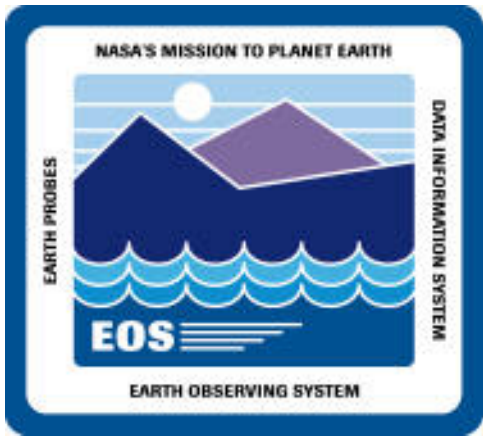


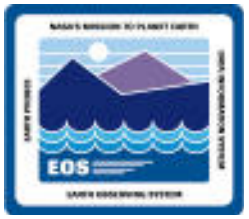
Ingest Software Design Review

LiLing Chao

lchao@eos.hitc.com

15 April 1996

Ingest Software Design Features



Based on the “Object Factory” concept

Utilize ECS provided framework and common services

- **Consistent Client/Server implementation**
- **Encapsulate infrastructure details**
- **Easier future technology extension**

Ingest Granule Processing

Ingest Request Control/Update capabilities:

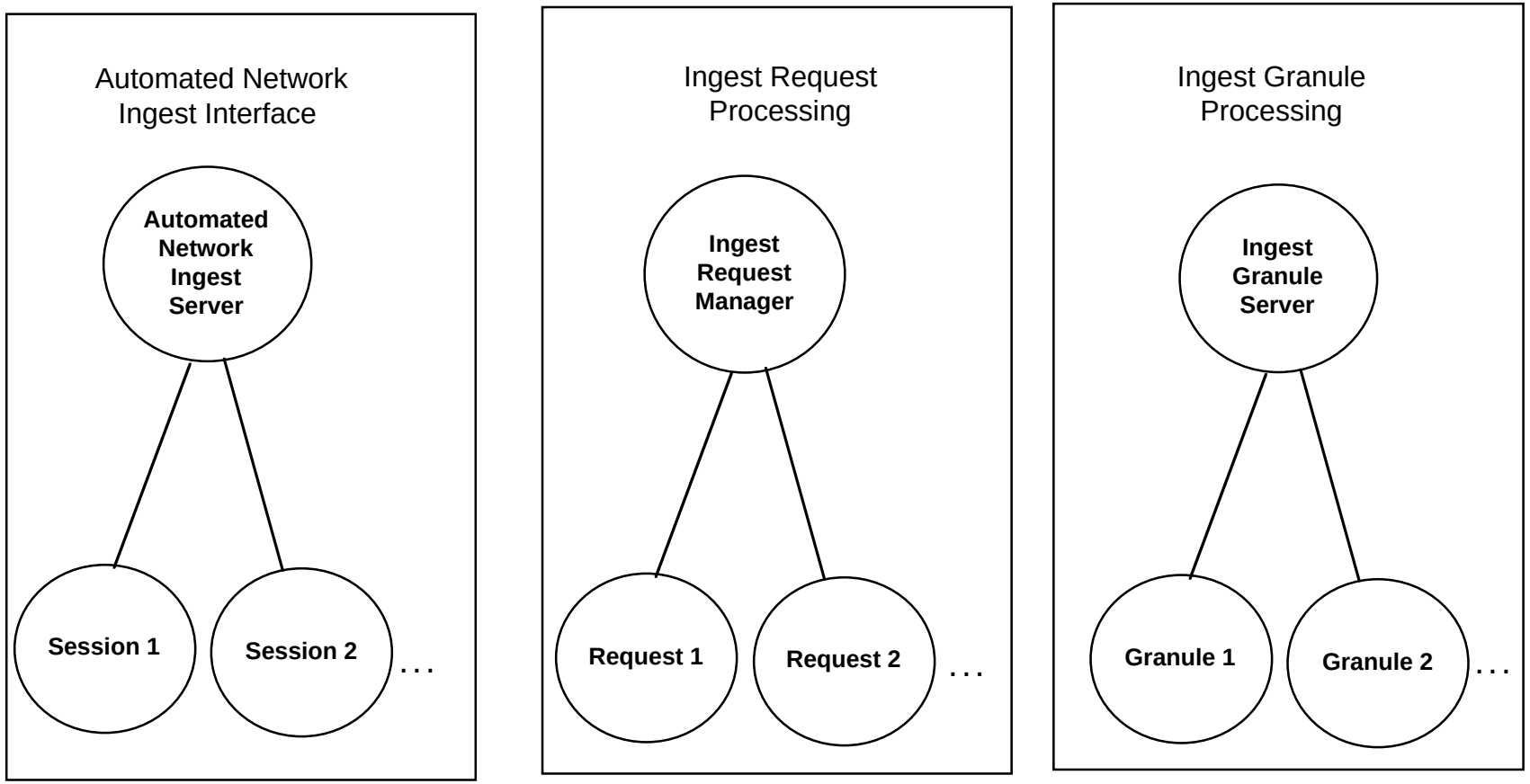
- **Suspend request**
- **Resume request**
- **Set request priority**

Request Throttling control

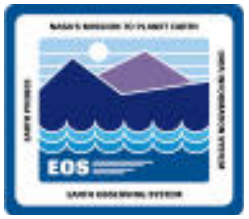
Performance and Accountability Parameters

Tunable Parameter control

Ingest Object Factory

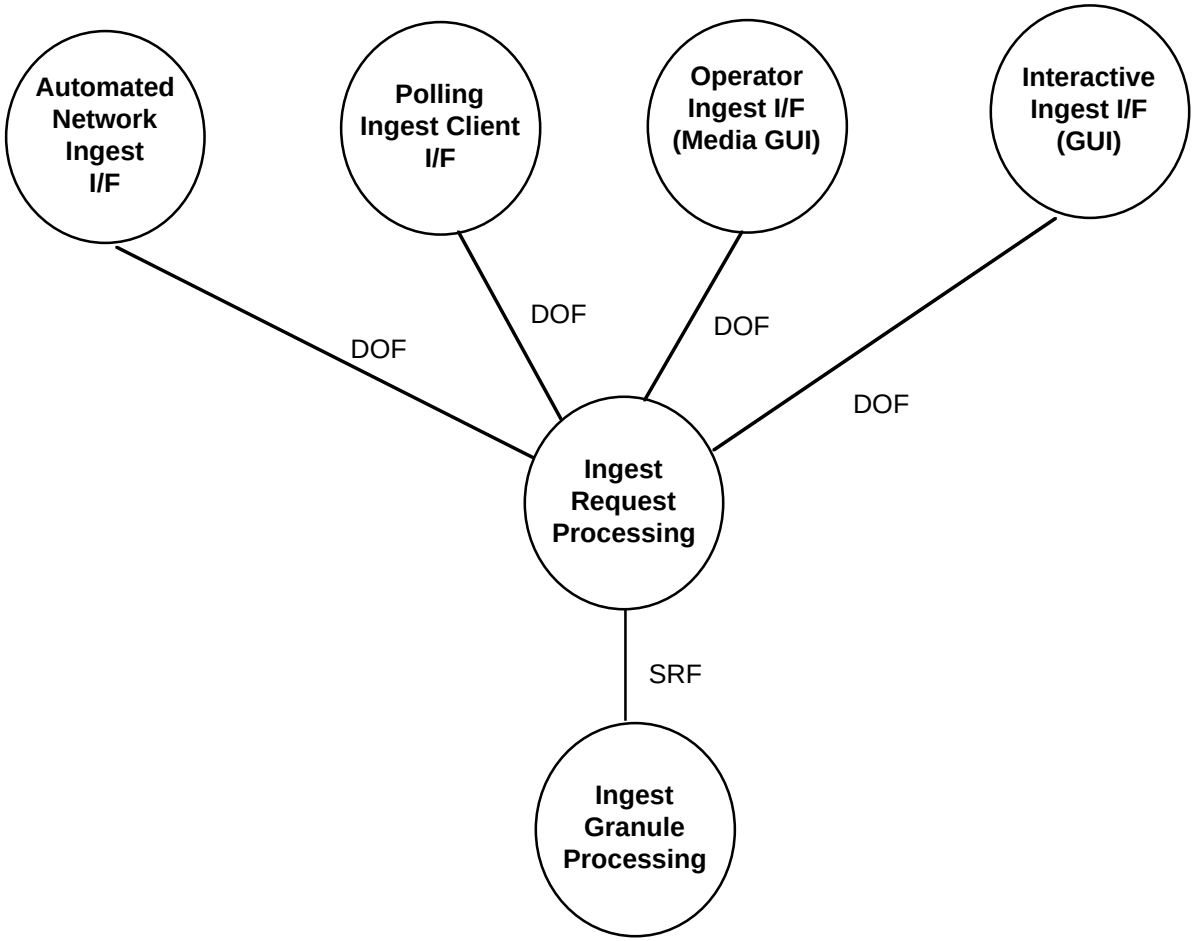


ECS Framework and Common Services Utilization

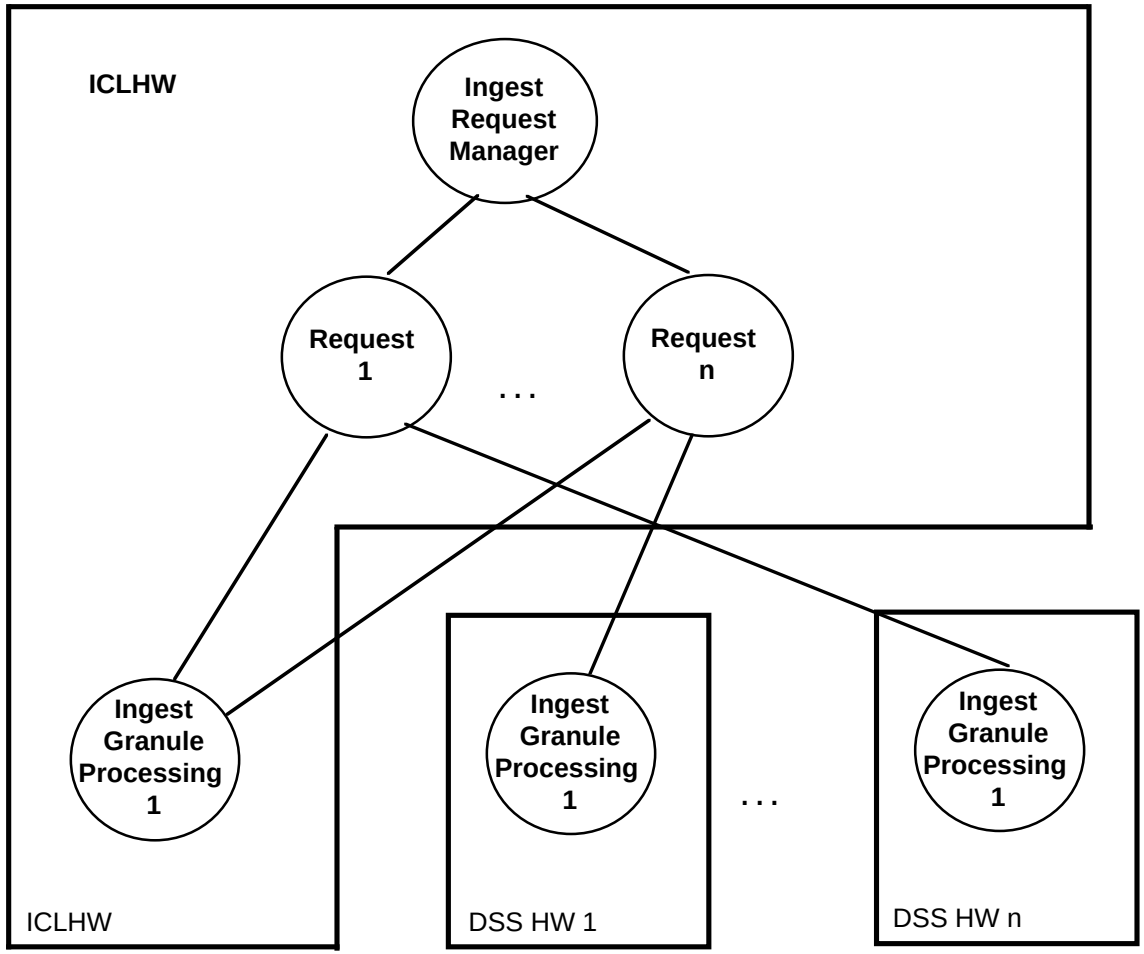


- **Use Managed Process Framework (MPF) for mode management and startup/shutdown**
- **Use Distributed Object Framework (DOF) for all Ingest Client OODCE interface**
- **Use Server Request Framework (SRF) for asynchronous request processing**
- **Use MSS Metrics for threshold configuration and for performance and accountability reporting**
- **Use MSS Logger for fault event reporting**
- **Use MSS Request Tracking Services to support MSS resource use accounting**
- **Use MSS User Profile for accessing user profile information**
- **Use Advertising UR to locate the appropriate Data Server Resource Manager and Science Data Server to handle the type of data being ingested**
- **Use Advertising to advertise Ingest HTML services**

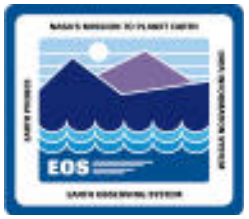
Ingest CSC Interface



Ingest/Data Server Hardware and Software Interactions



Ingest Granule Processing

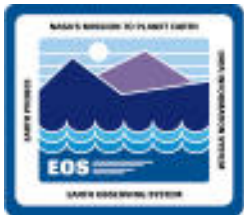


Uses SRF for asynchronous processing and message queuing

Distributes Ingest Granule Processing to efficiently process the data

- **Granule files are ftp'd to working storage where they will be archived**
  - **Level 0 data transferred to ICHLW RAID**
  - **Non-Level 0 data transferred to RAID associated with the appropriate EMASS Archive**
- **Preprocessing occurs on CPU controlling RAID working storage**

Ingest Request Control/Update Capabilities



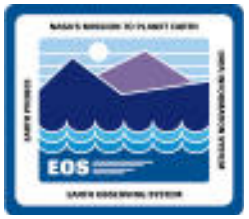
Control/update service consists of:

- **Cancel a request**
- **Suspend a request**
- **Resume a request**
- **Set priority of a request**

Enter the request via Ingest GUI for:

- **A single Ingest request**
- **All Ingest requests for an external data provider**
- **All Ingest requests in the system**

Ingest Request Suspend/Cancel Service

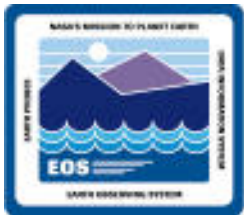


Similar Processing for Suspend Request and Cancel Request Except the Cleanup

Suspend/Cancel at the Request Level Ingest

- **Checks for Suspend/Cancel Before activating next granule**
- **Notifies all Remotely Processing Granules**
- **Waits for Currently Processing Granules to Process Suspend/Cancel Request**
- **Performs Cleanup Functions for Cancel Requests**
  - **Updates Checkpoint Data Base to Indicate Request Cancellation**
  - **Deletes Temporary Files**
  - **Deallocates Staging**
  - **Builds and Sends DDN**
  - **Exits Request Pthread**
- **Performs Cleanup Functions for Suspend Requests**
  - **Updates Checkpoint Data Base to Indicate request Suspension**
  - **Deletes Temporary Files**
  - **Exits Request Pthread**

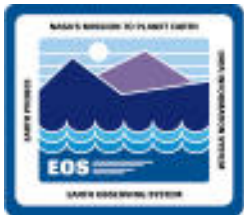
Ingest Request Suspend/Cancel Service (cont.)



Suspend/Cancel at the Granule Level Ingest

- Sends Cancel request to queued
  - Data Server Network Resource Requests (once a file transfer is started it can not be canceled)
  - Data Server Staging Disk Allocation Requests
  - Data Server Insert Requests (once an insert request is sent to EMASS, it can not be canceled)
- Checks for Suspend/Cancel
  - After each File Transfer has Completed
  - After the Preprocessing of the Entire Granule
  - When a Data Server Insert Request Completes

Ingest Request Resume Service



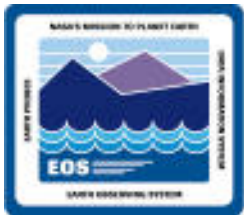
Resume at the Request Level

- Creates a pthread for the request
- Check checkpoint data base for each granule within the request to determine if resumption is needed
- Initiate granule processing

Resume at the Granule Level

- Check checkpoint data base to determine if granule processing should continue at the file transfer, granule preprocessing, or granule data server insertion level

Ingest Request Set-Priority Service



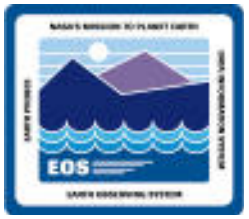
Set-Priority at the Request Level

- **Modify the priority of the request pthread**

Set-Priority at the Granule Level

- **Modify the priority of all remotely processing granules via SRF**

Ingest Request Throttling



The ingest request is throttled at two levels:

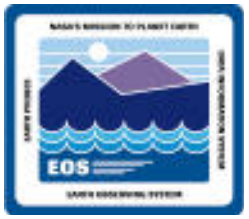
- **Request level**
- **Granule level**

No new requests will be accepted if the maximum request threshold or maximum granule threshold is reached

For granule throttling:

- **A volume threshold monitor is used; incremented when processing a new granule and decemented when a granule processing is complete**
- **Each granule server has a separate volume threshold monitor**
- **If volume threshold is reached, granule processing is delayed until sufficient volume is free**
- **Volume allocation to competing granules is based on priority**

Performance and Accountability Parameters



Provides request and granule summary events for fault and performance analysis

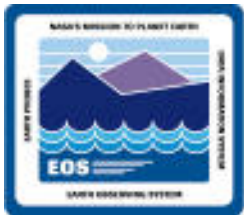
Allows operator to monitor granule

- **Request and granule thresholds**
- **Number of requests/granules being throttled**
- **Daily data quality faults**
- **Daily Ingest errors**
- **Daily request totals**

Performance and Accountability Parameters

Parameter / Event	Description	Recommended Sample Frequency / Event Frequency
Request Event	Per request: -Request ID -External Data Provider -Mission -Total processing time -Total data volume -Total number of files -Total Time to Transfer -Total Time to Preprocess -Total Time to Archive -Warm Start Flag -Completion Status (successful or with errors) -Ingest type(automated, interactive, polling files, polling delivery record, media)	Every Ingest request
Request Granule Event	Per granule: -Request ID -Data Type -Start Time -End Time -Data Volume -Number of files -Total time to transfer -Total time to preprocess -Total time to archive -Warm start flag -Final status (used to provide fault metrics)	Every request granule
System Request Maximum	Maximum number of requests allowed to be processed at one time in the Ingest system.	Every minute
Granule Server Data Volume Maximum	Maximum volume of data allowed to be processed across all ongoing ingest requests submitted to the Granule Server. A separate maximum volume exists for each Granule Server.	Every minute
Granule Server Data Volume Buffered	Total volume of data being throttled, by Ingest Request Processing, due to excessive system/external data provider volume loads	Every minute
Data Requests Buffered	Total number of requests being throttled, by Ingest Request Processing, due to excessive system/external data provider volume loads	Every minute
Daily Count of Data Quality Faults	Running count of data quality faults. Initialize to zero at Ingest startup. Reset to zero once a day by Ingest software. Data quality errors include metadata missing or invalid values, missing required files or DAN file validation failures.	Every fifteen minutes
Daily Count of Ingest Errors	Running count of Ingest errors. Initialize to zero at Ingest startup. Reset to zero once a day by Ingest software. Data quality errors include all failures, external communication errors, system failures, request rejections for security failures).	Every fifteen minutes
Daily Count of Request Processed (successfully/ unsuccessfully)	Running count of all Ingest requests. Initialize to zero at Ingest start. Reset to zero once a day by Ingest software.	Every fifteen minutes

Ingest Tunable Parameter Control



Three methods:

- **Loaded from the Ingest Configuration file to the MSS Management Information Block (MIB)**
- **Sent as a calling parameter during the program startup**
- **Stored in an Ingest data base table**

MIB parameters can be modified via MSS GUI

Calling parameters are changeable at program/system startup

Ingest data base parameters can be modified via the Ingest GUI

Ingest Tunable Parameters

Parameter / Parameter Category	Description	Change Control	Source
Polling Ingest Timer	The time period which indicates how often the Polling Ingest Session should check for the existence of ingest files for ingest processing. A separate polling timer exists for each external data provider using the Polling Ingest Mechanism.	Inget GUI	Ingest Data Base
External Data Provider Priority	Priority associated with ingest requests from the external data provider. A separate priority exists for each external data provider.	Ingest GUI	Ingest Data Base
System Request Maximum	Maximum number of requests allowed to be processed at one time in the Ingest system.	HpOpenView	Ingest Configuration Metrics File
Granule Server Data Volume Maximum	Maximum volume of data allowed to be processed across all ongoing ingest requests submitted to the Granule Server. A separate maximum volume exists for each Granule Server.	HpOpenView	Ingest Configuration Metrics File
Communication Retry Count	Number of retries to perform when a communication failure is encountered with the external data provider.	HpOpenView	Ingest Configuration Metrics File
Monitor Time For Completed Request	Time (in minutes) a completed request will remain on the operator's monitor screen.	HpOpenView	Ingest Configuration Metrics File
Mode Parameter	Determines in which mode (operational, test, training) to initiate an Ingest process. Supplied by operations personnel to the MSS Process Framework at system startup.	Process Framework	Calling Parameter
Startup Mode	Determines if Ingest software is warm of cold started.	Process Framework	Calling Parameter
Warm Start Time Limit	Time (in days) failed requests can be warm started from checkpointed state (checkpointed data exceeding this time limit will be automatically deleted from the Ingest checkpoint tables).	HpOpenView	Ingest Configuration Metrics File