Overview of the SDSC Storage Resource Broker

Wayne Schroeder
(and other SRB team members)

May, 2004

San Diego Supercomputer Center,
University of California San Diego
The SDSC Storage Resource Broker (SRB) is client-server middleware that provides a uniform interface for connecting to heterogeneous data resources over a network and accessing unique or replicated data objects.

SRB, in conjunction with the Metadata Catalog (MCAT), provides a way to access data sets and resources based on their logical names or attributes rather than their names and physical locations.
• The SDSC SRB system is a comprehensive distributed data management solution, with features to support the management, collaborative (and controlled) sharing, publication, and preservation of distributed data collections.

• The SRB also serves as middleware via a rich set of APIs available to higher-level applications and by providing a management layer on top of a wide variety of storage systems.
The SRB is an integrated solution which includes:
– a logical namespace,
– interfaces to a wide variety of storage systems,
– high performance data movement (including parallel I/O),
– fault-tolerance and fail-over,
– WAN-aware performance enhancements (bulk operations),
– storage-system-aware performance enhancements ('containers' to aggregate files),
– metadata ingestion and queries (a MetaData Catalog (MCAT)),
– user accounts, groups, access control, audit trails, GUI administration tool
– data management features, replication
– user tools (including a Windows GUI tool (inQ), a set of SRB Unix commands, and Web (mySRB)), and APIs (including C, C++, Java, and Python).

SRB Scales Well (many millions of files, terabytes)

Supports Multiple Administrative Domains / MCATs (srbZones)

And includes SDSC Matrix: SRB-based data grid workflow management system to create, access and manage workflow process pipelines.
SRB Projects

- **Digital Libraries**
  - UCB, Umich, UCSB, Stanford, CDL
  - NSF NSDL - UCAR / DLESE
- **NASA Information Power Grid**
- **Astronomy**
  - National Virtual Observatory
  - 2MASS Project (2 Micron All Sky Survey)
- **Particle Physics**
  - Particle Physics Data Grid (DOE)
  - GriPhyN
  - SLAC Synchrotron Data Repository
- **Medicine**
  - Digital Embryo (NLM)
- **Earth Systems Sciences**
  - ESIPS
  - LTER
- **Persistent Archives**
  - NARA
  - LOC
- **Neuro Science & Molecular Science**
  - TeleScience/NCMIR, BIRN
  - SLAC, AfCS, …

Over 90 Tera Bytes in 16 million files
## SRB Scalability

### Storage Resource Broker (SRB)

**Data brokered by SDSC instances of SRB**

<table>
<thead>
<tr>
<th>Project Instance</th>
<th>As of 7/24/2003</th>
<th>As of 9/12/2003</th>
<th>As of 10/01/2003</th>
<th>As of 11/14/2003</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count (files)</td>
<td>Users</td>
<td>Count (files)</td>
<td>Users</td>
</tr>
<tr>
<td>NPACI</td>
<td>6,050.00</td>
<td>2,317,368</td>
<td>367</td>
<td>8,350.00</td>
</tr>
<tr>
<td>Digsky</td>
<td>46,100.00</td>
<td>5,719,025</td>
<td>68</td>
<td>46,100.00</td>
</tr>
<tr>
<td>DigEmbryo</td>
<td>720.00</td>
<td>45,365</td>
<td>23</td>
<td>215.00</td>
</tr>
<tr>
<td>HyperLer</td>
<td>968.00</td>
<td>27,250</td>
<td>316</td>
<td>1,133.00</td>
</tr>
<tr>
<td>Hayden</td>
<td>7,078.00</td>
<td>59,399</td>
<td>142</td>
<td>1,790.00</td>
</tr>
<tr>
<td>Portal</td>
<td>232.00</td>
<td>15,809</td>
<td>23</td>
<td>92.00</td>
</tr>
<tr>
<td>SLAC</td>
<td>121.00</td>
<td>237,283</td>
<td>138</td>
<td>273.00</td>
</tr>
<tr>
<td>AICS</td>
<td>95.30</td>
<td>18,762</td>
<td>20</td>
<td>99.00</td>
</tr>
<tr>
<td>UCSDLib</td>
<td>1,084.00</td>
<td>138,415</td>
<td>29</td>
<td>1,085.00</td>
</tr>
<tr>
<td>NSDL/CI</td>
<td>278.00</td>
<td>993,886</td>
<td>113</td>
<td>379.00</td>
</tr>
<tr>
<td>SCEC</td>
<td>720.00</td>
<td>18,660</td>
<td>38</td>
<td>7,561.00</td>
</tr>
<tr>
<td>TeraGrid</td>
<td>623.00</td>
<td>36,508</td>
<td>1,978</td>
<td>1,644.00</td>
</tr>
<tr>
<td>TOTAL</td>
<td>66,008.30</td>
<td>9,979,239</td>
<td>3,461</td>
<td>77,696.10</td>
</tr>
</tbody>
</table>

**TOTAL**

| 66 TB           | 9.97 million    | 3 thousand      | 77 TB           | 13.9 million    | 3 thousand      | 81 TB           | 15.6 million    | 3 thousand      | 90 TB           | 16 million      | 3 thousand      |

**Does not cover data brokered by SRB spaces administered outside SDSC.**

**Does not cover databases; covers only files stored in file systems and archival storage systems**

**Does not cover shadow-**
Case Study: SRB in BIRN
SRB History

- A DataGrid since SRB 1.0, Production 1997
- SDSC Started by General Atomics, 1985
  - GA/UCSD Staff
  - On UCSD Campus
  - SRB by GA Employees
- Today, SDSC no longer GA, all UCSD
  - All staff UCSD employees
- GA Commercial SRB Version (Nirvana)
  - Based on SRB 1.1.8 (2001)
  - Nirvana and SDSC versions diverged
  - SDSC SRB free to academic organizations
  - License from Nirvana for commercial
SRB – A Data Grid Solution

• Storage Resource Broker
  – Collaborative client-server system that federates distributed heterogeneous resources using uniform interfaces and metadata
  – Provides a simple tool to integrate data and metadata handling – attribute-based access
  – Blends browsing and searching
  – Developed at SDSC
    - Operational for 5+ years;
    - Under continual development since 1997;
    - Customer-driven;
    - Brokering over 90 TeraBytes in over 16 million files at SDSC
• Data Grid has arbitrary number of servers
• Complexity is hidden from users
## SDSC Storage Resource Broker & Meta-data Catalog

### Application

<table>
<thead>
<tr>
<th>Resource, User</th>
<th>User Defined</th>
<th>Resource, User</th>
<th>User Defined</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCAT</td>
<td></td>
<td>MCAT</td>
<td></td>
</tr>
<tr>
<td>Dublin Core</td>
<td></td>
<td>Dublin Core</td>
<td></td>
</tr>
<tr>
<td>Application Meta-data</td>
<td></td>
<td>Application Meta-data</td>
<td></td>
</tr>
<tr>
<td>Archives</td>
<td></td>
<td>Archives</td>
<td></td>
</tr>
<tr>
<td>HPSS, ADSM, UniTree, DMF</td>
<td></td>
<td>HPSS, ADSM, UniTree, DMF</td>
<td></td>
</tr>
<tr>
<td>File Systems</td>
<td></td>
<td>File Systems</td>
<td></td>
</tr>
<tr>
<td>Unix, NT, Mac OSX</td>
<td></td>
<td>Unix, NT, Mac OSX</td>
<td></td>
</tr>
<tr>
<td>Databases</td>
<td></td>
<td>Databases</td>
<td></td>
</tr>
<tr>
<td>DB2, Oracle, Sybase</td>
<td></td>
<td>DB2, Oracle, Sybase</td>
<td></td>
</tr>
<tr>
<td>C, C++, Linux I/O</td>
<td></td>
<td>C, C++, Linux I/O</td>
<td></td>
</tr>
<tr>
<td>Unix Shell</td>
<td></td>
<td>Unix Shell</td>
<td></td>
</tr>
<tr>
<td>Java, NT Browsers</td>
<td></td>
<td>Java, NT Browsers</td>
<td></td>
</tr>
<tr>
<td>Prolog Predicate</td>
<td></td>
<td>Prolog Predicate</td>
<td></td>
</tr>
<tr>
<td>Web</td>
<td></td>
<td>Web</td>
<td></td>
</tr>
<tr>
<td>Third-party copy</td>
<td></td>
<td>Third-party copy</td>
<td></td>
</tr>
<tr>
<td>Remote Proxies</td>
<td></td>
<td>Remote Proxies</td>
<td></td>
</tr>
<tr>
<td>DataCutter</td>
<td></td>
<td>DataCutter</td>
<td></td>
</tr>
</tbody>
</table>
Federated SRB Operation

1. Logical-to-Physical mapping
2. Identification of Replicas
3. Access & Audit Control

Server(s) Spawning

Parallel Data Access

Peer-to-peer Brokering

Data Access

Logical Name Or Attribute Condition

Read Application in Boston

1. Logical-to-Physical mapping
2. Identification of Replicas
3. Access & Audit Control

R1

MCAT

R2

SRB server

SRB server

SRB server

SRB agent

SRB agent

SRB agent

San Diego

Durham
Virtual Hierarchical Collection Management
Attributes

- **SRB metadata**
  - Location, protocol
  - Unix semantics
  - Authorization, authentication
  - Latency management
  - Container aggregation
- **Administrative**
  - Dublin core, provenance
  - Annotations, comments
- **Discipline specific attributes**
  - Collection
  - User defined
- Grid Security Infrastructure (GSI)
- Encrypted Password
- GSS-API for Kerberos or DCE
- Collection-owned Data
  - Collection ID installed at each storage system
  - Users authenticate themselves to the SRB
  - SRB authenticates to local server
  - Or GSI Delegation (Ananta Manandhar, CCLRC)
One of the major functions of SRB is the mapping between a logical file name and its physical file. The mapped info of a logical filename includes:

- Location of name in collection hierarchy
- Physical file location: host name and path
- Protocol: for fetching ‘local’ file
- Unix semantics for file manipulation
- Location in container
- Audit trail
- Access control list
- Locking status
Replica Management

- Files can be replicated into any valid physical storage resource registered in SRB.
- Each replica is managed by the same logical filename as the original one and a unique replication number. Each replica can have unique metadata.
- 1-to-many Replication: A logical resource can contain several physical storage resources.
- Multiple replicas can be made to the same storage resource.
- Many Modes of Replication:
  - Synchronous Replication; Sput to a logical resource
  - Asynchronous Replication; Sput then later Sreplicate
  - Out of Band Replication; Outside SRB, then register
Containers

• Physical Grouping of Objects
• Similar to tar but has significant differences
• Multiple Uses:
  – To take advantage of resource characteristics
  – To aid access patterns
  – Move data sets together
  – Tie together logically different files
  – Automatic Archiving/Caching
• Chaining of Containers
• Sharing of metadata
• Containers for Collections
Proxy Operation

• Proxy operation -
  – server performs operations on behalf of client
  – performs operations where the data are located
  – subset and filter operations – datacutter
  – Metadata extraction and ingestion checks
  – srbExecCommand() API and Spcommand utility -
    • request a specific server to execute a specific command and stream the result to stdin
    • used by the NVO(national virtual observatory) cutout service
SRB – More Features

• Client Support
  – Pure Java Client
  – Web Services - WSDL, Matrix workflow system
  – Web Support - MySRB Extensions
  – Pure Java Client & Browser
  – inQ Version 3.1 and more Windows Support

• Administrative Support
  – GUI-based Administration
  – More Features - Resource, User, Method Management
  – User-friendly Installation Procedures
Metadata Management

- Metadata Insertion Through User Interfaces
- Bulk Metadata Insertion
- Template Based Metadata Extraction
- Metadata Search
  - system data
  - user-defined metadata
  - File Content Search: Key words are pre-extracted by a template and saved as user-defined metadata.
SRB wears many hats:
- It is a distributed but unified file system
- It is a database access interface
- It is a digital library
- It is a semantic web
- It is a data grid system
- It is an advanced archival system
Criticisms of SRB

• Not completely open source
  – But semi-open and available to academics
• Not standards-based
  – But internal protocols need not be
• Monolithic
  – Integrated
  – And well partitioned
Some SRB Weaknesses (my view)

• Difficult to explain and understand
  – SRB does so much, people tend to learn subsets and are often unaware of useful features
  – Different groups are interested in different sets of features
  – An “elevator speech” is either vague or incomplete

• Not completely open source

• Collaborations difficult
  – Need to expand

• Limited Staff
  – Feature-focused projects (+/-): docs, error messages
Some SRB Strengths

- Integrated solution
  - High performance
  - Highly functional
  - Relatively easy to enhance
- Middle-ware and Complete-ware
- Customer driven
- Sound architecture
- Mature, but also being actively developed
- Growing user base
- Highly coordinated centralized team
TeamSRB, San Diego

- Reagan Moore (Program Director, DAKS)
- Arcot Rajasekar (Director)
- Michael Wan (Chief Architect)
- Wayne Schroeder
- George Kremenek
- Bing Zhu
- Sheau-Yen Chen
- Charles Cowart
- Arun Jagatheesan (GriPhyN)
- Lucas Gilbert
- Roman Olsachnowsky (BIRN)
- Tim Warnock (BIRN)
Contacts

• For Additional Information:
  Web:  http://www.npaci.edu/dice/srb
  Mail:  srb@sdsc.edu
  Mailing-list: srb-chat@sdsc.edu