basho

www.basho.com



Technical Overview

Simple, Scalable, Open Source Storage

Table of Contents

Table of Contents 1

Introduction & Overview 1

Architecture 2

- **How it Works** 2
- **APIs and Interfaces** 3
  - Administration and Users 3
  - Storage API 3
  - Multipart Upload 3
  - Storage and Usage Statistics 3
- **Operations** 4
  - Monitoring and Metrics 4
  - Scaling Out 4

Riak CS Enterprise: 4

Multi-Datacenter Replication 4

Next Steps 6

- **Getting Started** 6

All Content © Basho Technologies, Inc. | All Rights Reserved

Introduction & Overview

Riak CS (Cloud Storage) is enterprise cloud storage software. It is open source and built on top of Riak, offering S3-compatible, multi-tenant large object storage. It can be used for building public or private clouds, or as reliable storage for applications and platforms.

Use cases for Riak CS include:

- Storage for images, text, documents, videos and other large objects
- Internal storage providing on-demand capacity for business units
- A foundational storage layer for public clouds and cloud services
- Infrastructure for moving applications off of Amazon S3, or for hybrid solutions
- Redundant storage for backup and disaster recovery scenarios

The primary features of Riak CS are:

- Highly available, fault-tolerant storage
- Large object support
- S3-compatible API, authentication, and authorization
- Multi-tenancy with per-user reporting on usage and network I/O
- Simple operational model for adding capacity

Riak CS Enterprise, a commercial offering, adds:

- Multi-datacenter replication for active backups, disaster recover, and data locality
- Per-node or capacity-based pricing, inclusive of 24/7 support

This technical brief covers Riak CS architecture, APIs, and operations with a brief note on Riak CS Enterprise pricing. Additional technical details on Riak CS can be found in the online documentation.

All Content © Basho Technologies, Inc. | All Rights Reserved

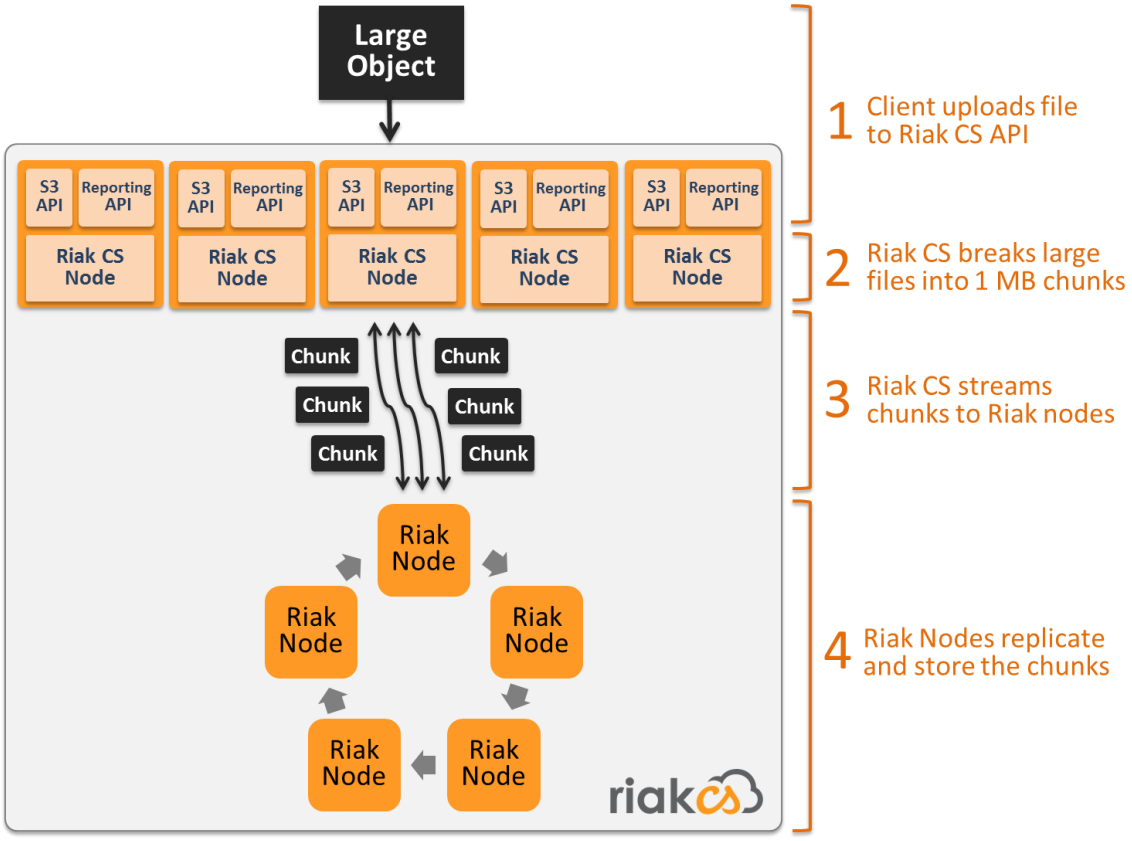
Architecture

Riak CS is a cloud storage product built on top of the open source, distributed database Riak. Riak CS handles large file uploads and exposes an S3-compatible API, user and administrative functions, and usage reporting. The underlying Riak cluster stores objects and metadata for retrieval - providing replication, fault tolerance and high availability for data stored in the system. We recommend a one-to-one mapping of Riak to Riak CS nodes, so there should be one Riak node for every Riak CS node. Ideally, Riak and Riak CS nodes should be run on the same physical machine to minimize unnecessary network traffic.

How it Works

In a Riak CS system, any node can respond to client requests - there is no master node and each node has the same responsibilities. Since data is replicated (three replicas per object by default) and other nodes automatically take over the responsibility of failed or non-communicative nodes, data remains available even in the event of node failure or network partition. More information on how Riak handles failure conditions can be found in the Riak documentation.

When an object is uploaded via the storage API, Riak CS breaks the object into smaller chunks that are streamed, written, and replicated in Riak. Each chunk is associated with metadata for later retrieval. The diagram below provides a visualization.



All Content © Basho Technologies, Inc. | All Rights Reserved

APIs and Interfaces

Riak CS has HTTP interfaces for handling storage operations, authentication, administration, and per-user storage and bandwidth statistics. Below is an overview and additional details can be found in the Riak CS documentation.

Administration and Users

Riak CS exposes an interface for user creation, disablement, and credential management. Once created and issued credentials, users are able to authenticate, create buckets, upload and download files to their account, retrieve account information, obtain new credentials, or disable their account through the API. Data access can also be defined based on IP address. Riak CS can be configured so only administrators can create users, or so that anyone can create a user directly. Administrator accounts have additional special privileges: they can retrieve a list of all users in the system and query the user account information of any user.

Riak CS uses a request serializer to ensure that global entities (Riak CS users and bucket names) are unique in the system. One instance of this serializer (Stanchion) must be installed in every Riak CS system.

Riak CS supports the standard S3 authentication scheme, with support for header and query parameter authorization. In addition, integration with OpenStack's Keystone authentication is available.

Storage API

The Riak CS storage interface is S3-compatible and can be configured for use with existing S3 tools, including s3cmd, official S3 libraries, and clients such as Fog and Boto. Like S3, the API has simple, RESTful GET, PUT, Copy, and DELETE operations for objects and buckets in addition to support GET range queries. Riak CS also supports the addition of arbitrary metadata to objects so users can store more useful information about their objects. Additionally, S3-style ACLs are provided for managing object and bucket permissions. Riak CS also supports integration with OpenStack's Object Storage API.

Multipart Upload

As of the Riak CS 1.3 release, Riak CS supports multipart uploads. At its simplest, multipart upload allows the uploading of a single large object as a series of parts. Once the individual parts have been uploaded, Riak CS presents the data as a single object. The parts can be between 5MB and 5GB in size.

Storage and Usage Statistics

Operators can use the Riak CS storage, usage, and network statistics to support use cases like accounting, subscription, billing, or multi-group utilization for public or private clouds. Riak CS will report information on how much storage a named user is consuming and the network operations related to accessing it. This data can be queried on the default timespan “now,” or as a range from start time through end time. Access statistics are reported as bytes in and bytes out for both object and bucket operations. Reporting of this information can be scheduled for a set interval or manually triggered. Additional details billing and usage reporting are available in the docs.

All Content © Basho Technologies, Inc. | All Rights Reserved

Operations

Riak CS is designed to be easy to operate and scale, with monitoring and metrics solutions and a simple model for adding capacity as needed.

Monitoring and Metrics

Riak CS exposes stats on critical operations that are accessible via HTTP request. Additionally, Riak CS and Riak nodes both have DTrace support for analysis of running systems. In version 1.5 Syslog support was added to facilitate log aggregation.

Scaling Out

Adding new capacity to a Riak CS cluster requires installing both Riak CS and Riak on a new physical node and joining the node to the cluster using several simple management commands. Riak automatically redistributes data in the system so all nodes have equal responsibility, preventing storage hot spots and decreasing the operational burden of adding new nodes.

As part of the Riak CS Enterprise license, Basho offers set up, configuration, and operational support for the implementation. Basho also provides Chef recipes to automatically install Riak CS on all nodes in the system.



Riak CS Enterprise: Multi-Datacenter Replication

Riak CS Enterprise is a commercial extension of Riak CS that provides multi-datacenter replication and 24/7 support.

Multi-datacenter replication for Riak CS Enterprise provides two modes of object replication: full-sync and real-time sync. Data is streamed over a TCP connection, and multi-datacenter replication has support for SSL so data can be securely replicated between sites.

In Riak CS, large objects are broken into blocks and streamed to the underlying Riak cluster on write, where they are replicated for high availability (three replicas by default). A manifest for each object is maintained so that blocks can be retrieved from the cluster and the full object presented to clients. For multi-datacenter replication in Riak CS Enterprise, global information for users, bucket information, and manifests are streamed in real-time from a primary implementation to a secondary site, so global state is maintained across locations. Objects can then be replicated in either full-sync or real-time sync mode.

In full-sync, objects are replicated from a primary Riak CS Enterprise implementation to a secondary site on a configurable interval – the default is six hours. During full-sync replication, each cluster computes a hash for each key's block value. Key/block pairs are compared, and the primary site streams any missing blocks or updates needed to the secondary site.

All Content © Basho Technologies, Inc. | All Rights Reserved

Real-time sync is triggered when an update is sent from a client to a primary Riak CS Enterprise implementation. Once replicated in the first location, the updates are streamed in real-time to the secondary site. But what if a client requests an object from the secondary cluster and not all of its blocks have been replicated to that cluster? With Riak CS Enterprise multi-datacenter replication, the secondary cluster will request any missing blocks from the primary cluster so the client request can be served quickly and accurately.

Riak CS Enterprise has two pricing models. Riak CS Enterprise can be purchased on a per-node, annual subscription or perpetual license. The per-node price is inclusive of the Riak CS nodes as well as the underlying Riak nodes required. Riak CS Enterprise can also be purchased in blocks of raw storage. Under this model, at higher storage tiers, the effective per-gig price of the license is lower. Both models include 24/7 support. Request a developer trial of Riak CS Enterprise to try it on your own infrastructure.

All Content © Basho Technologies, Inc. | All Rights Reserved

Next Steps

Getting Started

Riak CS packages are available on our downloads page and source code is available on Github.

If you are interested in Riak CS Enterprise, please contact us. We would love to talk to you about your possible use case and needs.

ABOUT BASHO

Basho Technologies is the leader in highly-available, distributed database technologies used to power scalable, data-intensive Web, mobile, and e-commerce applications and large cloud computing platforms. Basho customers, including fast-growing Web businesses and large Fortune 500 enterprises, use Riak™ to implement content delivery platforms and global session stores, to aggregate large amounts of data for logging, search, and analytics, to manage, store and stream unstructured data, and to build scalable cloud computing platforms. Riak is available open source for download. Riak Enterprise is available with advanced replication, services and 24/7 support. Riak CS enables multi-tenant object storage with advanced reporting and an Amazon S3 compatible API. For more information visit http://www.basho.com.

All Content © Basho Technologies, Inc. | All Rights Reserved