Policy-based data management

Reagan Moore,[1] Arcot Rajasekar,[1] Mike Wan,[2] Wayne Schroeder,[2] Chien-Yi Hou[1]

Data Intensive Cyber Environments Center
[1]University of North Carolina at Chapel Hill
[2]University of California at San Diego

Abstract. Next generation data grid technologies automate policy enforcement, as exemplified by the integrated Rule-Oriented Data System (iRODS). The combination of data grid virtualization mechanisms with policy based data management makes it possible to automate administrative functions, enforce policies such as distribution, retention, disposition, replication, and periodically validate assessment criteria. In this paper, we provide brief outlines of production-level use cases for iRODS being undertaken around the world.

1. Introduction

The **integrated Rule-Oriented Data System** (iRODS) [1–2] is data grid middleware. It provides mechanisms for managing, querying, accessing, and preserving data that are distributed across multiple storage resources. The iRODS system applies policy-based controls when performing these functions. Specifically, the iRODS system provides the following capabilities:

- Global persistent identifiers for naming objects.
- Support for metadata to identify system-level physical properties of the stored data.
- Support for descriptive metadata to enable discovery through simple query mechanisms.
- Standard access mechanisms such as browsers, shell commands, load libraries, Java, C library calls, file system interfaces, digital libraries, workflows, and web services.
- Storage repository abstraction for storing files on tape, disk, and cloud storage.
- Inter-realm authentication system for secure access to remote storage systems.
- Support for replication and synchronization of files between resource sites.
- Support for caching copies of files onto a local storage system.
- Support for aggregating files to optimize management of large numbers of small files.
- Access controls and audit trails to control and track data usage.
- Support for execution of remote operations for data sub-setting, metadata extraction, indexing, remote data movement, etc. using workflows composed from micro-services.
- Support for diverse I/O models for files.
- Support for federation of data grids.

The iRODS data grid system consists of four types of peer-to-peer servers. A metadata catalog server, called the iCAT server, provides the metadata and abstraction services. Resource servers provide access to one or more storage resources through a built-in distributed rule engine. The rules can be viewed as defining processing pipelines or workflows. The building blocks for the iRODS rules are micro-services—small, well-defined procedures or functions that perform a certain task. Extensibility and customizability are achieved by encoding server-side operations (including the main

access APIs) into sequences of micro-services. The micro-services are controlled by user-defined and/or administrator-defined Event-Condition-Action rules similar to those found in active databases. For example, one may encode a rule that when accessing a data object from a collection, additional authorization checks need to be made. These authorization checks can be encoded as a set of micro-services with triggers that fire based on current operating conditions. In this way, one can control access to sensitive data based on rules and can escalate or reduce authorization levels dynamically as the situation warrants. The iRODS rule engine design builds upon the application of theories and concepts from a wide range of well-known paradigms from fields such as active databases, transactional systems, logic programming, business rule systems, constraint-management systems, workflows, service-oriented architecture and program verification. Apart from iRES servers and an iCAT server, iRODS also has two other servers: iSEC for scheduling and executing queued rules, and iXMS for providing a message-passing framework between micro-services. Figure 1 shows the various components of the iRODS system as well as some of its user interfaces.

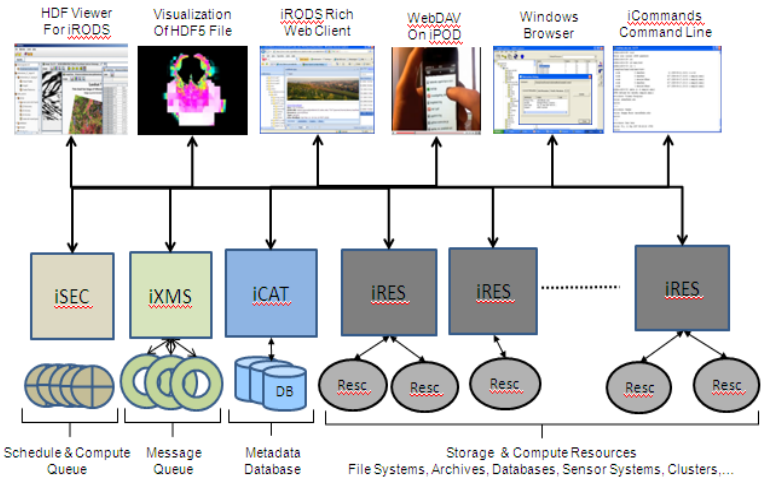

Figure 1. iRODS System Architecture.

The iRODS system is in production use in multiple projects including the US National Archives Transcontinental Persistent Archive Prototype (TPAP) [3], the NSF Science of Learning Centers [4], the Australian Research Collaboration Services [**Error! Reference source not found.**], and the SHAMAN project in UK [6]. More than 2 PetaBytes are managed by iRODS and its forerunner the Storage Resource Broker at the IN2P3 supercomputing center in Nice France [7]. Several large-scale projects such as the iPlant Collaborative [8], the Ocean Observatories Initiative [9], and the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) [10] are starting to use the iRODS system for their production data sharing and archiving environment.

2. Use Cases from Ongoing Large-Scale Production Environments

One can categorize the usage models for iRODS into different types:

a) Uniform name space for a large wide-area file system for use by users who perform analyses from multiple computing resources,
b) Sharing of files among cooperating small discipline-centric groups,
c) Large-scale data sharing across large groups of users, who share data from multiple sensors or share data from computer simulations,
d) Publication of data in a digital library environment with curation, organization, and standardized metadata capabilities, and
e) Long-term storage and access to data in a preservation environment.

The use cases that we describe are from existing production systems run by several organizations that are not part of the iRODS DICE group. The feedback from these users provides us with a set of assurances about the usability of the iRODS data grid middleware, and indeed provides a strong indication of its extensibility and ability to span requirements across multiple disciplines.

2.1. PetaShare

(*Uniform Name Space*) PetaShare [11] is an NSF funded large-scale data management project with the goal of making data available for multi-disciplinary scientists across the state of Louisiana for data-intensive analysis and visualization. The project provides data cyber infrastructure to transparently handle all low-level details of file and data management across multiple sites and provide transparent data sharing, retrieval and archival mechanisms to promote scientific computing. Petashare uses an enhanced version of iRODS to provide a global name space and efficient data access among geographically distributed storage resources. PetaShare applies the Stork [12] data placement scheduler that takes the responsibility of managing data resources and scheduling data tasks from the user and performs these tasks transparently. For the users, PetaShare provides light-weight interfaces called PetaFs (Fuse-based file interface), Petashell (based on Parrot as a unix-type shell), and Pcommands (based on iRODS iCommands). The PetaFs and Petashell interfaces provide Unix-type standardized access as if the data were located on local disk. The Pcommands provide a richer metadata-based discovery and access mechanism.

PetaShare is currently deployed across five state universities and two health sciences centers in Louisiana. These institutions include Louisiana State University (LSU), Tulane University, University of New Orleans, University of Louisiana at Lafayette, Louisiana Tech University, and LSU Health Sciences Centers in New Orleans and Shreveport. PetaShare manages approximately 300 Terabytes of disk storage distributed across the PetaShare network sites as well as 400 Terabytes of tape storage centrally located at the LSU campus. PetaShare uses the LONI communication network, a statewide 40 Gbps fiber-optic network in Louisiana. The PetaShare project currently supports more than 25 multi-disciplinary projects from coastal hazard prediction to DNA sequencing to computational fluid dynamics.

2.2. CC-IN2P3 Data Grids

(*Small-Scale Data Sharing, Large-Scale Data Sharing*) CC-IN2P3 [13] is a French national supercomputing center located at Lyon. It provides computing and storage facilities for French and international scientists in the fields of high energy physics, nuclear physics, astrophysics, biology and biomedical applications as well as in arts and humanities. iRODS is being used in production at CC-IN2P3 on multiple projects including:

a) TIDRA (Traitement Informatique Distribué en Rhône-Alpes) provides support for five major laboratories located in the Rhône-Alpes area, and uses iRODS to serve biology applications (phylogeny), biomedical imaging applications (mice) and human data (heart and lung studies). The growing set of applications is expected to reach 20 TB by 2010.
b) Adonis is a nationally funded project providing computing and storage facilities for Arts and Humanities and is linked to other European projects such as DARIA. iRODS provides archival functionality, data access through the web, access to data across batch processing farms for riverbed studies in geography, movie simulations of ancient monuments, etc. The iRODS rule-based capability is used for automating management, distribution and extraction processes. With currently more than 20 TB, the Adonis project expects to grow to 100 TB by 2010.

The iRODS service at CC-IN2P3 is supported by 10 servers, two used for running iCAT services, 6 for storage services and two for running the Oracle databases needed by iCAT. To mitigate single-point failure and provide redundancy, a DNS load-balancing alias (developed originally at SLAC) is used. The two iCAT servers in this DNS alias provide redundancy and fault-tolerance. An iRODS Health monitoring system developed at IN2P3 (and shared with the iRODS community) is used to check the performance and working of the various components of the distributed iRODS data grid. Several other projects are also supported by CC-IN2P3 and expect to host more than 300TB of data by 2010. CC-IN2P3 also expects to migrate its 2 Petabytes of data holdings (from BaBar and Lattice QCD) on the Storage Resource Broker (SRB) into iRODS with expected completion of this migration in 2012.

2.3. Dataverse Archive

(*Long-Term Preservation*) The archivists at the H. W. Odum Institute for Research in Social Science at the University of North Carolina at Chapel Hill, use an open-source web publishing platform called Dataverse [14] to publish their extensive collection of files related to social science research studies. As a component of their preservation strategy, Odum has developed an inter-operability mechanism between Dataverse and iRODS that copies the contents of an entire Dataverse archive into iRODS. The transfer process also extracts descriptive metadata and stores them in the iCAT metadata catalog. The transfer of contents between Dataverse and iRODS is accomplished by using methods based on OAI-PMH and HTTP protocols. An additional advantage of this transfer is that the users now have the rich set of interfaces of iRODS to access data including keyword-based search and discovery on the serialized copy of the Dataverse archive. Moreover, the integration allowed the Dataverse administrators to enforce data preservation policies across their holdings including redundancy, integrity checking, technology migration, etc.

2.4. The ARCS Data Fabric

(*Large-Scale Data Sharing*) The Australian Research Collaboration Service (ARCS) Data Fabric [15] is national data infrastructure deployed in Australia for researchers to easily store their research data, and share them across institutional boundaries. ARCS Data Fabric is conceived as a generic service crossing multiple research disciplines that is available to every Australian researcher and their international collaborators. ARCS supports two types of access (both built with iRODS as an underlying infrastructure): webDrive is a two-layered access interface, whose back-end is iRODS which brokers storage at distributed member sites and whose front-end is a WebDAV server-browser system which allows transparent access to and ingestion of files across the ARCS Data Fabric. Apart from uploading and downloading functionalities, the WebDAV interface also provides access control mechanisms, metadata ingestion and discovery for files and collections, as well as a 'trash can' feature. The authentication mechanism of the ARCS webDrive supports the Australian Access Federation (AAF) standardized methods and technologies. Apart from the webDrive interface, ARCS also provides an access mechanism through the OPeNDAP protocol based on THREDDS Data Server. IRODS is used at the back-end to provide data access. Currently, the ARCS Data Fabric runs a seven-node iRODS zone with nodes set up at each of the capital cities in Australia. Currently more than 18TB of data is registered into the ARCS DF and is used by more than 280 users.

2.5. TELDAP

(*Long-Term Preservation*) Taiwan e-Learning and Digital Archives Program (TELDAP) [16] collects and integrates digital material from archives from different disciplinary institutes, museums and research groups from across Taiwan. As part of a Digital Archives Remote-Backup (DARB) project, a Sinica Data preservation System (SIDPS) is being deployed across Taiwan to capture and store digital archives and content from diverse fields including history, biology, ethnology, education and language. The SIDPS system uses the iRODS as a backbone for providing rule-based policy mechanisms, optimized data transfer functions, and to flexibly coordinate data stored within distributed environments in different institutes across Taiwan. To provide a better user experience for its users, the DARB project has developed a specialized interface called UrSpace that provides a GUI data management tool for data manipulation, data transfer, encryption, search and metadata management. The DARB also supports a synchronization mechanism that can be used for synchronizing files from user spaces with iRODS. These tools provide international language support so that files and metadata with Chinese characters will be handled correctly. The DARB project also uses a monitoring system called SIMS for detecting problems in the iRODS systems in order to maintain high production quality.

3. Conclusion

The NSF funded iRODS system has been under deployment for more than two years and has been adopted by several large-scale users. We have briefly described a few of these usage models that support different data sharing models, from name space management to data sharing to preservation and digital library maintenance. The extensibility and modularity of iRODS makes it a sustainable software model powered with policy-guided data management using micro-service oriented architecture. More information on iRODS can be found at www.irods.org

Acknowledgement

The research results in this paper were funded by the NSF Office of Cyberinfrastructure OCI-0848296 grant, "NARA Transcontinental Persistent Archive Prototype." The iRODS technology development has been funded by NSF SDCI 0721400, "SDCI Data Improvement: Data Grids for Community Driven Applications."

References

[1] Integrated Rule Oriented Data Systems, https://www.irods.org/index.php.

[2] Rajasekar A., R. Moore, C-Y. Hou, L. Christopher, R. Marciano, A. de Torcy, M. Wan, W. Schroeder, S-Y. Chen, L. Gilbert, P. Tooby, and B. Zhu, "iRODS Primer: integrated Rule-Oriented Data Systems," ISBN: 9781608453337, Book published by Morgan-Claypool Publishers, San Rafael, CA., January 2010.

[3] NARA Transcontinental Persistent Archive Prototype, http://www.dlib.org/dlib/july07/07inbrief.html.

[4] NSF Science of Learning Centers, http://silccenter.org/nsf_slcs_index.html.

[5] S. Zhang, P. Coddington, and A. Wendelborn, "Davis: A Generic Interface for SRB and iRODS," www.dhpc.adelaide.edu.au/reports/197/dhpc-197.pdf.

[6] EU SHAMAN Project, http://shaman-ip.eu/shaman.

[7] J-Y Nief, "SRB fête ses 2 péta-octets au CC-IN2P3," in IN2P3 Information Letters, No 9. January 2010, http://informatique.in2p3.fr/li/spip.php?article43.

[8] NSF iPlant Collaborative, http://www.iplantcollaborative.org/.

[9] NSF Oceans Observatories Initiative, http://www.whoi.edu/programs/ooi/.

[10] NSF CUAHSI Project, http://www.cuahsi.org/.

[11] Petashare, http://www.petashare.org/.

[12] Stork, http://stork.cct.lsu.edu/index.php.

[13] IN2P3, Institut National de Physique Nucleaire et de Physique des Particules, http://www.in2p3.fr/#.

[14] Dataverse, http://thedata.org/.

[15] ARCS Australian Research Collaboration Service, http://www.arcs.org.au/.

[16] TELDAP, http://teldap.tw/en/.