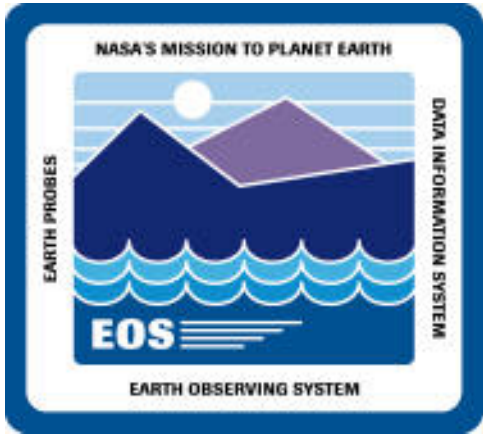


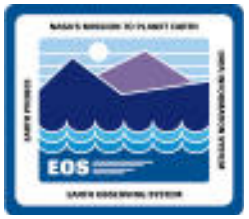
Science Data Processing Sizing

Eric Dodge

eric@eos.hitc.com

31 October 1995

Overview



Science Data Processing sizing / capacity analysis

- Will cover Science Processor capacity sizing: standard & SSI&T only
  - Reprocessing discussion follows this presentation
- Topics for Science Processing:
  - Design and sizing drivers
  - Process overview
  - Status for IDR and sample numerical results
  - The next steps for the CDR time frame

Scenario Context

- Applies to all Processing “Push” scenarios

Design & Sizing Drivers

Science Processing H/W Class & Capacity Based on August 1995 ECS Technical Baseline Includes Core Requirements as well as...

- **AHWGP ("Push") baseline**
  - **Input/Output File Descriptions, PGE activations & processing needs**
  - **Production dependencies**
- **User Modeling ("Pull") baseline (e.g., on demand processing)**
- **Processing / Site Allocations (e.g. MISR, MOPITT, CERES at LaRC, ASTER & MODIS land at EDC, etc.)**

Operational Constraints and Needs at the DAAC Sites

- **Operations hours are "24hours x 7days-a-week" for EDC, GSFC and LaRC**
- **Reduced operations hours at: JPL - "8 x 5", MSFC and NSIDC - "8 x 7"**
- **Transition from SSI&T only at GSFC and EDC, and at LaRC for AM-1 SSI&T and TRMM support to AM-1 operations**
  - **Migration to operations**
  - **Parallel support requirements**

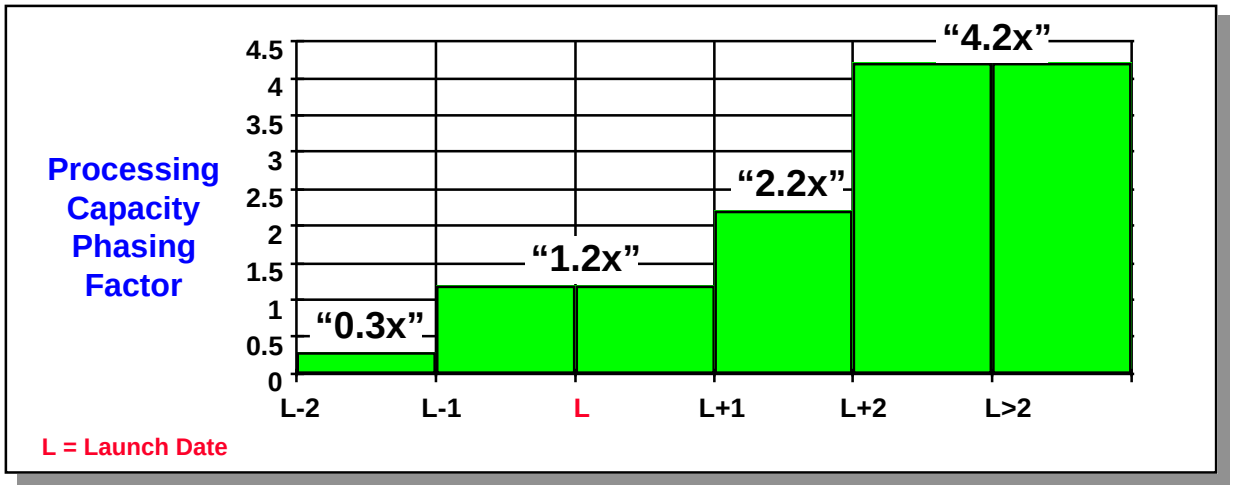
Current Technology and Technology Projections

Design Drivers
(Continued)

Separation of SSI&T Environment from Operations Environment

- Algorithm development, integration and maintenance
- Provision for backup and failsoft processing requirements

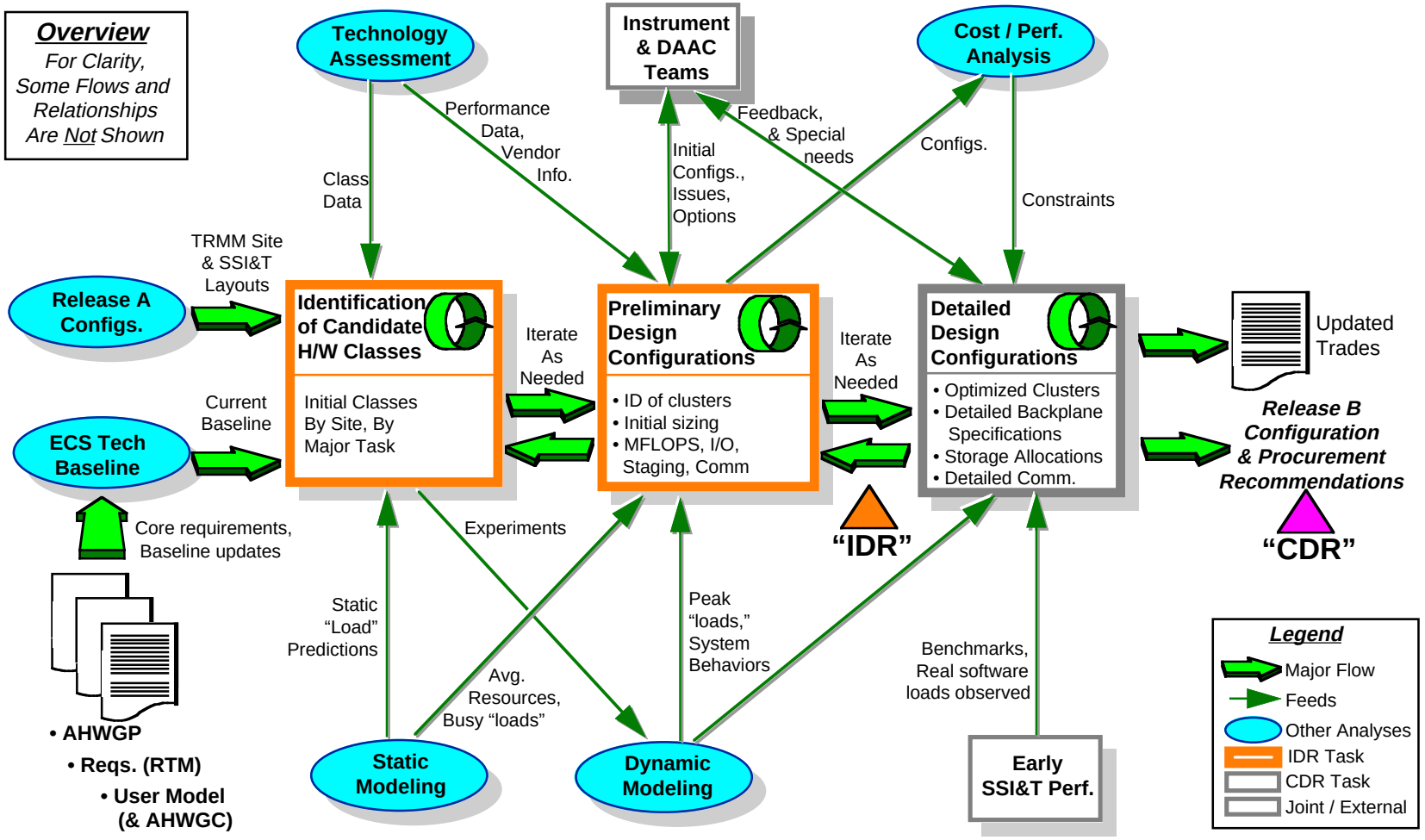
ESDIS Provided Capacity Phasing Factors



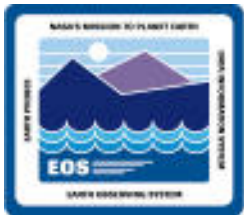
Sizing Analysis Addresses Full Mission Needs

- Scalability & Evolution
- IDR sizing figures for AM-1 based on epoch "k" (3rd Q 1999)

Process Overview ("How")



Status for IDR



Analysis Uses Release A CDR Configurations as a Foundation

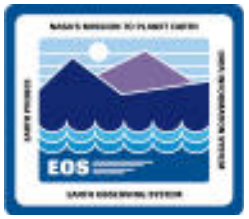
Results for Release B IDR:

- Recommended classes of processing hardware (CPU, I/O, Staging, Comm.)
  - Classes identified for IDR (Details: see DAAC Specific Volumes -- DID 305)

- Static and Dynamic analysis of Push requirements baseline used
  - Static (Average & Busy Day cases) for all sites
    (for IDR: average CPU, host I/O, staging/destaging I/O, network I/O utilization)
  - Initial dynamic model outputs factored into configuration analysis
    (for IDR: peak & average staging disk utilization, average CPU utilization)

- Initial On Demand Processing loads

IDR Results Sample For LaRC

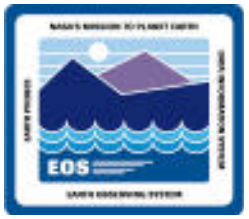


LaRC DAAC ***("1x" Processing)***		**Required Phased Capacity**			**Provided Capacity**				
Release	**Activity**	**MFLOPs**	**I/O**	**Disk Volume**	**Platform Class**	**MFLOPs**	**I/O Bandwidth**	**Disk Volume**	**RAM**
Ir1	Early AI&T	1,100	25 MB/sec	30 GB	SMP (4 CPU)	1,200	320 MB/sec	68.8 GB	.5 GB
A	AM-1 AI&T	6,148	4 MB/sec	104 GB	SMP (18 CPU)	6,480	320 MB/sec	104 GB	.5 GB
B	CERES AM	6,836	13.7 MB/sec	407.8 GB	SMP (14 CPU)	10,080	320 MB/sec	408 GB	(CDR)
	MISR	12,443	28.7 MB/sec	209.4 GB	SMP (22 CPU)	15,840	320 MB/sec	210 GB	(CDR)
	MOPITT	26	0.6 MB/sec	18.3 GB	Uni-Proc WS	125	100 MB/sec	19 GB	(CDR)
	On Demand	1,260	25 MB/sec	105 GB	SMP (2 CPU)	1,440	320 MB/sec	105 GB	(CDR)

TABLE NOTES

- Capacity figures do not show TRMM operations capabilities.
- "Required" MFLOPS based on averages (Dynamic model), "provided" MFLOPS based on peaks, as configured (interim results)
- Release B capacities shown above do not include reprocessing sizing (1x at epoch "k"), and AI&T and AM-1 instrument capacities are cumulative.
- Early AI&T hardware delivered as part of Ir1 capabilities to DAACs (Ir1 host transitions to TRMM operations support at Release A).
- AM-1 AI&T capacity delivered as part of the Release A hardware procurement.
- MOPITT processor class is uni-processor based on derived requirements, but will likely be combined and run on excess SMP capacity (will be revisited for CDR).
- On Demand processing derived from User Model and Dynamic analysis (e.g. "pull" side / user subsetting on demand).

Next Steps / Plan for CDR



More detailed Dynamic Model runs and subsequent analysis

- Factor in refined AHWGP data as required

Complete and Detailed Physical Topologies for the Processing Clusters

- Optimized communications topology (high speed comm. subnetworks)
  - Inter-cluster/Inter-machine: Physical HiPPI clustered resources
  - Communications paths to key external resources (e.g. DS FSMS host, working storage)
- Optimized processing and I/O
  - Intra-machine: CPU to chassis allocations, I/O subsystem allocations, comm. paths
- Staging disk allocations and any required cross-strapping
- RAM requirements through joint work with instrument teams
- Incorporate results from observed loads (SSI&T, early software releases)

Incorporate WAN Traffic Reduction Methods for Production Side

- Subsetting and compression/decompression methods for large AM-1 flows