

Data Release DR5 Overview and Operating Experience

Svetlana Lebedeva

Contents

- The SkyServer
- SkyServer Production Cluster
- SDSS Web Server Cluster
- SDSS Data Server Cluster
- Security
- Suggested Configuration for mirror DR5
- Summary

SDSS Mirror Archive
Planning Meeting
The John Crerar Library,
University of Chicago
May 23, 2007

The SkyServer

- The SkyServer is an Internet portal to the multi-Terabytes Sloan Digital Sky Survey Catalog Archive Server (CAS).
 - ◆ The CAS is a collection of SQL Server databases each storing a particular “release” of the SDSS data.
 - ◆ It's accessible online to astronomers and the general public via two Web portals:
 - SkyServer.sdss.org
 - or cas.sdss.org.
 - ◆ The SkyServer is deployed on MS Windows servers at Fermilab.

The SkyServer (continue)

- The SkyServer is basically a federation of Websites that serve different communities and functions:
 - ◆ **cas.sdss.org:**
 - public Website offering access to the SDSS data, documentation on the data, and online-astronomy education in six languages (English, Japanese, German, Portuguese, Spanish, and Hungarian).
 - Access to DR5: <http://cas.sdss.org/dr5/en/>
 - ◆ **Collaboration and Astronomer portals:**
 - separate Websites operated for members of the SDSS collaboration with restricted access that allow longer-running queries on dedicated hardware.
 - The user interface is streamlined for use by professional astronomers, and collaboration members usually have exclusive access to each data release for a few months prior to its public availability.
 - DR5 Astro URL: <http://cas.sdss.org/astrodr5/en/>
 - DR5 Collab URL: <http://cas.sdss.org/collabdr5/en/>

[The SkyServer \(continue\)](#)

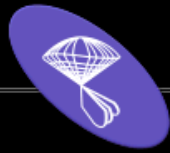
- ◆ **Imagecutout:**
 - Visual exploration tools built for the SkyServer to visualize and explore detailed astronomical data in regions of the sky covered by the SDSS.
 - The regions of interest are specified by central position and size. An underlying web service then combines the relevant images to form a JPEG mosaic* at the requested wide range of resolutions.
 - DR5 Image List Tool: <http://cas.sdss.org/dr5/en/tools/chart/list.asp>

The SkyServer (continue)

- ◆ QBuilder: There are several tools for searching SDSS database:
 - Radial: Search the sky around a given point (EASY)
 - Rectangular: Search the sky in a rectangular region (EASY)
 - Search Form: Search the sky by selecting choices from an online form (EASY)
 - Query Builder: Build a SQL query by clicking on various menus (HARDER)
 - SQL: Type in a SQL query directly (HARDEST)
 - Query Builder is different from other search tools because it let to build up a SQL query interactively step-by-step clicking on different part of the query. It is more powerful and flexible than the Search Form tool, but can't do as much as the SQL Search tool.
 - All queries from SkyServer search tools are restricted by timeouts and row limits.
 - CasJobs batch query service allows running a query without such restrictions.

[The SkyServer \(continue\)](#)

- Virtual Observatory:
 - ◆ A collection of Web services being developed by the Astronomy community as part of their efforts to build the World-Wide Telescope. It is not part of the SkyServer, it's deployed on the servers at JHU, but VO traffic appears in our Web logs and in the web configuration.



The SkyServer (continue)

■ The Web and SQL logs:

- ◆ The Web and SQL activity logs have been archiving from the Fermilab and JHU servers. A collector running at JHU harvests the logs every few hours from across the Internet using a Web services interface offered by each SkyServer and CasJobs server. The harvested logs are aggregated in a publicly accessible database along with an activity summary. Collaboration queries are hidden from public view but are also included in Web and SQL summary.

■ CasJobs - batch jobs for the CAS:

- ◆ A public Web service that allows users to create a personal database (MyDB) on a server at Fermilab, upload personal datasets to it, and submit long-running programs and SQL queries that convolve MyDB data with the CAS datasets
- ◆ CasJobs deployed on its own portal: casjobs.sdss.org.

SkyServer Production Cluster

- The skyServer is built in the Windows environment, using MS SQL Server database.
- It contains the Full Object Catalog, with redshifts and spectral lines
- Full set of color images in compressed JPEG format
- GIF images of all the spectra, with detected spectral lines graphically marked.
- This server was developed jointly by MS Bay Area Research Center and JHU.
- Very early release started with 80GB with 14M objects, 50K spectra.
- Now DR5 is 2.15TB with 215M photo objects, 0.9M spectra, and ~10B rows spread among ~400 tables.



SkyServer Production Cluster (continue)

- Types of data:
 - ◆ SkyServer offers access to many different types of data, but most users focus on four types: images, spectra, photometric data, and spectroscopic data.
 - ◆ Images: Pictures of celestial objects, stored as .jpg or .fits files
 - ◆ Spectra: Graphs of the amount of light emitted by an object at different wavelengths. Spectra can be used to study the compositions and temperatures of objects, as well as their distances from Earth.
 - ◆ Photometric Data: Various quantities measured from an object's image, such as its brightness or size
 - ◆ Spectroscopic Data: Various quantities measured from an object's spectrum, such as its spectral type or redshift.



SDSS Web Server Cluster

- OS: Windows 2003 Server Standard Edition, SP1
- IIS 6.0 with 2 pools: asp2.net pool and default 1.1.43
- Disk space for C:\drive ~ 100GB to deploy web front-end configuration data for SDSS Web portals
- Network Bandwidth - 1 Gbps
- RAM - 2 GB
- RAID configuration - 10

[SDSS Web Server Cluster \(continue\)](#)

SDSS front-end configuration for web interface.

WebService	Configuration file
Astro	\\webserver\C:\skyserver\asrto.DR5\global.inc
Collab	\\webserver\C:\skyserver\collab.DR5\global.inc
Public	\\webserver\C:\skyserver\public.DR5\global.inc
VO Service	\\webserver\C:\Inetpub\wwwroot\vo\ Web.config files in DR5CONE, DR5Fields, DR5ISP
QBuilder	\\webserver\C:\skyserver\public.DR5\en\tools\qbuilder\web.config
Imagecutout	\\webserver\C:\Inetpub\wwwroot\DR5_imagecutout\web.config
CasJobs	\\webserver\C:\Inetpub\wwwroot\CasJobsFNAL\web.config \\webserver\C:\Inetpub\wwwroot\CasJobsFNAL\ JobService\bin\Release\jobservice.config

SDSS Data Server Cluster

- OS - Windows 2003 Server Standard Edition, SP1
- Network Bandwidth - 1 Gbps
- RAM - 4 GB, 3GB RAM for SQL Server
- RAID configuration - 10
- SQL Server 2005, SP2
- Disk space - minimum 4.5TB:
 - ◆ 100GB for Tempdb and 20GB to Tempdb log file
 - ◆ 2.2 TB for BestDR5
 - ◆ 1.8 TB for TargDR5
 - ◆ 3GB for Weblog



Security

- Computer security policies and procedures of the institution.
- Protection against corruption by SDSS users
 - ◆ Use of commercial MS SQL Server database ensures data protection
- Protection against unauthorized access by non-SDSS participants
 - ◆ DB server not directly accessible
 - ◆ Access only via web (HTTP) interface
 - ◆ Web pages IP-restricted or password-protected
- SQL Server security audit on monthly basis.
- Support: 24 by 7



Suggested configuration for DR5 Web and Data Servers

Web Server

MS Windows 2003

Raid 10

IIS 6.0

Astro, Collab

*Public, Imagecutouts,
qBuilder, VO service*

SDSS Data Server 1

MS Windows 2003

Raid10 Capacity 5.6 TB

SqlServer 2005

BestDR5 - 2.15TB

TargDR5 - 1.80TB

Weblog - 3GB

SDSS Data Server 2

MS Windows 2003

Raid10 Capacity 5.6 TB

SqlServer 2005

BestDR5 - 2.15TB

TargDR5 - 1.80TB

Weblog - 3GB



Summary

- .Over last few years we successfully operate and support SDSS Data Releases from DR2 to DR6.
- The major fractions of our efforts are directed to preparing the Data Releases
- As Data Release became public there is no need in major upgrades on our system
- Support requires MS Security patches on regular basis, monitoring tools and SqlServer maintenance plans to overview the performance.

Resources:

<http://www.cas.sdss.org/>

<http://research.microsoft.com/research/pubs/view.aspx?type=Technical%20Report&id=1236>