iRODS A Large-Scale Rule-Oriented Data Management System

Wayne Schroeder

Data Intensive Computing Environments,

San Diego Supercomputer Center,

University of California San Diego

schroede@sdsc.edu

http://diceresearch.org

http://www.irods.org

SDSC

Topics

- **Who We Are**
- **Our Software**
  - Storage Resource Broker (SRB)
  - Integrated Rule Oriented Data management System (iRODS)
- **How we use DBMS**
- **Informal Comparison of PostgreSQL and Oracle**

DICE @ SDSC @ UCSD

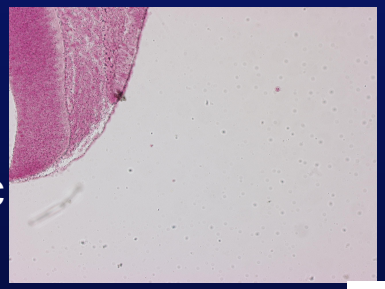
- **Team of about a dozen**
  - Dr Reagan Moore, Dr Arcot Rajasekar, Dr Richard Marciano
  - Michael Wan, Wayne Schroeder, other software engineers
  - Software Engineering is Key; Must be Useful and Work Well
- **Data Intensive Computing Environments (DICE)**
  - 1997 DARPA
  - Series of awards NARA, NSF
  - National and International Uses
  - Customer Driven
- **San Diego Supercomputer Center**
  - NSF Funded, Series of initiatives
  - National Resource
  - Started 1985 under General Atomics at UCSD
  - 2000 as part of University of California San Diego
  - High Performance Computing

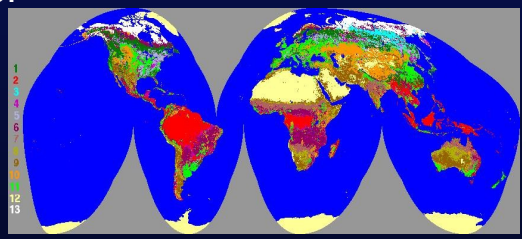
My Own Background

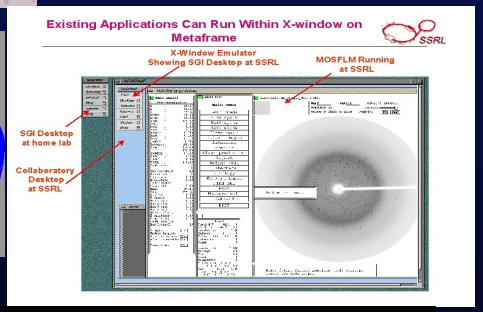
- **Software Developer (BS CS 1976)**
- **SDSC at Start, 1985**
  - Enthused to Support Science, etc
  - LLNL (Fusion Energy Center, NMFECC) before SDSC
- **Entropia (startup) 2000-2002**
- **DICE 2002**
  - SRB Installation/Testing, Java GUI Admin, etc
  - iRODS Co-Developer
    - Michael Wan, Arcot Rajasekar (Raja), myself
  - Catalog (DBMS) Interface (ICAT)
  - Administration
  - Installation/Testing
  - Authentication (password, GSI)
  - Etc

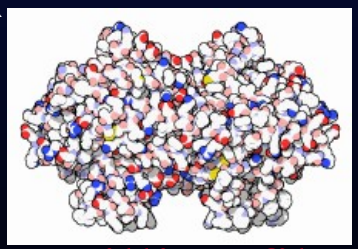
SRB Projects (Old Slide)

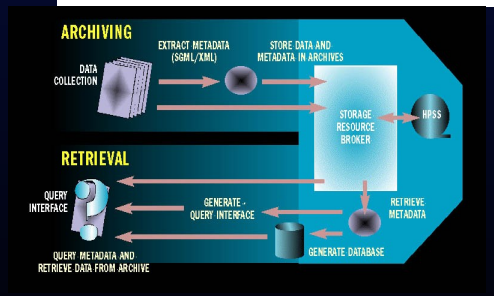
- **Astronomy**
  - National Virtual Observatory
- **Data Grids**
  - UK e-Science CCLRC
  - Teragrid
- **Digital Libraries and Archives**
  - National Archives and Records Administration
  - National Science Digital Library
  - Persistent Archive Testbed
- **Ecological, Environmental, Oceanographic**
  - ROADnet
  - Southern California Earthquake Center
  - SIO Digital Libraries
- **Molecular Sciences**
  - Synchrotron Data Repository
  - Alliance for Cellular Signaling
- **Neuro Sciences**
  - Biomedical Information Research Network
- **Physics and Chemistry**
  - BaBar
- **Many others**











Over 650 Tera Bytes in 106 million files

UCSD iRODS SRB SDSC

Sampling of Funded Projects

Massive Data Analysis System (MDAS)	1995-1997	DARPA
Distributed Object Computation Testbed	1996-1999	DOD, USPTO
National Partnership for Advanced Computational Infrastructure	1997-2004	NSF
Information Power Grid	1998-2004	NASA
Data Visualization Corridor	1998-2001	DOE ASCI
Persistent Archive Research	1999-	NARA
(20 + more, see SRB Web site)	2000 -	Various

Extremely Successful

- **Storage Resource Broker (SRB) manages 2 PBs of data in internationally shared collections**
- **Data collections for NSF, NARA, NASA, DOE, DOD, NIH, LC, NHPRC, IMLS: APAC, UK e-Science, IN2P3, WUNgrid**

Astronomy	Data grid
Bio-informatics	Digital library
Earth Sciences	Data grid
Ecology	Collection
Education	Persistent archive
Engineering	Digital library
Environmental science	Data grid
High energy physics	Data grid
Humanities	Data Grid
Medical community	Digital library
Oceanography	Real time sensor data, persistent archive
Seismology	Digital library, real-time sensor data

- **Goal has been generic infrastructure for distributed data**

Date	5/17/02		6/30/04			11/29/07		
Project	GBs of data stored	1000Õs of files	GBs of data stored	1000Õs of files	Users with ACLs	GBs of data stored	1000Õs of files	Users with ACLs
Data Grid								
NSF / NVO	17,800	5,139	51,380	8,690	80	88,216	14,550	100
NSF / NPACI	1,972	1,083	17,578	4,694	380	39,697	7,590	380
Hayden	6,800	41	7,201	113	178	8,013	161	227
Pzone	438	31	812	47	49	28,799	17,640	68
NSF / LDAS-SALK	239	1	4,562	16	66	207,018	169	67
NSF / SLAC-JCSG	514	77	4,317	563	47	23,854	2,493	55
NSF / TeraGrid			80,354	685	2,962	282,536	7,257	3,267
NIH / BIRN			5,416	3,366	148	20,400	40,747	445
NCAR						70,334	325	2
LCA						3,787	77	2
Digital Library								
NSF / LTER	158	3	233	6	35	260	42	36
NSF / Portal	33	5	1,745	48	384	2,620	53	460
NIH / AfCS	27	4	462	49	21	733	94	21
NSF / SIO Explorer	19	1	1,734	601	27	2,750	1,202	27
NSF / SCEC			15,246	1,737	52	168,931	3,545	73
LLNL						18,934	2,338	5
CHRON						12,863	6,443	5
Persistent Archive								
NARA	7	2	63	81	58	5,023	6,430	58
NSF / NSDL			2,785	20,054	119	7,499	84,984	136
UCSD Libraries			127	202	29	5,205	1,328	29
NHPRC / PAT						2,576	966	28
RoadNet						3,557	1,569	30
UCTV						7,140	2	5
LOC						6,644	192	8
Earth Sci						6,136	652	5
TOTAL	28 TB	6 mil	194 TB	40 mil	4,635	1,023 TB	200 mil	5,539

iRODS Tutorials - 2008

- January 31, SDSC
- April 8 - ISGC, Taipei
- May 13 - China, National Academy of Science
- May 27-30 - UK eScience, Edinburgh
- June 5 - OGF23, Barcelona
- July 7-11 - SAA, SDSC
- August 4-8 - SAA, SDSC
- August 25 - SAA, San Francisco

iRODS Development

- **NSF - SDCI grant Adaptive Middleware for Community Shared Collections**
  - iRODS development, SRB maintenance
- **NARA - Transcontinental Persistent Archive Prototype**
  - Trusted repository assessment criteria
- **NSF - Ocean Research Interactive Observatory Network (ORION)**
  - Real-time sensor data stream management
- **NSF - Temporal Dynamics of Learning Center data grid**
  - Management of IRB approval

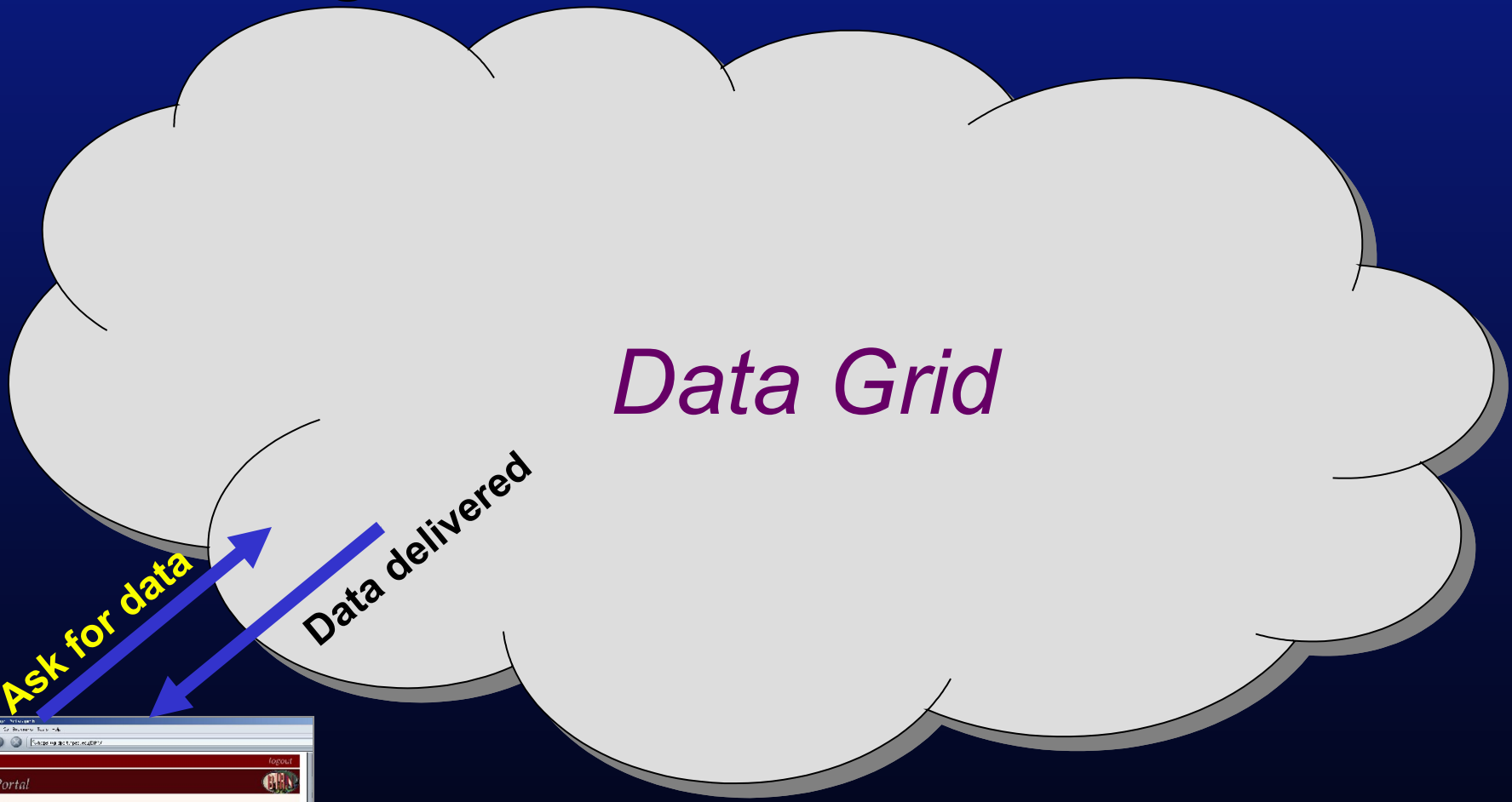
iRODS Development

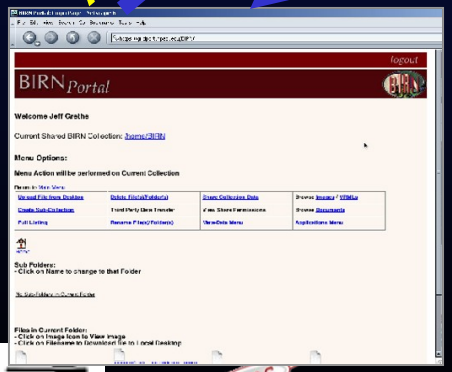
- 2005: Planning, Some Initial Development
- 2006, December: iRODS .5 Released
- 2007, June: iRODS .9 Released
- 2008, January: iRODS 1.0 Released
- Soon: iRODS 1.1

iRODS/SRB Flavors

- **Data grids**
  - **Share data** - organize distributed data as a collection
- **Digital libraries**
  - **Publish data** - support browsing and discovery
- **Persistent archives**
  - **Preserve data** - manage technology evolution
- **Real-time sensor systems**
  - **Federate sensor data** - integrate across sensor streams
- **Workflow systems**
  - **Analyze data** - integrate client- & server-side workflows

Using a Data Grid *in Abstract*

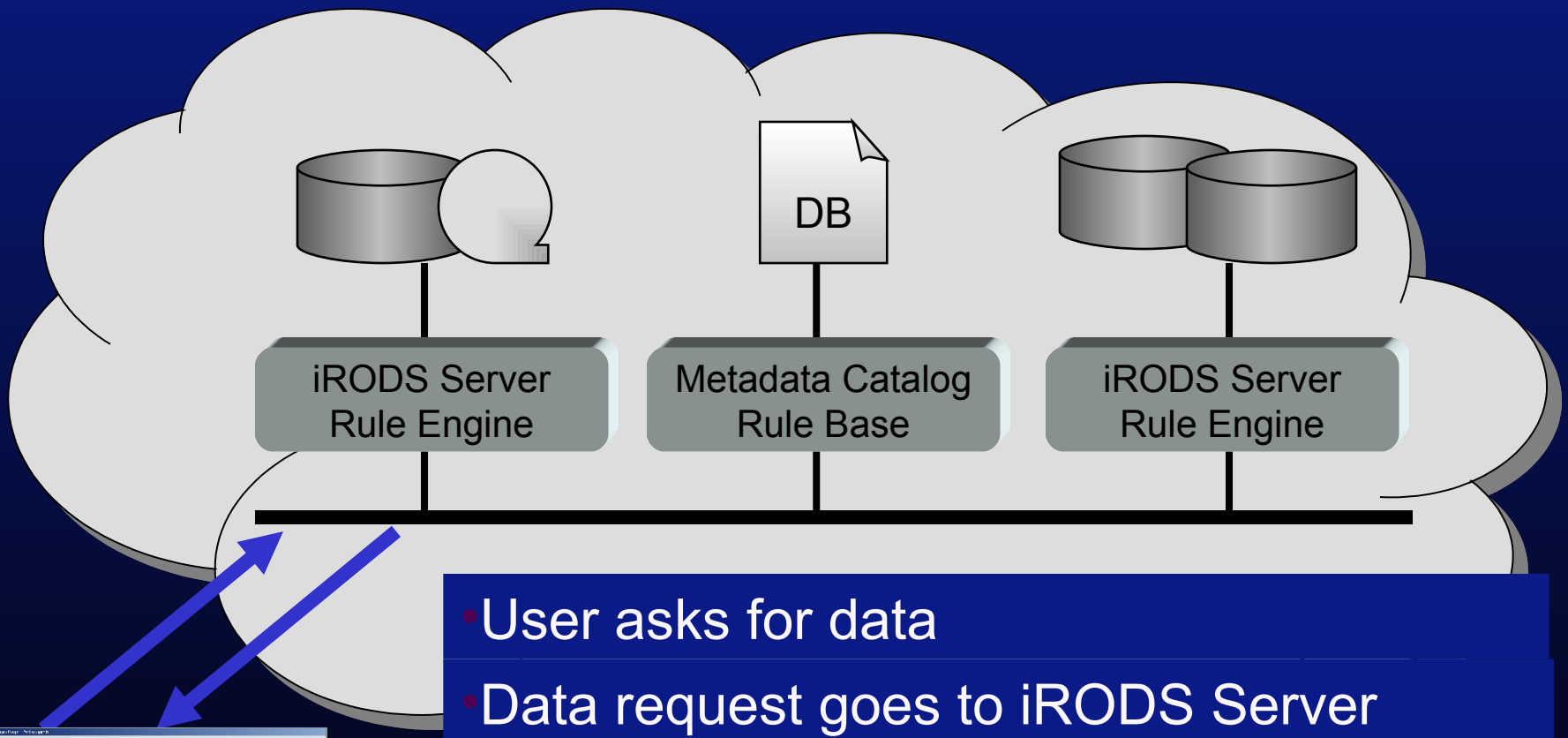


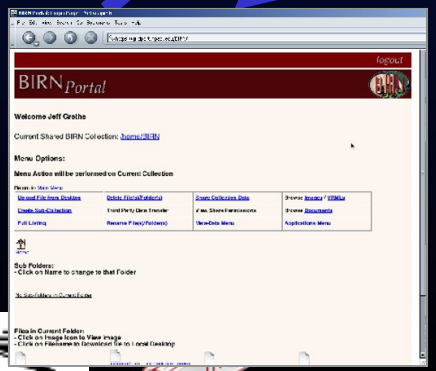


- User asks for data from the data grid
- The data is found and returned
  - Where & how details are hidden

UCSD iRODS SRB SDSC

Using a Data Grid - *Details*





- User asks for data
- Data request goes to iRODS Server
- Server looks up information in DB catalog
- Catalog tells which iRODS server has data
- 1st server asks 2nd for data
- The 2nd iRODS server applies rules

UCSD

SDSC

Data Grid State Information

State Information in DBMS

- Files (DataObjects)
- Directories (Collections)
- Users
- Resources, etc

For Each File DBMS information includes:

- Location: Host and Directory
- Other System Metadata
- User-defined Metadata
- Replica, etc

Data Grid Capabilities

- **Logical file name space**
  - Directory hierarchy / soft links
  - Versions / backups / replicas
  - Aggregation / containers
  - Descriptive metadata
  - Digital entities
- **Physically Distributed on Network**
- **Authentication and authorization**
  - GSI, challenge-response, Shibboleth
  - ACLs, audit trails
  - Checksums, synchronization
  - Logical user name space
  - Aggregation / groups

Generic Infrastructure

- **Data grids manage data distributed across multiple types of storage systems**
  - File systems, tape archives, object ring buffers
- **Data grids manage collection attributes**
  - Provenance, descriptive, system metadata
- **Data grids manage technology evolution**
  - At the point in time when new technology is available, both the old and new systems can be integrated

Tension between Common and Unique Components

- **Synergism - common infrastructure**
  - Distributed data
    - Sources, users, performance, reliability, analysis
  - Technology management
    - Incorporate new technology

- **Unique components - extensibility**
  - Information management
    - Semantics, formats, services
  - Management policies
    - Integrity, authenticity, availability, authorization

Storage Resource Broker
A Data Grid Solution

- 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 11+ years;
  - Under continual development since 1997;

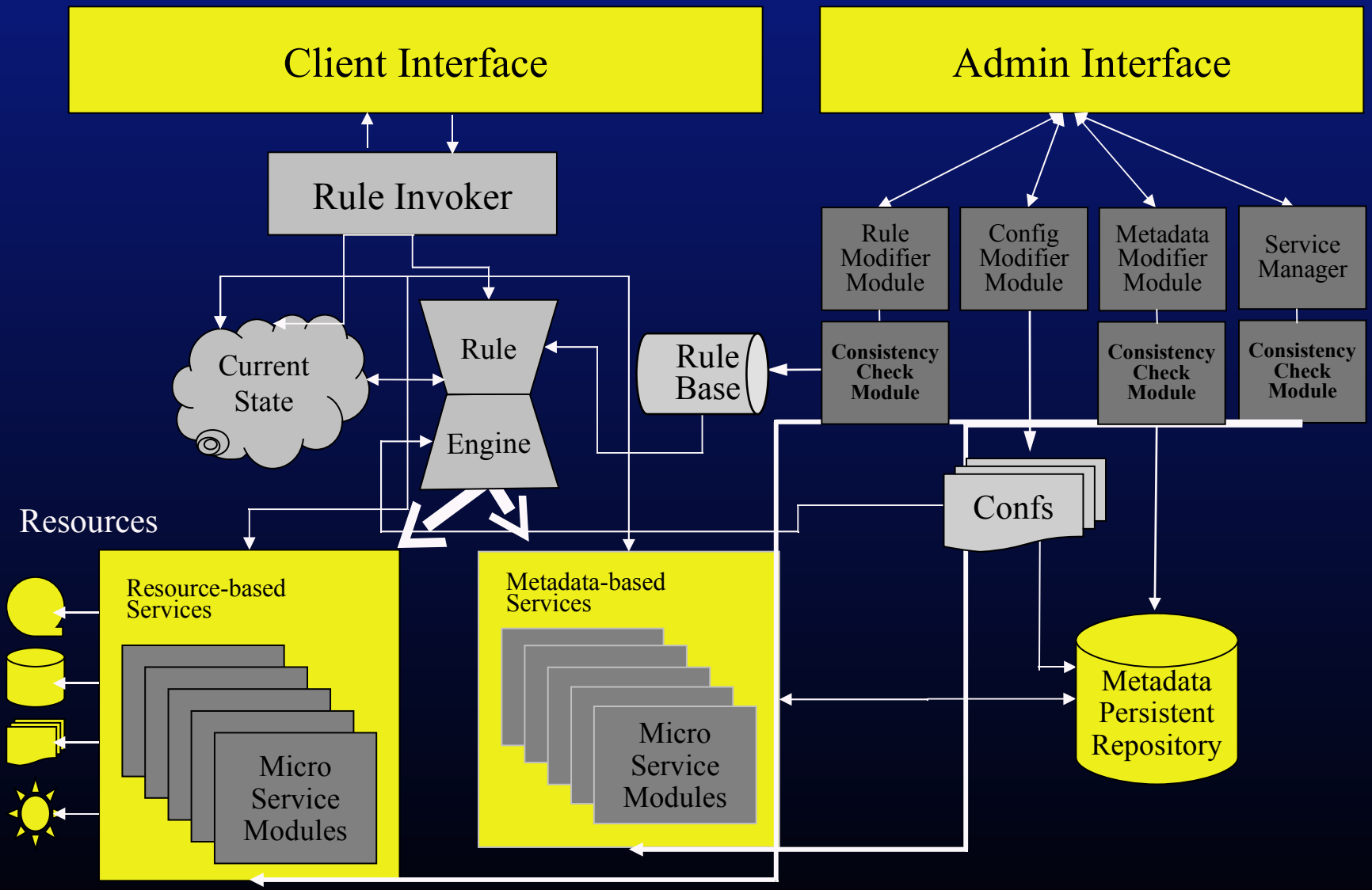
UCSD iRODS SRB SDSC


IRODS - the Next Generation
of Data Grid Technology

iRODS

- **Rule-based**
  - Rules Engine at core
  - Our own implementation (Raja)
- **Rules invoke microservices and/or rules**
- **Complete rewrite, but based on experience with SRB**
- **Client/Server, Server-Server**
- **Open Source (BSD) (SRB is available to edu and gov sites)**

integrated Rule-Oriented Data System



UCSD iRODS SRB SDSC

Data Grids

- **SRB - Storage Resource Broker**
  - Persistent naming of distributed data
  - Management of data stored in multiple types of storage systems
  - Organization of data as a shared collection with descriptive metadata, access controls, audit trails
- **iRODS - integrated Rule-Oriented Data System**
  - Rules control execution of remote micro-services
  - Manage persistent state information
  - Validate assertions about collection
  - Automate execution of management policies

iRODS Clients

- **Currently seven clients**
  - iRODS rich web client
    - https://rt.sdsc.edu:8443/irods/index.php
  - Unix shell commands
    - iRODS/clients/icommands/bin
  - FUSE user level file system
    - iRODS/clients/fuse/bin/irodsFs fmount
  - Jargon Java I/O class library
    - iRODS/java/jargon
  - PHP web browser and PHP client library
    - http://irods.sdsc.edu
  - C library calls
  - Parrot user level file system
    - Douglas Thain, Notre Dame University

iCommands

~/irods/clients/icommands/bin

- icd
- ichmod
- icp
- ils
- imkdir
- imv
- ipwd
- irm
- ienv
- ierror
- iget
- iput
- ireg
- irepl
- itrim
- irsync
- ilsresc
- iphymv
- irmtrash
- ichksum
- iinit
- iexit
- iqdel
- iqmod
- iqstat
- iexecmd
- irule
- iuserinfo
- isysmeta
- imeta
- iquest
- imiscsvrinfo
- iadmin

irodssetup: Installation

- **Linux, Mac, Mac/Intel, Solaris, AIX, 32/64 bit**
- **Prompt User**
- **Download, Configure, Build, Install, Run**
  - PostgreSQL
  - ODBC (Unix or PostgreSQL)
- **Configure, Build, Install, Run iRODS**
- **Install ICAT Database**
- **Bring Up System**
- **Basic Tests, Optional Advanced Tests**

Testing

- **iCommand test suite from IN2P3, France**
  - Thomas Kachelhoffer, Jean-Yves Nief
- **ICAT test suite all 204 SQL Forms**
- **Layers of Scripts**
  - Tinderbox
  - installation (rewritten by Dave Nadeau)
  - irodsctl test the above two test suites
- **NMI Build & Test Facility, U of Wisc**

iRODS Development Status

- **Production release is version 1.0**
  - January 24, 2008
- **Version 1.1 Soon**
- **International collaborations**
  - SHAMAN - University of Liverpool
    - Sustaining Heritage Access through Multivalent ArchiviNg
  - UK e-Science data grid
  - IN2P3 in Lyon, France
  - DSpace policy management

iRODS Data Grid Capabilities

- **Logical Name Space**
- **Logical Storage Space**
  - Dynamic resource creation
  - Standard operations
  - Heterogeneous storage systems
  - Trash
  - Collective operations / storage groups
- **Data transport**
  - Parallel I/O
  - Small file transport
  - Message engine
  - Containers / tar files / HDF5
  - Aggregation of I/O commands - remote procedures

iRODS Data Grid Capabilities

- **Remote procedures**
  - Atomic / deferred / periodic
  - Procedure execution / chaining
  - Structured information
- **Structured information**
  - Metadata catalog interactions / 204 SQL forms
  - Information transmission
  - Template parsing
  - Memory structures
  - Report generation / audit trail parsing

SRB DBMS

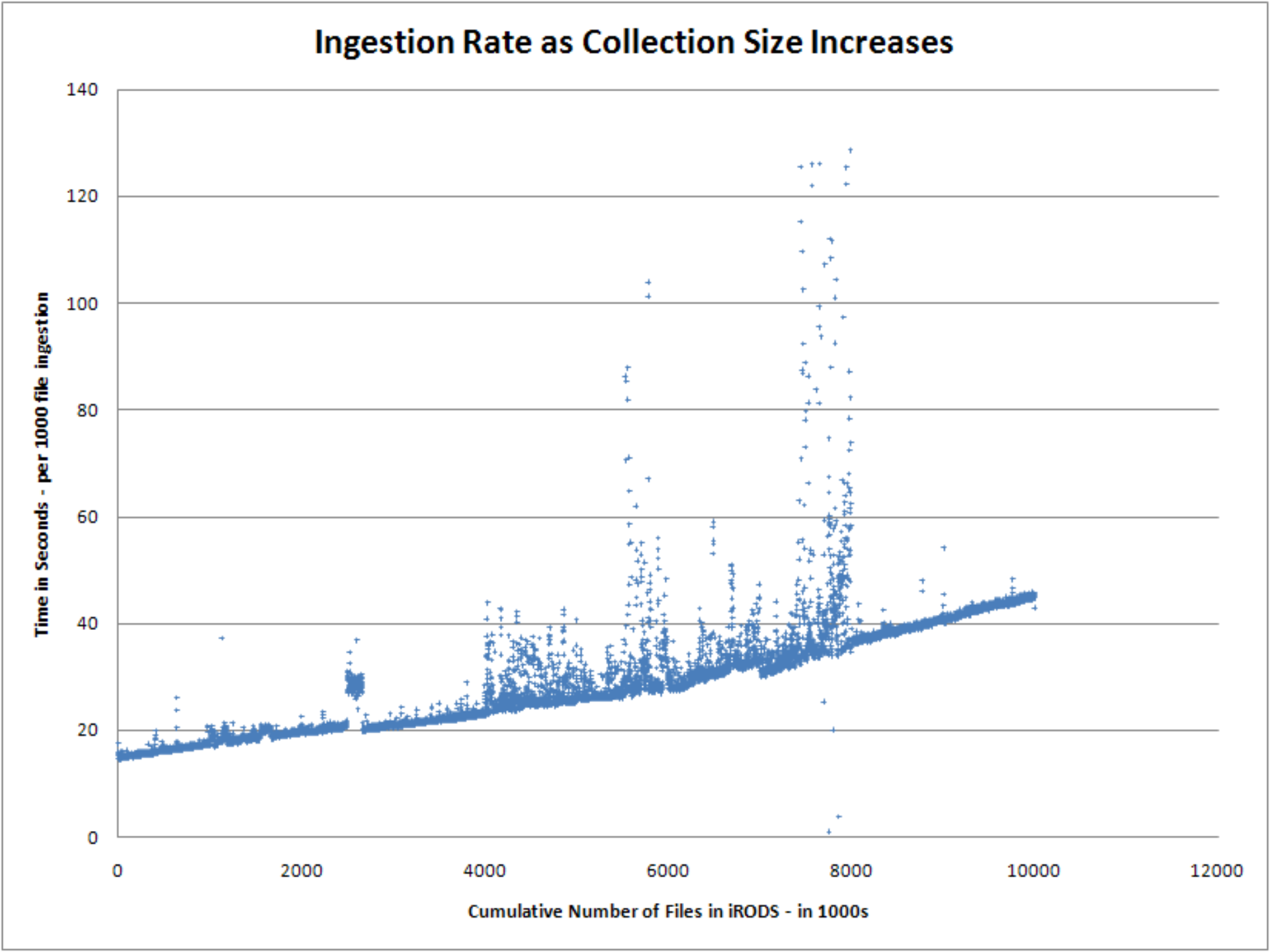
- **SRB CATALOG (MCAT)**
  - Oracle, DB2, Sybase, PostgreSQL, Informix, or MySQL4 (primarily Oracle and PostgreSQL)
- **Binary Large Objects**
  - DB2, Oracle, Illustra
- **Oracle in Production**
  - SDSC and Elsewhere
- **PostgreSQL for Testing/Demos**

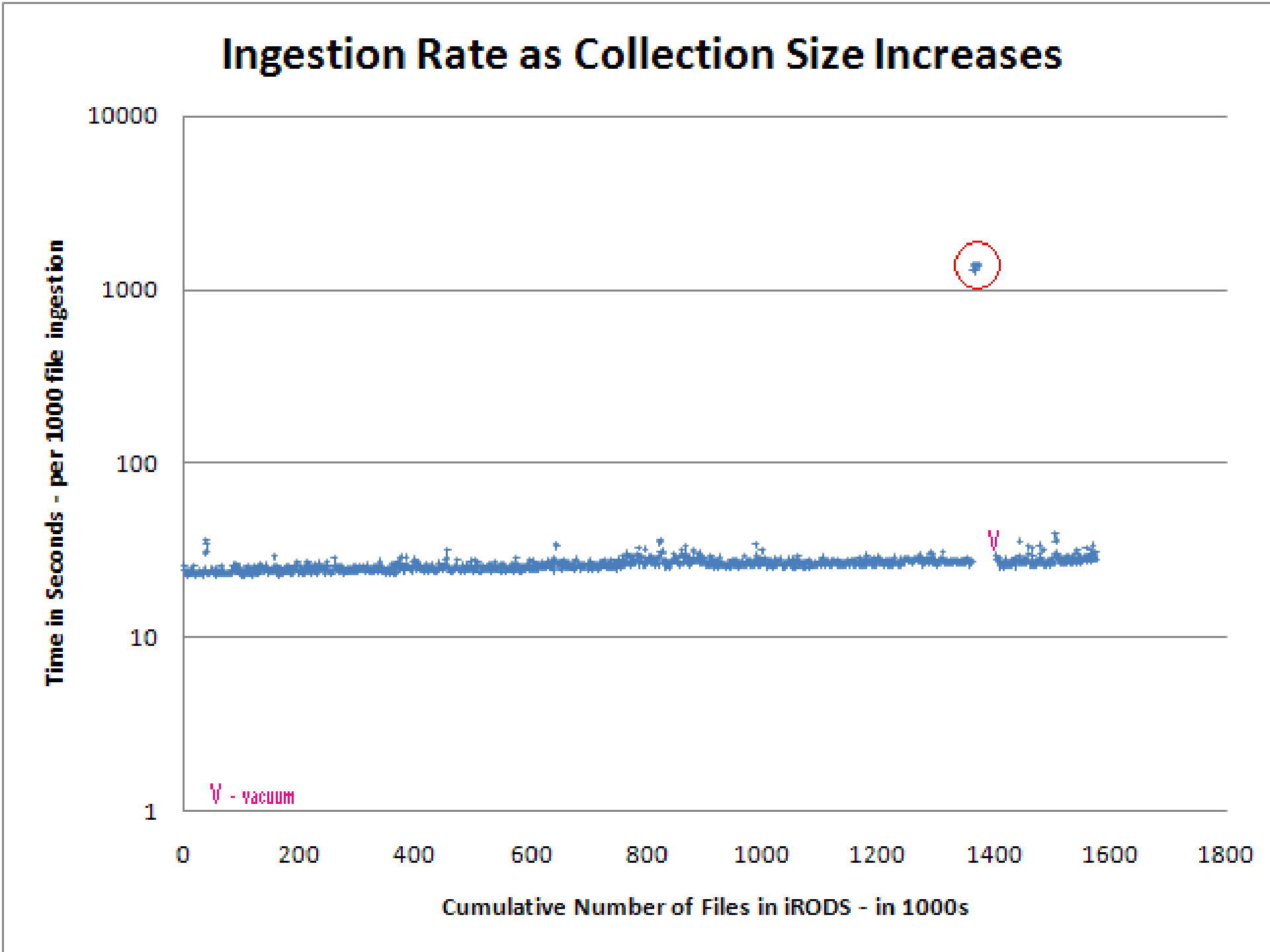
iRODS DBMS

- **Catalog (ICAT)**
  - PostgreSQL or Oracle (primarily PostgreSQL)
  - MySQL Planned
- **PostgreSQL In Production (soon)**
- **PostgreSQL for Test/Demo**

iRODS ICAT

- Interface to RDBMS
  - iRODS State Information
- Simplified Schema (Raja)
- Bind Variables for Performance/Security
- Three levels:
  - API - High Level calls (~45)
  - Mid-level/Helpers
  - PostgreSQL/ODBC or Oracle/OCI
- Called by
  - MicroServices/Rules, Server Code, Client/Server calls
- GeneralQuery, GeneralAdmin, SimpleQuery
- iadmin interface for Administration


Ingestion Rate as Collection Size Increases
Time in Seconds - per 1000 file ingestion
0
20
40
60
80
100
120
140
0
2000
4000
6000
8000
10000
12000
Cumulative Number of Files in iRODS - in 1000s


Ingestion Rate as Collection Size Increases
10000
1000
100
10
1
Time in Seconds - per 1000 file ingestion
V - vacuum
0
200
400
600
800
1000
1200
1400
1600
1800
Cumulative Number of Files in iRODS - in 1000s

PostgreSQL Advantages

- **Freely Downloaded/Installed for:**
  - Testing, SRB/iRODS
  - Integrated Installation
    - SRB Demos/Tutorials
    - SRB in a Box (Shipboard Environmental Science)
    - iRODS Demos/Tutorials/Production Use
- **Faster**
  - i-cmd/ICAT test suite >2x Oracle
  - Same Host, Small DB
- **Open Source**
- **psql vs sqlplus**

iRODS WebSite-Wiki

- http://irods.sdsc.edu
- Descriptions of the technology
- Publications / presentations
- Download
- Performance tests
- Tinderbox system (continual build/test)
- irods-chat page

Planned Development

- GSI support (1)
- Time-limited sessions via a one-way hash authentication
- Python Client library
- GUI Browser (AJAX in development)
- Driver for HPSS (in development)
- Driver for SAM-QFS
- Porting to additional versions of Unix/Linux
- Porting to Windows
- Support for MySQL as the metadata catalog
- API support packages based on existing mounted collection driver
- MCAT to ICAT migration tools (2)
- Extensible Metadata including Databases Access Interface (6)
- Zones/Federation (4)
- Auditing - mechanisms to record and track iRODS metadata changes

For More Information

Wayne Schroeder
San Diego Supercomputer Center
schroede@sdsc.edu

http://diceresearch.org
http://www.irods.org