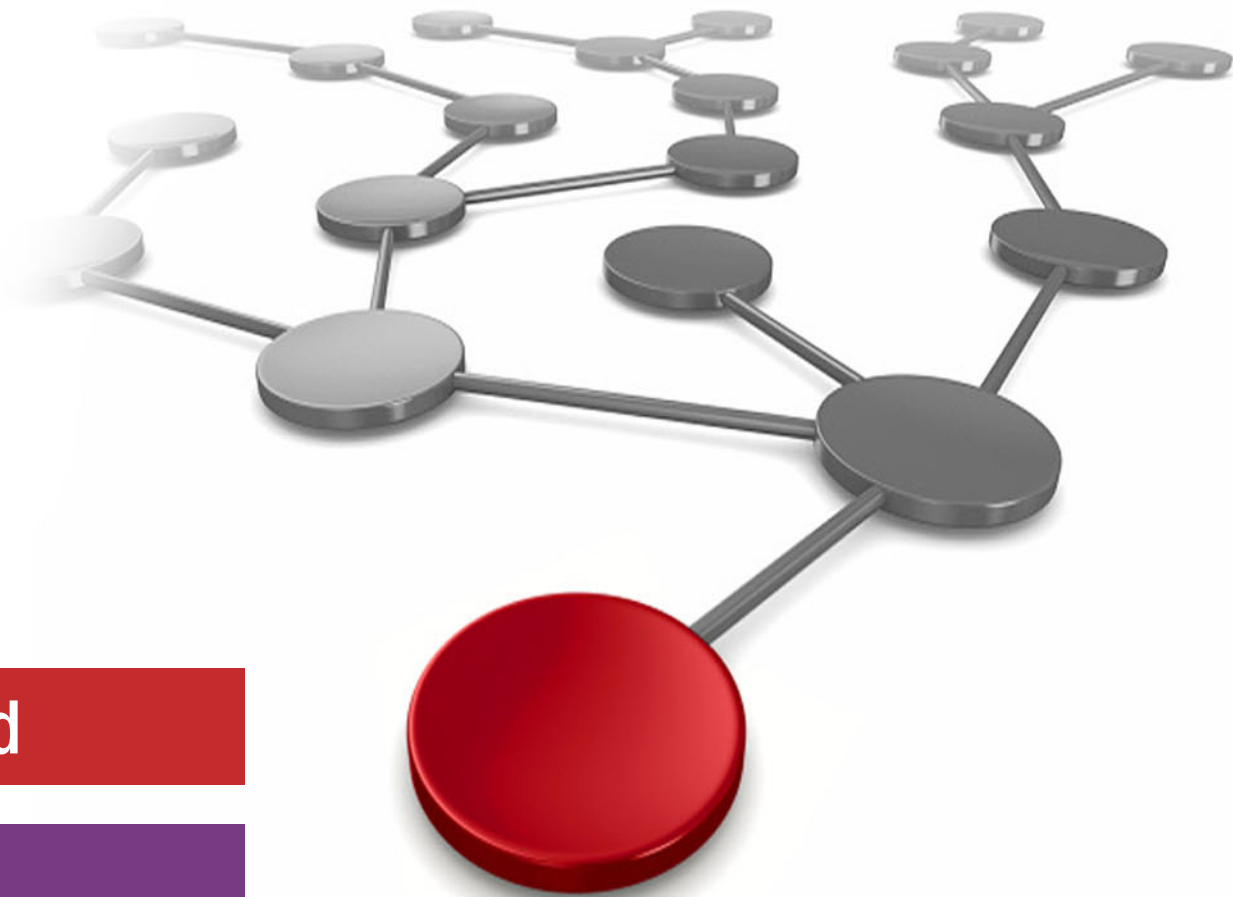


IBM Cloud Object Storage Concepts and Architecture

An under-the-hood guide for IBM Cloud Object Storage

Deepak Rangarao

Vasfi Gucer



 Cloud

Storage



Introduction

Object storage is the primary data resource used in the cloud, and it is also increasingly used for on-premise solutions. Object storage is growing for the following reasons:

- ▶ It is designed for scale in many ways (multi-site, multi-tenant, massive amounts of data).
- ▶ It is easy to use and yet meets the growing demands of enterprises for a broad expanse of applications and workloads.
- ▶ It allows users to balance storage cost, location, and compliance control requirements across data sets and essential applications.

Due to its characteristics, object storage is becoming a significant storage repository for active archive of unstructured data, both for public and private clouds.

IBM® Cloud Object Storage (IBM COS) provides industry leading flexibility enabling your organization to handle unpredictable but always changing needs of business and evolving workloads.

This IBM Redpaper™ explains the architecture of IBM Cloud Object Storage and the technology behind the product. In other words, it is an *under-the-hood guide for IBM Cloud Object Storage*.

The target audience for this paper is IBM COS architects, IT specialists and technologists.

Flexible deployment options

IBM Cloud Object Storage is available in the following modes:

- ▶ Private on-premises object storage
- ▶ Dedicated object storage (single-tenant)
- ▶ Public object storage (multi-tenant)
- ▶ Hybrid object storage (a mix of on-premises, dedicated or public offerings)

IBM COS gives you the choice deploy object storage on-prem, in the public cloud or both on-prem and in the cloud, in a hybrid solution. In addition, public cloud services (Standard Object Storage, Vault Object Storage) can be configured in either a Regional or Cross Regional model – giving you even more choice when it comes to the level of data protection and resiliency you need for your workloads.

Figure 1 on page 2 shows the IBM COS deployment options.

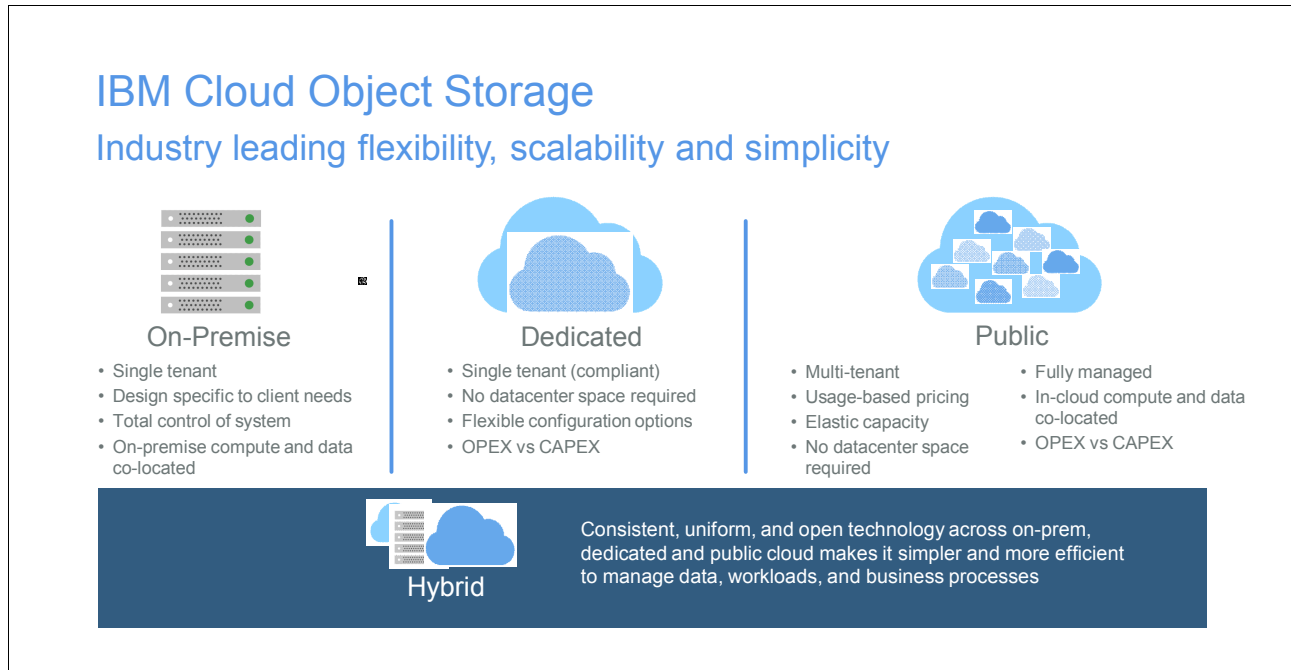


Figure 1 IBM COS deployment options

In addition, IBM COS is available in several licensing models, including perpetual, subscription, or consumption.

Additional reference: You a detailed discussion on IBM Cloud Object Storage use cases scenarios and deployment options you can refer to the *Cloud Object Storage as a Service: IBM Cloud Object Storage from Theory to Practice*, SG24-8385 IBM Redbooks.

IBM Cloud Object Storage architecture

IBM COS is a dispersed storage mechanism that leverages a cluster of storages nodes to store pieces of the data across the available nodes. IBM COS uses an *Information Dispersal Algorithm (IDA)* to break files into unrecognizable slices that are then distributed to the storage nodes. No single node has all the data, which makes it safe and less susceptible to data breaches while needing only a subset of the storage nodes to be available to fully retrieve the stored data. This ability to reassemble all the data from a subset of the chunks dramatically increases the tolerance to node and disk failures.

The IBM COS architecture is composed of three functional components. Each of these components runs ClevOS software that can be deployed on compatible, industry-standard hardware. The three components include:

- ▶ IBM Cloud Object Storage Manager
- ▶ IBM Cloud Object Storage Accesser
- ▶ IBM Cloud Object Storage Slicestor

IBM Cloud Object Storage Manager provides an out of band management interface that is used for administrative tasks such as system configuration, storage provisioning, and monitoring the health and performance of the system.

IBM Cloud Object Storage Accesser imports and reads data, encrypting/encoding data on import and decrypting/decoding data on read. It is a stateless component that presents the storage interfaces to the client applications and transforms data by using an IDA.

The IBM Cloud Object Storage Slicestor node is primarily responsible for storage of the data slices. It receives data from the Accesser on import and returns data to the Accesser as required by reads. Figure 2 illustrates the most simplistic architecture layout of the different components in IBM COS.

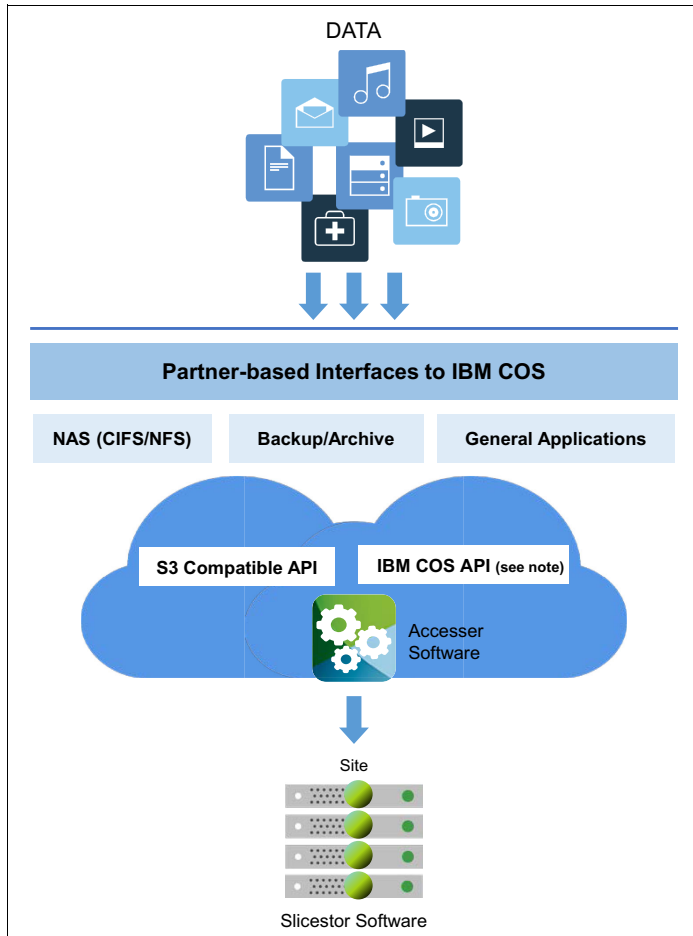


Figure 2 IBM Cloud Object Store architecture

IBM COS API: At the time of writing this book, IBM COS API is not available and expected to be available in 1H 2017.

The IBM COS API will be geared towards consistency and integration with the rest of the IBM Cloud features. This API will be consistent with IBM Cloud API guidelines, including cosmetic aspects, such as JSON encoding, and functional items, such as Identity and Access Management (IAM) compliance and OAuth2 authentication for IBM ID.

Objects that are created by using the S3 API will be able to be accessed by using the IBM COS API and vice versa.

Note that the final name of the API, as well as the API features described here are subject to change at the time of general availability of the API.

Core concepts

In this section we discuss IBM COS core concepts

Device Sets

IBM COS uses the concept of *Device Sets* to group Slicestor devices (Figure 3). Each device set is consists of the *width* number of Slicestor devices, as explained above.

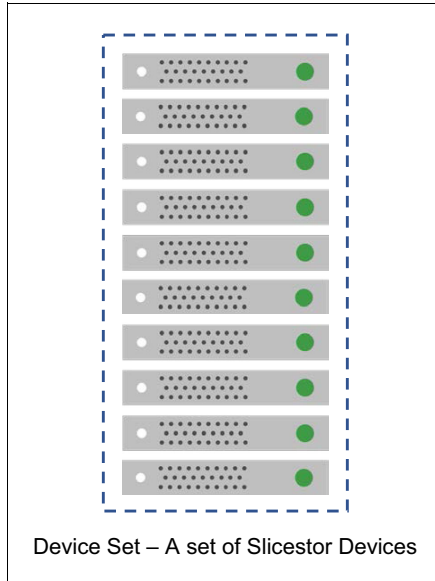


Figure 3 IBM Cloud Object Storage: Device Sets

Device Sets may be spread across one or multiple data centers and regions.

Width

The width of a system is the number of slicestore devices that the data is striped across in a device set and storage pool. For example a storage pool that has 30 storage devices is a '30 wide' storage pool. As the storage pool grows additional device sets of 30 more devices are added, however the width of the storage pool will stay at 30.

Threshold

The threshold of an IBM COS system is the number of devices that need to be available for the data to be transparently readable to the end user. For example, '30 wide' system with a threshold of 18 means that any 18 of the devices need to be up for the data to be readable. Conversely this means that 12 of the 32 devices could be down or unavailable without any impact to data accessibility.

Storage pools

Storage pools are a set of one or more device sets, as shown in Figure 4 on page 5.

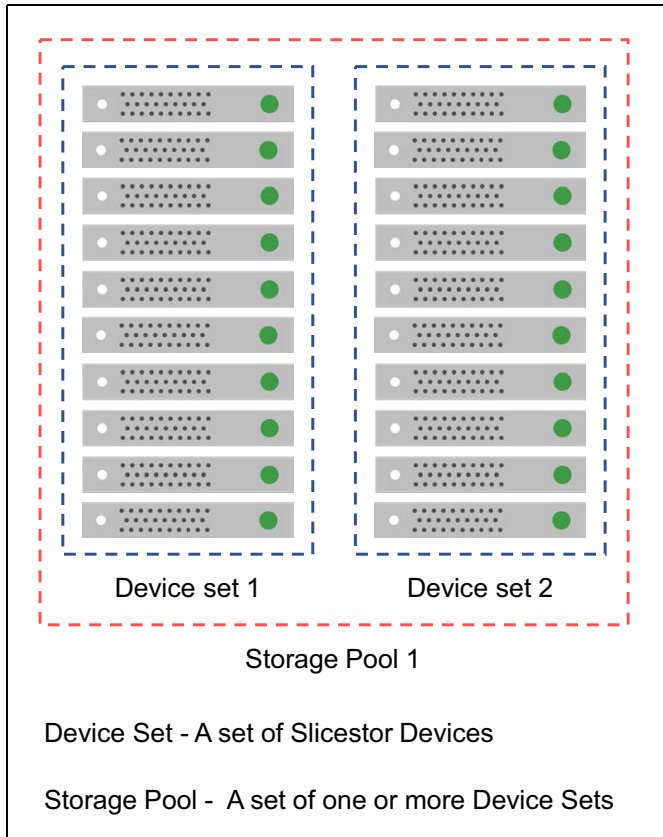


Figure 4 IBM Cloud Object Storage storage pools

Storage pools may be spread across one or multiple data centers and regions as they consists of one or many *device sets*.

Vaults

Vaults are logical storage containers for data objects that are contained in a storage pool, as shown in Figure 5 on page 6.

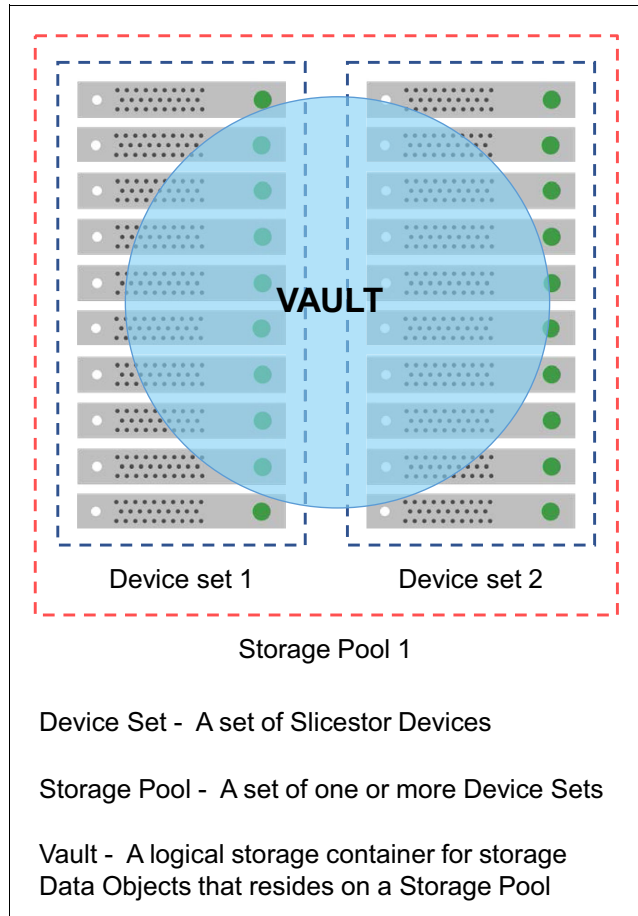


Figure 5 IBM Cloud Object Storage vaults

Vaults span multiple device sets and are automatically spread across all the device sets in the storage pool so as to optimize access speeds.

Geo dispersal

Geo dispersal allows IBM COS Accesser and IBM COS Slicestor nodes to be distributed across multiple sites and regions for scalability, accessibility, and security, as shown in Figure 6 on page 7.

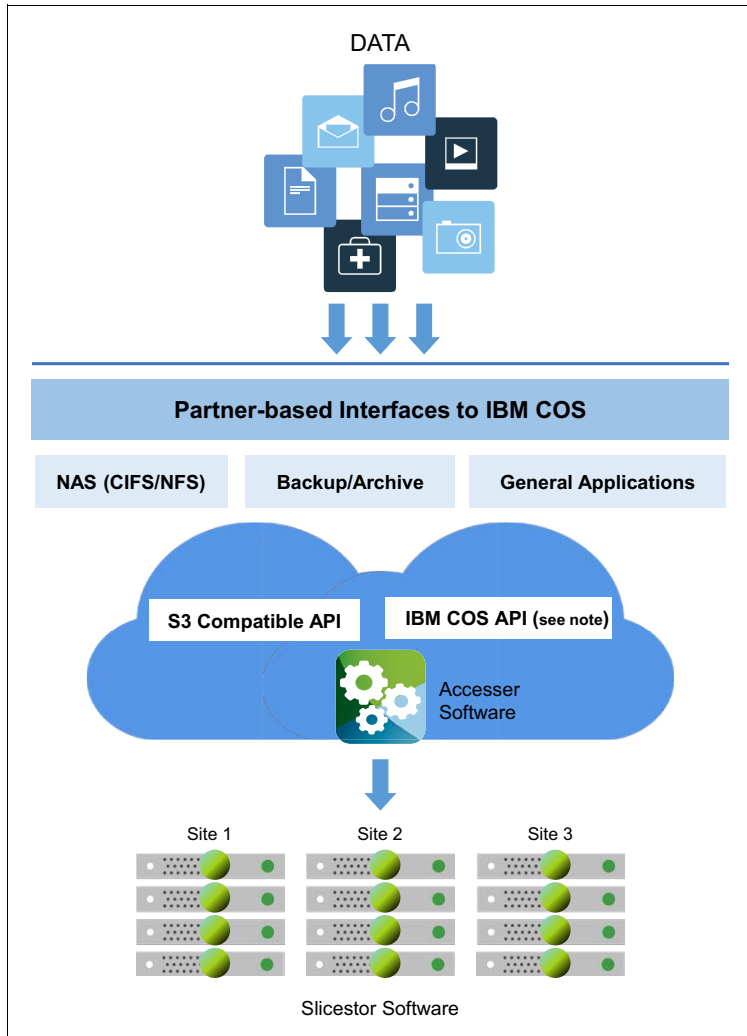


Figure 6 IBM Cloud Object Store multi-site architecture

IBM COS API: At the time of writing this book, IBM COS API is not available and expected to be available in 1H 2017.

How IBM Cloud Object Storage works

The copying and storage of data with IBM COS goes through several steps across the IBM COS Accesser and IBM COS Slicestor components. The following sections show the individual steps for reading and writing data to IBM COS.

Additional reference on how IBM COS works: You can refer to the following website for more information on the internals of IBM COS:

<https://www.ibm.com/cloud-computing/products/storage/object-storage/how-it-works/>

Client application pushes data to IBM Cloud Object Storage

As shown in Figure 7, client applications use one of the available API interfaces for IBM COS to push data to the Accessers.

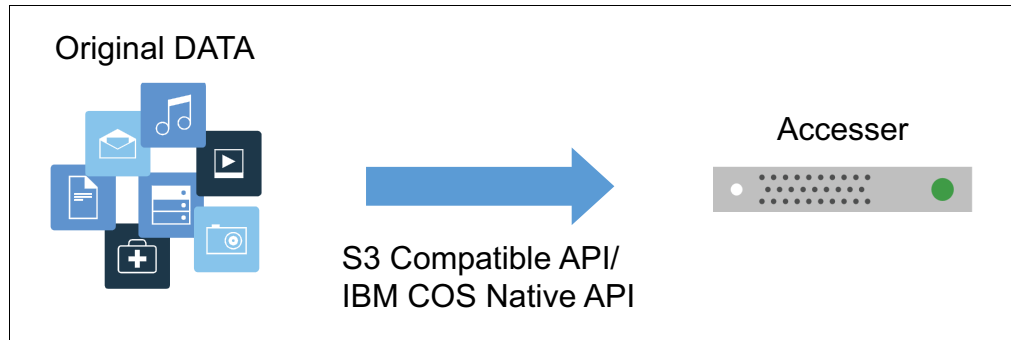


Figure 7 Client application copying data to IBM Cloud Object Storage

The IBM COS Accesser component splits the data into 4 MB segments. For example, a 1 GB object is split into 250 4 MB segments, as shown in Figure 8.

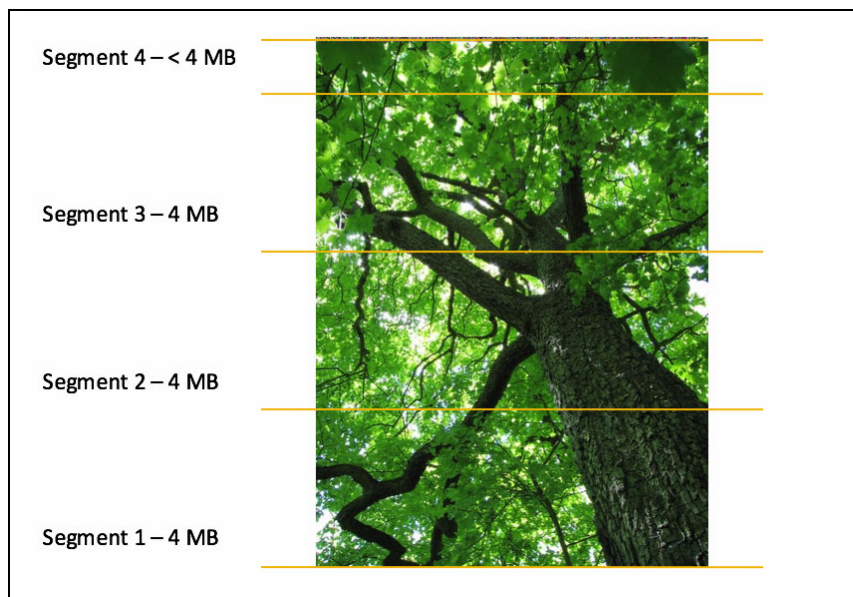


Figure 8 IBM Cloud Object Storage Accesser creating a 4 MB segments

Accesser slicing

The IBM COS Accesser component slices the individual segments of the input data based on the defined or default width, as shown in Figure 9 on page 9.

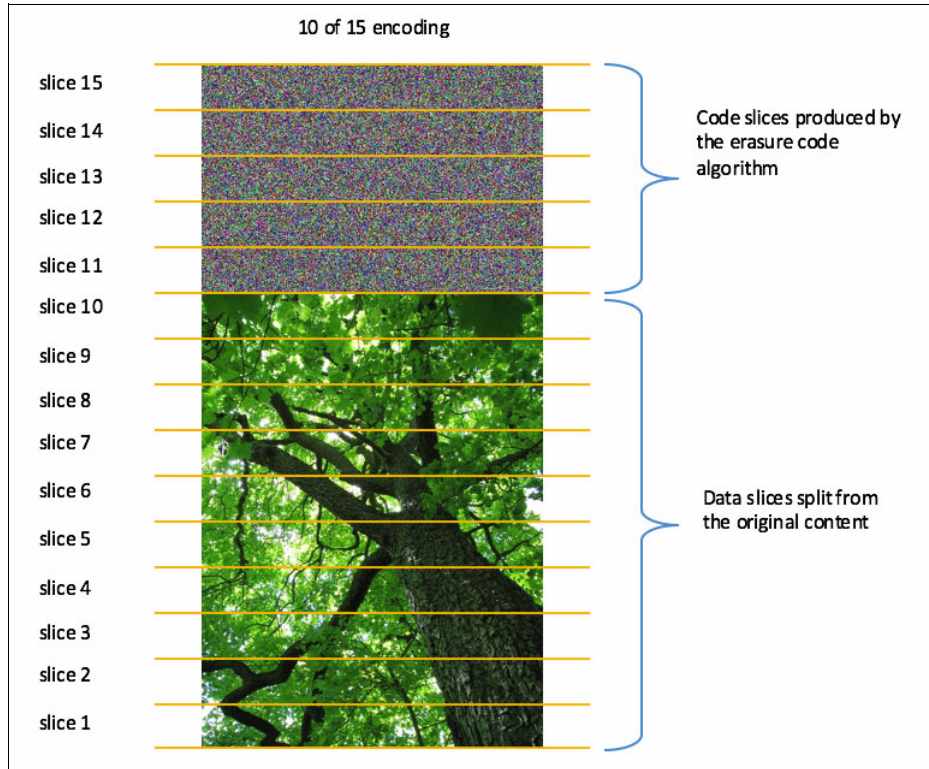


Figure 9 IBM Cloud Object Storage Accesser slicing individual segments

Accesser and SecureSlice to expand and transform data

This section describes how Accesser and SecureSlice expand and transform data.

SecureSlice

SecureSlice combines All Or Nothing (AONT) encryption with IDA to form a computational secret sharing scheme. AONT is applied as a preprocessing step to IDA.

Internals of AONT encoding

Here are the internals of AONT encoding:

1. Append an integrity check value.
2. Generate random encryption key = R.
3. Encrypt data by using encryption key R.
4. Calculate the hash of encrypted data = H.
5. Calculate H+R and append the result to the encrypted data to create the AONT package.

Figure 10 on page 10 shows the IBM COS SecureSlice AONT encryption.

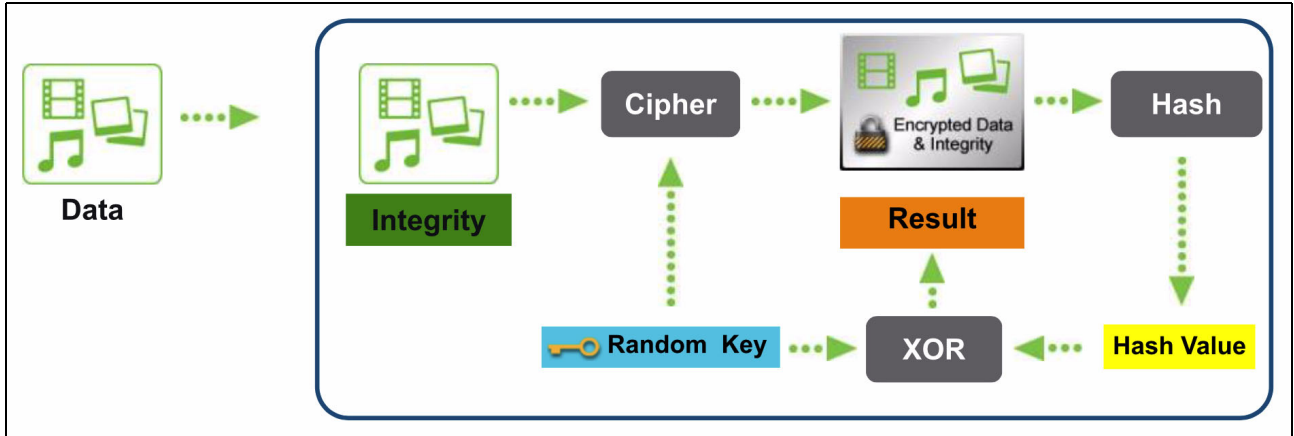


Figure 10 IBM Cloud Object Storage SecureSlice AONT encryption

The IBM COS Accesser component uses SecureSlice to transform the data into segments and slicing data based on defined width. Erasure coding is used as an additional layer of security where the data is not compromised even when the required number of slices are obtained, as shown in Figure 11 on page 10.

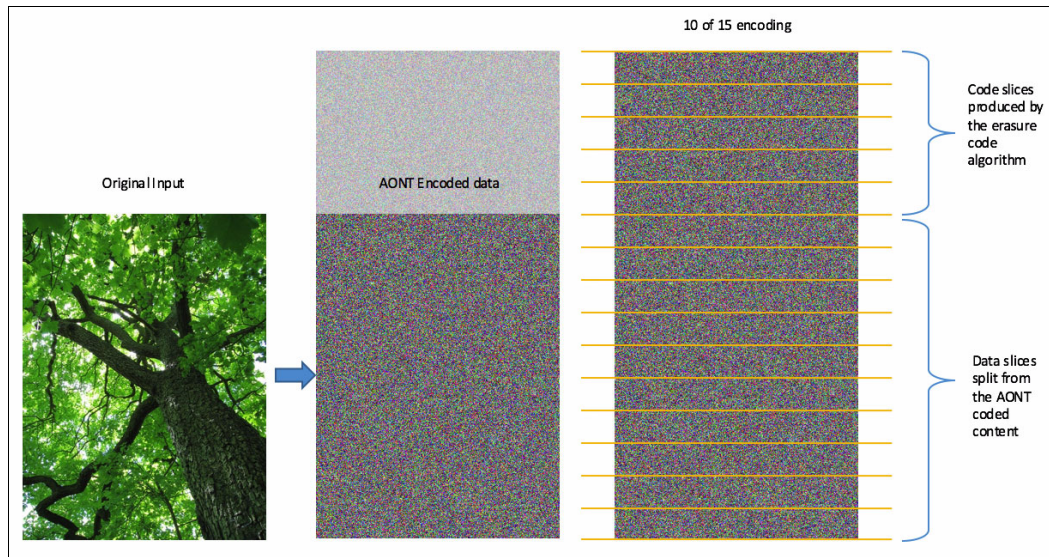


Figure 11 IBM COS Accesser SecureSlice

What this means for you: If security is compromised in a region, the full content will not be exposed. If one region is offline, your applications continue to run without disruption and without you having to intervene. This is called *always-on availability*, which is major benefit for IBM COS customers.

Storage management with Slicestor

The IBM COS Slicestor is responsible for distributing the slices (potentially encrypted by using SecureSlice) to the storage nodes, as shown in Figure 12.

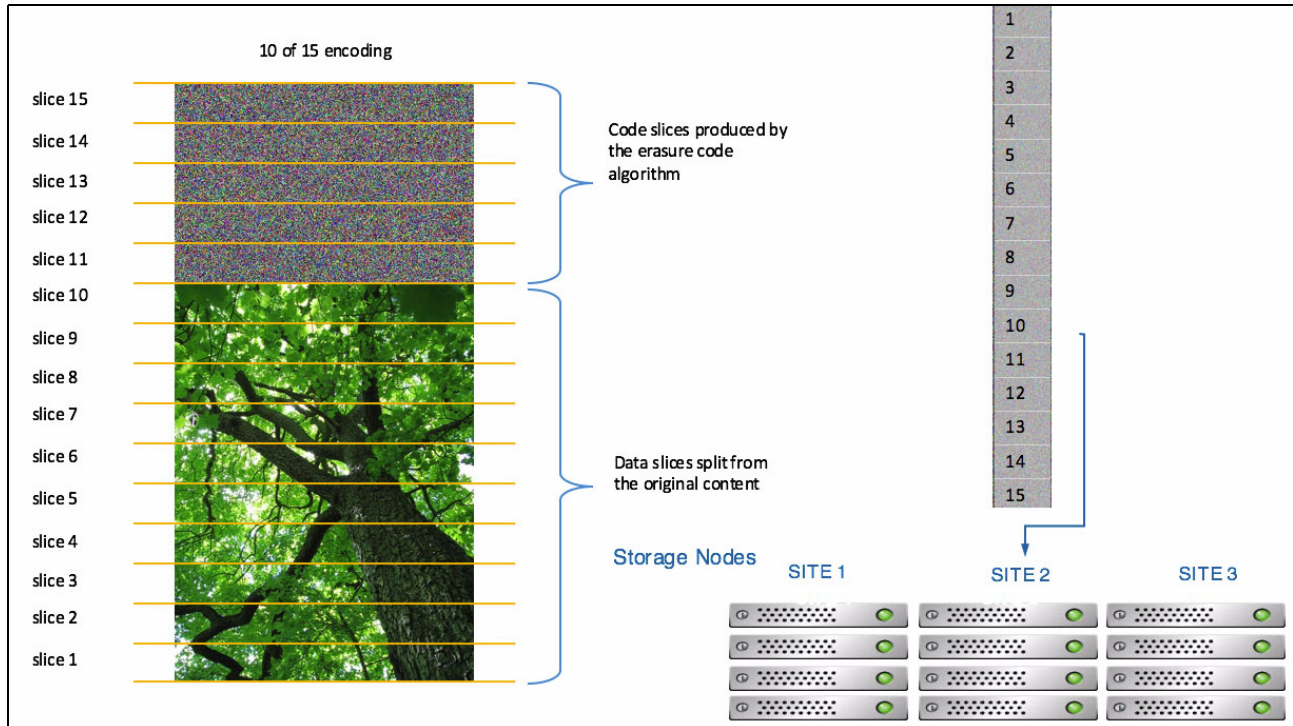


Figure 12 IBM Cloud Object Storage Slicestor copying data to storage nodes

What this means for you: IBM COS encrypts and slices up data as soon as it comes in, with the slices dispersed across multiple regions automatically. If a write operation is confirmed, data is protected *immediately*. This means that if a region goes down, data can still be delivered from the slices that exist in remaining regions. Applications that rely on that data remain up and running. They can survive regional outages.

Reading data from IBM Cloud Object Storage with IBM Cloud Object Storage Accesser

This section describes how data is read from IBM COS with IBM COS Accesser.

SecureSlice

SecureSlice combines AONT encryption with IDA to form a computational secret sharing scheme.

Internals of AONT decoding

Here are the internals of AONT encoding:

1. Strip the appended result from the end of the encrypted data.
2. Calculate the hash of the encrypted data = H.
3. Exclusive OR(XOR) the hash with the result to recover the random key = R.
4. Use the key to decrypt the data.
5. Verify the integrity of the data.

Figure 13 on page 12 show IBM COS SecureSlice decryption.

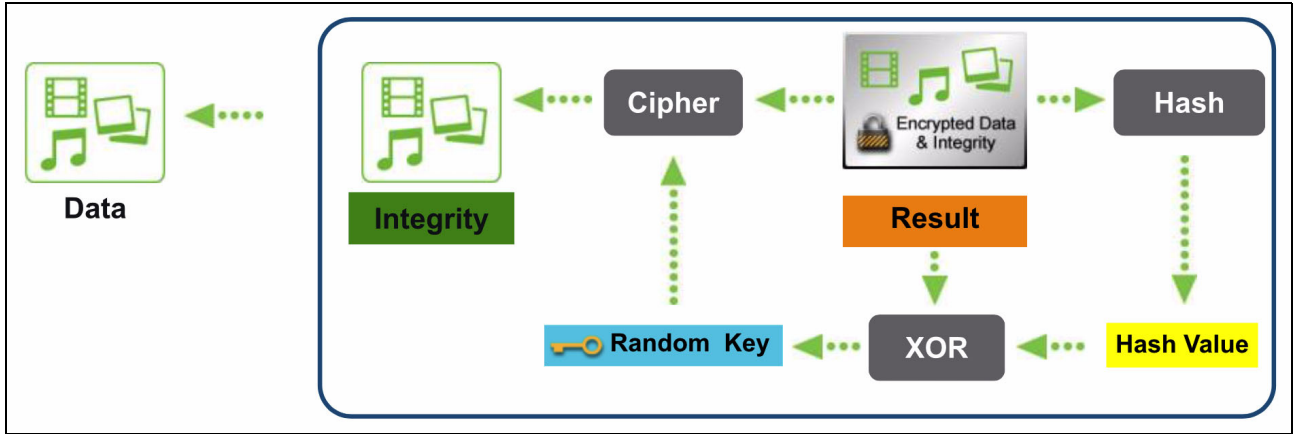


Figure 13 IBM Cloud Object Storage SecureSlice decryption

Figure 14 on page 12 shows reading data from IBM COS.

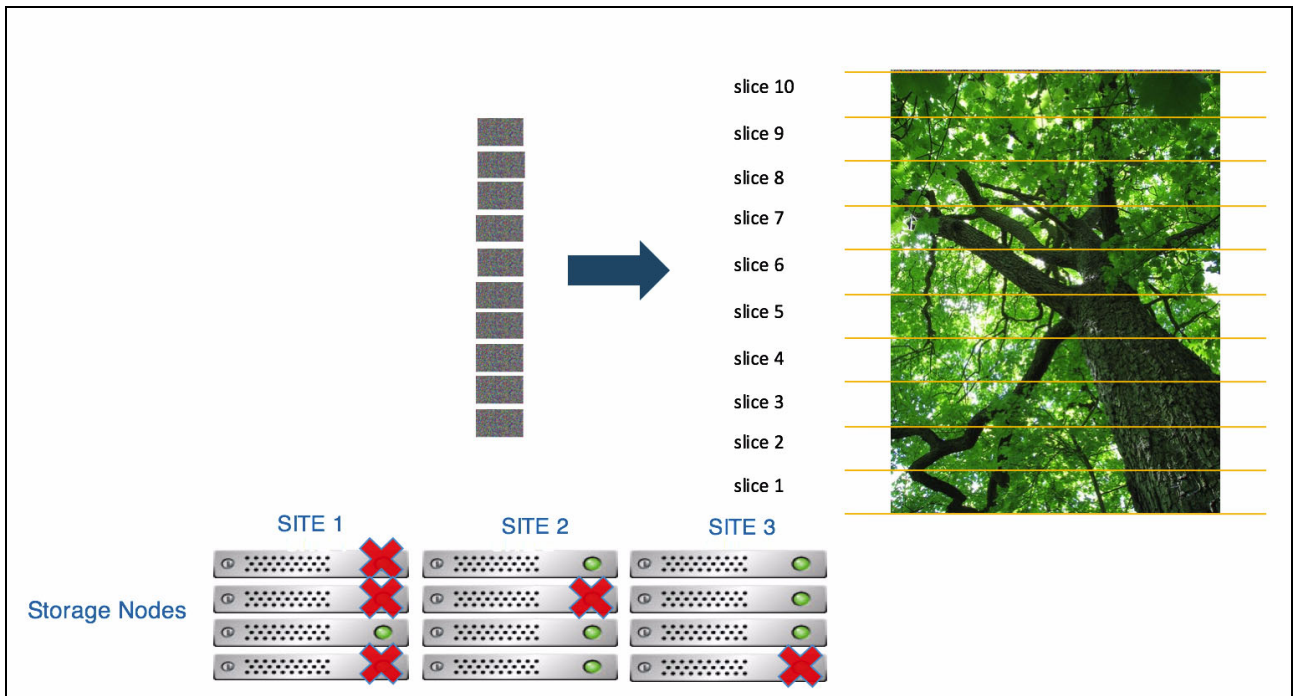


Figure 14 Reading data from IBM Cloud Object Storage

Service features

This section describes the service features of IBM COS.

Security

The underlying IBM COS Manager, IBM COS Accesser, and IBM COS Slicestor appliances have multiple levels of security:

- ▶ Operating system level security: Uses a paired-down Linux distribution that has only the minimal essential functions for the individual components. NSA hardening guidelines are applied.
- ▶ Firewall: Each appliance has its own firewall.
- ▶ Monitoring: Uses AES-encrypted SNMP messages and encrypted system logs.

Object-level authentication is done by using an Public Key and Secret Access Key mechanism.

Key characteristics

Here are the key characteristics of IBM COS security:

- ▶ All crucial configuration information is digitally signed to avoid being compromised.
- ▶ Certificate-based authentication of every node (Manager, Accesser, and Slicestor).
- ▶ Transmission and storage of data is inherently private and secure.
- ▶ No copy of data is present in any single disk, node, or location.
- ▶ TLS is supported for network connections within IBM COS for data-in-motion protection.
- ▶ TLS is supported on Client to Accesser network connections for data-in-motion protection.

Encryption

IDAs are used to separate the data into unrecognizable slices. These individual slices are distributed across a network of data centers making transmission and storage of data inherently private and secure. No complete copy of the object is stored in any single storage node.

Key characteristics

Here are the key characteristics of IBM COS encryption:

- ▶ SecureSlice is a standard product feature.
- ▶ SecureSlice encryption ensures the confidentiality of data at rest on Slicestor storage nodes if no more than N Slicestor nodes have their data exposed, where $N = \text{IDA Read Threshold} - 1$.
- ▶ SecureSlice does not require a key management system.
- ▶ SecureSlice can be configured to use any of the following combinations of encryption and data integrity algorithms:
 - RC4-128 encryption with MD5-128 hash for data integrity.
 - AES-128 encryption with MD5-128 hash for data integrity.
 - AES-256 encryption with SHA-256 hash for data integrity.

By default, encryption is enabled and all data is RC4 128-bit encrypted. There is an option to enable AES 256-bit encryption.

What this means for you: IBM COS is designed with a high level of security in mind to protect your data from security breaches. From built-in encryption of data at rest and in motion to a range of authentication and access control options, the IBM COS solution includes a wide range of capabilities designed to help you meet your security requirements. These security capabilities have been implemented to help enable better security without compromising scalability, availability, ease of management, or economic efficiency.

Scalability

IBM COS software has been tested at web-scale with production deployments that exceed hundreds of petabytes of capacity, and can scale to exabytes (EB).

Key characteristics

Here are the key characteristics of IBM COS scalability:

- ▶ Internet-style, distributed, shared-nothing, and peer-to-peer architecture.
- ▶ Yottabyte-scale global namespace with 10^{38} object IDs available per vault.
- ▶ Increased storage capacity and performance by adding Slicestor nodes.
- ▶ Scale to thousands of Slicestor storage nodes in a single system.
- ▶ No limit on the number of Accesser nodes. You can deploy the nodes as needed based on your performance requirements.
- ▶ Near-linear increase in system throughput and HTTP operations per second as the system grows.

What this means for you: IBM COS is built for cloud scale and can scale to exabytes (EB) while maintaining availability, reliability, manageability, and cost-effective options, without any compromise. The upward scalability is virtually unlimited with this offering.

Availability

IBM COS can operate with no downtime during software upgrades, hardware refreshes, and in the face of hardware failures, such as disk failure or node failure, or even when an entire site is unavailable. This availability is achieved due to geo-dispersed erasure coding.

Key characteristics

Here are the key characteristics of IBM COS availability:

- ▶ Non-disruptive code upgrades are initiated and managed by the IBM COS Manager component as rolling upgrades.
- ▶ Slicestor hardware refreshes involve installing new hardware, evacuating data from individual Slicestor nodes, and performing writes simultaneously to the new Slicestor nodes.

What this means for you: IBM COS provides *always-on availability*, which is major means it can tolerate even a catastrophic regional outage without down time or intervention. Continuous availability is built into the architecture.

Authors

This paper was produced by a team of specialists from around the world working at the International Technical Support Organization, San Jose Center.

Deepak Rangarao is an Executive IT Specialist with the Global Analytics CTO Office at IBM. He has more than 18 years experience in the telecommunications, public sector, finance and insurance industries in both pre-sales and post-sales capacities. Deepak's key consulting experience includes hybrid cloud, data management, data warehousing, advanced analytics and business intelligence solutions.

Deepak is a member of the IT Certification board at IBM mentoring IT Specialist / Architects and is also a member of the Global Cloud Expertise Council at IBM, helping our field technical staff and our customers understand the value of IBM Cloud. Deepak has a Masters Degree in Information Technology from RMIT, Australia and several developer certifications around Visual Basic, MS SQL Server (Microsoft Certified Professional), IBM Cloud Developer Certification, Apache Spark (O'Reilly Media).

Vasfi Gucer is an IBM Technical Content Services Project Leader with the Digital Services Group. He has more than 20 years of experience in the areas of systems management, networking hardware, and software. He writes extensively and teaches IBM classes worldwide about IBM products. His focus has been primarily on cloud computing for the last 6 years. Vasfi is also an IBM Certified Senior IT Specialist, Project Management Professional (PMP), IT Infrastructure Library (ITIL) V2 Manager, and ITIL V3 Expert

Thanks to the following people for their contributions to this project:

Bert Dufranse
International Technical Support Organization

Riz Amanuddin, Nicholas Lange, Wesley Leggette, Laura Noonan
IBM USA

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Stay connected to IBM Redbooks

- ▶ Find us on Facebook:
<http://www.facebook.com/IBMRedbooks>
- ▶ Follow us on Twitter:
<http://twitter.com/ibmredbooks>
- ▶ Look for us on LinkedIn:
<http://www.linkedin.com/groups?home=&gid=2130806>
- ▶ Explore new Redbooks® publications, residencies, and workshops with the IBM Redbooks weekly newsletter:
<https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm>
- ▶ Stay current on recent Redbooks publications with RSS Feeds:
<http://www.redbooks.ibm.com/rss.html>

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

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.

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.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.


COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

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

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

Redbooks (logo) ®
IBM®

Redbooks®
Redpaper™

The following terms are trademarks of other companies:

ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.

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

Other company, product, or service names may be trademarks or service marks of others.



REDP-5435-00

ISBN DocISBN

Printed in U.S.A.

Get connected

