Making massive data sets nimble and flexible

High performance, high availability iRODS

Benefits

- Scales file-sharing capacity to exabyte and beyond without the need for iRODS reconfiguration or disruption
- Reduces physical storage requirements by up to 50 percent[1]
- Offers a unified, single point of management to help you manage massive data sets with a small staff

Life sciences applications for genome sequencing and personalized medicine, finance, and healthcare applications for ensuring compliance with regulatory data retention create vast amounts of data in hundreds of millions, if not billions, of individual files. Getting the most value from this information usually requires that users securely collaborate, often discovering and sharing files around the globe using automated workflows. However, when storage capacity stretches into hundreds of petabytes, simple tasks such as managing secure sharing, ensuring provenance and archiving become impossible on a file-by-file basis.

Challenge: Managing millions of files

The Integrated Rules-Oriented Data System (iRODS) was created by the Data Intensive Computing Environments (DICE) group at the University of California, San Diego, (UCSD) Supercomputer Center. It is an open source solution developed to solve the problem of data management tasks such as virtualization, discovery, workflow automation and secure collaboration at scales of hundreds of petabytes, billions of files and decades of time. iRODS is middleware. At its core, iRODS is a rules engine providing automation of data management tasks based on predefined procedures. This makes it feasible to manage petabyte-scale data collections consisting of billions of files distributed on disparate storage systems worldwide.

iRODS provides consistent application programming interface (API) access to file data, regardless of the back-end storage device. The API uses storage resource drivers that are installed on the iRODS server in compound resources. Compound resources pool the necessary components for accessing file data, including external cache. The downside of the iRODS compound resource framework, however, is the need to allocate and manage cache for each driver.

Solution: IBM Cloud Object Storage and iRODS

IBM® Cloud Object Storage offers a software platform designed to provide massive scale for unstructured, object-based data storage. The solution's decentralized, shared-nothing architecture provides an ideal complement to iRODS data-management capabilities. By using IBM Cloud Object

Storage, iRODS users gain an information lifecycle management (ILM) solution with outstanding scalability, performance, security, availability and cost benefits.

IBM Cloud Object Storage gives iRODS users proven, web-scale storage scalability without the need for iRODS reconfiguration or disruption. The software architecture is designed for cost-effective scaling to exabytes of available capacity without compromising reliability, availability or manageability.

Like iRODS, IBM Cloud Object Storage uses a single global namespace, offering a unified, single point of management. This helps deliver enormous productivity gains so that administrators can manage up to 15 times as much storage as traditional storage systems.

iRODS performance and availability

IBM Cloud Object Storage provides configurable reliability and availability based on the individual needs of iRODS users and their sensitivity to cost. Unlike conventional storage, the IBM solution uses a distinctive information dispersal algorithm (IDA) that encodes data and disperses it across the system. Because there is no replication, IBM Cloud Object Storage reduces physical storage requirements by up to 50 percent and can maintain availability of the system even if multiple hard drives fail, servers fail or a site experiences an outage. For the most demanding applications, IBM Cloud Object Storage can deliver durability of "11-15 nines" and availability of "8 nines".[2]

IBM Cloud Object Storage simplifies iRODS deployment by eliminating the need for a compound resource (see Figure). The iRODS plug-in does not require additional cache resources. This speeds iRODS performance and reduces the hardware footprint of the iRODS server.

With more than a decade of award-winning research behind it, iRODS (which is funded by the National Science Foundation and the National Archives and Records Administration) is finding a home in industry, academia and government. With IBM Cloud Object Storage, iRODS can scale to hundreds of petabytes of file-sharing capacity, providing more assured and cost-effective reliability, availability and manageability.

How it works

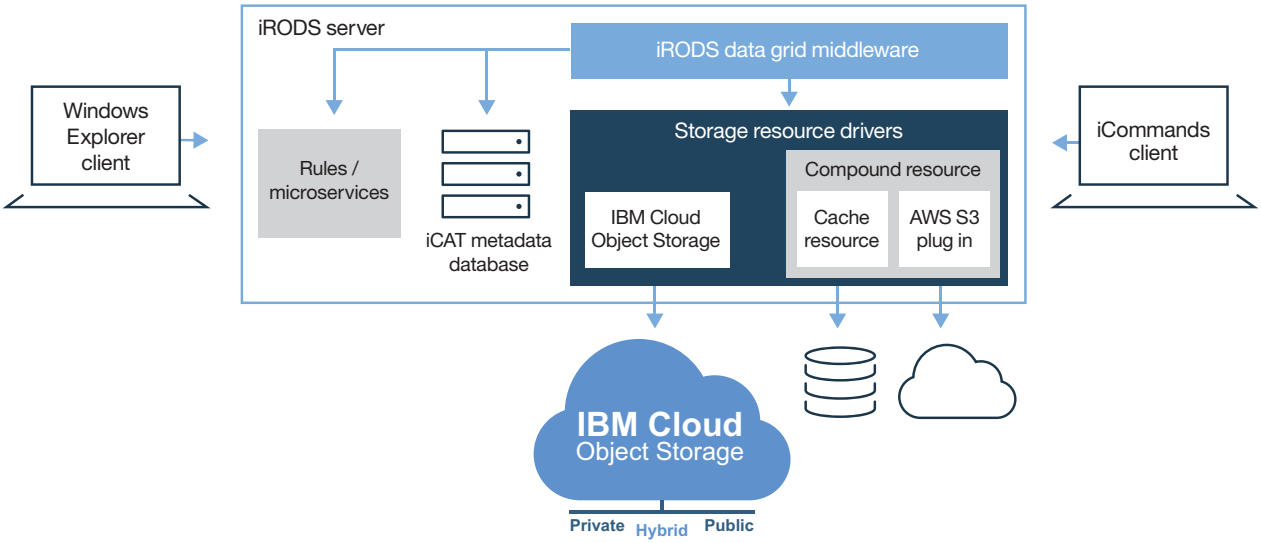

Figure 1: IBM Cloud Object Storage gives iRODS users an information lifecycle management solution, web-scale storage scalability, simplified management and deployment.

IBM Cloud Storage highlights

- Exabyte-scale storage capacity
- Configurable availability and reliability
- Strong data-at-rest security by combining encryption and information dispersal
- Shared-nothing architecture with strong consistency
- A unified, single point of management that uses fewer administrative resources than traditional storage

ILM features

- Policies to define storage tiers and data placement
- Sophisticated rules engine
- Preservation environments
- Audit trails

Use cases

- Content repositories
- Digital libraries
- Real-time data systems
- Preservation environments

About iRODS

iRODS is widely used in national scale data cyberinfrasctructure projects across a range of disciplines sponsored by federal agencies from NSF, NARA, NHPRC, and IMLS to NIH, DOE, NASA, NOAA, and more, as well as international projects and numerous federal, state, and other projects. iRODS core development is funded by the NSF and NARA. This project uses the open-source iRODs Data System, developed by the DICE Center at UNC and UCSD.

About IBM Cloud Object Storage

IBM Cloud Object Storage provides the flexibility, scalability and simplicity needed to store, manage and access today's rapidly growing volumes of unstructured data in a private, public or hybrid cloud environment. IBM's solutions transform storage challenges into business advantages by reducing storage costs while reliably supporting both traditional and emerging cloud-born workloads for enterprise mobile, social, analytics and cognitive computing.

1. *Based on IBM internal analysis of client engagements. Individual results will vary.*

2. *Numbers shown for availability and durability are for on-premises deployments and require specific supported configurations.*

© Copyright IBM Corporation 2017

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
June 2017

IBM, the IBM logo and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarksis available on the web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml.

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions. The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

Statement of Good Security Practices: IT system security involves protecting systems and information through prevention, detection and response to improper access from within and outside your enterprise. Improper access can result in information being altered, destroyed, misappropriated or misused or can result in damage to or misuse of your systems, including for use in attacks on others. No IT system or product should be considered completely secure and no single product, service or security measure can be completely effective in preventing improper use or access. IBM systems, products and services are designed to be part of a lawful, comprehensive security approach, which will necessarily involve additional operational procedures, and may require other systems, products or services to be most effective. IBM DOES NOT WARRANT THAT ANY SYSTEMS, PRODUCTS OR SERVICES ARE IMMUNE FROM, OR WILL MAKE YOUR ENTERPRISE IMMUNE FROM, THE MALICIOUS OR ILLEGAL CONDUCT OF ANY PARTY.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Please Recycle

WUS12372-USEN-00