottlenecks in the V1 system limited the volume of global MODIS products that could be produced, archived and distributed in a day. First, the single filesystem limited the number of simultaneous streams that could do I/O to the disk and the throughput to the disk was such that the system rarely achieved better than 50% cpu utilization. Second, the V1 database used a single table to hold all metadata about each product. Improving queries that used this table helped reduce deadlocks but the single table significantly slowed overall processing and data distribution. Third, the nearline Ampex tape libraries held only 12TB each necessitating tape transfers between the libraries and offline tape racks, which significantly increased the time to retrieve data files. Fourth, the Legato Networker software, which handled I/O to the tape libraries, would report bad savesets in the Ampex libraries which prevented file retrieves and necessitated reprocessing these “lost” files. Finally, a FDDI network connection used to distribute data product orders to the MODIS science team was saturated by the product volume shipped on a daily basis.

III. MODAPS V2: SYSTEM ARCHITECTURE AND PROCESSING PERFORMANCE

In the V2 MODAPS system, which took over processing in March 2001, all of the performance bottlenecks identified above have been addressed through changes in the system architecture and hardware components as shown in Figure 3. In addition to improving the overall throughput of the Origin 2000 through V2 software changes, the SGI Origin 3000 was installed to handle production of the Aqua MODIS products and reprocessing of the Terra products.

The changes implemented in V2 that were related to hardware, such as disk, tape or network interfaces, on the production system have yielded immediate and significant improvements in system performance and operability. The single 10TB file system has been replaced by 44 individual 1TB filesystems, which removed the I/O bottleneck to disk resulting better than 90% cpu utilization during Level 2 (scene based) product generation. The additional disk storage has also enabled the MODAPS system to hold more products online which significantly reduces the number of file retrieves from the tape libraries. Replacing the Ampex DST 812 tape libraries with ADIC Scalar 1000 libraries increased nearline tape storage from 36TB to 100TB and the number of tape drives from 9 to 12. The increase to 100TB of storage has eliminated the need for operators to look for tapes in offline storage both speeding up production and simplifying the operator’s task. With an individual AIT-2 tape drive costing $6,000 versus $100,000 for the Ampex drives, it also became affordable to add additional tape drives up to the Legato Networker limit of 32 drives per host. Finally, the FDDI network connection used to ship products to the science team has been replaced with a Gigabit Ethernet connection to a high speed science network, which has significantly improved performance of the ftp transfers to the science team computing facilities.

The MODAPS development team also redesigned the graphical user interfaces (GUIs) that operators use to control production, redesigned database tables to reduce deadlocks and improve query performance and replaced the SeaWiFS job scheduler with a scheduler better suited to handle complex MODIS production and one which would support running without operations support on the graveyard shift from midnight to 8am. As one might expect after such an extensive redesign, some parts of the system are working well while others require tuning or redesign to function effectively.
under a processing load of 4,000 jobs and a distribution volume of 16,000 files/day. Some positive changes were:
1) Improved I/O performance resulting from the division of a single large filesystem into small 1TB filesystems.
2) A job planner that holds back jobs until their inputs are available thus eliminating holes in production.
3) Better GUIs for operators to monitor jobs running in the system.

Problems observed included:
1) Production rules for some PGEs which ran too slowly or staged the wrong files for processing runs,
2) Database queries which were inefficient caused deadlocks or temporarily blocked many other queries, and
3) Difficulties producing or exporting products due to database slowdown arising from inefficient queries.

At present the development team and the database administrator are examining queries in portions of the software which are running much more slowly than was observed during early testing with relatively empty database tables. Most of the severe performance problems result from queries that give rise to row by row scans of large tables. Problems have been fixed by indexing the tables or rewriting queries to eliminate table scans. The same type of database query changes were required when V1 MODAPS was first placed into production and reflect the need to examine database performance under realistic processing and distribution loads as early as possible.

IV. MODAPS V3: SCALABLE SERVERS

MODIS processing will continue through 2008 to finish a full reprocessing of all MODIS data after the Aqua mission ends. MODAPS is also being generalized to serve as a production system for the Ozone Monitoring Instrument (OMI), which is scheduled to launch in 2003 on the EOS Aura spacecraft, and has been proposed as a production system for upcoming earth remote sensing instruments. Based on our experience so far with Linux processors and inexpensive disk storage we envision a production system based upon processing, archive, database and storage servers linked by high speed network connections as shown in Figure 5.

In May 2001, 40 dual Pentium 3 processor systems each with 72-200GB of local disk will be connected to the Origin 3000 through fast network links as shown in Figure 3.

Testing to date, MODIS PGEs ran 3x faster on the 850Mhz Pentium processors than on the 250Mhz MIPS R10000 processors in the Origin 2000. In the Origin 2000, the cost per processor including memory and disk storage is $20,000 versus $3,000 per processor for the Linux systems. Given that the V2 system supports processing on the network attached Linux boxes and many of the MODIS Level 2 PGEs are already ported to them, we expect to encounter little difficulty in adding them to the MODAPS as an inexpensive way to significantly increase processing throughput. This will be needed to achieve the science team’s goal for both MODIS instruments of reprocessing data at the rate quadruple real time plus keeping up with current production.

Fig. 5 Scalable processing servers in V3 MODAPS

In the above system, bandwidth between any production server and a storage server is 2 Gbps with distribution bandwidth from any storage server of up to 2Gbps through 2 Gigabit routers connected to a high-speed science network. At current prices the production servers cost $3,000 per processor ($200,000 for a full rack of 64 processors), a 4 processor database server costs $25,000 and a storage server with 9TB of RAID disk costs $160,000 or approximately $18,000 per terabyte of storage. For systems that need to
archive large volumes of data, tape libraries like the ADIC Scalar 10000, which can hold up to two petabytes of compressed data on AIT-3 tapes, can be attached to the storage servers via Fibre Channel. Advantages of the above architecture are that it is highly scalable, low cost and individual components are easily upgraded or replaced while the rest of the system continues to produce, archive and distribute products.

Some software changes will be needed to support processing on a distributed collection of production, storage and archive servers though much of the functionality required to do so was implemented in V2 with support for Linux processors and multiple filesystems.

V. CONCLUSIONS

The MODAPS will continue to evolve to achieve higher processing rates over the life of the MODIS missions and to support the production of data products for other Earth observing instruments. The V2 delivery provided enhanced functionality, operability and improved performance over that of V1 and will serve as a foundation for future development. Components of the V3 architecture will be added to the V2 system to reduce costs for storage and processing. However, the decision on whether to build the full server based distributed production system will depend on the requirements of the instrument teams supported by MODAPS, available funds, the development team’s enthusiasm for adding new functions to support the cluster of servers and the relative priority of this change versus preparing an open sourced version of MODAPS.

REFERENCES


[3] Information about the SeaWiFS processing system is available at: http://seawifs.gsfc.nasa.gov/