

CERES Production TigerTeam

Membership

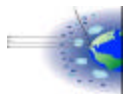
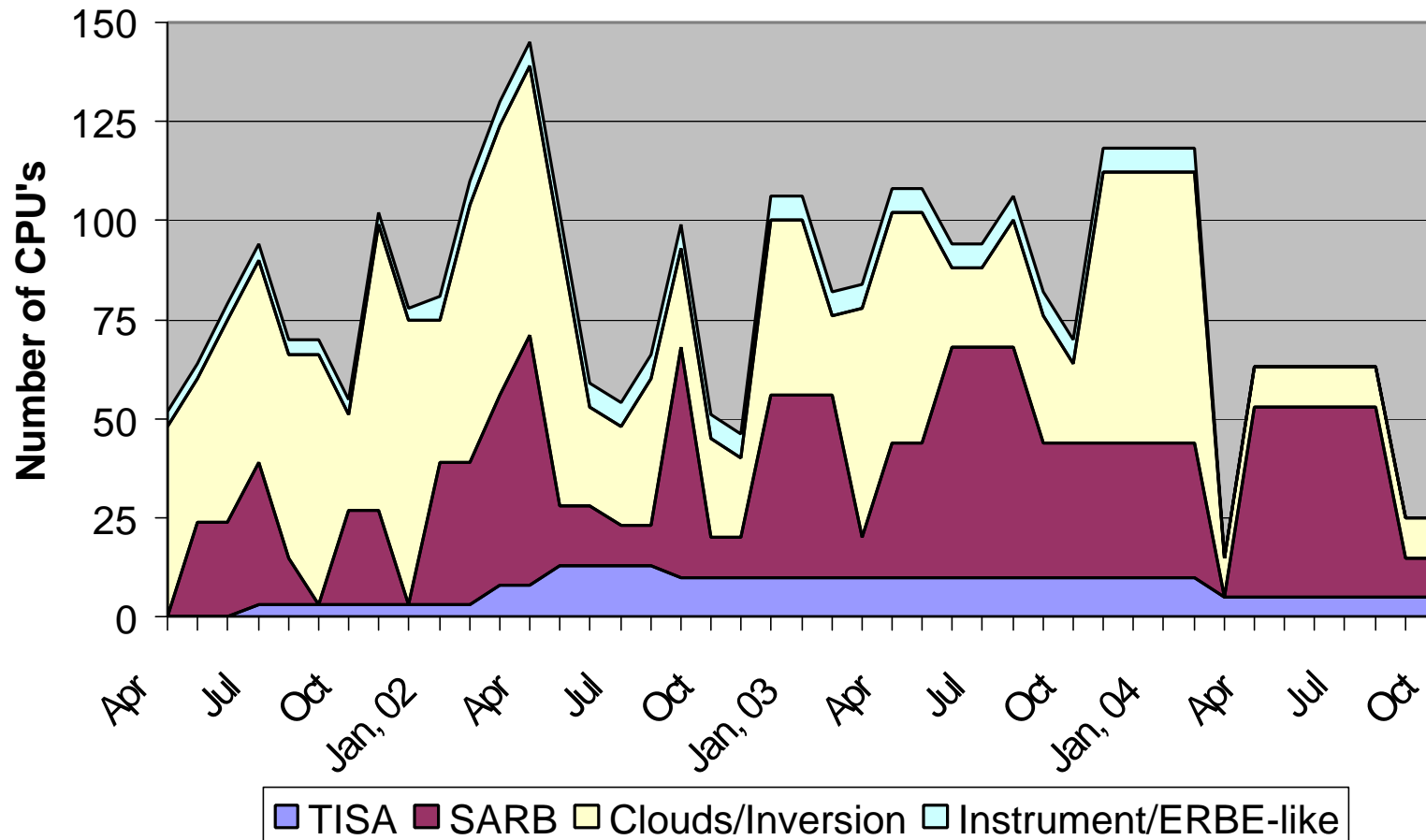
Henry Flippo	CERES SCF lead
Paul Garn	SGI SSE
Chris Harris	ASDC LaTIS Lead
Vertley Hopson	ASDC SSI&T
Calvin Mackey	Facilitator/Chair
Walt Miller	CERES DMT
Sue Sorlie	ASDC LaTIS production
Tom Regan	ASDC APGS
William Seufzer	ASDC System Engineering



Calvin Mackey
Langley ASDC
757 864 7464

c.e.mackey@larc.nasa.gov

CERES Processor Usage (summary of BAW's 35 job streams)



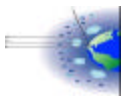
Charter: Improve the production and archival of CERES SSFs by factor of six. We will get a factor of 4 from 2x2 sub-sampling. So, If we can get a cumulative 50% increase from production, we're there.

Due Date: MODIS SSF Edition1 production goal is September 2001

Scope: End-to-end SSF production.

Assumptions

- **PGE restructure: currently combined Clouds/Convolution. Can we gain with parallel PGE's?**
- **SSF & Output staging: more than 30GB/day. Produce 5-minute CloudVIS on demand for diagnostics?**
- **MODIS ingest: more than 100GB/day. If we subset, change interface with GSFC DAAC to reduce volume?**

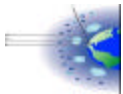


ARCHIVE GROWTH

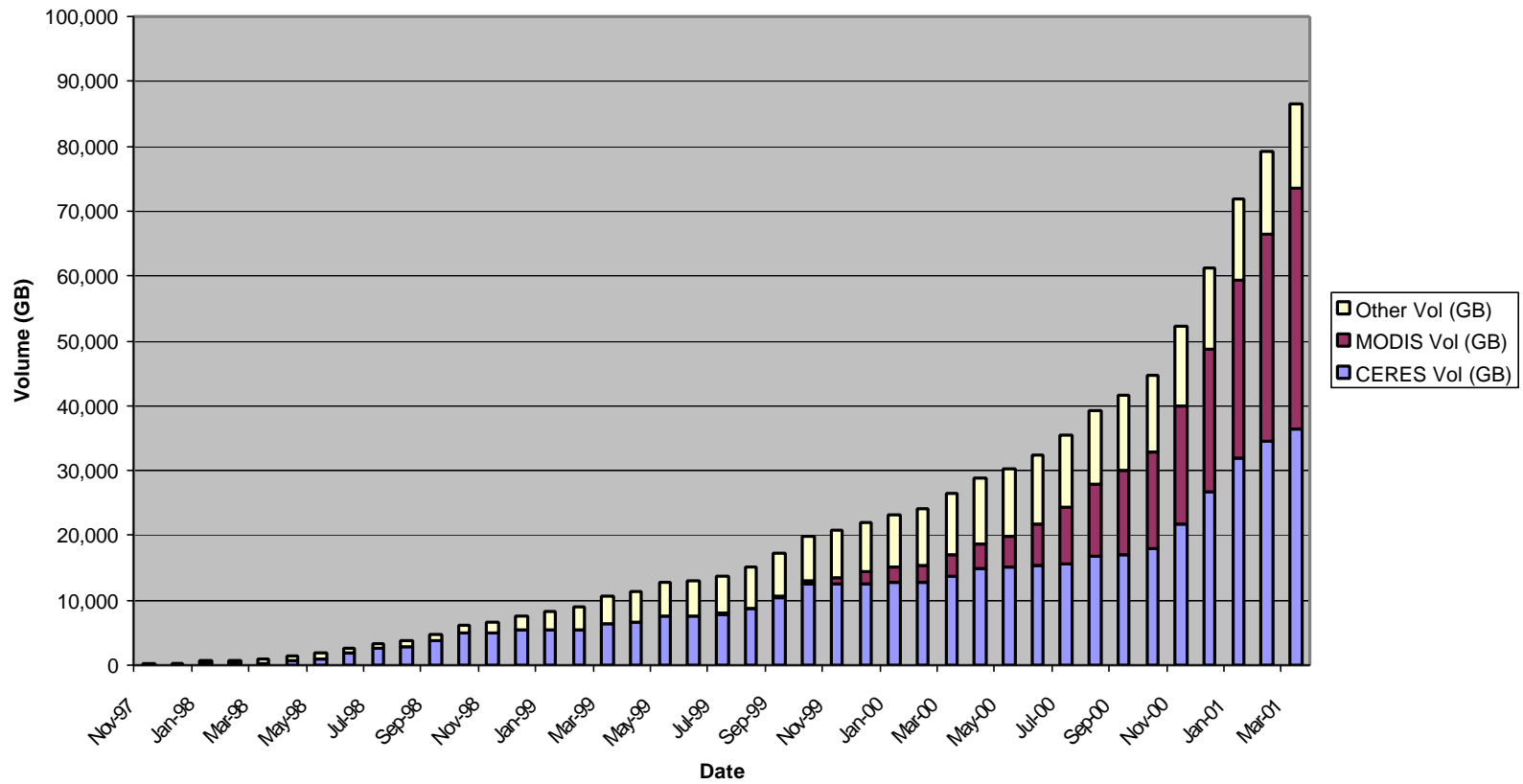
- **Currently (End of March 2001):**
 - **Archive volume: 86 TB**
 - **Fill rate: 273 GB/Day**
 - **Assuming 3:1 mix of 20 GB:60 GB tapes: silo full in 300 days**

- **With MODIS divided by 4:**
 - **Archive volume 60 TB**
 - **Fill rate: 170 GB/day**
 - **with same tape size ratio: silo full in 2 years**
 - **with doubling of tape capacity: silo full in 4 years**
 - **with removal of "deep archive" (duplicate): silo full in 8 years**

- **Current stats on CERES: (average daily volume from March 2000 to March 2001)**
 - **41.5 GB/day archived by CERES production**
 - **6.1 GB/day read by CERES production**
 - **50 GB/day read by Atmospheric Sciences (local servers et al.)**
 - **25.6 GB/day read by ALL others**



Archive Volume

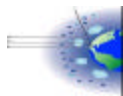
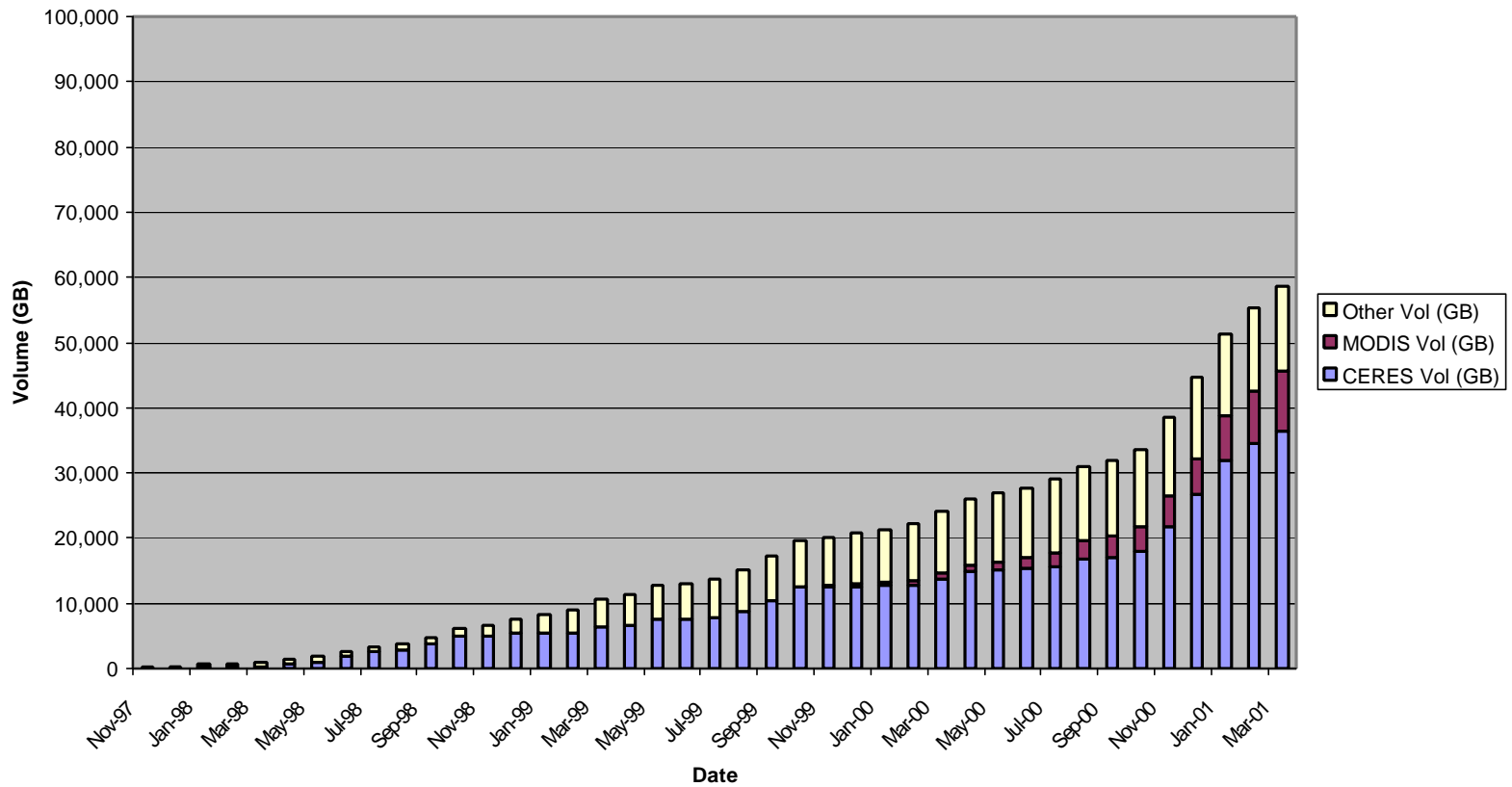


May 01, 2001

Calvin Mackey Langley ASDC



Archive Volume (MOD/4)



May 01, 2001

Calvin Mackey Langley ASDC



ARCHIVE PERFORMANCE

Archive Results

1) Volume distribution

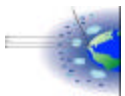
Finding: 90% of LaTIS files are of size 5 MB or less. Total volume of these files is less than 2 TB.

Recommendation: Place small files on disk cache.

2) Archive activity

Finding: Small file reads from the archive dominate total of read requests. Large file reads from the archive dominate total volume. Archive system has been tuned based on small file performance.

Recommendation: If Recommendation(1) is followed, optimize archive system performance for large files.



ARCHIVE PERFORMANCE

3) Daily volume

FINDING: Total daily volume of read and write activity has grown from ca. 100 GB/day to nearly 500 GB/day. The maximum throughput of the system to date is around 500 GB/day.

Recommendation: Add more tape drives to the system.

4) Data Transfer - daac_get command (LaRC campus ONLY)

FINDING: daac_get is the primary means of access for Atmospheric Sciences. This process ALWAYS pulls the file(s) directly from the archive.

Recommendation: Create a daac_get disk cache similar to the media and ftp caches.



CLOUD PGE OPTIMIZATION

Software Results

PROBLEM:

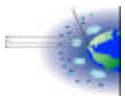
- Software processing time for one data hour (two instruments) averaged 30 hours (19-39)
- Data staging time was 16 hours

LIMITATIONS:

- Clear Sky Processing has required finishing one day before the next day can start (24 CPU limit)

IMPLEMENTED:

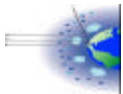
- Interface change with MODIS for 2x2 sampled data
- Reduces processing time by 75%
- Reduces data staging and storage requirements



CLOUD PGE OPTIMIZATION

STUDIED:

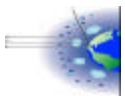
- **Modifying Clear Sky cycle to two days**
 - **Gain only an additional 24 CPU**
 - **Additional complications to ASDC operations**
- **Modify code to work in parallel**
 - **Gain use of CPU without modifying Clear Sky**
Scalable to available CPUs
 - **Compiler automated option does not perform correctly**
 - **Investigate compiler options to improve speed**
Optimization level 3 did not improve performance, but results are real now
 - **Fast Math library didn't produce noticeable improvement**
- **Samantha system tuning – no major bottlenecks found**



SUMMARY/PLANS for near-term action.

Archive and Processing improvements

- **Network upgrade to GIGE complete to achieve faster data flow between servers**
- **Subset MODIS input**
- **Optimize or restructure Cloud algorithm**
- **Network upgrade to GIGE complete to achieve faster data flow between servers**
- **Move small files from the archive to disk cache in June to free up drives for large files**
- **Configure additional Science Processing server (56P 400Mhz) in June to allow for independent reprocessing**
- **Migrate existing archive data to new media in June**



May 01, 2001

Calvin Mackey Langley ASDC

