440-TP-014-001

ECS Ingest Subsystem Topology Analysis

Technical Paper

Technical Paper--Not intended for formal review or government approval.

August 1995

Prepared Under Contract NAS5-60000

RESPONSIBLE ENGINEER

Paul Roycraft /s/	8/16/95
Paul Roycraft, Systems Engineer EOSDIS Core System Project	Date

SUBMITTED BY

Parag N. Ambardekar /s/	8/16/95
Parag Ambardekar, Release A CCB Chairman EOSDIS Core System Project	Date

Hughes Applied Information Systems
Landover, Maryland

This page intentionally left blank.

Abstract

The ECS Ingest Subsystem Topology Analysis analyzes and documents the configuration of the Ingest Subsystem based on the subsystem architecture identified in the ECS Ingest Subsystem Design Analysis (440-TP-002-001) delivered at the ECS Preliminary Design Review (PDR). This topology analysis focuses on factors that affect the sizing requirements of Ingest Subsystem components, as well as the capability of the configuration to meet reliability and availability requirements.

Keywords: Ingest, Client host, Working storage, RMA, Topology, Sizing, Architecture

This page intentionally left blank.

Contents

Abstract iii

Contents v

1. Introduction 1

1.1 Purpose 1

1.2 Organization 1

1.3 Review and Approval 1

2. Ingest Subsystem Sizing 3

2.1 Ingest Subsystem Overview 3

2.2 Ingest Data Flows 3

2.3 Ingest Queuing Model Description 4

2.3.1 Ingest Subsystem Simplified Data Flows 5

2.3.2 Queues Defined 6

2.3.3 Analysis 6

2.3.4 Conclusions 9

3. Ingest RMA Architecture 10

3.1 Data Ingest RMA Requirements 10

3.2 Ingest RMA Architecture 10

Appendix A - Ingest Queuing Model Calculations 13

This page intentionally left blank.

1. Introduction

1.1 Purpose

The ECS Ingest Subsystem Topology Analysis analyzes and documents the configuration of the Ingest Subsystem based on the subsystem architecture identified in the ECS Ingest Subsystem Design Analysis (440-TP-002-001) delivered at the ECS Preliminary Design Review (PDR). This topology analysis focuses on factors that affect the sizing requirements of Ingest Subsystem components, as well as the capability of the configuration to meet reliability and availability requirements. Methods for assuring acceptable system reliability, maintainability, and availability (RMA), as well as the flexibility to evolve the ingest configuration as mission requirements change are key Ingest Subsystem design drivers. This technical paper has been developed in response to these issues as identified in the System Design Review (SDR) version of the ECS System Design Specification.

1.2 Organization

This paper is organized as follows:

Section 1 introduces the purpose and organization of this technical paper and describes the review and approval process.

Section 2 provides an overview of the ingest subsystem and issues associated with sizing of ingest subsystem components. Data rates and volumes to be supported at each Distributed Active Archive Center (DAAC) are characterized, and subsystem resources required to support these volumes are identified.

Section 3 discusses methods for assuring that ingest RMA requirements can be met with the architecture and components discussed in Section 2.

1.3 Review and Approval

This technical paper is an informal document approved at the Office Manager level. It does not require formal Government review or approval; however, it is submitted with the intent that review and comments will be forthcoming.

The ideas expressed in this study are valid through the time period up until the SCDO Release A Critical Design Review (CDR). The concepts presented here are expected to migrate into the following formal Contract Data Requirements List (CDRL) deliveries:

Table 1-1. Technical Paper to CDRL Migration

Technical Paper Section	CDRL DID/Document Number
Entire paper	DID305/DV2

Questions regarding technical information contained within this paper should be addressed to the following ECS contact:

- ECS Contact
  - Paul S. Roycraft, 301-925-0623, proycraf@eos.hitc.com

Questions concerning distribution or control of this document should be addressed to:

Data Management Office
The ECS Project Office
Hughes Applied Information Systems
1616A McCormick Dr.
Landover, MD 20785

2. Ingest Subsystem Sizing

This section introduces the ingest subsystem and addresses the trade study issues identified in the SDR version of the ECS System Design Specification. This topology analysis focuses on factors that affect the sizing requirements of Ingest Subsystem components.

2.1 Ingest Subsystem Overview

The ingest subsystem is a collection of hardware and software that supports the ingest of data into ECS repositories on a routine and ad-hoc basis, and triggers subsequent archiving and/or processing of the data. The ingest subsystem configuration must be flexible to support a variety of data formats and structures, external interfaces, and ad-hoc ingest tasks. Data processing and archiving functions to be performed by the ingest subsystem and ingest clients vary according to attributes of the ingested data such as data type, data format, and the level to which the ingested data has been processed. From a software perspective, the ingest subsystem is organized into a collection of tools from which those required for a specific situation can be configured. The resultant configuration is called an ingest client. Ingest clients may exist in a static configuration to service a routine external interface, or they may be specially configured and exist only for the duration of a specific ad-hoc ingest task. The ingest clients provide a single virtual interface point for the receipt of all external data to be archived within the SDPS. Individual ingest clients are established to support each unique interface, allowing the interface parameters to be modified as interface and mission requirements evolve. Ingest data preprocessing, metadata validation, and metadata extraction is performed by the ingest clients on any incoming data, as required.

The hardware components of the ingest subsystem are similar to those of the data server, but are specialized to meet the ingest requirements at a given site. Specialized forms of ingest clients may be incorporated into site unique architectures, and additional processing hardware may also be incorporated at those sites where special transformations must be accomplished on ingest data sets. The Ingest Subsystem hardware configuration uses virtually identical components at each Release A operational site, with minor variances based on data volume to be processed and stored. Additionally, this configuration is planned to continue into Release B at the smaller DAAC sites, with additional equipment required at the sites (EDC, GSFC, and LaRC) supporting the receipt of EDOS L0 and Landsat 7 L0R data.

The following sections discuss issues related to ingest subsystem sizing.

2.2 Ingest Data Flows

Data is staged to one of two areas depending on the data level, data type, and other data set specific characteristics:

- Level 0 data from ongoing missions will be staged to the Ingest Subsystem working storage area, where the data will be ingested and stored in the Level 0 (L0) rolling store. The staged data will also be accessible by the SDPS Processing Subsystem for that data which must be processed to higher levels.

- Level 1A-4 data will be staged directly to the working storage area in the Data Server Subsystem. Ingest client functionality such as quality checking and reading of metadata will be performed on this data upon the Data Server Subsystem processor hardware. The data server will then archive the data in the logical and physical data server to which the particular data has been assigned.

The ingest subsystem must support different interface mechanisms in receiving the data from ingest sources. Some ingest data sources will provide data via different types of hard media, but this is generally limited to contingency transfers and possibly some Version 0 migration data in Release A. The majority of the Release A data transfers will occur through an electronic interface. The mechanism for electronic transfer also varies depending on the data source. Some data sources will notify ECS of the availability of data, which the Ingest Subsystem will get from the source. Others will put data in a predefined location which is then periodically polled by the ingest client software. At Release A the principal driver in terms of daily electronic L0 data ingest is the data received through the SDPF for the TRMM platform. At Release B the principal driver of daily electronic L0 data ingest is the data received through EDOS at GSFC and LaRC for the AM-1 platform and Landsat-7 data received at EDC. The Moderate Resolution Imaging Spectrometer (MODIS) and Multi-angle Imaging Spectro-Radiometer (MISR) AM-1 instrument data sent to GSFC and LaRC, respectively, comprise the majority of the L0 data to be archived within the Ingest Subsystem for the first several years of ECS operations. The reliable ingest and archiving of this data, and the associated management of Ingest Subsystem resources are the principal drivers in the Ingest Subsystem design. Release A daily Level 0, annual Level 0, and Version 0 migration data volumes are shown in Table 2-1.

Table 2-1. Release A DAAC Data Volumes

DAAC	Daily L0 Volume (MB/day)	Annual L0 Volume (GB/year)	Version 0 Data Migration Volume (GB)
LaRC	249	90.1	785
MSFC	262	95.6	152
GSFC	None	None	2000

2.3 Ingest Queuing Model Description

An Ingest Queuing Model (Imodel) was developed to assist in the sizing of Ingest Subsystem components, and is a high level look at the data flows in the Ingest Subsystem. It is implemented as a EXCEL workbook and uses the M/M/1 queuing equations (see Appendix A). This analysis is valid for systems where each queue has only one server, and the inter arrival times and service times have exponential probability densities. In addition, the results are only valid for steady-state conditions, where the probability of finding the system in a given state does not change with time. The Release A Ingest Subsystem approximates this configuration sufficiently for these assumptions to be valid. The Release B configuration, with its greater number of data sources and more complicated interactions, will be sized with the aid of a dynamic model that can more effectively take these additional factors into account.

2.3.1 Ingest Subsystem Simplified Data Flows

Figure 2-1 presents the Ingest Subsystem as seven interconnected items.

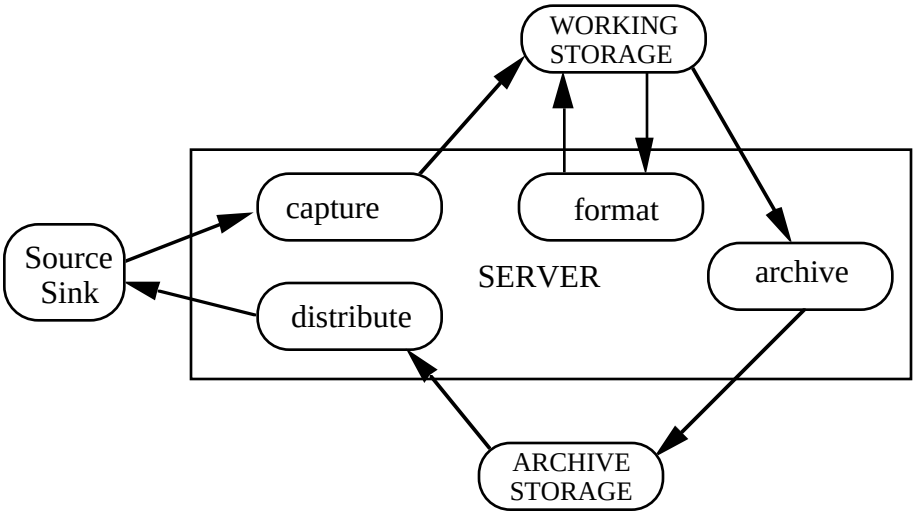

Figure 2-1: Ingest Subsystem Flow Diagram

Data enters and exits this subsystem through the "Source/Sink" item which models the external network links to the Ingest Subsystem. The data then flows into the "SERVER" during the "capture" process. It is assumed that the amount of data created by this, and the other processes, is very small compared with the input data flow "I", therefore the same volume of data "I" is sent to the "WORKING STORAGE" item.

When the capture of a data set is complete the data is sent back to the "SERVER" to the "format (which includes metadata extraction/validation)" process. It is assumed that all the data "I" is sent through "format" and back to "WORKING STORAGE" where it is stored again. This assumption is valid if "format" is performing a true reformat on the data. It is recognized that for Release A "format" will only be doing a simple metadata extraction and validation based on the data in the file headers so it is generally not necessary to move the whole data volume back and forth. The GRIB to HDF-EOS conversion that is being performed on the TRMM ancillary data is small enough in volume that, even if it is necessary to move the entire file or group of files, the volume is small compared with other data transfers that are taking place. The assumption that the full data set is moved is a simplification of the process and yields conservative results.

When "format" is done with a data set it is sent from "WORKING STORAGE" back to the "SERVER" to the "archive" process which sends the data "I" to the "ARCHIVE STORAGE".

It is assumed that sometime later all of the data will be requested by some external process, this will cause the process "distribute" on the "SERVER" to retrieve the data from "ARCHIVE STORAGE" and send it to the "Source/Sink" item. It is further assumed that while a data set may sit in the "ARCHIVE STORAGE" for an indeterminate amount of time (up to one year) on average one copy (the number of copies is a parameter) of the data must be exported from the "ARCHIVE STORAGE" as soon as and fast as it is stored into it.

2.3.2 Queues Defined

Figure 2-2 illustrates the four queues that were used to model the Ingest Subsystem.

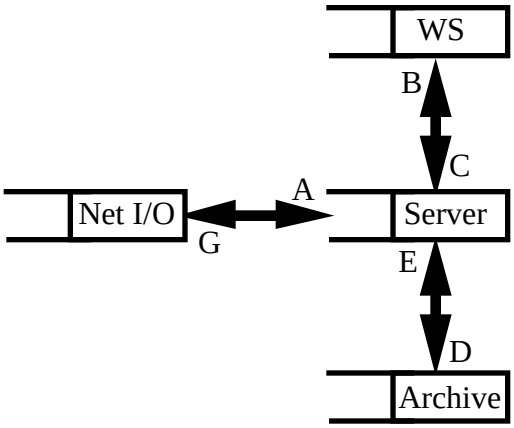

Figure 2-2: Queue Definition

The "Source/Sink" is modeled with the Net I/O queue, with "A" equal to the ingest data and "G" the exported data. Because of our conservation of data assumption (no significant amount data is created or destroyed) "A" equals "B".

The amount of data flowing to WS (WORKING STORAGE) equals two times "A", one "A" is the output of "capture" and the second from the "format" process (all of the data ingest must go through "format" so it must equals "A" also).

The data flowing into and out of "ARCHIVE STORAGE" equals the input data plus an insignificant amount of status that is ignored.

The amount of data going into the "SERVER" is four times the ingest rate to simulate the four process running (capture, format, archive, and distribute). The CPU required for each process is a function of "A" times a separate parameter for each process.

<u>Equation Summary</u>

$A = D = E = G$

$B = C = 2 * A$

$F = f(cap*A + for*C + arc*C + dst*E)$

2.3.3 Analysis

Model output is dependent on a series of model input parameters that may be varied depending on characteristics of the data to be ingested, processed, and stored, as well as network and Ingest Subsystem component capabilities and performance. Parameters such as server CPU and I/O performance, disk I/O, network performance, and the number of operations/byte associated with each server process (e.g., capture, format, archive, distribute) may all be varied to analyze the sensitivity of changes in data flows and system architecture. Tables 2-2, 2-3, and 2-4 show select

model output parameters generated by the model for a specific configuration at LaRC, MSFC, and GSFC, respectively. The load presented by each flow in pkts/sec is a function of the number

Table 2-2: Queuing Model Derived Requirements for the LaRC Configuration at Release A

Flow	Flow Utilization (% of maximum available)	Time in System (Range for 10 MB - 160 MB file size, in seconds)
Source	4.14	.0006
Working store in	11.67	2.60 - 41.62
Working store out	11.67	2.60 - 41.62
Archive in	7.8	2.60 - 41.62
Archive out	31.14	2.60 - 41.62
Distribution	16.62	.0006

Table 2-3: Queuing Model Derived Requirements for the MSFC Configuration at Release A

Flow	Flow Utilization (% of maximum available)	Time in System (Range for 10 MB - 160 MB file size, in seconds)
Source	0.9	.0006
Working store in	2.55	2.52 - 40.34
Working store out	2.55	2.52 - 40.34
Archive in	1.68	2.52 - 40.34
Archive out	6.78	2.52 - 40.34
Distribution	3.6	.0006

Table 2-4: Queuing Model Derived Requirements for the GSFC Configuration at Release A

Flow	Flow Utilization (% of maximum available)	Time in System (Range for 10 MB - 160 MB file size, in seconds)
Source	10.29	.0006
Working store in	28.92	2.76 - 44.26
Working store out	28.92	2.76 - 44.26
Distribution	41.1	.0007

of bits/sec input from the previous process and the mean size of packet/data set that this process expects. The “Source” item sends packets whose size and rate depend on the network technology chosen. FDDI is baselined at the Release A sites, with an assumed network efficiency of 60% from the maximum 100 Mbps clock rate. The "capture" item includes server functions that receive data from the network or hard media device. Capture rate is a function of

the input bandwidth and server I/O capabilities. The "work.store IN" includes the writing of the ingested data to working storage and "work.store OUT" models the writing of data back to the server to perform the "format" operation. Both read and write operations have associated transfer rate and access time estimates for each data transfer. Conservative estimates of 2 MB/sec are used based on results from Data Server prototyping efforts. Writing of data to the L0 rolling store archive involves the capabilities of the working storage disks, server I/O, and rolling store devices. Finally, the number of copies of data read from the archive and sent to a data sink (e.g., Processing Subsystem) may be varied to determine the additional load of reprocessing on Ingest Subsystem resources. Further description of the calculations used to determine the model output parameters may be found in Appendix A.

Analysis of the ingest queuing model developed for CDR confirms the results of the paper analysis conducted prior to PDR supporting the ability of the Release A Ingest Subsystem configurations to support Level 0 data ingest requirements. Projected maximum individual component utilization in support of Level 0 data ingest is approximately 10% under nominal L0 data ingest operations at LaRC and MSFC. The additional subsystem capacity is available for contingency (e.g., ingest of more than one days worth of data in one day or resolving difficulties encountered during reformatting or metadata validation activities), Version 0 data ingest, and subsystem testing requirements. Inclusion of support for Version 0 ingest functions increases subsystem component utilization to the 20% to 40% range at LaRC, the 10% to 30% range at MSFC, and the 40% to 70% range at GSFC, depending on the required daily ingest volume. Note that one or more individual flow utilization figures shown in the tables may sum to comprise the total utilization for a particular component.

As can be seen from the flow utilization estimates in Tables 2-2, 2-3, and 2-4, Ingest Subsystem components have significant reserve capacity in the Release A timeframe. In the case of working storage, the specified RAID devices are required to meet RMA and capacity requirements, but also provide transfer rates that are greater than estimated needs. This is does not imply, however, that the drives provide exceptional performance. Rather, it is simply a function of the relatively low daily volume of data to be processed. Note that the stringent RMA requirements that require the redundant RAID devices at LaRC and MSFC are not in place at GSFC at Release A. The early interface testing and Version 0 migration to be supported at GSFC do not require the same RMA capabilities as the more time-critical L0 data ingest at the other sites.

Ingest client hosts also provide I/O and processing capabilities at Release A at LaRC and MSFC that are somewhat greater than required based solely on the daily data ingest volume. Two design factors contribute to the selection of the client hosts. The first is that the RMA demands of the Ingest Subsystem dictate that a backup client host and RAID disk be incorporated in the design. The client hosts and RAID disk banks must be dual-homed so that either processor can transfer data to either RAID bank in the event of failure of a single component. Workstation-class machines typically do not support this function, requiring upgrade of the client hosts to a higher class machine than is required solely for performance reasons. A second factor is that the required volume of data to be ingested from Version 0 in the Release B timeframe increases significantly at MSFC. The planned 772+ GB to be migrated at Release B is approximately 5 times the amount to be migrated at Release A. This additional load on the client hosts would require an upgrade to the class of machine that has been specified for Release A.

Competitive procurement restrictions constrain this paper to identification of the class of component and general performance characteristics used in the analysis, rather than actual

candidate components. The client host CPUs at LaRC and MSFC are low to mid-level Symmetric Multi-Processing (SMP) 32-bit machines, capable of supporting multiple network (FDDI) and direct-connect (SCSI II) devices. The client host at GSFC is a mid-level workstation with 6 GB of local disk. Working storage devices at LaRC and MSFC are RAID 5 units with a minimum of 2.0 days worth of space allocated to ingest working storage required to support the functions of acquiring, processing, validation, and archiving L0 data. This volume of working storage allows for one days worth of L0 data to be staged for processing, an additional days worth available for subsequent ingest, and an additional 25% available to service additional Ingest Subsystem needs (e.g., L0 archive retrieval support, pre-processing, quality checking). Additional magnetic disk resources are supplied within the Ingest Subsystem to support items such as client host operating system and application software and L0 archive database directory information. The L0 rolling store may be implemented in several ways, subject to additional analysis prior to the delivery of the final version of this technical paper. Additional RAID was originally planned to function as the L0 rolling store, as the annual volumes at LaRC and MSFC were in the 30-35 GB range. Recent inclusion of TRMM spacecraft housekeeping packets in the volume to be stored increases the annual volumes to those shown in Table 2-1 (90-95 GB). A study has been initiated to compare cost and performance issues associated with adding more RAID versus the use of an alternate archive device (e.g., 3590 drives in a small stacker configuration).

2.3.4 Conclusions

Analysis of the ingest queuing model developed for CDR confirms the results of the paper analysis conducted prior to PDR supporting the ability of the Release A Ingest Subsystem configurations to support Level 0 data ingest requirements. Projected maximum individual component utilization in support of Level 0 data ingest is less than ten percent under nominal operations. The additional subsystem capacity is available for contingency (e.g., ingest of more than one days worth of data in one day or resolving difficulties encountered during reformatting or metadata validation activities), Version 0 data ingest, and subsystem testing requirements.

3. Ingest RMA Architecture

A principal objective of separating ingest from other SDPS functions is to assure the high reliability and availability of the system for L0 data ingest. Ingest availability requirements are met through the use of high reliability individual components in hot and warm spare configurations, as necessary. The following paragraphs provide additional detail on ingest RMA requirements and how the ingest architecture ensures that these requirements will be met.

3.1 Data Ingest RMA Requirements

The principal L0 data sources (e.g., EDOS, SDPF) each support a data driven architecture that processes data within 24 hours or less from receipt of the data from the spacecraft. Once the data is processed into Production Data Sets (PDSs) of TBD size (may be based on clock time, number of packets, or Tracking and Data Relay Satellite System [TDRSS] service session) the data must be transferred to ECS in a timely fashion for archiving and any required higher level processing. Typically, data transfer must be completed within several hours to free up resources at the L0 processing sites, as new data sets are being received on a nearly continuous basis. The L0 processing sites typically do provide long term archiving of either the raw data or the PDSs in the event that data is corrupted or lost in the transfer to ECS or within ECS itself. While it is possible to access these archived data sets if they are needed, it is unattractive to do so from an operational standpoint due to the added time and complexity of pulling the data from the deep archive and staging it for transfer. Therefore, the SDPS function of receiving science data availability requirement of 0.999 must be met to reduce the required number of data retransmissions and ensure that the needs of the overall data system are met. The following section describes how this requirement and others which place RMA requirements on the ingest subsystem of 2000 hours mean time between failures (MTBF) and 15 minutes mean time to restore (MTTR) from a failure to operational capability will be met. A more detailed availability analysis may be found in the Availability Models/Predictions for the ECS Project document (515-CD-001-002). That document presents a summary of the preliminary availability assessments of the ECS Segments, as well as a complete set of Excel spreadsheets for all required RMA functions and their availability results.

It should be noted that certain of the sites at each ECS release are not staffed 24 hours per day, and are subject to somewhat different operational requirements during unstaffed periods. The Ingest Subsystem is designed to operate with minimal operator involvement, and is planned to continue the function of ingesting data during unstaffed periods. Moreover, faults may require human involvement to reconfigure and reboot the ingest client hosts. Therefore, the 15 minute MTTR requirement for support of the receipt of science data will not be supported during unstaffed periods. In order to eliminate any potential system overload, Ingest Subsystem capabilities are being sized to satisfy full 24 hour L0 data ingest requirements during staffed periods.

3.2 Ingest RMA Architecture

The instantiation of the ingest subsystem will vary at each site, but will all be based on the same architecture concepts and classes of hardware and software. The generic ingest subsystem

architecture is shown in Figure 3-1. The ingest client software required for a specific L0 interface at a given site will run on a client host computer residing in the ingest subsystem. Multiple ingest clients may run on a single client host, or may be divided among multiple client hosts, depending on the data load supported by each interface. At least one spare client host is provided at each site in order to provide a warm backup failover capability in the event of a primary host failure. A hot backup approach was initially considered, but was determined to be unnecessary and undesirable for several reasons. Current EDOS requirements and architecture concepts do not support the transmission of data to multiple targets with different application ids (APIDs). This makes the implementation of a prime and hot spare ingest client host configuration more complicated, but a warm spare configuration was determined to be sufficient to meet requirements. Commercially available workstations of the class required for the client hosts typically support MTBFs in the 20,000 to 40,000 hour range. Preliminary calculations indicate that the use of this hardware in conjunction with redundant working storage and deep archive components will meet data ingest RMA requirements. Final calculations will be completed when specific components are chosen. Working storage and L0 archive data server media drive and robotics devices are sized to accommodate redundant devices or components at each location, as necessary. A combination of paper analysis and system modeling efforts as discussed in Section 3.1 will determine the configuration necessary to meet system performance and RMA requirements.

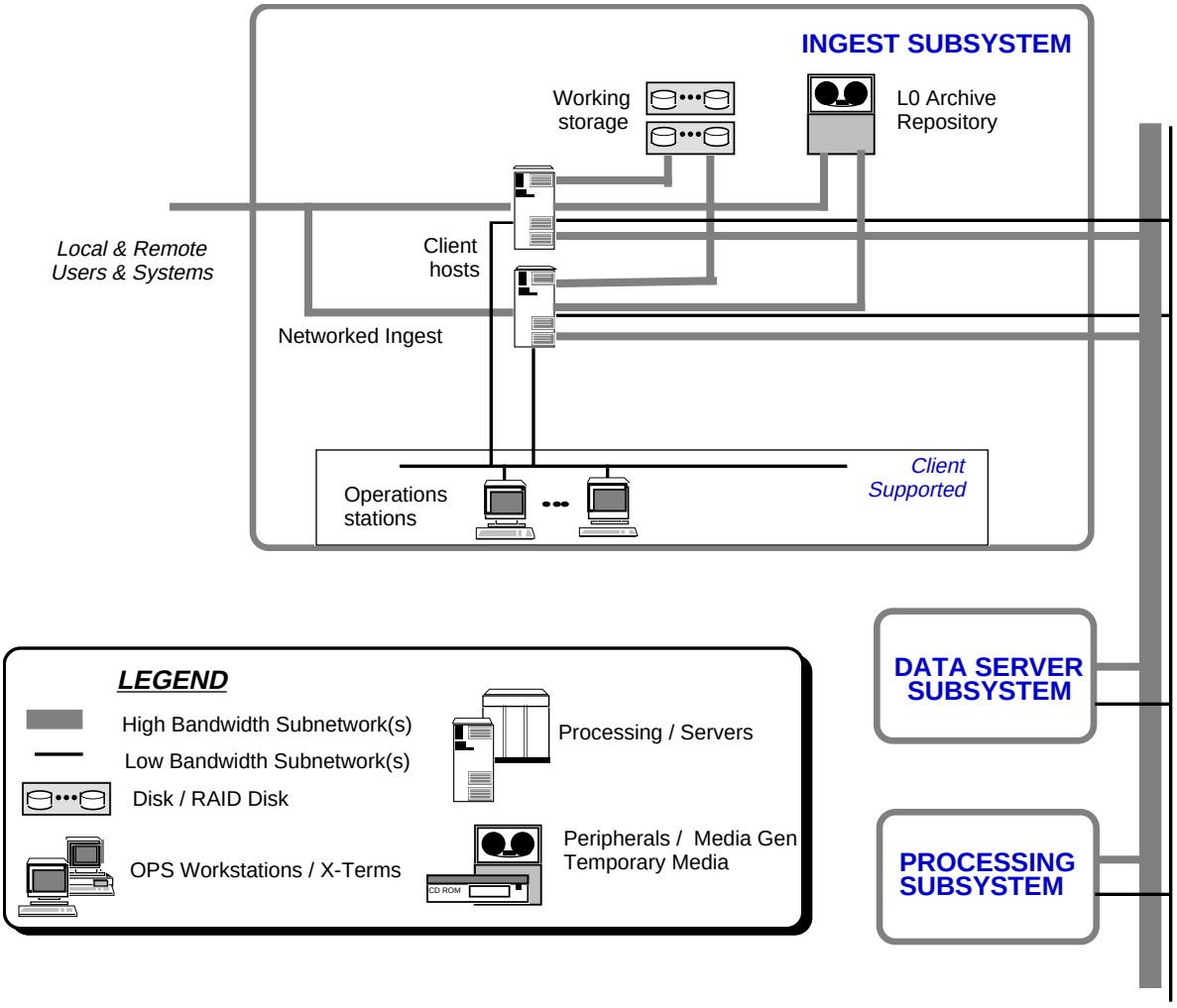

Figure 3-1: Generic Ingest Subsystem Configuration

Appendix A - Ingest Queuing Model Calculations

FOR A M/M/1 SYSTEM

exponential probabilty density for arrival, service & 1 server

n number customers

λ (lamda) mean arrival rate = customer / sec

P probabilty Pn(t) = Probabilty that exactly n customers arrive during a interval length t

τ (tau) time interval

Poisson Law

$$Pn(t)=((\lambda\tau)^n * e^{-\lambda\tau}) / n!$$

μ (mu) mean service rate = customer / sec

ρ (rho) Utilization of server = traffic intensity, it must be less then 1

$$\rho = \lambda / \mu$$

N mean # of customers in the system = # in que + in server

$$N = \rho / (1-\rho)$$

T total wait time, mean time between a customer's arrival and depature

Little's results

$$T = \frac{N}{\lambda} = \boxed{1/(\mu - \lambda)}$$

In a network made up of a series of M/M/1 queues the following two findings are very useful:

1) If the output of several M/M/1 servers feed the input of another server, the resulting input process is also a Poisson process, with mean equal to the sum of the means of the feeding processes. (“The Outputs of a Queuing System” Burke, 1956)

2) In an open network of M/M/1 servers queues can be analyzed as though each one were isolated from all the others. (“Networks of Waiting Lines” Jackson, 1957)

For NETWORKS in a M/M/1 system

λ mean arrival rate = packets / sec

Ci channel i's capacity in bits/sec

μ packets/bit

$$\mu * Ci = \text{service in packets / sec}$$

$$T_i = \boxed{1/(\mu Ci - \lambda_i)}$$