ADASS XIII
ASP Conference Series, Vol. XXX, 2004
F. Ochsenbein, M. Allen and D. Egret eds.

High Availability Architecture for the Chandra Data Archive

P. Zografou, P. Harbo, K. McCusker, J. Moran, A. Patz,
P. Ramadurai, D. Van Stone

Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138

Abstract. The Chandra Data Archive is distributed at three physically remote locations, two of them in Cambridge, MA and a third in Leicester, UK. Each installation operates local hardware and a locally configured software release. The data are stored at a single location or in synchronized copies at multiple locations. The architecture enables processes to access the installation that is closest to the user or another installation if the first becomes overloaded or unavailable. This paper presents the archive architecture for the multiple installations. We explain the mechanisms that synchronize the data and we analyze the differences in data holdings across sites. We discuss how the software release is configured to operate at each installation and how users are routed to an installation depending on their profile. Finally, we describe the load balancing and failover mechanisms built into the archive.

1. Introduction

The Chandra Data Archive contains data from observations in the form of telemetry and data products. It also contains catalogs and operational data like observing proposals, mission planning schedules and Chandra users' information. The archive serves both as a storage area for existing data and as an active data store interacting with daily Chandra X-ray Center (CXC) operations. The large number of archive users may generate a heavy load at peak times or may request large data transfers to remote, slow networked locations. Operational processes that access the archive may require continuous availability of the data regardless of potential heavy load or system downtime. In order to address these needs, the archive was designed to operate at multiple installations, each of which is configurable to the needs of different groups of users.

2. Archive Installations

A single archive installation consists of an RDBMS server that manages databases and one or more archive servers that manage files. The server software is configurable at runtime to operate at a specific installation. An installation can operate with all or a subset of the archive data holdings. A number of archive

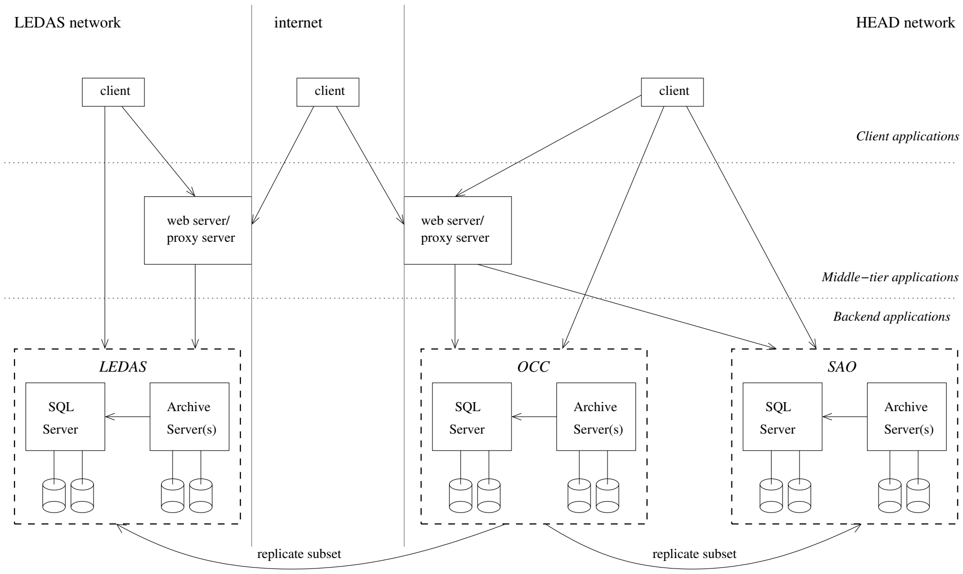

Figure 1. Archive Installations

installations can operate simultaneously at the same or at remote locations. They differ by the IP address or the port of the servers.

Archive installations are currently in production at three different locations (Figure 1). The *SAO* and *OCC* installations are at the Center for Astrophysics (CfA) and at the Chandra Operations Control Center (OCC), within the High Energy Astrophysics Division (HEAD) network in Cambridge, MA. The *LEDAS* installation is at the Leicester Database and Archive Service (LEDAS), in the UK.

3. Replication Mechanism

The replication is one-directional. Data are entered in the archive at a primary installation and are replicated to secondary installations. The primary installation is the one nearest to the data production site.

The Sybase ASE 12.5 SQL and replication servers are used for the storage and replication of relational data. The replication server monitors the activity of the SQL server at the primary installation and repeats recorded transactions at the subscribing secondary installations.

Data files entered in the primary archive server cause entry of metadata in the primary SQL server. When a row of metadata is replicated to a secondary SQL server it triggers a call to the secondary archive server to transfer the data file from the primary archive server (Figure 2). The archive server is a Sybase 12.5 Open Server application and can receive RPC calls from the SQL server. It is also a Sybase 12.5 Open Client application, which allows it to connect to the primary archive server and retrieve files.

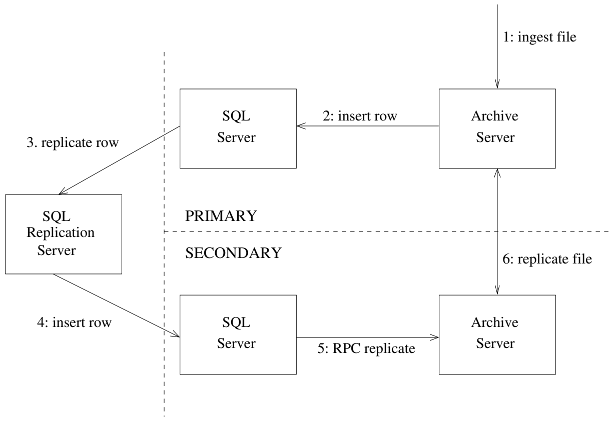

Figure 2. Replication Data Flow

4. Data Location

Data are stored at single or multiple locations according to their type.

All data are stored at their primary installation where they are ingested in the archive by data production operations. For example, the *OCC* installation at the site where telemetry is received and processed, is the primary for telemetry and pipeline products; the *SAO* installation at the site where the User Support group is located, is the primary for proposals and user information.

Frequently accessed data are replicated to one or more installations in order to provide better network access to users and decouple them from CXC operations. In this category belong the high-level data products which are accessed by remote clients and are replicated to *SAO*. The public subset is also replicated to *LEDAS*.

Mission-critical data are replicated for continuous availability, in case the primary installation fails. Mission Planning data can be accessed by operations at both the *OCC* and *SAO* installations.

A list of datatypes and their location is available to clients connecting to the archive servers. The list points to two installations for each replicated datatype or a single one for non-replicated data. The installation names are parameters that are set for each client at runtime.

5. Software Configuration

All client, server and middle-tier components are bundled in the same archive software release. All backend server installations need to run the same release version for components that affect replicated data.

The software, including the backend and middle-tier servers and all their clients, is configured at runtime by a number of parameters. The server parameters determine the archive installation where the server runs. The client parameters determine the archive installations where the client connects. The configuration for all known archive installations is included with the release runtime environment. Each server or client process is automatically assigned a set of parameter values. The values are stored in environment variables and are determined by the user ID for the backend servers and by the IP address for clients. For middle-tier Java servers, the values are stored as properties when a release is installed. The same properties are also used to configure remote Java clients.

6. Failover and Load Balancing

Clients browse and retrieve data at their nearest available installation. If this is not available because of failure or heavy usage, they failover to a different installation, if it exists for the requested data. A client may failover at the beginning or at any time during a session and fallback to the first installation if it becomes available during the same session.

7. Conclusion

The *OCC* and *SAO* archive installations, both within the CfA subnet, have been operating successfully since the Chandra launch, serving end-users and operations. More recently, the *LEDAS* installation was added to provide closer access for users in Europe.

With the growing size of the public archive, there is interest in establishing more installations to reach users in Asia and other parts of the world. While the mirror archives fulfill all the initial requirements, they are costly to establish and operate. Furthermore, they include features like proprietary rights checking, user authentication and operations support which are not needed in a public data archive. A simplified approach for remote archive installations is currently under development. In this approach, a remote archive is a public FTP directory. Clients use the primary installation to browse the contents of the archive and to submit retrieval requests. The primary installation forwards the request to the FTP server installation nearest to the user. The first FTP archive installation already exists at the CfA. More FTP archives in Europe and Asia are planned for the near future.

Acknowledgments. This project is supported by the Chandra X-ray Center under NASA contract NAS8-39073.

References

Patz et al 2002, in APS Conf. Ser., Vol. 295, Astronomical Data Analysis Software and Systems XII, ed. H. E. Payne, R. I. Jedrzejewski, & R. N. Hook (San Francisco: ASP), 249

Estes et al 1999, in ASP Conf. Ser., Vol. 216, Astronomical Data Analysis Software and Systems IX, ed. N. Manset, C. Veillet, & D. Crabtree (San Francisco: ASP), 457

Zografou et al 1997, in ASP Conf. Ser., Vol. 145, Astronomical Data Analysis Software and Systems VII, ed. R. Albrecht, R. N. Hook, & H. A. Bushouse (San Francisco: ASP), 391