Implementing and Using SRB
Introduction

• What Is SRB:  
  – **Mike Doherty**

• Using SRB in the e-Minerals Mini-Grid: 
  – **Lisa Blanshard**

• Installing SRB: 
  – **Ananta Manandhar**

• Federated MCAT and SRB Futures: 
  – **Mike Wan**

• CCLRC and SRB: **Mike Doherty**

Michael Doherty  
RAL
Managing Data

• Historically data has been **STORED** rather than **MANAGED**
• Problems arising from this include
  – Scaling
  – Distribution
  – Access Control, Authentication, Security
  – Data Migration
  – Data Curation
What is SRB?

• Storage Resource Broker (SRB) is a software product developed by the San Diego Supercomputing Centre (SDSC).
• Allows users to access files and database objects across a distributed environment.
• Actual physical location and way the data is stored is abstracted from the user
• Allows the user to add user defined metadata describing the scientific content of the information
How SRB Works

- **4 major components:**
  - The Metadata Catalogue (MCAT)
  - The MCAT SRB Server
  - The SRB Server
  - The SRB Client
SRB in Detail

Application

C, C++, Linux I/O  Unix Shell  Java, NT  Browsers  Prolog Python  Web

SRB

Archives
HPSS, ADSM, UniTree, DMF

File Systems
Unix, NT, Mac OSX

Databases
DB2, Oracle, PostgreSQL

MCAT

Resource, User

User Defined

Dublin Core

Application Meta-data

Third-party copy

Remote Proxies

DataCutter

Michael Doherty
RAL
The MCAT Database

- MCAT database is a metadata repository that provides a mechanism for storing information used by the SRB system.
- Includes both
  - Internal system data required for running the system
  - Application (user) metadata regarding data sets being brokered by SRB.
- SRB makes a clear distinction between these two types of data
The MCAT Server

- At least one SRB Server must be installed on the node that can access the MCAT database. This is known as the MCAT SRB Server.
- MCAT SRB Server works directly against the MCAT database to provide SRB Services
- All other SRB Servers interact through the MCAT
The SRB Server

- The SRB Server is a middleware application that accepts requests from clients and obtains/queries/manages the necessary data sets.
- It queries the MCAT SRB Server to gather information on datasets and supplies this back to the SRB client.
SRB Clients

• Provides a user interface to send requests to the SRB server.
• There are 4 main implementations of this:
  – Command line
  – MS Windows (InQ)
  – Web based (MySRB).
  – Java
Concepts

- **Location**: A physical node running an SRB Server
- **Physical Resource**: A storage area managed by an SRB Server
- **Logical Resource**: On or more Physical Resources – can be distributed
- **Collection**: Data abstraction of resources
SRB Client Examples

- Ingesting a file
- Replicating a file
- Adding Metadata
- Querying Metadata
- Downloading a file
Administration

- Users/Locations/Resources must be managed
- Two methods for doing this
  - Java MCAT Admin Tool
  - Command line tools
Administration Example

- Adding a new Location
- Creating a new Physical Resource
- Adding a new Logical Resource
Questions?

Any at this point?
SRB Example: CMS

• Largest project using CCLRC SRB services at present is the CERN CMS experiment.
• SRB chosen by the CMS Production team for the PCP
• SRB Chosen for ‘Pre-Challenge Production’, producing data for Data Challenge 2004 (DC04)
CMS: Why

- Nothing available CMS-wide on the timescale (end of year) for DC2003
- Needed stable and supported product for 6 months continuously
CMS Results (So far)

- How long does it take to register and replicate 500GB between to widely separated locations?
  - *Tested 200GB which took approximately 6 hours, almost completely network limited*

- How long does it take to register and replicate 50k 10kB files?
  - *SRB has a bulk file registration mode which they have clocked at 400 files per second. Registered and replicated 1000 files in a few seconds*

- What is the maximum sustainable transfer rate out of a single server and what is the maximum rate a server can accept data from three servers?
  - *About 80-90% of network speed for 5 streams x number of servers*

- How many files can be registered in a day?
  - *No inherent limits*

- How many parallel streams can a server accept?
  - *Unknown, very small load on CPU with 10 streams*
SRB Example: e-Minerals
CCLRC and SRB

• Have established links with SRB community and SDSC
• Implemented SRB on projects
• Set up test systems for new projects
• Can help community with
  – SRB Test Systems
  – SRB Productions Systems
  – SRB Local Support