

NGAO KOTN: xxxx

**NGAO
Data Server Preliminary Design**

**Version 1.0
Date: 4/5/2010**

Authors:
Jimmy Johnson

Contributors:

W. M. Keck Observatory
California Association for Research in Astronomy
65-1120 Mamalahoa Highway
Kamuela, Hawaii 96743
808-885-7887

Table of Contents

1	INTRODUCTION.....	3
1.1	Project Description.....	3
1.1.1	Clients	4
1.1.2	Databases	4
2	HARDWARE	6
2.1	Optimal Solution	6
2.1.1	Specifications.....	7
2.2	Traditional solution.....	8
3	SOFTWARE	10
3.1	Relational database	10

List of Figures

Figure 1 Data Server Overview	3
Figure 2 Data Server Hardware Architecture	8
Figure 3 Sun Fire X4140	9
Figure 4 SurfRaid Triton 14FA4	9

List of Tables

Table 1 MySql and Postgres comparison	10
---	----

1 Introduction

The Data Handling System for the AO Controls has been called the “Data Server”. The NGAO Data Server is defined as a computer in the network, with associated software, that is dedicated to the data storage and retrieval needs required by the AO Controls System. It holds the database management system, the databases and is part file server capable of housing non-database files such as FITS. In addition to providing data persistence, retrieval and data management there will be a telemetry data recorder, an ad-hoc data recorder, an alarm recorder, general logger, and configuration. High speed RTC data will be stored in its own data recording system.

1.1 Project Description

Figure 1 shows a high level overview of the data server. The data server hardware consists of expandable and reliable (RAID) hardware capable of supporting multiple databases and file formats.

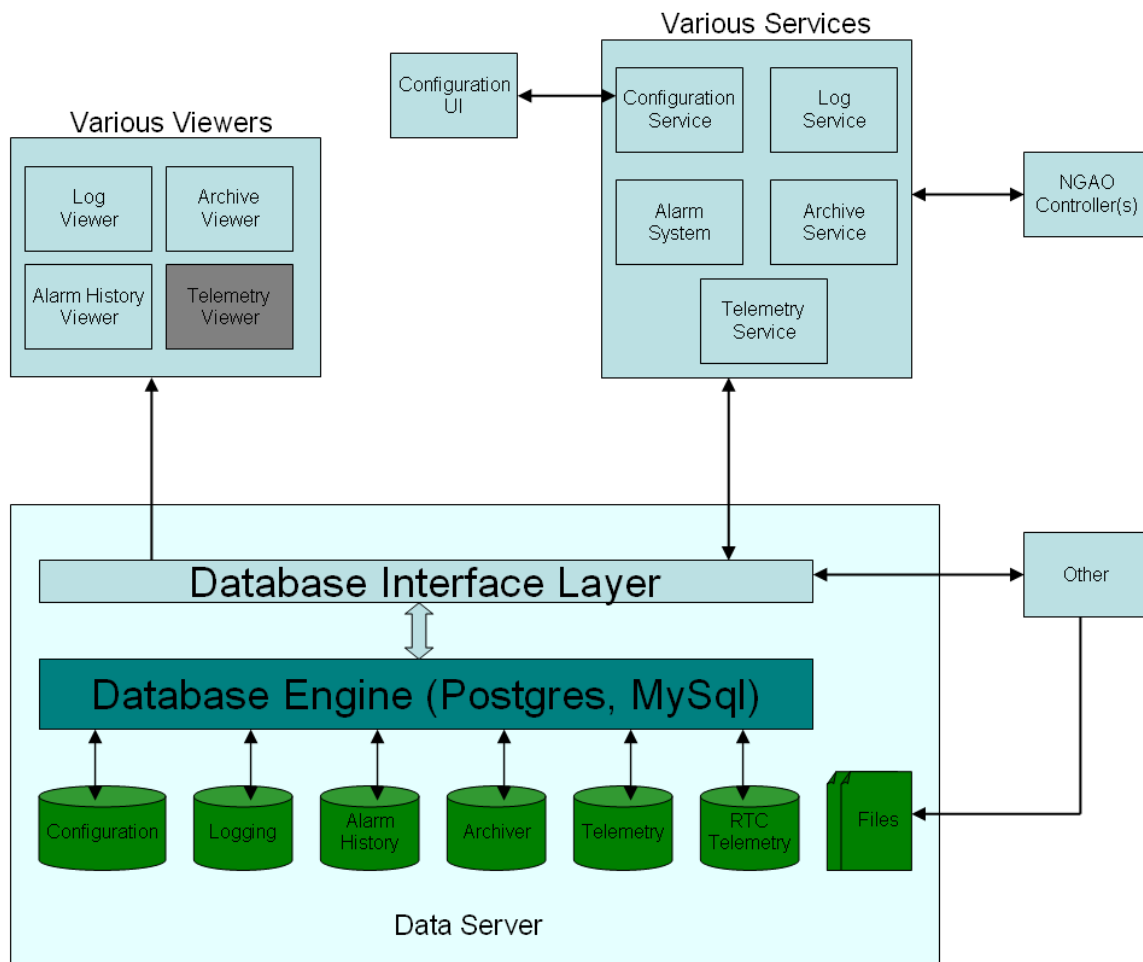


Figure 1 Data Server Overview

The database engine is expected to be either PostgreSQL or MySQL. Both are fairly comparable at this stage though some analysis should be done to compare the two in detail. As a frame of reference the telemetry server in NGWFC uses PostgreSQL as does the Advanced Technology

Solar Telescope. Regardless of which database engine is used there will be a database interface layer to provide a suitable abstraction to the database clients.

Three distinct sets of clients have been identified:

- Viewers
- NGAO Components/Controllers/Applications (via services)
- Other

Five (possibly six) distinct databases have been identified:

- Configuration
- Logging
- Alarm History
- Archive
- Telemetry (polled RTC data may go here or may go to its own database)

1.1.1 Clients

1.1.1.1 Viewers

It is expected that there will be user interfaces (viewers) to allow easy viewing of logged information, archived data, alarm history and possible telemetry. KCSF will provide viewers for logged and archived data as well as alarm history. The need for a standardized telemetry viewer is TBD.

1.1.1.2 NGAO Components/Controllers/Applications (via services)

KCSF will provide services to all components, controllers and applications that use to framework. Many of these services interact with the data server. Each component will have the ability to retrieve its configuration from the configuration database. A configuration user interface will provide the ability to configure and re-configure the overall system. The log, archive and telemetry service will allow components to write appropriately formatted data to the corresponding service databases.

1.1.1.3 Other

It is TBD as to what these clients will be but it is expected that applications will be developed outside of the framework that want the ability to read and write to one or more databases and to be able to store and retrieve files form the server.

1.1.2 Databases

1.1.2.1 Configuration Database

The configuration database and associated support will be provided by KCSF. The database is quite complex and is covered in detail under KAON xxxxx.

1.1.2.2 Log Database

The log database and associated support will be provided by KCSF. The log database is simple and will have the following schema:

Field	Type	Description
Time_Stamp	timestamp	UT Timestamp

NGAO: Data Server Preliminary Design

Mode	VCHAR	Message type: e.g. Warning, Info, Error
Source	VCHAR	Who generated the message
Message	TEXT	The actual message

1.1.2.3 Alarm History Database

The alarm history database and associated support will be provided by KCSF. The alarm history schema is still TBD.

1.1.2.4 Archive Database

The archive database and associated support will be provided by KCSF. The archive database is simple and will have the following schema:

Field	Type	Description
Time_Stamp	timestamp	UT Timestamp
Source	VCHAR	Source of the archived data
Name	VCHAR	Attribute/Variable name
Value	Text	Archived value

1.1.2.5 Telemetry Database

The telemetry database schema is dynamic in that each telemetry topic will have its own table with column/field information specific to that specific telemetry item.

2 Hardware

Two hardware solutions are presented. The first is what is considered to be the optimal solution allowing for great performance, easy growth and a wide range of configuration options. The down side is that we are unfamiliar with the product and the potential increased initial cost. Option two is more a tried and tested solution that can do the job but is not quite a flexible or expandable. Regardless of the solution the expectation would be that the hardware will be located in the telescope computer room and will, most likely, be mounted in a standard 19" rack.

2.1 Optimal Solution

The optimal solution is to use Suns (Oracles) Unified Storage Solution which is a family of storage solutions designed for high performance and high availability.. The solution uses an optimized storage hierarchy with Hybrid Storage Pools containing DRAM, SSD, and HDD drives. To deliver high performance using cost-effective components, the Sun Storage 7000 series file system, ZFS, seamlessly optimizes access to the different types of media in the Hybrid Storage Pools. ZFS was designed to automatically recognize different I/O patterns and place data in the best storage media for optimal performance.

ZFS transparently executes writes to low-latency SSD media so that writes can be quickly acknowledged, allowing the application to continue processing. ZFS then automatically flushes the data to high-capacity drives as a background task. Another type of SSD media acts as a cache to reduce read latency, and ZFS also transparently manages the process of copying frequently accessed data into this cache to seamlessly satisfy read requests from clients.

The systems can be scaled as needed by adding additional RAM, SSD or hard disks. Scalability is available in multiple dimensions by increasing compute power, storage capacity, or performance independently. As application requirements change, it is possible to

- Increase computational power by adding more CPUs and cache
- Expand total capacity by adding enterprise-class SATA drive expansion units
- Increase performance using additional SSD to buffer storage reads or writes

The system is touted as having very low energy consumption with up to 75 percent less than the cost of competitive solutions. This is due to the reduced power consumption of SSD and HDD disk drives rather than the use of high RPM drives.

All systems come with a management interface that guides installation, configuration and tuning as well as DTrace Analytics software to identify and diagnose and issues. Real-time analysis and monitoring is supported.

To meet a variety of needs for capacity, price, and performance, the Sun Storage 7000 series family comes in three different configurations plus a cluster configuration that offers maximum availability. These are known as Sun Storage 7110, 7210, 7310 and 7410.

All non-clustered systems come bundled with the same software including data protocols, replication, compression, and DTrace Analytics software for system troubleshooting and performance optimization. Clustered systems include an additional software module for cluster software features.

2.1.1 Specifications

Sun Storage 7000 Series Family Configurations						
	Key Customer Requirement	Maximum Storage Capacity	Space (Rack Units)	Write Optimized SSD	Read Optimized SSD	Cluster Option
Sun Storage 7110 system	Low-priced entry-level system with all software features	2 TB or 4.2 TB (16 x 2.5 in. 146 GB or 300 GB disks)	2U	N	N	N
Sun Storage 7210 system	Midrange scalability and performance in a compact energy-efficient 4U system	48 TB (48 x 3.5 in. 1 TB disks) and up to 142 TB with two J4500 arrays*	4U	Up to 2 SSDs	N	N
Sun Storage 7310 system	Entry-level cluster option for high availability	96 TB	1U per node, 4U per storage array	Up to 16 SSDs for cluster configuration	Up to 6 per node (600 GB total)	Y
Sun Storage 7410 system	Best price/performance and maximum growth path	288 TB	2U per node, 4U per storage array	Up to 16 SSDs for cluster configuration	Up to 6 per node (600 GB total)	Y

*There are two specific configurations of the Sun Storage J4500 array that can be used with the Sun Storage 7210 system: a 24 TB array with 48 x 500 GB 7,200 RPM drives and a 48 TB array with 48 x 1 TB 7,200 RPM drives.

Product



[Sun Storage 7110 Unified Storage System](#) Updated!



[Sun Storage 7210 Unified Storage System](#)



[Sun Storage 7310 Unified Storage System](#) Updated!

Key Benefits

- Enterprise-class storage features at entry-level prices
- Up to 4 TB capacity, one of the most dense in its class
- Easy-to-use management interface and DTrace Analytics reduce management costs
- Enterprise-class data services with no additional fees

Availability

From \$10,995

- High-density unified storage solution
- 46 TB in energy-efficient 4U, scales up to 142 TB
- Optimize performance with DTrace Analytics for increased productivity
- Hybrid Storage Pool and Enterprise Flash technology deliver breakthrough performance while reducing power, cooling, and rack space

From \$34,995

- Entry-level cluster option for high availability
- Easy-to-use management interface and DTrace Analytics reduce management costs
- Hybrid Storage Pool and Enterprise Flash technology deliver breakthrough

From \$40,140

Product	Key Benefits	Availability
 Sun Storage 7410 Unified Storage System Updated!	performance while reducing power, cooling, and rack space	
	- Scale up to 96 TB to grow with your business - Highest performance with highly available cluster option - Easy-to-use Dtrace Analytics increases uptime while maximizing performance - Hybrid Storage Pool and Enterprise Flash improve application response times - Scales up to 288 TB (576 TB when 2 TB drive is available)	From \$73,165

2.2 Traditional solution

A more traditional approach (similar to what is used for NGWFC TRS) is to use a rack-mounted server such as the Sun Fire X4140 and a RAID storage solution such as the SurfRaid Triton 16FA4 RAID controller. They will be located in the telescope computer room and will be mounted in a standard 19" rack.

This proposes two main components, a storage server and a disk array. The storage server can be linked to the rest of the system through Fibre Channel or Gbit Ethernet. The preferred way to connect the disk array to the storage server will be through Fibre Channel, with the Fibre Channel configured as a point-to-point topology.

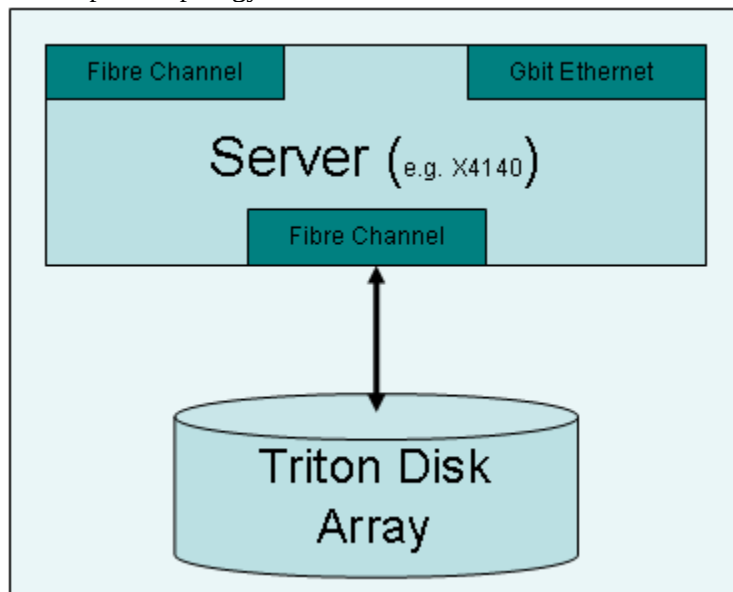


Figure 2 Data Server Hardware Architecture

TRS uses the discontinued Sun Fire X4100 and SurfRaid Triton 16FA (2Gb Fibre Channel). The Sun Fire x4140 and SurfRaid Triton 16FA4 are the current equivalents. For reference the performance tests carried out by Microgate achieved a sustained rate of 35.4 MB/sec on the TRS hardware.

Without a more detailed analysis of the AO Controls data handling needs these may not meet the requirements though there are other similar more performant hardware options. For example there are a complete line of servers with AMD in the Sun Fire X4x50 range and Intel in the Sun Fire X4x70 range. Likewise there are alternatives to the Triton offering 8 Gb Fibre Channel and greater storage capacity. If additional processor performance was needed a blade solution such as the Sun Blade X6270 could be used instead of the Sun Fire X4140.

The Sun Fire X4140 supports two 6 core Opteron processors (older model supported quad core) and significantly more RAM and the Triton 16FA4 supports up to 32 TB (older model supported 6 TB).

Pictures of the Sun Fire X4140 and the Triton 16FA4 are shown in Figure 3 and Figure 4 respectively. The maximum power dissipation is 650 watts for the X4140 and 460 watts for the Triton 16FA, for a total of 1100 watts maximum. This power will be dissipated in the telescope computer room, which is air conditioned.



Figure 3 Sun Fire X4140

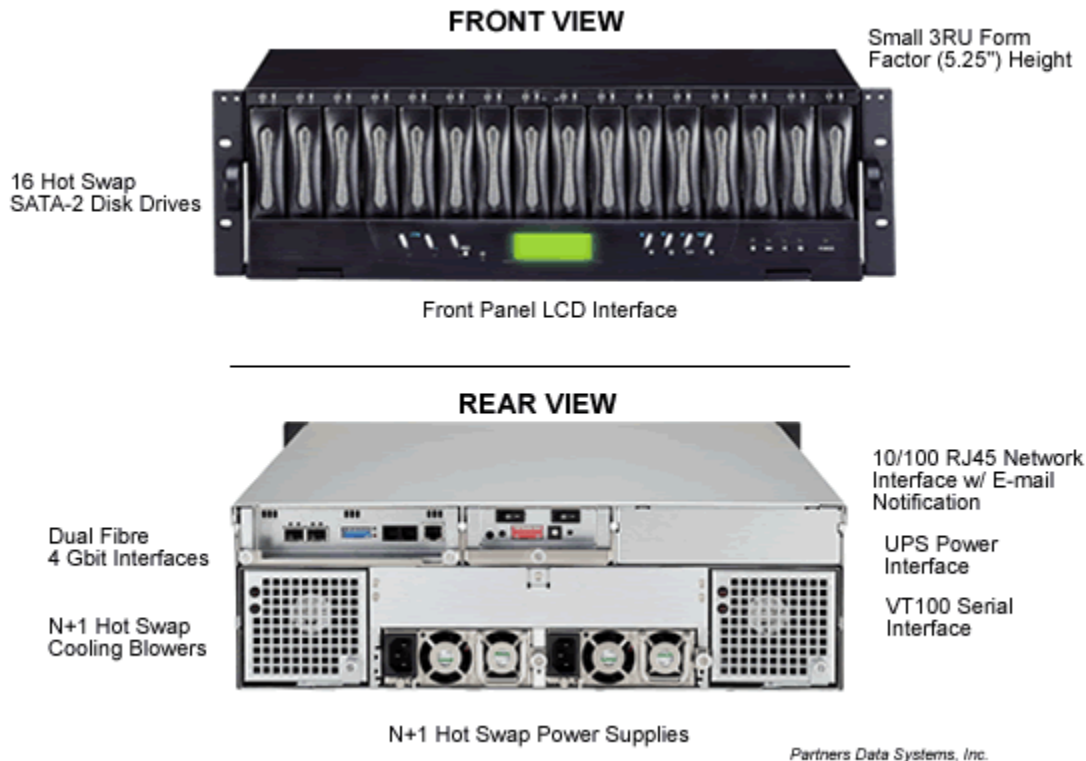


Figure 4 SurfRaid Triton 14FA4

3 Software

3.1 Relational database

The following shows a high level comparison between MySQL and Postgres.

Feature	MySQL 5.x	PostgreSQL
SubSelects	Yes	Yes
Views	Yes	Yes
Foreign Key relationships	Yes	Yes
Foreign Key constraints	Yes	Yes
Triggers	Yes	Yes
Indexing on non trivial types	Yes	Yes
Sequences	Some ¹	Yes
Transactions	Yes	Yes
Transactional DDL statements	No	Yes
OO (Inheritance of tables)	No	Yes
Async Notifications	No	Yes
Constraints	Yes	Yes
INSERT ... SELECT	Yes	Yes
Stored Procedures	Yes	Yes
Row level locking	Yes ²	Yes
Table level locking	Yes	Yes
Geospatial Datatypes (ie WKT)	Yes	No ³
Error! Hyperlink reference not valid.	No	Yes ⁴
Native replication (M/M, M/S, Ring)	Yes	No ⁵

Table 1 MySQL and Postgres comparison

In general the performance of the two databases is now comparable. Each has unique areas that they are stronger in. Postgres has better transaction support, MySQL has better large text area support.