

Kristina Sprietzer* and Paul Haggerty
Science and Technology Corporation, Suitland, Maryland

Gene Legg
National Oceanic and Atmospheric Administration, Suitland, Maryland

1. INTRODUCTION

The Near Real Time MODIS processing system was conceived as a proof of concept program, and implemented under the Computer Sciences Corporation (CSC) Central Satellite Data Processing (CSDP) contract. It was designed to provide NOAA with rapid access to environmental data from the TERRA and AQUA satellites, and to provide risk reduction for future high data rate satellite systems (NPP and NPOESS) by identifying the risks and problems associated with high data volumes and Near Real Time (NRT) operations. This project began in 2000, and serves products to NOAA, NASA and military customers for production use. This poster will discuss the challenges in processing voluminous data sets, designing ftp servers to distribute data, and complex system networks.

2. Processing data

The NOAA MODIS NRT project has discovered that processing large data sets does not solely rely on the number of, or the processing rate of CPUs, but also on a machine's architecture. The NOAA MODIS NRT project was able to cut processing time in half by changing the architecture of the shared drives of the SGI machines. This change in architecture increased the capacity of the internal network in the server and reconfigured the drive caches. This change resulted in a decrease in I/O wait and decreased processing time. This I/O wait occurs not only when converting raw data to products, but also during FTPing products.

3. Designing FTP servers

The solution to freeing up I/O wait during FTPing of products was to push the data once to an FTP server rather than using multiple CPUs to push the same data to multiple locations. This freed processing time on the production server to convert raw data to products more efficiently. The FTP server is currently set up with a limited number of users to manually pull data. Current investigations involve creating automated systems to

push data to customers from this server and test the capacity of the FTP server network.

4. System Networks

Since the team is still developing methods for sending out these voluminous data sets without taking down the production server or FTP server, there is only limited access to the FTP server. The biggest problem of shipping this data are networks that are internal and external to organizations. Even though there may be dedicated lines that can accept data to organizations, they are still dependent on the network internal to that organization. These dependencies are: bandwidth being used by other users in the organization, the amount of bandwidth behind the router of the dedicated line connecting directly to the receiving server, and the I/O wait on the receiving machine. These factors can all affect the performance rate at which the data is received by the customer.

5. Conclusions

The solution to processing and transporting voluminous data sets lies in working with a team of individuals which implement more efficient processing servers with large bandwidth networks for both the sender and receiver of the data. This project is considered to be risk reduction for NPP and NPOESS, since the knowledge learned at processing these large volume data sets can be applied to handling the even larger expected data files from NPP and NPOESS.

* Corresponding author address: Kristina Sprietzer,
STC, NOAA/NESDIS/IPD Federal Center FB-4, 4401
Suitland Road, Suitland, MD 20746;
e-mail: Kristina.Sprietzer@noaa.gov