

The SuperCOSMOS Science Archive

Nigel Hambly, Mike Read, Bob Mann, Eckhard Sutorius, Ian Bond,
Harvey MacGillivray, Peredur Williams, Andrew Lawrence

*Wide Field Astronomy Unit, Institute for Astronomy, University of
Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK*

Abstract. The SuperCOSMOS Sky Survey (SSS¹; Hambly et al., 2001) consists of digitised scans of Schmidt photographic survey material in a multi-colour (BRI), multi-epoch, uniformly calibrated product. It covers the whole southern hemisphere, with an extension into the north currently underway. Public online access to the 2 Tbytes of SSS pixel data and object catalogues has been available for some time; data are being downloaded at a rate of several gigabytes per week, and many new science results are emerging from community use of the data. In this poster we describe the terabyte-scale SuperCOSMOS Science Archive² (SSA), which is a recasting of the SSS object catalogue system from flat files into an RDBMS, with an enhanced user interface. We describe some aspects of the hardware and schema design of the SSA, which aims to produce a high performance, VO-compatible database, suitable for data mining by ‘power users’, while maintaining the ease of use praised in the old SSS system. Initially, the SSA will allow access through web forms and a flexible SQL interface. It acts as the prototype for the next generation survey archives to be hosted by the University of Edinburgh’s Wide Field Astronomy Unit, such as the WFCAM Science Archive of infrared sky survey data, as well as being a scalability testbed for use by AstroGrid, the UK’s Virtual Observatory project. As a result of these roles, it will display subsequently an expanding functionality, as web – and later, Grid – services are deployed on it.

1. Relational Model for the SSA Data

The photographic material used in the SSA comprise the SERC J/EJ, ER/AAO–R and I original surveys along with first epoch R data from the ESO–R and POSS–I E copies in the southern hemisphere; ultimately, the northern hemisphere will be included via POSS–II J, R and I copies with POSS–I E copies again providing early epoch R data. Hence, 8 single colour surveys on 3 different field systems (ESO/SRC, POSS–I and POSS–II) make up the source material. The rest of the schema follows from the several individual colour/epoch plates in

¹<http://www-wfau.roe.ac.uk/sss>

²<http://thoth.roe.ac.uk/ssa>

each field, each plate giving rise to a set of detections, some of which are flagged as spurious (eg. Storkey et al. 2003), and all of which are merged into multi-colour, multi-epoch sources (merging allows computation of colour-corrected magnitudes and proper motions). In this way, all information concerning the source photographic material, its measurement on SuperCOSMOS and subsequent processing by in-house software is available as metadata in the schema, in addition to the parameterised detection and source attributes. The raw data volume is dominated by two tables: the Detection table contains 3.7 billion rows; with 228 bytes per row, it is 0.86 Tbyte in size; the Source table contains just over 1 billion rows; with 246 bytes per row, it is 0.26 Tbyte in size. So the raw data volume is a little over 1.1 Tbyte before adding in DBMS overheads such as indexing.

2. Hardware Design

The hardware design of the catalogue server for the terabyte scale SSA is based around maximising IO bandwidth in order to service user queries as fast as possible. The SSA catalogue server uses a Tyan Thunder PC motherboard based on dual Xeon processors (2.8 GHz) and an associated Intel chipset. The bus architecture of the motherboard incorporates three independent PCI-X (64 bit, 133 MHz) busses for the highest potential aggregate IO bandwidth. Other aspects of the science archive hardware design that are relevant to maintenance of Tbyte data volumes include 1 Gbit/s LAN connectivity and an Ultrium-II LTO tape backup facility (encapsulated in an Overland 30-slot library system) which is capable of backing up Tbytes in several hours.

Design of the disk subsystem is important for high aggregate IO applications. We have experimented with a number of disk array configurations along with tests using different interfaces (IDE and SCSI) and different controllers (hardware RAID, fibre-to-IDE, Ultra-SCSI etc). Our findings were in agreement with those of others (eg. Gray et al. 2002) in that hardware RAID fault tolerance comes at a significant performance cost – a simple design based around software striping over multiple Ultra-SCSI channels yields the best IO performance at reasonable cost. Hence, our SSA catalogue server employs four dual channel Adaptec Ultra320 SCSI controllers on each of the three external and one on-board PCI-X motherboard interface slots. We have attached four Seagate 15 krpm, 150 Gbyte Ultra320 SCSI disks to each of the eight controller channels – measurements using MemSpeed show that individual disks sustain nearly 60 Mbyte/s read/write speed on a single channel, but also that software stripe IO performance saturates at around 200 Mbyte/s (ie. 4 disks). The 32 disks yield 4.8 Tbyte of storage that is arranged in four logical volumes of 1.2 Tbyte each with 8 disks (one on each independent SCSI channel) used in each stripe set. Because there is no fault tolerance in this design (ie. this is a RAID0 configuration), we are mirroring database files on a separate system in addition to backing up on LTO-2.

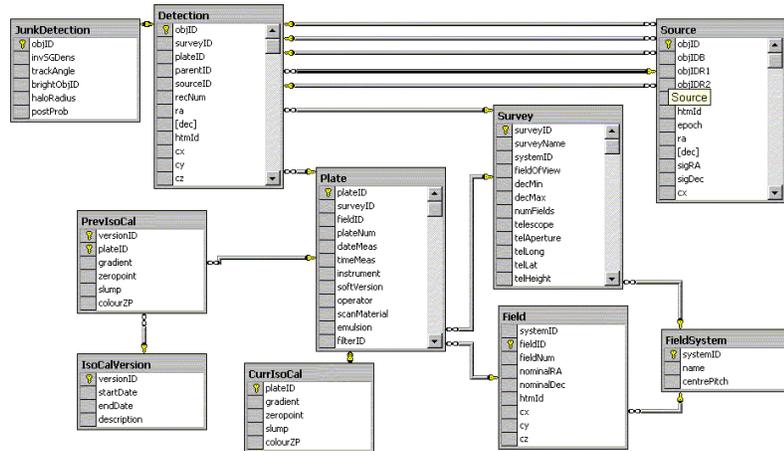


Figure 1. Database picture of the SSA tables generated in MS SQL Server.

3. OS, RDBMS and Software

Implementation of the SSA leans heavily on developments in science archiving for the Sloan Digital Sky Survey at Johns Hopkins (Gray et al. 2002 and references therein). The SSA is deployed within Microsoft SQL Server running on the Windows 2003 Server operating system. Figure 1 illustrates the full SSA schema as implemented within SQL Server; primary keys and foreign key relationships between the tables are shown. Some software from the SDSS SkyServer design have been used, most notably the 2d spatial indexing scheme known as Hierarchical Triangular Mesh³. We have used Windows software striping as described previously to set up high IO bandwidth disk arrays, and have additionally used file groups to distribute all large database files across the logical devices.

4. Data Loading

Small data files are easily imported into RDBMSs like SQL Server using comma-separated ASCII files. However, for terabyte volumes, numerical data files formatted in this way become impractically large. We have designed a binary loading scheme for the SSA that uses native format files (ie. 4 byte floats occupy 4 bytes in the ingest files; short integers occupy 2 bytes etc). This has enabled efficient transformation, transfer and ingest of the data stored in the pre-existing SSS flat files. The native binary files are typically less than 1/3rd the size of their ASCII equivalents making staging and network transfer significantly easier. Furthermore, loading of the native files is much faster. We employed a minimally logged, heaped load using BULK INSERT within SQL Server, attaching primary key constraints after the loading process has finished.

³<http://www.sdss.jhu.edu/htm>

5. User Interfaces

Currently, user access to the SSA is via browsable web forms. These action Java servlets which parse the query, connect to the MS SQL Server and retrieve and format the results set. Features of the interface include choice of traditional radial search, form-filled and free-form SQL queries and a catalogue cross-match facility; choice of output formats (HTML table summary, comma-separated ASCII, FITS binary and VOTable) with complete control over returned attribute sets; and links to the existing SSS pixel image facilities. Future enhancements to the SSA will include deployment of web services as alternative client access points, and ultimately Grid services as we enter the next phase of science archiving at WFAU: implementation of the WFCAM Science Archive.

6. The WFCAM Science Archive

The next large scale imaging datasets that will be curated and archived at WFAU will be the survey programmes undertaken with WFCAM, a large format infrared imager for the UK Infrared Telescope UKIRT. Because the UKIDSS survey programme⁴ is more complex than the legacy Schmidt surveys, the relational model for the associated science archive (hereafter, the WSA⁵) is far more complicated. Furthermore, curation of the programmes (eg. source association, final photometric calibration, computation of derived quantities like proper motion) will be database-driven and will take place within the RDBMS (unlike analogous procedures for the SSA). Other features of the WSA relational model include provision for database driven products, eg. difference images, image stacks and image mosaics; and provision for pixel re-analysis given master lists of positions and apertures (ie. SDSS-like consistent image analysis across all colours available in a given field). WSA development also includes investigation of scale-out issues for 10s to 100s of Tbytes of data, eg. choice of horizontal (clustered) versus vertical (multiprocessor) hardware scaling; spatial indexing, eg. HTM versus others, eg. HEALPix; and investigation of other enterprise class RDBMSs, ie. DB2 and Oracle.

Acknowledgments. We would like to thank Jim Gray and Alex Szalay for much advice and help concerning design and implementation of the SSA. Funding for WFAU is provided by the UK PPARC. We acknowledge software and support from the Microsoft Development Network Academic Alliance.

References

- Gray, J. et al. 2002, *Data Mining the SDSS SkyServer Database*, Microsoft Technical Report MSR-TR-2002-01
Hambly, N. et al. 2001, MNRAS, 326, 1279
Storkey, A. et al. 2003, MNRAS, in press (astro-ph/0309565)

⁴<http://www.ukidss.org/>

⁵<http://www.roe.ac.uk/~nch/wfcam>