



# Multi-Datacenter Replication

A Technical Overview & Use Cases

# Table of Contents

---

- Table of Contents** ..... 1
- Introduction**..... 1
- How It Works** ..... 1
  - Default Mode**..... 1
  - Advanced Mode** ..... 2
- Architectural Strategies**..... 3
  - Primary Cluster With Failover** ..... 3
  - Active-Active Cluster Configuration** ..... 3
  - Availability Zones** ..... 4
  - Secondary Analytics Clusters** ..... 4
  - Public Cloud Use Cases** ..... 5
- Business Goals** ..... 6
  - Disaster Recovery**..... 6
  - Data Locality** ..... 6
  - Compliance With Regulatory Requirements** ..... 6
  - Secondary Analytics Clusters** ..... 6
  - Cloud Bursting and Multi-Region Cloud Deployments**..... 7
- Next Steps** ..... 8

# Introduction

Multi-datacenter replication is a critical part of modern infrastructure, providing essential business benefits for enterprise applications, platforms, and services. Riak Enterprise offers multi-datacenter replication features so that data stored in Riak can be replicated to multiple sites. With Riak Enterprise, data can be replicated across locations and geographic areas, providing disaster recovery, data locality, compliance with regulatory requirements, the ability to “burst” peak loads into public cloud infrastructure, and more.

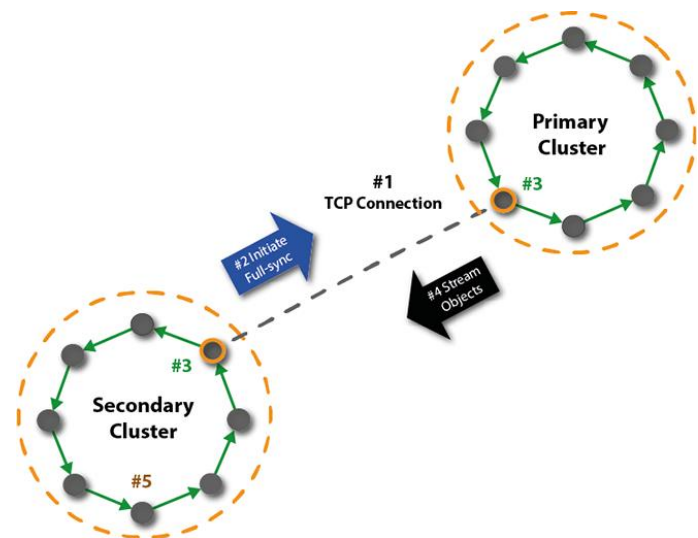
This technical brief introduces Riak Enterprise’s multi-datacenter replication and common use cases. It provides an overview of features and architecture, as well as considerations for developers and operators. It also looks at some common configurations for multi-datacenter replication, including backup/failover clusters, data geo-location, availability zones, secondary analytics clusters, and bursting from private to public cloud environments for peak loads or disaster recovery.

## How It Works

With Riak Enterprise, users have access to multi-datacenter replication features, as well as 24/7 support and other benefits. Riak Enterprise currently offers two approaches to multi-datacenter replication: a default mode and an advanced mode that offers newly upgraded replication features (available as of the Riak Enterprise 1.3 release, February 2013). While the advanced mode is already being run by Riak Enterprise customers and is ready for production usage, it will not replace the default mode until it is extended to support Secure Sockets Layer (SSL) and Network Address Translation (NAT), which are already supported with the current default mode. These support features are scheduled for release in mid-2013.

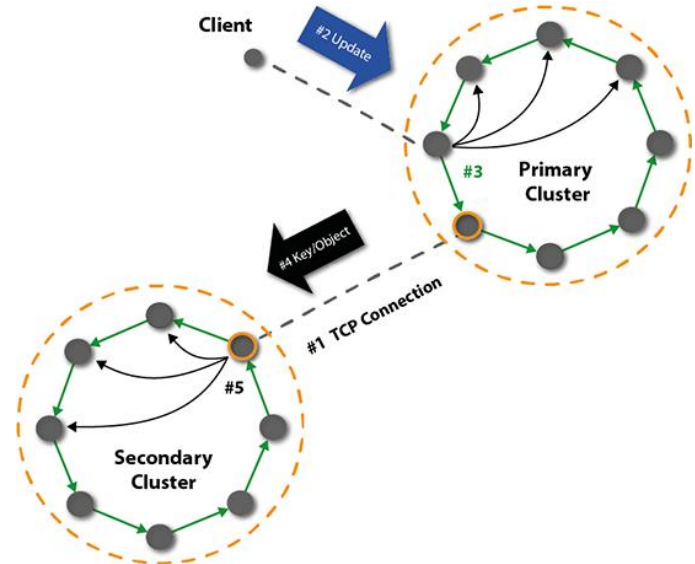
### Default Mode

Riak Enterprise’s default mode features two options for multi-datacenter replication: full-sync and real-time sync. These modes are complementary and designed to run simultaneously. With full-sync, replication of data occurs at scheduled intervals (default interval is six hours) between two clusters. When full-sync is initiated, clusters generate and compare hashes for all of their objects. During the comparison process, the “primary” cluster detects missing or out-of-date objects in the “secondary” cluster(s). It then streams any new objects or updates so they all have the same information.



With real-time sync, replication to the secondary datacenter(s) is triggered by updates to the primary datacenter. After writing an object to the primary cluster, writes are sent to the secondary cluster(s) via post-commit hook. Maintaining continuous real-time sync operation minimizes the amount of data exchange needed when full-sync is run and increases availability of replicated data on the secondary cluster.

All multi-datacenter replication occurs over a TCP connection. By default, this connection is unidirectional, however, bidirectional replication can be achieved by establishing two unidirectional connections between clusters. Replication can be configured for all data in the cluster or on a per-bucket basis, which allows for replication of a subset of data. Default mode supports SSL as well as NAT. For more information on the default mode for multi-datacenter replication, check out the [full documentation](#).



## Advanced Mode

With the default mode, there is a single TCP connection over which data is streamed from one cluster to another. Due to bandwidth limitations between datacenters, this single connection can cause a performance bottleneck, especially when users have significant amounts of data, frequently update data, or when the connection is run on nodes constrained by per-instance bandwidth limits (such as in a cloud environment). In the advanced multi-datacenter replication mode, multiple concurrent TCP connections (approximately one per physical node) and processes are used to maximize performance and network utilization. As with default mode, advanced mode supports full-sync and real-time sync between clusters, as well as their simultaneous operation.

Major improvements to the implementation of the advanced mode replication include full concurrency of TCP channels between nodes, graceful shutdown and handoff of the real-time queues when nodes are taken down, additional stats for monitoring performance, flags to indicate when full-sync is needed, and much easier replication setup and operation. Additionally, all concurrent connections use a single TCP port, which reduces network setup, while still giving each connection full independent bandwidth.

As of the Riak Enterprise 1.3 release, the advanced mode does not yet support SSL or NAT. Additionally, it does not yet support full-sync scheduling, meaning that full-sync must be triggered manually rather than occurring automatically at a set interval. In the next Riak Enterprise release, scheduled for mid-2013, these features will be added and the advanced mode will become the default mode for multi-datacenter operations. For more information on advanced mode, visit the [full documentation](#). In the interim, full-sync can be scheduled using the command line interface and a cron table, as suggested.

# Architectural Strategies

This section reviews the common use cases and architectural strategies for Riak Enterprise multi-datacenter replication.

## Primary Cluster With Failover

One of the most common architectural patterns for multi-datacenter replication is maintaining a primary cluster that serves traffic and a backup cluster for emergency failover. Maintaining a backup cluster can also be an important component of compliance with regulatory requirements, ensuring business continuity and access to data even in serious failure modes.

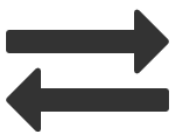
In this configuration, a primary cluster serves as the production cluster from which all read and write operations are served. A backup cluster is maintained in another datacenter. In the event of a datacenter outage or critical failures at the primary site, requests can be directed to the backup cluster either by changing DNS configuration or rules for routing via a load balancer.

Depending on business requirements, operators may use either full-sync or real-time sync for replicating data from the primary to the backup cluster. For some use cases, keeping data up-to-date within a 24-hour period is sufficient, however, sometimes more frequent syncs may be required. If “hot” failover is required, real-time sync can be used. Maintaining continuous real-time sync both keeps the data more up to date on the secondary, and reduces the load/time required for a full-sync operation.

Operationally, it is recommended that the failover strategy is tested periodically so any potential issues can be resolved in advance of a crisis. The failover strategy should also be fully defined upfront - know the conditions in which a failover mode will be invoked, decide how traffic will be directed to the backup, and document and test the failover strategy to ensure success.

**Note:** Additional backup clusters can be maintained by replicating the primary cluster to two or more backup sites, or configuring a “chain” of datacenters so one backup replicates to additional backups in other datacenters. However, the chaining of real-time sync data over multiple cluster “hops” is not planned until a release in mid-2013 (known as cascading writes feature).

## Active-Active Cluster Configuration



For some use cases, it may be desirable to maintain two (or more) active, synced clusters that are both responsible for serving data to clients. This approach is generally used to achieve data locality - when clients are served at low latency by whatever datacenter is nearest to them. An added benefit of this approach is, in the event of a datacenter failure where one of the clusters is hosted, all traffic can be served from the other, still-functional site for business continuity.

Bidirectional replication is often desirable for this scenario, so both clusters have all data and updates. For data locality, requests can be load balanced across geographies, with geo-based client requests directed to the appropriate datacenter. For example, US-based requests can be served out of a US-based datacenter while EU-

based requests can be served out of a regional site. For situations where not all data needs to be shared across all datacenters (or if certain data, such as user data, must only be stored in a specific geographic region to meet privacy regulations), multi-datacenter replication can be configured on a per-bucket basis so only shared assets, popular assets, etc. are replicated.

## Availability Zones



Availability zones are common for enterprises providing cloud services, either as a public platform or for internal consumption. Availability zones provide efficient multi-datacenter replication and data redundancy within a geographic region (such as a coast or a country). In this configuration, data is replicated within an availability zone's series of datacenters. In the event that one of datacenters experiences an outage or serious failure, data can still be served from other datacenters within the same region.

There are multiple approaches to setting up availability zones using Riak Enterprise. One approach is to have a "primary" site in a region to which all reads and writes for specific users, applications, or data sets are directed. This primary cluster can then be replicated to one or more proximal secondary clusters, either using real-time sync or full-sync, depending on the business case.

In other approaches, data can be replicated in real-time from one cluster to both another datacenter and other cold backups maintained for emergency conditions. The right approach is highly dependent on the requirements of users, availability, expense of bandwidth, and other constraints. If this structure is relevant to your business use case, please [contact us](#) to further discuss the best way to meet your needs.

## Secondary Analytics Clusters



As mentioned earlier, many Riak Enterprise users need to serve heavy production traffic and perform other computationally intensive tasks such as MapReduce. Since the request patterns of writing and reading data differ significantly from distributed search, analytics, and aggregation tasks, performing both types of computation on the same cluster isn't ideal. Analytics workloads can cause degraded performance and periods of higher latency for regular GET/PUT/DELETE operations.

An alternative to running these workloads on the same cluster is to replicate data from the primary cluster, which is responsible for serving all production requests, to a secondary cluster on which analytic and other computations can be performed. Replication can be configured to occur on a given interval depending on the nature and temporal requirements of the analytics tasks. Additionally, all of the data or only some of the data (via per-bucket replication settings) can be replicated depending on your needs.

To decide if a secondary analytics cluster is right for your use case, it is necessary to determine the profile of production traffic on your system as well as the profile of the analytic/aggregation tasks you must perform. This will determine if both workloads can be handled on the same cluster with the desired performance and latency results. If necessary, we offer a discount on Riak Enterprise pricing for secondary clusters that are used for analytics and other tasks. If you are interested in using Hadoop for your analytics requirements, there is also an alpha Hadoop connector for Riak. We are actively looking for users to help us test and develop this solution. Please get in [touch](#) to learn more and gain access.

## Public Cloud Use Cases



Public clouds are becoming a critical component of enterprise infrastructure. Riak is designed to be easy to use and operate on public clouds, and is partnered with many of the leading cloud providers. We provide virtual machine images on Amazon Web Services, Microsoft Azure, and Joyent. Hosted Riak is also available from Engine Yard and more explicit support for cloud platforms is being added all the time. Of course, Riak packages can be manually installed on any physical or virtual provider even if a machine image isn't explicitly supported.

There are several use cases for Riak Enterprise's multi-datacenter replication in the public cloud. Many enterprises want to maintain a cold or hot backup of their cluster in a public cloud for business continuity in the event of a datacenter outage in their private infrastructure. For this use case, please see the earlier section on primary clusters with failover for configuration options.

For other customers, the public cloud can provide a more cost-effective way of meeting peak loads, rather than building out private infrastructure to accommodate them. For example, many retailers and media providers need to offer increased capacity over the holiday season. Riak Enterprise is used to scale out capacity on public clouds over these periods, either with full-sync or real-time sync depending on the business needs.

Finally, some enterprises run certain applications or services entirely on public clouds. For these users, redundancy and data locality across public cloud availability zones is necessary for optimal performance and resiliency. Riak Enterprise's multi-datacenter replication allows clusters to be easily replicated across availability zones.

# Business Goals

---

Some common business goals that Riak Enterprise's multi-datacenter replication can help with include:

## Disaster Recovery

While your infrastructure must be designed to remain available despite intra-cluster failure like node failure and/or network partition, that isn't enough. Your business must also be resilient to broader-scale failures, including natural disasters, severe network outages, cascading failure conditions, and more. Riak Enterprise's multi-datacenter replication ensures data is replicated to geographically dispersed locations, allowing for efficient failover to a backup in the event of datacenter failure.

## Data Locality



Modern enterprises must be able to serve users, partners, and clients all over the world with predictable and low latency. Requiring requests to traverse large geographic distances to serve clients – whether across a country or the world – can lead to high latency and a degraded user experience. Riak Enterprise's multi-datacenter replication can be used to create a global data footprint, which, when combined with appropriate routing and load balancing strategies, allows client requests to be served from geographically closer sites. This can dramatically lower the latency of read and write operations for all users.

## Compliance With Regulatory Requirements

For some classes of data, regulatory requirements make it mandatory that data be replicated and available across multiple physical locations. Whether this applies to all of your data or a subset, such as certain types of sensitive information, Riak Enterprise's multi-datacenter replication makes it possible for enterprises to stay current with the latest industry regulatory requirements, while providing a low operational footprint and scalable design.

## Secondary Analytics Clusters



Many users have data in Riak that must serve heavy production traffic. Some of these users also need to perform other types of intensive computation on the data, i.e., using MapReduce to perform aggregation and analytics. Because the request patterns of serving traffic (setting and retrieving data) tends to differ significantly from the request patterns of distributed analytics and search workloads (aggregation/analytics tasks), running these workloads on the same cluster can result in degraded performance and an unpredictable latency profile. For customers who need to both serve low latency data at scale and perform bulk or analytics processing, having a primary cluster for serving clients and a secondary cluster for other workloads results in a much more uniform and performant architecture. Riak Enterprise's replication technology can be used to maintain this architecture, whether in a single datacenter or spanning geographically dispersed sites.

## Cloud Bursting and Multi-Region Cloud Deployments



More and more enterprises are using public cloud infrastructure as their primary IT platform or as part of a hybrid strategy that includes private datacenters as well as hosting providers. For companies operating entirely in a cloud environment, Riak Enterprise’s multi-datacenter replication allows for straightforward data redundancy and locality between availability zones, something that is otherwise often operationally intensive and can be prohibitively expensive on public clouds. This allows companies to withstand public cloud failures in a specific region or zone, and/or simply ensure that clusters are serving traffic at low latency, no matter where the clients are. Other companies may want to maintain hot or cold backups in a public cloud provider in case of a datacenter outage, or to replicate data to the cloud in times of peak load (i.e. the holiday season, major events, or other times of elevated traffic).

## Next Steps

---

Full documentation for Riak Enterprise is available at <http://docs.basho.com>.

If you are evaluating Riak Enterprise and are interested in multi-datacenter replication features, please [contact us](#). We would love to arrange a tech talk with your team, or answer any questions about our product and how customers are using it in production to meet their business goals. If you want to try Riak Enterprise, we can provide a free developer trial that you can set up on your own hardware and evaluate on your own time. Finally, our Professional Services Team can assist you in planning, setting up and optimizing your multi-datacenter strategy.

### 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>.