

Tencent Cloud Distributed Cache (Redis OSS-Compatible)

Product Introduction

Product Documentation



Tencent Cloud

Copyright Notice

©2013–2026 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

Trademark Notice



All trademarks associated with Tencent Cloud and its services are owned by the Tencent corporate group, including its parent, subsidiaries and affiliated companies, as the case may be. Trademarks of third parties referred to in this document are owned by their respective proprietors.

Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

Contents

Product Introduction

Overview

Product Strengths

Use Cases

Storage Engine

Product Series

Proxy Features

Standard Architecture

Cluster Architecture

Product Versions

Major version feature list

Minor Version Update History

Valkey

Redis

Proxy Version Update History

Version Lifecycle Plan

Specifications and Performance

Read/Write Separation

Multi-AZ Deployment

Regions and AZs

Terms

Service Regions and Service Providers

Product Introduction

Overview

Last updated: 2026-03-18 11:44:48

Tencent Cloud Distributed Cache Database is compatible with both Redis and Memcached protocols, featuring high availability, reliability, and scalability.

Storage Engine

Tencent Cloud Distributed Cache supports the following three storage engines. For more information, see [storage engine](#).

Storage Engine	Compatible Version	Description
Redis Edition	4.0,5.0,6.2,7.0	Fully compatible with the Redis protocol, it provides a mature and stable caching solution.
Valkey Edition	8.0	Compatible with the Redis protocol, as an open-source branch of Redis, it provides performance optimizations.
Memcached Edition	1.6	Compatible with the Memcached protocol, designed specifically for caching scenarios.

Architecture Version

It provides both Standard Edition and Cluster Edition architectures to meet business needs of different scales. For more information, see [Product Series](#).

Architecture	Feature	Scenarios
Standard architecture	<ul style="list-style-type: none">● Master-Slave Architecture: Using a two-node master-slave architecture, with the primary node and secondary node deployed on different physical servers.● Replica Configuration: Supports 1 to 5 replicas (secondary nodes).● Proxy Component: Built-in Proxy component responsible for request routing and load balancing.● Capacity Specifications: A single instance supports 0.25GB to 64GB.	General caching services requiring data reliability and availability.

	<ul style="list-style-type: none"> ● Cluster Mode: Automatic data synchronization between primary and secondary nodes. 	
Cluster Architecture	<ul style="list-style-type: none"> ● Distributed Architecture: Built on the community edition Redis Cluster. ● Shard Scaling: Supports horizontal scaling from 1 to 128 shards. ● Replica Scaling: Supports 1 to 5 replica. ● Proxy Component: Built-in Proxy component responsible for request routing, load balancing, and cross-shard command processing. ● Shard Specifications: Provides multiple specifications ranging from 1GB to 64GB. ● Cluster Mode: Automatic sharding, data balancing, and data migration. 	Distributed caching services requiring large-capacity storage and high-concurrency performance.

Key Strengths

Tencent Cloud Distributed Cache offers high availability, elastic scaling, and intelligent O&M as core capabilities, ensuring stable and reliable caching for businesses.

High Availability

Based on Tencent Cloud's self-developed high-availability system, it provides cross-device and cross-AZ disaster recovery capabilities, enabling automatic switch within seconds during failures to ensure stable business operations.

- Standard architecture: A multi-replica mode with one primary node and multiple replicas (such as 1 primary node and 5 replicas) is supported, providing auto-failover to effectively prevent a single point of failure and ensure business continuity.
- Cluster architecture: Each shard independently implements primary-secondary replication, with automatic isolation and failover for shard-level failures, ensuring other shards remain unaffected.

Elastic Scale-Out

Both horizontal and vertical elastic scaling are supported, with a smooth scaling process that ensures continuous business operations and helps you easily handle business traffic peaks and sustained data growth.

- Multi-dimensional scale-out: It supports horizontal scale-out of shards (for performance improvement) and vertical scale-out of replicas (for read capability enhancement), with built-in automatic read-write separation.
- Automated Ops: During shard scaling, data automatically migrates between nodes, significantly reducing the complexity and risks of Ops.

Deep Intelligent Monitoring

It provides multi-dimensional monitoring metrics from the Proxy layer to the engine layer, along with built-in diagnostic analysis tools, helping you move from reactive troubleshooting to proactive performance optimization, making it easier to identify and resolve performance bottlenecks.

- Multi-dimensional monitoring: Through multi-channel collection at the Proxy and engine layers, it provides comprehensive monitoring metrics, including performance, latency, network, capacity, and hit rate.
- Visual insights: It presents complex performance data through intuitive charts, enabling you to easily grasp the health status and operational trends of instances.
- Autonomous diagnostic analysis: It features built-in advanced capabilities such as hot Key collection, big Key analysis, and latency analysis to automatically identify common performance issues.

Key Concepts

The following are the core terms of Tencent Cloud Distributed Cache, helping you quickly grasp the product architecture and features.

Concept	Description
Instance.	A database environment that runs independently in Tencent Cloud, where an instance can contain multiple user-created databases.
Replica	A replica is a data backup node for the primary node, used for data redundancy and failover.
Shard	A shard is the smallest data storage unit in a cluster, consisting of a set of primary/secondary nodes, used for data distribution and horizontal scaling.
Node	The basic operational unit within an instance, categorized into master nodes (Master) and replica nodes (Replica).
Proxy	The proxy layer is responsible for request forwarding, routing, and load balancing.
Region	The instance is located on the Tencent Cloud region.
Availability Zone	A physical area within the same region with independent power and network.
VPC	A user-defined virtual network space that is logically isolated from other resources.
Security Group	A virtual firewall that performs security access control for instances, specifying IP address, protocol, and port rules.

Console	Tencent Cloud provides a visual management interface that enables users to perform management tasks, including instance creation, configuration, monitoring, and operations, via a web page.
Horizontal scaling (shard scaling)	Increasing the number of shards enhances the overall performance and capacity of the instance.
Vertical scaling (replica scaling)	By increasing the number of replicas to enhance the read capacity of the instance.
Read–write separation	In scenarios with more reads than writes, the primary node handles write requests while replica nodes offload read requests.
Project	Tencent Cloud provides a resource management approach to group cloud products by business or team, facilitating unified management and permission control.

Start Now

1. Configure a Tencent Cloud Distributed Cache database instance and specify the following key parameters based on your business requirements:
 - Select a region and an AZ.
 - Select a storage engine (Redis/Valkey/Memcached).
 - Select the architecture edition (Standard Architecture/Cluster Architecture).
 - Set the memory capacity and the number of shards.
 - Configure the network and security group.

For specific operations, see [Quick Instance Creation](#).

2. Obtain the instance connection address, connect to the database instance via a client, and perform basic read and write operations. For specific operations, see [Connecting to an Instance](#).
3. Monitor the instance status and test database performance. For specific testing methods, see [Specifications and Performance](#).

Product Strengths

Last updated: 2026-03-17 16:36:34

Ease of Use

- **Open-source compatibility:** 100% compatible with open source Redis and Memcached protocols. You do not need to modify any code and can directly use open source clients to communicate with Tencent Cloud Distributed Cache database instances.
- **Installation-free use:** You can directly request a Tencent Cloud Distributed Cache database instance on the Tencent Cloud platform purchase page, select the desired system architecture, with no need to install the instance manually, enabling convenient and fast use.
- **Cloud migration:** Tencent Cloud distributed cache database provides comprehensive support for data migration in self-built scenarios, including Tencent Cloud self-built environments, VPN self-built, dedicated line self-built, IDC self-built, and other scenarios, supporting both cold migration and hot migration capabilities.

Ultra-High Performance

Note:

For performance data reference of Tencent Cloud distributed cache database, see [Specifications and Performance](#).

- **High throughput:** The Standard Edition has a performance of up to 100,000+ QPS for concurrent response, and the Cluster Edition supports up to tens of millions of QPS as the number of shards increases. Their ultra-high performance can perfectly meet the needs in most business scenarios ranging from gaming, mobile apps to advertising and ecommerce.
- **Low latency:** Tencent Cloud Distributed Cache delivers stable low-latency response, where 99% requests can be responded to within 1 ms or 10 ms under 50% or 90% load respectively.
- **Read expansion:** In hot key scenarios, Tencent Cloud Distributed Cache allows dynamically increasing replicas to expand the read performance. It supports up to 5 replicas to read up to 500,000 hot keys.
- **Write expansion:** The performance of the Cluster Edition equals to the shard quantity multiplied by the performance per shard and increases linearly as the shard quantity increases between 3 and 128.

Flexible Architecture

Tencent Cloud Distributed Cache supports multiple system architectures to meet the needs of various business scenarios. You can deploy the appropriate architecture based on your actual business requirements.

- **Standard architecture:** Refers to the version that supports one or more replicas (where replicas denote non-master nodes), representing the most common distributed cache database version.
- **Cluster architecture:** Adopting a distributed architecture, it supports scaling out and scaling down shards and replicas, offering high flexibility, availability, and performance of up to tens of millions of QPS.
- **Architecture upgrade:** Supports upgrading from the standard architecture to the cluster architecture to meet the need of transitioning when the standard architecture faces insufficient performance and capacity.
- **Read/Write separation:** Automatic read/write separation is implemented at the proxy layer. The proxy writes data only to the master node, while read requests are automatically routed to the configured read-only nodes by load balancing.
- **Multi-AZ deployment:** It refers to the disaster recovery architecture where the master and replica nodes of an instance are deployed in different AZs in the same region to deliver a higher availability.
- **Global replication:** Based on the original master-replica replication scheme, a new log file is added for remote replication to ensure the eventual data consistency for instances in different regions in the replication group. This solves the problem of cross-region data inconsistency and offers region-level disaster recovery capabilities.
- **Automatic disaster recovery:** Tencent Cloud Distributed Cache adopts a hot standby architecture with dual servers. Upon host failure, access switches to the standby server within seconds without any user intervention. This saves the manpower and time costs required for developing master-replica systems.

Diverse Specifications

Tencent Cloud Distributed Cache offers capacity specifications of 0.25GB – 8TB, available in both standard and cluster editions.

- **Standard architecture:** Supporting 1 – 9 replicas to meet varying requirements for availability and performance across different scenarios. When the number of replicas is ≥ 1 , read/write separation can be enabled to improve read performance through replica nodes.
- **Cluster architecture:** Supports horizontal scaling of 1–128 shards and 1–9 replicas. Scaling out, scaling in, and specification changes can be performed with minimal service disruption, ensuring maximum service availability.

Data Security

- **RDB persistence:** The TencentDB backend service periodically creates snapshots for the data stored in the storage engine according to the backup policy to generate RDB files and then saves them in disks for data persistence.
- **Network security protection:** Tencent Cloud Distributed Cache supports configuring security groups in VPCs to implement allowlist-enabled network access control, which ensures the security and reliability of network environments.

- **Account authentication:** Tencent Cloud Distributed Cache allows you to authorize root accounts, sub-accounts, or across accounts, so you can manage your resources at a fine-grained level and benefit from enterprise-grade security protection.
- **Data isolation:** Tencent Cloud Distributed Cache supports multiple layers of data isolation between different regions, AZs, networks, and account levels to ensure the data security and integrity.

Intelligent Ops

- **System monitoring:** The instance usage process is fully transparent. It supports configuring alarm rules through Tencent Cloud Observability Platform (TCOP), providing over 30 automated monitoring metrics. This helps you maintain full visibility into the operational status of your cache database service and quickly troubleshoot issues.
- **TencentDB for DBbrain (DBbrain):** It monitors and diagnoses database instance exceptions in real time (including slow logs, big keys, hot keys, and delays), automatically generates health reports, and gives expert optimization suggestions. This helps you promptly optimize the database performance.
- **Management and Control API Interfaces:** Tencent Cloud Distributed Cache database provides a complete set of management and control API interfaces for implementing a series of resource self-management and Ops features.

Comparison with Self-Built Databases

Tencent Cloud Distributed Cache, provided as a service, offers significant advantages in flexibility and ease of use, high availability, fully managed Ops, and data security and reliability.

Item	Self-Built Cache Database	Tencent Cloud Distributed Cache Database
Performance	80,000–100,000/shard.	80,000–100,000/shard.
Cost	You need to purchase hardware on your own, and the cost of a single storage server is relatively high. Building a high-availability master-slave architecture usually requires purchasing three devices, which can easily lead to resource redundancy. Additionally, you need to hire professional database administrators, resulting in high labor costs.	No hardware or software investment is required. Multiple instance specifications are provided, allowing you to request resources on demand with high resource utilization. Free Ops features including management, backup, cloning, monitoring, and alarms are offered, delivering higher overall cost-effectiveness.
High Availability (HA)	Fault recovery time generally exceeds 60 seconds.	Failover time is less than 60 seconds.

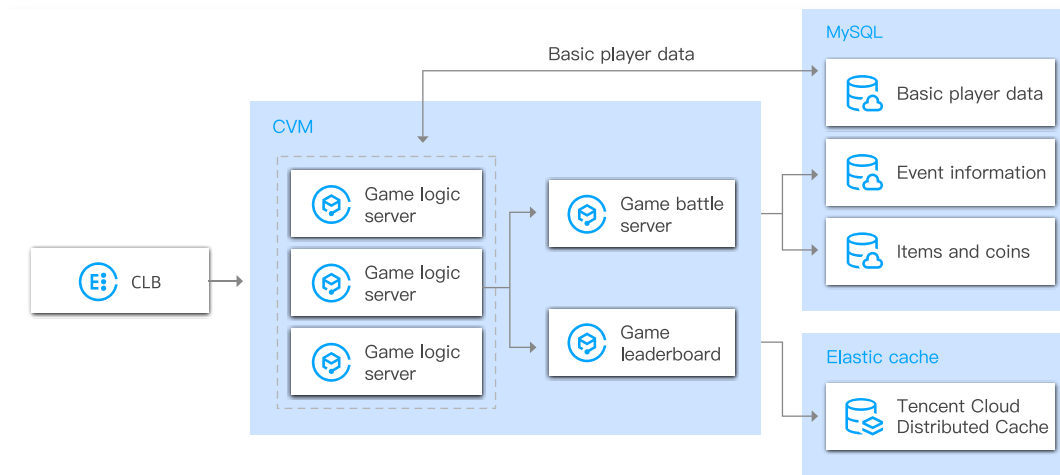
Availability	You need to build the primary–secondary replication architecture on your own and handle failures, resulting in high Ops complexity.	It supports multi–node hot standby, automatic disaster recovery and failover that are transparent to services, achieving 99.95% availability. It also supports prioritized reading from secondary nodes to ensure high–concurrency read performance.
Scale–out/Scale–in.	Requires purchasing hardware, hosting data centers, and redeploying applications, resulting in a lengthy process. During data migration, brief outages and migration risks may occur.	One–click online scale–out/scale–in with rapid deployment, supporting no service interruption, no service suspension, and not read–only.
Read–write separation	Requires implementation at the business code level.	Supports automatic read/write separation, which can be flexibly enabled or disabled based on business requirements.
Resource isolation	Supports CPU and memory isolation.	Supports isolation across multiple dimensions such as CPU, memory, traffic, and number of connections.
Multi–account	Not supported.	Supports multi–account management.
Security Control	Not supported.	Supports a blocklist mechanism for disabling high–risk commands.
Monitoring	Relies on Server collection with limited monitoring metrics.	Provides over 30 cluster and node monitoring metrics, including specialized monitoring for latency, traffic, large keys, and hot keys. Supports alarm reporting to facilitate autonomous Ops and risk prevention.

Use Cases

Last updated: 2026-03-17 16:36:34

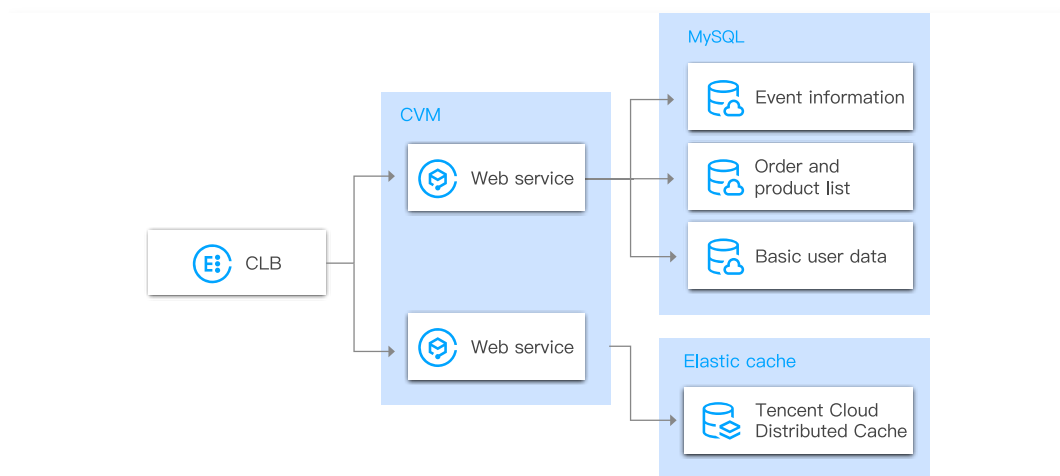
Gaming

In gaming scenarios, non-player data such as leaderboards can be stored in cache databases for rapid access. The SortedSet data type natively supported by the Redis protocol helps sort player data.



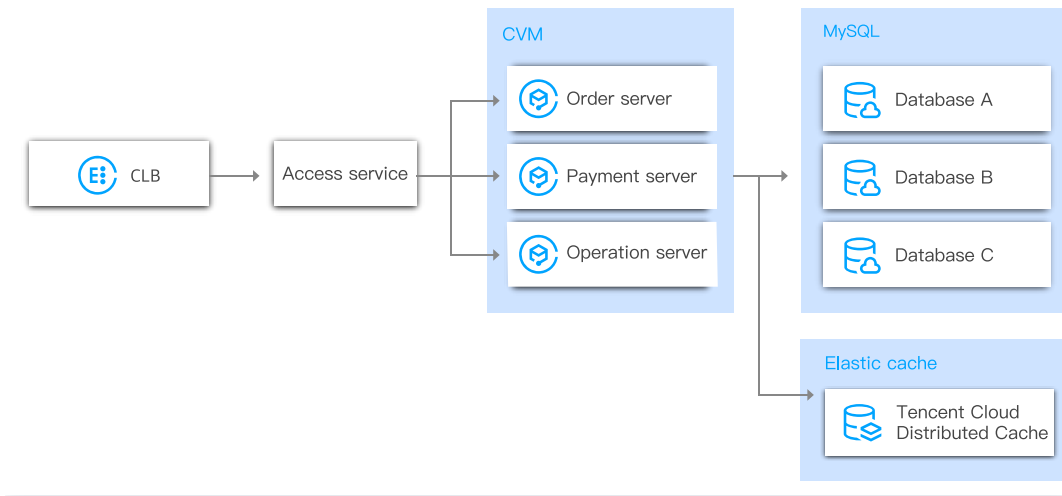
Internet and App

In Internet and App applications, user profile data can be cached in the database to improve read/write performance. Additionally, static images and resources can be cached in the database to enhance application loading speed.



Ecommerce Display

In ecommerce display, data such as product images and recommendations can be stored in the database for quick access. In addition, the distributed caching database with tens of millions of QPS can easily sustain high concurrent requests during large-scale campaigns (for example, flash sales).



Storage Engine

Last updated: 2026-03-17 22:19:41

Tencent Cloud Distributed Cache provides multiple high-performance in-memory storage solutions, including Redis Edition engine and Valkey Edition engine as its two core product lines. These solutions provide users with highly available, high-performance caching services that are fully compatible with the Redis protocol.

Redis Edition

Redis Edition is a high-performance, in-memory database service that is fully compatible with the Redis standard protocol. It supports multiple versions and a flexible architecture, and is designed to help you build a caching solution with millisecond-level response time for your applications.

Core Features

- Protocol compatibility: Fully support the standard Redis protocol; existing applications can be used directly without any code modifications.
- Multi-version support: Fully compatible with Redis 4.0, 5.0, 6.2, 7.0, and other mainstream versions to meet the needs of different businesses at various development stages.
- Data structures: Fully support native data structures of Redis, such as String, Hash, List, Set, and Sorted Set.

Architecture and Performance

- Standard architecture: Adopt the classic primary-secondary replication architecture, supporting 1 to 5 replica nodes with specifications ranging from 0.25 GB to 64 GB. It is an ideal choice for small and medium-sized applications.
- Cluster architecture: Adopt a distributed cluster design, supporting horizontal scaling of 1 – 128 shards with specifications up to 1GB – 16TB. It is specifically designed for massive data and large-scale concurrency scenarios.
- High-performance: The standard architecture delivers a processing capability of up to 100000 QPS, while the cluster architecture supports tens of millions of QPS, effortlessly handling high-concurrency challenges.

Note:

Redis Edition, officially promoted as a caching service proven in large-scale production environments, provides comprehensive features. The specific features available in each version are subject to display in the console.

Valkey Edition

Tencent Cloud Valkey Edition is a high-performance, in-memory database service that is fully compatible with the Redis standard protocol. It offers 100% compatibility with Redis 7.0 and earlier versions, supports a flexible architecture, and is designed to help you build a caching solution with millisecond-level response time for your applications.

Core Features

- Protocol compatibility: Support the Redis standard protocol, ensuring that existing applications can migrate seamlessly at zero cost.
- Version support: Currently support Valkey 8.0, integrating cutting-edge community technologies and performance optimizations.
- Data structures: Fully support native data structures of Redis, such as String, Hash, List, Set, and Sorted Set, ensuring application compatibility.

Architecture and Performance

- Standard architecture: Adopt the classic primary-secondary replication architecture, supporting 1 to 5 replica nodes with specifications ranging from 0.25 GB to 64 GB. It is an ideal choice for small and medium-sized applications.
- Cluster architecture: Adopt a distributed cluster design, supporting horizontal scale-out with 1 to 256 shards and specifications ranging from 2 GB to 8 TB. It is specifically designed for massive data and large-scale concurrency scenarios.
- High-performance: The standard architecture delivers a processing capability of up to 100000 QPS, while the cluster architecture supports tens of millions of QPS, effortlessly handling high-concurrency challenges.

Note:

Valkey Edition is currently available for purchase only through the console. More features will be released soon. Stay tuned for more updates.

Memcached Edition

Memcached Edition is a high-performance, in-memory distributed caching service based on the open-source Memcached protocol. It is specifically designed for large-scale, high-speed caching scenarios.

Core Features

- Protocol compatibility: Fully compatible with the Memcached standard protocol. Existing applications developed based on Memcached can be used directly without any code modifications.
- Version support: Currently support Memcached 1.6, integrating the latest stable features and security patches from the community.

Architecture and Features

Memcached Edition adopts a native distributed cluster architecture, supporting seamless online scale-out and scale-in. The entire process is transparent and imperceptible to your business. The supported features of Memcached Edition are as follows.

Feature	Supported or Not	Description
Connection method	Telnet method supported	<ul style="list-style-type: none"> • Connection to Memcached Edition databases through Database Management Center (DMC) is not supported. • Access to Memcached Edition database instances through redis-cli is not supported.
Version upgrade	Not supported	Currently, only version 1.6 of the Memcached protocol is supported, and version upgrades are not involved.
Node management	Not supported	Memcached Edition does not support viewing instance node information, nor does it support actively changing node properties , including read-only replicas, replica promotion to primary, and failure simulation.
Backup and recovery	Supported	Memcached Edition currently supports automatic backup, but does not allow backup file downloads.. Data can only be recovered by cloning instances. For details, see Backup and Recovery .
Password authentication	Not supported	Only password-free access is supported.
Data migration	Not supported	Migrating instance data of Memcached Edition through Data Transfer Service (DTS) is not supported.
Parameter configuration	Supported	Only the following parameters can be set. For details, see Configurable Parameter List . <ul style="list-style-type: none"> • maxmemory-policy: data eviction policy when the memory cache is full. • proxy-slowlog-log-slower-than: recording threshold for slow logs.
Log shipping	Not supported	Memcached Edition currently does not support shipping slow logs to Cloud Log Service (CLS).
Slow query	Supported	For details, see Querying Memcached Edition Slow Queries .

Global replication	Not supported	Memcached Edition instances do not currently support data synchronization through global replication groups. For details, see Use Limits .
Performance optimization	Not supported	Memcached Edition instances have not been integrated with TencentDB for DBbrain (DBbrain) for performance optimization.
Monitoring and alarms	Supported	<ul style="list-style-type: none">• Memcached Edition provides a fully transparent monitoring service at the instance level and supports configuring monitoring metric alarms. For details, see Monitoring Memcached Edition Instances.• Configuring event alarms is supported, including primary–secondary switch, service unavailability, and server failure. For details, see Creating Event Rules.

Product Series

Proxy Features

Last updated: 2026-03-17 16:36:34

Tencent Cloud Distributed Cache's Proxy server (Proxy) undertakes core responsibilities such as routing and distribution, CLB, and automatic failover, aiming to simplify client-side processing logic. In-depth analysis of the Proxy's routing mechanism and its handling policies for specific commands will help you build more efficient and robust business architectures.

Proxy Overview

The proxy server (Proxy) operates as an independent component and does not consume the computing resources of data shards. By deploying multiple Proxy nodes, the system not only achieves load balancing (CLB) and failover for traffic but also significantly enhances the availability and scalability of distributed cache database instances.

Features	Description
Routing and Forwarding	The Proxy maintains persistent connections with backend data shards and plays a crucial role in evenly distributing requests and ensuring precise routing, guaranteeing efficient request processing and system stability.
Compatibility with Primary-Secundary Cluster Access Mode	When your business growth exceeds the capacity of the standard architecture, you can seamlessly migrate data to the cluster architecture equipped with Proxy. For non-multi-key operations, compatibility is maintained, allowing for smooth architecture transitions, significantly reducing business adjustment costs, and ensuring continuous business growth. For more information, see Upgrade Instance Architecture .
Seamless Shard Changes	Proxy itself does not directly adjust the number or size of data shards but works in coordination with the backend shard management mechanism to enable elastic scaling of data shards. When data shards need to be migrated (such as during scale-out or scale-in), Proxy supports a transparent data migration process, ensuring that client operations are not affected.
Multiple Databases (DB)	The native Redis protocol and its Cluster client typically support only single-database operations, using only database 0 by default and not providing support for the `SELECT` command. By using Proxy to connect to cluster instances, you can access multiple databases and switch between them using the `SELECT` command. In the cluster edition, each instance provides up to 256 databases by default, offering robust support for the multi-database feature.

Read-Only Replica	<p>In a read-write separation architecture, Proxy continuously monitors the status of read-only nodes to implement traffic management measures in specific scenes:</p> <ul style="list-style-type: none"> • When a read-only node exhibits abnormal behavior, Proxy reduces its weight, decreasing the proportion of requests it handles. If a connection attempt fails, Proxy refreshes routing and temporarily stops forwarding traffic to that node, rerouting business requests to other nodes until the issue is resolved and the node is reactivated. • If a read-only node is undergoing full data synchronization, Proxy pauses traffic to that node. This pause continues until synchronization is complete, ensuring the node has the latest and consistent data before handling user requests. Once synchronization is complete, Proxy resumes service to that node, allowing it to process requests normally.
Multi-Account Access Control	<p>Proxy can control the types of commands users can execute based on user-defined permissions. Different accounts are granted the ability to execute authorized commands:</p> <ul style="list-style-type: none"> • Read-only users: Can only execute read commands such as GET, MGET, HGET, SMEMBERS, and so on. Any attempt to execute write commands (such as SET, LPUSH, and SADD) will be rejected by the Redis server. • Read-write users: Users with read-write access can execute all commands, including both read and write operations.

Routing and Forwarding Rules

Forwarding Rules	Description
Basic Forwarding Rules	<ul style="list-style-type: none"> • Cluster Architecture <ul style="list-style-type: none"> ○ For commands that operate on a single key, each key is typically stored on one shard, and Proxy will directly forward the request to the slot where the key resides. ○ For commands involving multiple keys, if the keys are stored on different shards, Proxy will split the command into multiple sub-commands, each handling a single key, and forward each sub-command to the corresponding shard. • Read-Write Separation <ul style="list-style-type: none"> ○ Write requests: Since all data modifications should occur on the primary node, Proxy will directly forward write requests to the primary node (Master). ○ Read requests: Proxy dynamically adjusts the weight of read requests based on the performance and load of the replica nodes, balancing the requests across the replicas. Currently, specific node customization is not supported.

Specific Command Forwarding Rules	Proxy forwards transaction commands (MULTI or EXEC), Lua script commands (EVAL or EVALSHA), SCAN, INFO, and publish/subscribe commands (PUBLISH, SUBSCRIBE, etc.) to the primary node.
-----------------------------------	--

Connection Limit Explanation

Under normal circumstances, Proxy establishes persistent connections with data shards to handle requests. When certain commands are involved, Proxy will create additional connections to the corresponding data shards based on the command's processing requirements. For details on the maximum connection limit for a single data node shard, see [Adjust Connection Quantity](#). To avoid exceeding the connection limit, use the following commands with caution.

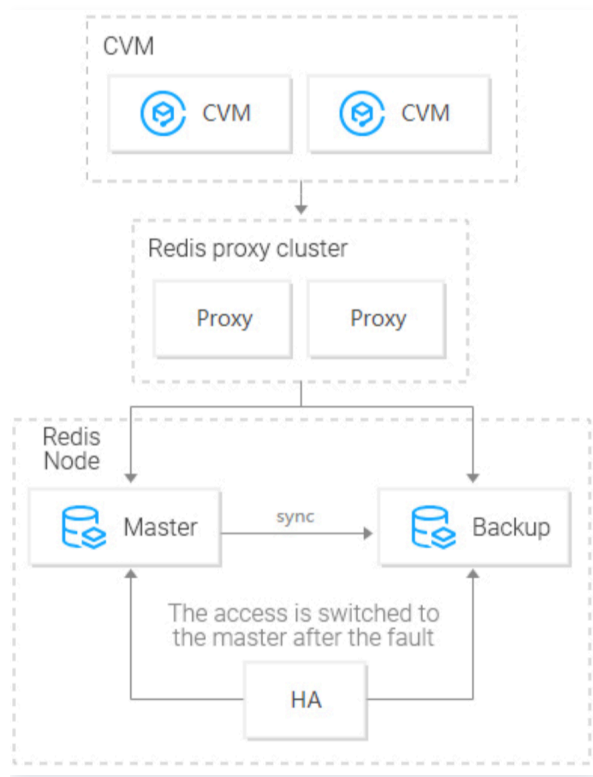
- Blocking commands: BRPOP, BRPOPLPUSH, BLPOP, BZPOPMAX, BZPOPMIN, BLMPOP, and BZMPOP.
- Transaction commands: MULTI, EXEC, and WATCH.
- MONITOR commands: MONITOR and IMONITOR.
- Subscription commands: SUBSCRIBE, UNSUBSCRIBE, PSUBSCRIBE, and PUNSUBSCRIBE.

Standard Architecture

Last updated: 2026-03-17 16:36:34

Architecture Overview

The standard architecture adopts the primary–secondary replication mode, supporting 1 to 5 replica nodes, and is compatible with Redis 4.0, 5.0, 6.2, and 7.0 as well as Valkey 8.0 protocols. The master node provides read/write services, while replica nodes provide high availability assurance. When the master node fails, the system automatically switches to a replica node, ensuring smooth business operation.



Core Components

Component	Description
Master node	The master node is responsible for handling read/write requests and synchronizing data to replica nodes. Each instance has one master node.
Replica node	A replica node is responsible for data redundancy and high availability. It synchronizes data from the master node in real time and automatically takes over as the new master node upon master node failure. The system supports 1 to 5 replicas, where more replicas result in higher availability.
Proxy	A built-in proxy component receives client requests and routes them to the corresponding

nodes, providing load balancing and read/write separation capabilities.

Replica Configuration

A larger number of replicas provides higher availability. All replica nodes participate in high-availability failover. When the number of replicas is greater than or equal to 1, you can enable the read/write separation feature to vertically scale read performance. For more information, see [Read/Write Separation](#).

Version	Supported Replica Quantity	Read/Write Separation
Redis 4.0, 5.0, 6.2, and 7.0	1–5	Supported
Valkey 8.0	1–5	Supported

Specification Limits

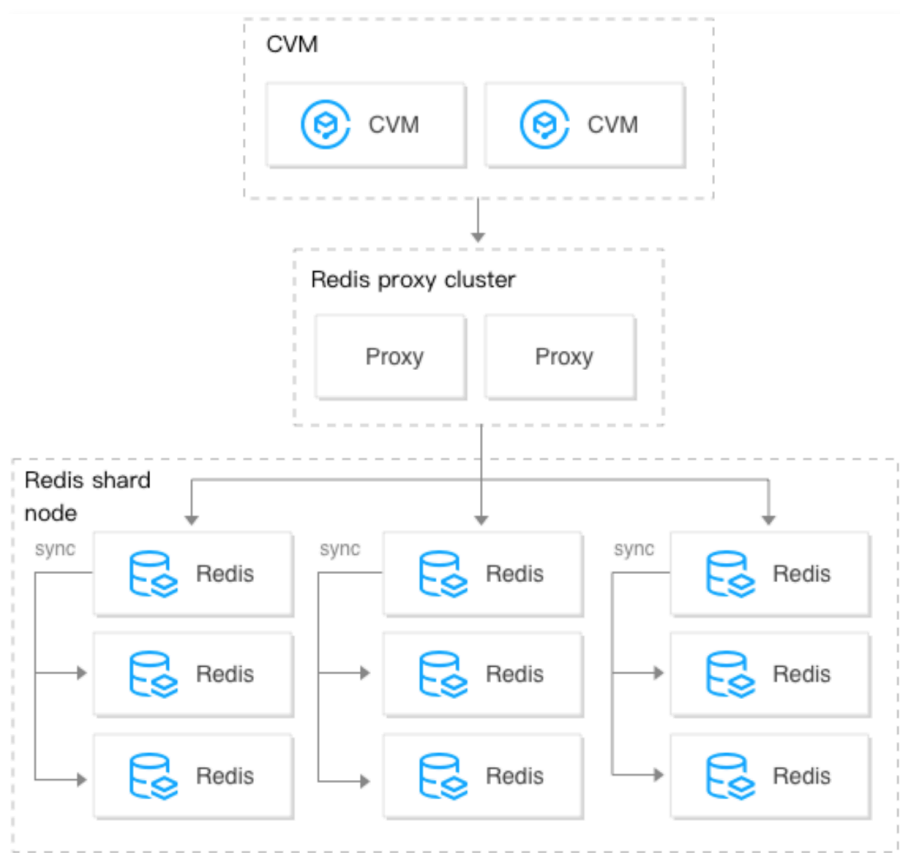
Item	Value	Description
Capacity	0.25–64 GB	Upgrade the architecture to the cluster architecture (up to 8 TB). For detailed instructions, see Upgrading the Instance Architecture .
Performance upper limit	100,000 QPS	Tencent Cloud conducts multiple stress tests in standard testing environments and uses stable values as official references. For details, see Specifications and Performance .

Cluster Architecture

Last updated: 2026-03-17 16:36:34

Architecture Overview

The cluster architecture is a distributed version built based on the Redis Community Edition. It is compatible with Redis 4.0, 5.0, 6.2, and 7.0, as well as Valkey 8.0 protocols. It adopts a distributed sharding design and supports horizontal scaling from 1 to 128 shards and vertical scaling from 1 to 5 replicas, delivering high performance of tens of millions of QPS. Capacity expansion, capacity reduction, and migration processes have minimal impact on businesses, ensuring high service availability. For specific performance data, see [Specifications and Performance](#).



Core Components

Component	Description
Shard	The smallest data storage unit, consisting of a group of master and replica nodes. The cluster evenly distributes 16,384 hash slots across shards. Each shard stores data for its assigned slots and supports online data balancing and migration.

Master node	The node within a shard responsible for handling read/write requests and synchronizing data to replica nodes. Each shard has one master node.
Replica node	A node within a shard responsible for data redundancy and high availability. It synchronizes data from the master node in real time and automatically takes over as the new master node upon master node failure.
Proxy	A built-in proxy component that receives client requests and routes them to the corresponding shards. It provides load balancing, read/write separation, and cross-shard command processing capabilities, shielding cluster topology details from clients.

Shard and Replica Configuration

- A larger number of shards provides higher capacity and performance. All shards participate in data storage and request processing. When the number of shards is greater than or equal to 1, data is automatically distributed based on hash slots, enabling horizontal scaling of storage capacity and throughput.
- A larger number of replicas provides higher availability. All replica nodes participate in high-availability failover. When the number of replicas is greater than or equal to 1, you can enable the read/write separation feature to vertically scale read performance. For more information, see [Read/Write Separation](#).

Note:

Commands that span multiple hash slots have certain compatibility limitations. For details, see [Command Reference Overview](#).

Configuration Item	Optional Specifications
Shard specifications (GB):	1, 2, 4, 8, 12, 16, 20, 24, 32, 40, 48, and 64
Shard quantity	1, 3, 5, 8, 12, 16, 24, 32, 40, 48, 64, 80, 96, and 128
Replica quantity	1, 2, 3, 4, and 5

Upper Limits of Specifications

Item	Upper Limit	Description
Total storage capacity	8TB	128 shards x 64 GB

Concurrent performance	Tens of millions of QPS	Tencent Cloud conducts multiple stress tests in standard testing environments and uses stable values as official references. For details, see Specifications and Performance .
------------------------	-------------------------	--

Product Versions

Major version feature list

Last updated: 2026-04-01 11:06:15

This document details the feature differences among the Redis edition, Valkey edition, and Memcached edition of the storage engine. It aims to help users clearly understand the technical features of each edition, so that they can make informed technical selections based on business requirements.

Feature Attribute	Feature Name	Redis Edition	Valkey Edition	Memcached Edition
Instance management	Create instances.	Supported	Supported	Supported
	Instance list/details			
	Changing instance specifications			
	Unsubscribe instances			
	Batch recovery	Recovery is supported within the retention period.	Recovery is supported within the retention period.	Recovery is supported within the retention period.
	Immediate deactivation of instances	Supported	Supported	Supported
	Version upgrade	Supported	Not supported	Not supported
	Architecture upgrade	Supported	Supported	Not supported
Instance restart	Not supported	Not supported	Not supported	

	Clearing an Instance	Supported	Supported	Supported
	Tag management	Supported	Supported	Supported
Sale	Monthly Subscription	Supported	Supported	Supported
	Pay-as-You-Go			
	Automatic renewal/Batch renewal			
	Switching from Pay-as-You-Go to Monthly Subscription			
	Reminder of expiration and arrears			
Node management	Node list	Supported	Supported	Not supported
	Read-Only Replicas			
	Promotion of replica nodes to primary nodes			
	Multi-AZ			Supported
	Modification of AZ			Supported
Backup and recovery	Manual backup	Supported	Supported	Supported
	Automatic backup			Supported

	Backup file download			Not supported
	Instance cloning			Supported
Account and password	Custom account and permissions	Supported	Supported	Not supported
	Reset Password			
	Authentication-free access			Supported
Network and Security	VPC Network	Supported	Supported	Supported
	Security Group			Supported
	Public Network Access			Not supported
	SSL authentication			Not supported
Monitoring and Ops	Maintenance window	Supported	Supported	Supported
	Modification of Specifications			Supported
	Switching Networks			Supported
	Adjusting Bandwidth			Supported
	Parameter configuration			Supported

Slow query		Supported
Log Delivery CLS		Not supported
DTS Migration		Not supported
Monitoring Metrics		Supported
Monitoring and alarms		Supported
Event alarm		Supported
Exception Diagnosis		Not supported
Performance Trend		
Slow Log Analysis		
Memory Analysis		
Latency Analysis		

Minor Version Update History

Valkey

Last updated: 2026-04-01 18:41:19

This article introduces the updates of the Tencent Cloud Distributed Cache Valkey edition, including new features, performance optimizations, security fixes, and more.

Minor Version	Category	New/Optimizations/Fixes
8.0.8	Security Fixes	Fixed CVE-2025-67733: RESP protocol injection vulnerability via Lua error_reply.
	Security Fixes	Fixed CVE-2026-21863: Remote Denial of Service vulnerability caused by malicious cluster bus messages.
8.0.7	New Features	Supports controlling the display precision of floating-point numbers; after you enable this feature, floating-point return values retain higher precision.
8.0.6	New Features	<p>Based on Valkey community edition 8.0.6, the first Valkey edition is fully compatible with Redis protocol and supports all existing management features of Tencent Cloud.</p> <ul style="list-style-type: none"> Supported the scale-out capability of 256 DBs in the cluster architecture. Optimized the root causes of latency jitter during rehash, ensuring stable performance in high-concurrency business scenarios. Supported the atomic migration mechanism for slots and asynchronous authentication capability for new nodes, and combined these with seamless node migration and the separation of arbitration nodes from data nodes to enhance high availability.

Redis

Last updated: 2026-04-01 11:10:25

This article introduces the updates of the Tencent Cloud Distributed Cache Redis edition, including new features, performance optimizations, security fixes, and more.

Redis 7.0

Minor Version	Category	New/Optimizations/Fixes
7.1.3	Performance optimization	Optimize the timeout handling mechanism of the cluster Slot scan task to improve the operational stability of the cluster in high load scenarios.
		Optimize the cluster state calculation logic when adding arbitration nodes to ensure the accuracy of cluster size information.
7.1.2	Security Fixes	Fixed CVE-2026-21863: Remote Denial of Service vulnerability caused by malicious cluster bus messages.
		Fixed CVE-2025-67733: RESP protocol injection vulnerability via Lua error_reply.
7.1.1	New Features	Support dual-channel replication, using independent data transmission channels between primary/secondary nodes. During full synchronization, it does not affect incremental data synchronization, significantly shortening the full synchronization time.
		Dual-channel replication supports disk logging mode. During full synchronization, replica nodes store incremental data on disk rather than in memory, effectively reducing the memory pressure on replica nodes.
7.1.0	New Features	Supports the client redirection feature. In scenarios such as instance version upgrade or data migration, the system can automatically redirect client connections from the old instance to the new instance, without the need for businesses to modify connection configurations or restart applications, achieving seamless and smooth migration and version upgrade.
7.0.26	New Features	Supports the safe shutdown feature, ensuring the completion of data persistence when an instance is disabled to mitigate the risk of data loss.

		Supports automatically preventing key expiration and memory eviction during data synchronization to ensure the integrity of data migration.
		Supports asynchronous handshake authentication for cluster nodes, enhancing cluster security.
7.0.25	Security Fixes	Fixed CVE-2025-49844: Lua Use-After-Free remote code execution vulnerability.
		Fixed CVE-2025-46817: Integer overflow in Lua library commands that could lead to remote code execution vulnerability.
		Fixed CVE-2025-46818: User authentication vulnerability in Lua functions.
		Fixed CVE-2025-46819: Lua out-of-bounds read vulnerability.
7.0.24	Security Fixes	Fixed CVE-2025-27151: AOF filename length verification vulnerability.
		Fixed CVE-2025-32023: Out-of-bounds write in HyperLogLog commands that could lead to remote code execution vulnerability.
		Fixed CVE-2025-48367: Improper handling of connection errors that could lead to denial of service vulnerability.
7.0.24	Security Fixes	Fixed CVE-2025-21605: Unauthenticated client abuse that could lead to unbounded growth of the output buffer, resulting in a denial of service vulnerability.
	New Features	Added OOM statistics for client input/output buffers to facilitate monitoring of client memory usage.
	Performance optimization	Optimized CLUSTER RESET to execute asynchronously, minimizing service blocking during reset.
7.0.22	Security Fixes	Fixed CVE-2024-46981: Lua use-after-free vulnerability that could lead to remote code execution.
7.0.20	Security Fixes	Fixed CVE-2024-31449: Stack-based buffer overflow in Lua bit.tohex that could lead to remote code execution.
		Fixed CVE-2024-31228: Abuse of string pattern matching that could lead to a denial of service vulnerability.
		Fixed CVE-2024-31227: Malicious selectors in ACL SETUSER that could lead to a denial of service vulnerability.

7.0.19	New Features	Supports FIFO LRU eviction policy for Lua scripts. When the number of script caches reaches the upper limit, it automatically evicts the earliest unused scripts to avoid excessive memory usage.
		Added OOM (out-of-memory) checks when Lua scripts are executed to prevent script execution from causing instance memory exhaustion.
		Supports reserving additional memory for internal buffers to prevent actual memory usage from exceeding the maximum limit.
	Prevent Lua error_reply abuse to avoid unbounded growth of errorstats, ensuring monitoring data accuracy.	
	Performance optimization	Optimize the time consumption for failover after master node downtime, reducing high availability switchover time.
7.0.18	New Features	Supports controlling log output during the rehash process.
7.0.16	Performance optimization	Reserves additional file descriptors for connection management to enhance stability in high-concurrency scenarios.
7.0.15	New Features	Tencent Cloud's first distributed caching cloud database based on Redis community edition and compatible with Redis 7.0.

Redis 6.0

Minor Version	Category	New/Optimizations/Fixes
6.2.16	New Features	Support dual-channel replication, using independent data transmission channels between primary/secondary nodes. During full synchronization, it does not affect incremental data synchronization, significantly shortening the full synchronization time.
		Supports staging replication data to disk during replica synchronization to significantly reduce memory pressure.
		The replication backlog buffer and replicas utilize a globally shared replication buffer, enabling multiple replicas to share the same backlog data, effectively reducing the memory usage of the master node.
6.2.15	Performance optimization	Optimizing the key eviction policy during data synchronization to reduce unnecessary data loss risks.

6.2.14	New Features	Supports the client eviction feature, which automatically evicts client connections that consume excessive memory when client memory usage exceeds the configured threshold, ensuring the availability of instance memory.
		Supports the safe shutdown feature, ensuring the completion of data persistence when an instance is disabled to mitigate the risk of data loss.
		Supports automatically avoiding key expiration and memory eviction during data synchronization to ensure the integrity of data migration.
		Supports asynchronous handshake authentication for cluster nodes, enhancing cluster security.
	Performance optimization	Optimizes the index positioning algorithm for the List data structure, supporting automatic selection to initiate the search from the nearest end based on the target index position, effectively reducing traversal steps and improving access performance for List-type data.
6.2.13	Security Fixes	Fixed CVE-2025-49844: Lua Use-After-Free remote code execution vulnerability.
		Fixed CVE-2025-46817: Integer overflow in Lua library commands that could lead to remote code execution vulnerability.
		Fixed CVE-2025-46818: User authentication vulnerability in Lua functions.
		Fixed CVE-2025-46819: Lua out-of-bounds read vulnerability.
6.2.12	Security Fixes	Fixed CVE-2025-32023: Out-of-bounds write in HyperLogLog commands that could lead to remote code execution vulnerability.
		Fixed CVE-2025-48367: Improper handling of connection errors that could lead to denial of service vulnerability.
6.2.11	Security Fixes	Fixed CVE-2025-21605: Unauthenticated client abuse that could lead to unbounded growth of the output buffer, resulting in a denial of service vulnerability.
		Fixed CVE-2023-25155: Integer overflow vulnerability in RAND commands.
	New Features	Improved the performance of progressive rehash in high-load scenarios.
		Added the <code>lazyfree-lazy-user-del</code> configuration item. After you enable it, the DEL command will asynchronously free memory to avoid

		blocking the service when big keys are deleted.
		Restrict the storage directory for RDB/AOF files to enhance data security.
		Support viewing the number of asynchronously freed objects via the INFO command.
		Enable security control for protected configurations and sensitive commands.
	Improved performance	Optimized the execution performance of the CLUSTER SLOTS command.
6.2.10	New Features	COPY command supports Multi-DB and removes the database limitation in cluster mode.
	Security Fixes	Fixed CVE-2024-46981: Lua use-after-free remote code execution vulnerability.
6.2.9	Security Fixes	Fixed CVE-2024-31449: Stack-based buffer overflow in Lua bit.tohex that could lead to remote code execution.
	Security Fixes	Fixed CVE-2022-36021: Denial-of-service vulnerability caused by crafted patterns in string matching commands exhausting CPU.
	Security Fixes	Fixed CVE-2024-31228: Abuse of string pattern matching that could lead to a denial of service vulnerability.
6.2.8	New Features	Supports FIFO LRU eviction policy for Lua scripts. When the number of script caches reaches the upper limit, it automatically evicts the earliest unused scripts to avoid excessive memory usage.
	New Features	Added OOM (out-of-memory) checks when Lua scripts are executed to prevent script execution from causing instance memory exhaustion.
	New Features	Supports reserving additional memory for internal buffers to prevent actual memory usage from exceeding the maximum limit.
	New Features	Prevent Lua error_reply abuse to avoid unbounded growth of errorstats, ensuring monitoring data accuracy.
	Performance optimization	Optimized memory defragmentation capability to improve defragmentation efficiency.

	Performance optimization	The <code>SCRIPT FLUSH</code> command has been changed to execute asynchronously, preventing blocking of business requests.
	Performance optimization	Optimize the time consumption for failover after master node downtime, reducing high availability switchover time.
6.2.7	New Features	Supports controlling log output during the rehash process.
6.2.5	New Features	Tencent Cloud's first-ever distributed caching cloud database based on Redis community edition and compatible with Redis 6.2.
		Supported the multi-DB feature. Up to 256 DBs were supported.
		Supported performing rehash operations during maintenance.

Redis 5.0

Minor Version	Category	New/Optimizations/Fixes
5.2.17	Performance optimization	Optimizing the key eviction policy during data synchronization to reduce unnecessary data loss risks.
5.2.16	New Features	Supports the client eviction feature, which automatically evicts client connections that consume excessive memory when client memory usage exceeds the configured threshold, ensuring the availability of instance memory.
	New Features	Supports the safe shutdown feature, ensuring the completion of data persistence when an instance is disabled to mitigate the risk of data loss.
	New Features	During data synchronization, automatic avoidance of key expiration and memory eviction is supported to ensure data migration integrity.
	New Features	Supports asynchronous handshake authentication for cluster nodes, enhancing cluster security.
	Performance	Optimizes the index positioning algorithm for the List data structure, supporting automatic selection to initiate the search from the nearest

	optimization	end based on the target index position, effectively reducing traversal steps and improving access performance for List-type data.
	Security Fixes	Fixed CVE-2023-45145: Race condition vulnerability in Unix socket listening permissions.
	Security Fixes	Fixed CVE-2021-32626: Memory write vulnerability caused by Lua script stack overflow.
	Security Fixes	Fixed CVE-2021-21309: Integer overflow vulnerability.
5.2.15	Security Fixes	Fixed CVE-2025-46819: Lua out-of-bounds read vulnerability.
	Security Fixes	Fixed CVE-2025-46818: User authentication vulnerability in Lua functions.
	Security Fixes	Fixed CVE-2025-46817: Integer overflow in Lua library commands that could lead to remote code execution vulnerability.
	Security Fixes	Fixed CVE-2025-49844: Lua Use-After-Free remote code execution vulnerability.
5.2.14	New Features	Supports controlling whether replica nodes respond to read requests when data synchronization is not complete.
		Supports performing BITFIELD operations on read-only replicas.
	Security Fixes	Fixed CVE-2025-48367: Improper handling of connection errors that could lead to denial-of-service vulnerability.
		Fixed CVE-2025-32023: Out-of-bounds write in HyperLogLog commands that could lead to remote code execution vulnerability.
5.2.13	New Features	Added the <code>lazyfree-lazy-user-del</code> configuration item. After you enable it, the DEL command will asynchronously free memory to avoid blocking the service when big keys are deleted.
	New Features	Restrict the storage directory for RDB/AOF files to enhance data security.
	New Features	Support viewing the number of asynchronously freed objects via the INFO command.
	New Features	Enable security control for protected configurations and sensitive commands.

	Performance optimization	Optimized the execution performance of the CLUSTER SLOTS command.
	Security Fixes	Fixed CVE-2021-3470: Vulnerability caused by incorrect assumption in memory allocation function.
	Security Fixes	Fixed CVE-2021-32761: Overflow vulnerability in the BITFIELD command family on 32-bit systems.
	Security Fixes	Fixed CVE-2023-25155: Integer overflow vulnerability in RAND commands.
	Security Fixes	Fixed CVE-2025-21605: Denial-of-service vulnerability caused by unauthenticated clients abusing the system, leading to unbounded growth of output buffers.
5.2.12	Security Fixes	Fixed CVE-2024-46981: Lua use-after-free remote code execution vulnerability.
5.2.11	Security Fixes	Fixed CVE-2024-31449: Stack-based buffer overflow in Lua bit.tohex that could lead to remote code execution.
	Security Fixes	Fixed CVE-2024-31228: Abuse of string pattern matching that could lead to a denial of service vulnerability.
5.2.10	New Features	Supports FIFO LRU eviction policy for Lua scripts. When the number of script caches reaches the upper limit, it automatically evicts the earliest unused scripts to avoid excessive memory usage.
	New Features	Added OOM (out-of-memory) checks when Lua scripts are executed to prevent script execution from causing instance memory exhaustion.
	Performance optimization	Optimized memory defragmentation capability to improve defragmentation efficiency.
	Performance optimization	The <code>SCRIPT FLUSH</code> command has been changed to execute asynchronously, preventing blocking of business requests.
	Performance	Optimize the time consumption for failover after master node downtime, reducing high availability switchover time.

	optimization	
5.2.9	New Features	Supports controlling log output during the rehash process.
5.2.8	Security Fixes	Fixed CVE-2022-36021: Denial-of-service vulnerability caused by crafted patterns in string matching commands exhausting CPU.
	Security Fixes	Fixed CVE-2022-35977: Out-of-memory (OOM) vulnerability caused by integer overflow in SETRANGE/SORT commands.
5.2.7	New Features	Supported performing rehash operations during maintenance.
5.2.0	New Features	Supports the multi-DB feature. A single instance can support up to 256 DBs (databases), facilitating data isolation and management.
	Performance optimization	Handling time-consuming memory statistics operations asynchronously to avoid blocking the main thread, effectively ensuring low-latency responses under high concurrency.

Redis 4.0

Minor Version	Category	New/Optimizations/Fixes
4.3.18	Performance optimization	Optimizing the key eviction policy during data synchronization to reduce unnecessary data loss risks.
4.3.17	New Features	Supports the client eviction feature, which automatically evicts client connections that consume excessive memory when client memory usage exceeds the configured threshold, ensuring the availability of instance memory.
		Supports the safe shutdown feature, ensuring the completion of data persistence when an instance is disabled to mitigate the risk of data loss.
		During data synchronization, automatic avoidance of key expiration and memory eviction is supported to ensure data migration integrity.
		Supports asynchronous handshake authentication for cluster nodes, enhancing cluster security.

	Performance optimization	Optimizes the index positioning algorithm for the List data structure, supporting automatic selection to initiate the search from the nearest end based on the target index position, effectively reducing traversal steps and improving access performance for List-type data.
	Security Fixes	Fixed CVE-2023-45145: Race condition vulnerability in Unix socket listening permissions.
		Fixed CVE-2021-32626: Memory write vulnerability caused by Lua script stack overflow.
		Fixed CVE-2021-21309: Integer overflow vulnerability.
4.3.16	Security Fixes	Fixed CVE-2025-46819: Lua out-of-bounds read vulnerability.
		Fixed CVE-2025-46818: User authentication vulnerability in Lua functions.
		Fixed CVE-2025-46817: Integer overflow in Lua library commands that could lead to remote code execution vulnerability.
		Fixed CVE-2025-49844: Lua Use-After-Free remote code execution vulnerability.
4.3.15	New Features	Added the <code>BITFIELD_RO</code> read-only operation command, which supports executing BITFIELD operations on read-only replicas.
	New Features	Supports controlling whether replica nodes respond to read requests when data synchronization is not complete.
	Security Fixes	Fixed CVE-2025-48367: Improper handling of connection errors that could lead to denial-of-service vulnerability.
	Security Fixes	Fixed CVE-2025-32023: Out-of-bounds write in HyperLogLog commands that could lead to remote code execution vulnerability.
4.3.14	New Features	Added the <code>lazyfree-lazy-user-del</code> configuration item. After you enable it, the DEL command will asynchronously free memory to avoid blocking the service when big keys are deleted.
	New Features	Restrict the storage directory for RDB/AOF files to enhance data security.
	New Features	Support viewing the number of asynchronously freed objects via the INFO command.
	New Features	Enable security control for protected configurations and sensitive commands.

	Performance optimization	Optimized the execution performance of the CLUSTER SLOTS command.
	Security Fixes	Fixed CVE-2021-3470: Vulnerability caused by incorrect assumption in memory allocation function.
	Security Fixes	Fixed CVE-2021-32761: Overflow vulnerability in the BITFIELD command family on 32-bit systems.
	Security Fixes	Fixed CVE-2023-25155: Integer overflow vulnerability in RAND commands.
	Security Fixes	Fixed CVE-2025-21605: Denial-of-service vulnerability caused by unauthenticated clients abusing the system, leading to unbounded growth of output buffers.
4.3.13	Security Fixes	Fixed CVE-2024-46981: Lua use-after-free remote code execution vulnerability.
4.3.12	Security Fixes	Fixed CVE-2024-31449: Stack-based buffer overflow in Lua bit.tohex that could lead to remote code execution.
	Security Fixes	Fixed CVE-2024-31228: Abuse of string pattern matching that could lead to a denial of service vulnerability.
4.3.11	New Features	Supports FIFO LRU eviction policy for Lua scripts. When the number of script caches reaches the upper limit, it automatically evicts the earliest unused scripts to avoid excessive memory usage.
	New Features	Added OOM (out-of-memory) checks when Lua scripts are executed to prevent script execution from causing instance memory exhaustion.
	Performance optimization	Optimized memory defragmentation capability to improve defragmentation efficiency.
	Performance optimization	The <code>SCRIPT FLUSH</code> command has been changed to execute asynchronously, preventing blocking of business requests.
	Performance	Optimize the time consumption for failover after master node downtime, reducing high availability switchover time.

	optimization	
4.3.10	New Features	Supports controlling log output during the rehash process.
4.3.9	Security Fixes	Fixed CVE-2022-36021: A denial-of-service vulnerability where crafted patterns in string matching commands (SCAN/KEYS) could exhaust CPU resources.
	Security Fixes	Fixed CVE-2022-35977: Out-of-memory (OOM) vulnerability caused by integer overflow in SETRANGE/SORT commands.
4.3.8	New Features	Supported performing rehash operations during maintenance.
4.3.0	New Features	Supports the multi-DB feature, with a single instance supporting up to 256 DBs for data isolation and management.
		Changed the default broadcast mode to non-broadcast mode, effectively reducing network communication overhead and resource consumption within the cluster.
	Performance optimization	Optimized the fault discovery mechanism. When a replica node detects an abnormal condition, it can proactively send a notification to the cluster, significantly shortening the discovery and switchover time of faulty nodes.
	Performance optimization	Handling time-consuming memory statistics operations asynchronously to avoid blocking the main thread, effectively ensuring low-latency responses under high concurrency.

Proxy Version Update History

Last updated: 2026-04-01 11:04:33

Tencent Cloud Distributed Cache releases minor versions of Proxy irregularly to provide more database features or fix known issues.

Proxy 5.0

Version	Change Type	Feature Description
5.8.20	New Features	Supports the Valkey storage engine, providing users with more storage engine choices to meet the performance and compatibility requirements of different business scenarios.
5.8.19	Performance optimization	Optimize the security processing mechanism for protocol data transmission to ensure data integrity during transmission and prevent data abnormalities caused by special characters.
5.8.18	Performance optimization	Optimized the connection consistency check mechanism during hot version upgrades to ensure the stability of subscription-based connections during the upgrade process and reduce the impact of hot upgrades on business operations.
5.8.17	New Features	Added session health check capability, supporting automatic detection and cleanup of abnormal session connections to enhance session management efficiency in long-running scenarios.
	Performance optimization	Optimized the monitoring of connection status and abnormality detection capabilities to enhance system security and connection reliability.
5.8.16	Performance optimization	Enhanced access routing policies to improve system security compliance.
		Optimized the load balancing capability for read-only replicas to enhance resource utilization and response efficiency for read requests.
5.8.15	Issue Fixes	Fixed compatibility issues with the seamless migration feature in extreme scenarios.
		Fixed resource management issues in long-running traversal operations to enhance operational stability.

5.8.14	Issue Fixes	Fixed communication link issues for management connections, enforcing management commands to be transmitted through internal secure channels to enhance security compliance.
	Issue Fixes	Fixed a status synchronization issue during seamless migration in broadcast mode, ensuring data consistency across all nodes throughout the migration process.
5.8.13	Performance optimization	Upgraded the asynchronous refresh mechanism for routing tables, significantly reducing routing update latency and enhancing the stability and response speed of proxy services.
	Performance optimization	Enhanced the robustness of the instance hot upgrade process to eliminate potential stability risks during upgrades between specific versions.
5.8.12	Issue Fixes	Resolved the issue of sporadic increased latency caused by resource contention during configuration synchronization, enhancing system operational stability.
5.8.11	Performance optimization	Slow logs now include request trace IDs and command type tags to help users more efficiently locate performance bottlenecks.
	Performance optimization	Comprehensively upgraded the logging framework to enhance the stability of log writing.
5.8.10	Security Enhancement	Enhanced database operation security auditing capabilities by incorporating high-risk operations (such as full database scans and batch deletions) into the sensitive operation monitoring system, with automatic execution record reporting.
5.8.9	Issue Fixes	Enhanced the parameter validation capability for replica node queries to ensure correct request parameter formats, thereby improving system robustness.
5.8.8	New Features	Added multiple data operation commands, expanding data structure manipulation capabilities, including key copy (COPY), list blocking move (BLMOVE/LMOVE), sorted set range storage and difference operations (ZRANGESTORE/ZDIFF/ZUNION/ZDIFFSTORE), and more.

5.8.1	New Features	The shared connection feature is supported to effectively mitigate cluster overload caused by sudden surges of connections (connection storms), enhancing connection stability in high-concurrency scenarios.
5.8.0	New Features	Automatically disables write operations on replicas in global replication scenarios, ensuring cross-region data consistency and stability at the architectural level.
5.7.2	Issue Fixes	Fixed compatibility issues with the RESP3 protocol handshake command to prevent returning exception error messages.
5.7.1	New Features	Compatible with Redis 6.2's new commands, including key reading with expiration time (GETEX), username/password authentication, automatic claiming of message streams, and so on, enhancing protocol compatibility.
5.7.0	New Features	Slow log and big Key analysis logs have added client port information to facilitate precise locating of the client where the issue originates.
	New Features	Redis 6.2 version adds support for multiple new commands (GETDEL, HELLO, and so on).
	New Features	The cluster edition DBSIZE command supports aggregating the total number of keys across all shard nodes while retaining the ability to query individual nodes.
	New Features	Added the multiple AZs local read feature to reduce cross-AZ access latency.
5.6.6	New Features	Supporting seamless migration capabilities, scaling out/in shards is imperceptible to services during transactions, message subscriptions, and blocking operations, delivering a smooth scaling experience.
5.6.5	Issue Fixes	Fixed issues related to the password-free access feature to ensure normal operation in passwordless connection scenarios.
5.6.3	New Features	Redis 6.2 adds support for commonly used commands (GETEX, username/password authentication, automatic message claiming), enhancing version compatibility.
5.6.0	New Features	The cluster architecture supports the WAIT command, which can wait for data synchronization to a specified number of replicas before returning, ensuring reliable data writes.
	New Features	Supports SSL/TLS encrypted transmission to ensure secure data transfer between clients and servers.

	Performance optimization	Optimizes cluster topology query performance and enhances the speed of obtaining node information.
5.5.0	New Features	The cluster architecture supports the WAIT command, ensuring the reliability of data writes.
		Supports the nearby access feature to reduce cross-AZ latency.
		The Cluster Edition supports the DBSIZE command to return the total number of keys across all shards.
		Slow log supports viewing client port information.
		Supports cluster-level batch flush operations, automatically distributed to all shard master nodes for execution.
		Adds monitoring metrics for large Value requests to help users identify potential large Value issues.
		The Cluster Edition supports global traversal capability to traverse data across all shard nodes.
	Issue Fixes	Fixed an issue where switching databases after a transaction operation may result in an exception error.
Fixed an issue where connections locked by transactions in batch request scenarios were not released in a timely manner, causing request routing exceptions.		
5.4.0	Performance optimization	The statistical policy for P99 latency monitoring metrics has been optimized to cover all command types, providing more comprehensive and accurate performance monitoring data.
5.2.0	New Features	Monitoring data supports collection at a 5-second granularity, helping users observe instance performance fluctuations more precisely.
5.1.0	New Features	Slow logs now display the client IP address to facilitate locating the source of slow queries.
	Issue Fixes	Fixed an issue where nested transaction calls returned exception errors.
5.0.0	New Features	The cluster architecture features have been significantly enhanced, improving overall performance and stability in distributed scenarios.

Proxy 4.0

Version	Change Type	Feature Description
3.5.0	New Features	The command analysis feature is now supported, allowing users to view performance metrics such as QPS, P99 execution latency, average execution latency, and maximum execution latency for individual commands, helping users precisely analyze business access patterns and performance bottlenecks.
3.3.0	Performance optimization	The granularity of system monitoring data collection supports 5-second statistics, providing a more detailed performance monitoring view.
3.2.0	New Features	Slow logs now display the client IP address to facilitate locating the source of slow queries.
	Issue Fixes	Fixed an issue where nested transaction calls returned exception errors.

Version Lifecycle Plan

Last updated: 2026-03-17 16:36:35

Lifecycle Stages and Limits of Tencent Cloud Products

The table below lists the limits regarding purchase, usage, and service support in different lifecycle stages of products or features.

Lifecycle Stage	Supported Item	Limit	Suggestion
General Availability (GA)	<ul style="list-style-type: none"> All features of cloud products can be used normally. Renewal is supported. Configuration adjustments, including upgrades and downgrades, are supported. Return, termination, and deactivation are supported. Version upgrades are supported. Security vulnerability fixing is supported. Feature issue fixing is supported. The after-sales service is supported. 	None	None
End of Marketing (EOM)	<ul style="list-style-type: none"> All features of cloud products can be used normally. Renewal is supported. Configuration adjustments, including upgrades and downgrades, are supported. Return, termination, and deactivation are supported. Security vulnerability fixing is supported. 	<ul style="list-style-type: none"> Service purchase is stopped. Version upgrade is stopped. 	Upgrade to a later version. For specific operations, see Upgrading Instance Version .

	<ul style="list-style-type: none"> ● Only feature issue fixing is supported. ● The after-sales service is supported. 		
End of Full Support (EOFS)	<ul style="list-style-type: none"> ● All features can be used normally before EOFS. ● Return, termination, and deactivation are supported. 	<ul style="list-style-type: none"> ● Service purchase is stopped. ● Version upgrade is stopped. ● Renewal is stopped. ● Configuration adjustment is stopped. ● The after-sales service is stopped. ● Issue fixing and vulnerability fixing are stopped. 	<p>Maintenance is about to stop for all instances at this stage. Security vulnerabilities and issues will no longer be fixed, and related after-sales support requests will no longer be accepted. Please upgrade to a later version promptly. For specific operations, see Upgrading Instance Version.</p>
End of Service (EOS)	All services and supports are stopped.	<p>All features are unavailable.</p> <ul style="list-style-type: none"> ● Service purchase is stopped. ● Version upgrade is stopped. ● Renewal is stopped. ● Configuration adjustment is stopped. ● Return, termination, and deactivation are stopped. 	None.

- The after-sales service is stopped.
- Issue fixing and security vulnerability fixing are stopped.
- Product consoles cannot be accessed.
- Instance resources are released.

Lifecycle Plan of Tencent Cloud Distributed Cache

Tencent Cloud Distributed Cache is a cache database with a code branch independently maintained by Tencent Cloud and is 100% compatible with the Redis protocol. It features high performance, high availability, high reliability, high elasticity, and high security. The lifecycle of all its versions is not limited by the lifecycle in the open-source community but is independently maintained and managed by the Tencent Cloud database team. To ensure business stability, please plan your upgrade schedule according to the product lifecycle and upgrade to supported versions promptly.

Version	Lifecycle Stage	Estimated EOM Date	Estimated EOFS Date	Estimated EOS Date	Recommended Version for Upgrade
Redis 2.8	EOM	2021-12-30	2025-06-30	2025-12-30	Redis 5.0 or later versions
Redis 4.0	GA	2025-12-30	2026-06-30	2026-12-30	
Redis 5.0	GA	2026-06-30	2026-12-30	2027-06-30	Redis 6.0 or later versions
Redis 6.2	GA	None			
Redis 7.0	GA				
ValKey 8.0	GA				

Memcached 1.6 version	GA	
-----------------------	----	--

Specifications and Performance

Last updated: 2026-03-17 16:36:35

Specifications

Note:

As a trial version, the 256 MB specification on v4.0 or v5.0 is only suitable for product verification in testing environments. We don't recommend that you use it in production environments. It is available only in the following AZs:

Guangzhou (Zones 6 and 7), Shanghai (Zones 2, 3, 4 and 5), Beijing (Zones 1, 2, 3, 4, 5, 6, and 7), and Shenzhen Finance (Zones 1, 2, and 3). Other 1 GB and above specifications can be smoothly downgraded to the 256 MB specification.

Feature	Standard Architecture	Cluster Architecture
Memory specifications	256MB – 64GB	Single shard: 1 GB to 64 GB
Number of replicas	1 – 5	1 – 5
Number of shards	N/A	1, 3, 5, 8, 12, 16, 24, 32, 40, 48, 64, 80, 96, 128, and 256
QPS	80,000–100,000	80000 to 100000 per shard
Maximum number of connections	10000 by default, adjustable up to 40000	10000 per shard by default, adjustable up to 40000
Traffic limit	528MB/s – 608MB/s	288MB/s – 72GB/s
Multiple databases	Supported	Supported
Mget and Mset	Supported	Supported
lua	Supported	Supported (cross-slot access not supported)

Horizontal scale-out	Not supported	Supported
Replica scale-out	Supported	Supported
Read-write separation	Supported	Supported
Support for GEO	Supported	Supported

Description of the CPU core quantity

To ensure the stable operations of the system, each node is assigned with 2-core CPU, one of which is used to process the backend tasks.

Number of Connections and Traffic

Distributed Cache offers tiered options with customizable core performance metrics including storage capacity, connections, QPS, and bandwidth.

Content Capacity (GB)	Maximum number of connections	Maximum throughput (MB/s)
0.25	3,000	10
1	40,000	16
2	40,000	24
4	40,000	24
8	40,000	24
12	40,000	32
16	40,000	32
20	40,000	48
24	40,000	48
32	40,000	48
40	40,000	64
48	40,000	64

64	40,000	64
----	--------	----

The number of connections in the Cluster Edition = Shard connections * Number of shards.

For the throughput of the Cluster Edition, see [Bandwidth Adjustment](#).

Note:

After scaling, legacy instances capable of up to 9,000 connections will be capable of 10,000 connections.

Performance Data

Performance references

As the execution time of commands varies, and database operation commands differ across business scenarios in production environments, the corresponding performance will also vary. The test data provided in this document is a reference value obtained under specified parameters and shall only serve as a reference. Actual performance needs to be verified through tests based on actual business scenarios.

Single-node test performance

Instance Specifications	Connections	QPS
standard architecture 8GB	10,000	80,000–100,000
cluster architecture 8GB (single shard)	10,000	80,000–100,000

Cluster architecture test performance

Memory Edition (Cluster Architecture) performance = Memory Edition (Standard Architecture) performance * *number of shards*

Test method

Test environment

Test Client CVM Instances	CVM Cores	CVM Memory	Region	Instance Size
3	2	8 GB	Guangzhou Zone 2	Memory Edition (Standard Architecture) 8 GB

Test parameters

```
redis-benchmark -h 10.66.187.x -p 6379 -a crs-1znib6aw:chen2016 -t set -  
c 3500 -d 128 -n 25000000 -r 5000000  
redis-benchmark -h 10.66.187.x -p 6379 -a crs-1z5536aw:chen2016 -t set -  
c 3500 -d 128 -n 25000000 -r 5000000  
redis-benchmark -h 10.66.187.x -p 6379 -a crs-090rjlih:1234567 -t set -c  
3500 -d 128 -n 25000000 -r 5000000
```

QPS calculation

Sum of the QPS values of three pressure test clients (tested by redis-benchmark).

Read/Write Separation

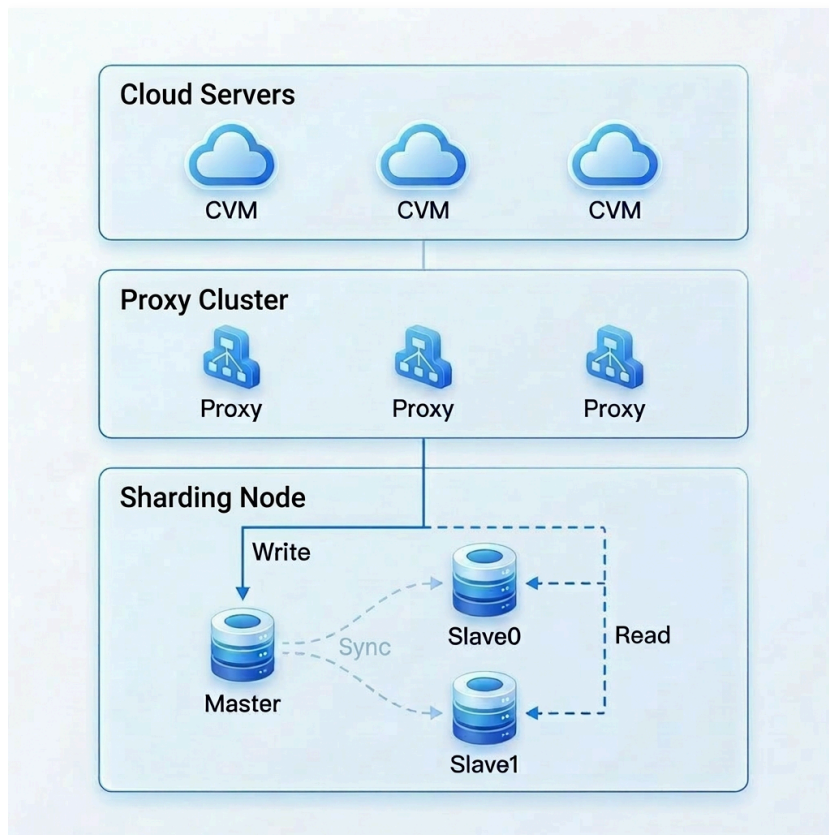
Last updated: 2026-03-17 16:36:35

Overview

Read/write separation is a performance scaling solution for read-intensive, write-light workloads. By distributing read requests to replica nodes, it alleviates pressure from concentrated reads on hotspot data. It supports up to one master node and five replica nodes, delivering up to 5x read performance scaling capability.

Architecture Principles

Read/write separation is implemented through a built-in proxy component for automatic routing and distribution of requests. Client requests are uniformly sent to the proxy, which automatically identifies the request type, routes write requests to the master node for processing, and distributes read requests to replica nodes according to the load balancing policy. After read/write separation is enabled, the master node only handles write requests, while read requests are handled by replica nodes, thus achieving linear scaling of read performance.



Supported Versions

Read/write separation is compatible with Redis 4.0, 5.0, 6.2, and 7.0 as well as Valkey 8.0. It applies to both standard and cluster architectures, and can be enabled when the number of replicas is greater than or equal to 1.

Enabling Methods

- Console: On the Node Management page, you can enable the read/write separation feature. For specific operations, see [Read-Only Replica](#).



- TencentCloud API: You can enable read/write separation through the API [EnableReplicaReadonly](#) and disable read/write separation through the API [DisableReplicaReadonly](#).

Multi-AZ Deployment

Last updated: 2026-03-17 21:34:11

Region

Overview

A region is the physical location of an IDC. In Tencent Cloud, regions are fully isolated from each other, ensuring cross-region stability and fault tolerance. When purchasing Tencent Cloud services, we recommend selecting the region closest to your end users to minimize access latency and improve download speed.

Characteristics

- The networks of different regions are fully isolated from each other, and Tencent Cloud services in different regions cannot communicate using private networks by default.
- Tencent Cloud services in different VPCs can communicate with each other through [Cloud Connect Network](#) that is faster and steadier.

Note:

Shenzhen/Shanghai Finance Zones are tailor-made AZs for compliance with regulatory requirements in the finance industry and feature high security and isolation. Currently, they are available for CVM, finance database, Redis storage, and face recognition services. They can be activated by verified financial customers by [contacting customer service](#).

AZ

Overview

An availability zone (AZ) is a physical IDC of Tencent Cloud with independent power supply and network in the same region. It can ensure business stability, as failures (except for major disasters or power failures) in one AZ are isolated without affecting other AZs in the same region. By starting an instance in an independent AZ, users can protect their applications from being affected by a single point of failure.

Characteristics

Tencent Cloud services in the same VPC are interconnected via the private network, which means they can communicate using [private IPs](#), even if they are in different AZs of the same region.

Note:

Private network interconnection refers to the interconnection of resources under the same account. Resources under different accounts are completely isolated on the private network.

List of Regions and AZs

Note:

- Resources available in different regions and AZs may be sold out and become unavailable, and previously sold-out resources may be replenished. The resource availability will be assessed and adjusted based on the actual business usage as displayed on the purchase page in the console.
- The Memcached Edition is currently available in the following regions: Qingyuan, Chengdu, Beijing, Shanghai, Guangzhou, Shenzhen Finance, Shanghai Finance, Singapore, and Hong Kong (China).

China

Region	AZ	ZoneId
South China (Guangzhou) ap-guangzhou	Guangzhou Zone 1 ap-guangzhou-1	100001
	Guangzhou Zone 2 ap-guangzhou-2	100002
	Guangzhou Zone 3 ap-guangzhou-3	100003
	Guangzhou Zone 4 ap-guangzhou-4	100004
	Guangzhou Zone 6 ap-guangzhou-6	100006
	Guangzhou Zone 7 ap-guangzhou-7	100007
South China (Shenzhen Finance) ap-shenzhen-fsi	Shenzhen Finance Zone 1 (only financial institutions and enterprises can contact us to apply for activation) ap-shenzhen-fsi-1	110001
	Shenzhen Finance Zone 2 (only financial institutions and enterprises can contact us to apply for activation) ap-shenzhen-fsi-2	110002
	Shenzhen Finance Zone 3 (only financial institutions and enterprises can contact us to apply for activation) ap-shenzhen-fsi-3	110003

East China (Shanghai) ap-shanghai	Shanghai Zone 2 ap-shanghai-2	200002
	Shanghai Zone 3 ap-shanghai-3	200003
	Shanghai Zone 4 ap-shanghai-4	200004
	Shanghai Zone 5 ap-shanghai-5	200005
	Shanghai Zone 7 ap-shanghai-7	200007
	Shanghai Zone 8 ap-shanghai-8	200008
East China (Shanghai Finance) ap-shanghai-fsi	Shanghai Finance Zone 1 (only financial institutions and enterprises can contact us to apply for activation) ap-shanghai-fsi-1	700001
	Shanghai Finance Zone 2 (only financial institutions and enterprises can contact us to apply for activation) ap-shanghai-fsi-2	700002
	Shanghai Finance Zone 3 (only financial institutions and enterprises can contact us to apply for activation) ap-shanghai-fsi-3	700003
East China (Nanjing) ap-nanjing	Nanjing Zone 1 ap-nanjing-1	330001
	Nanjing Zone 2 ap-nanjing-2	330002
	Nanjing Zone 3 ap-nanjing-3	330003
North China (Beijing) ap-beijing	Beijing Zone 1 ap-beijing-1	800001
	Beijing Zone 3 ap-beijing-3	800003
	Beijing Zone 4 ap-beijing-4	800004

	Beijing Zone 5 ap-beijing-5	800005
	Beijing Zone 6 ap-beijing-6	800006
	Beijing Zone 7 ap-beijing-7	800007
	Beijing Zone 8 ap-beijing-8	800008
North China (Beijing Finance) ap-beijing-fsi	Beijing Finance Zone 1 (only financial institutions and enterprises can contact us to apply for activation) ap-beijing-fsi-1	460001
Southwest (Chengdu) ap-chengdu	Chengdu Zone 1 ap-chengdu-1	160001
	Chengdu Zone 2 ap-chengdu-2	160002
Southwest (Chongqing) ap-chongqing	Chongqing Zone 1 ap-chongqing-1	190001
Hong Kong/Macao/Taiwan (China Region) (Hong Kong, China) ap-hongkong	Hong Kong Zone 1 (Hong Kong nodes cover services in the China regions of Hong Kong, Macao, and Taiwan) ap-hongkong-1	300001
	Hong Kong Zone 2 (Hong Kong nodes cover services in the China regions of Hong Kong, Macao, and Taiwan) ap-hongkong-2	300002
	Hong Kong Zone 3 (Hong Kong nodes cover services in the China regions of Hong Kong, Macao, and Taiwan) ap-hongkong-3	300003

Other countries and regions

Region	AZ	Zoneld
Southeast Asia Pacific (Singapore) ap-singapore	Singapore Zone 1 (Singapore nodes cover services in Southeast Asia) ap-singapore-1	900001

	Singapore Zone 2 (Singapore nodes cover services in Southeast Asia) ap-singapore-2	900002
	Singapore Zone 3 (Singapore nodes cover services in Southeast Asia) ap-singapore-3	900003
	Singapore Zone 4 (Singapore nodes cover services in Southeast Asia) ap-singapore-4	900004
Southeast Asia (Jakarta) ap-jakarta	Jakarta Zone 1 (Jakarta nodes cover services in Southeast Asia) ap-jakarta-1	720001
	Jakarta Zone 2 (Jakarta nodes cover services in Southeast Asia) ap-jakarta-2	720002
Southeast Asia Pacific (Bangkok) ap-bangkok	Bangkok Zone 1 (Bangkok nodes cover services in Southeast Asia) ap-bangkok-1	230001
	Bangkok Zone 2 (Bangkok nodes cover services in Southeast Asia) ap-bangkok-2	230002
Northeast Asia (Seoul) ap-seoul	Seoul Zone 1 (Seoul nodes cover services in Northeast Asia) ap-seoul-1	180001
	Seoul Zone 2 (Seoul nodes cover services in Northeast Asia) ap-seoul-2	180002
Northeast Asia (Tokyo) ap-tokyo	Tokyo Zone 1 (Tokyo nodes cover services in Northeast Asia) ap-tokyo-1	250001
	Tokyo Zone 2 (Tokyo nodes cover services in Northeast Asia) ap-tokyo-2	250002
West US (Silicon Valley) na-siliconvalley	Silicon Valley Zone 1 (Silicon Valley nodes cover services in West US) na-siliconvalley-1	150001

	Silicon Valley Zone 2 (Silicon Valley nodes cover services in West US) na-siliconvalley-2	150002
East US (Virginia) na-ashburn	Virginia Zone 1 (Virginia nodes cover services in East US) na-ashburn-1	220001
	Virginia Zone 2 (Virginia nodes cover services in East US) na-ashburn-2	220002
Europe (Frankfurt) eu-frankfurt	Frankfurt Zone 1 (Frankfurt nodes cover services in Europe) eu-frankfurt-1	170001
	Frankfurt Zone 2 (Frankfurt nodes cover services in Europe) eu-frankfurt-2	170002
South America (São Paulo) sa-saopaulo	São Paulo Zone 1 (São Paulo nodes cover services in South America) sa-saopaulo-1	740001
Middle East(Riyadh) me-saudi-arabia	Saudi Arabia Zone 1 me-saudi-arabia-1	101001
	Saudi Arabia Zone 2 me-saudi-arabia-2	101002

Selection of Regions and AZs

When you purchase Tencent Cloud services, we recommend that you select the region closest to your end users to minimize access latency and improve download speed.

Regions and AZs

Last updated: 2026-03-17 22:17:44

Region

Overview

A region is the physical location of an IDC. In Tencent Cloud, regions are fully isolated from each other, ensuring cross-region stability and fault tolerance. When purchasing Tencent Cloud services, we recommend selecting the region closest to your end users to minimize access latency and improve download speed.

Characteristics

- The networks of different regions are fully isolated from each other, and Tencent Cloud services in different regions cannot communicate using private networks by default.
- Tencent Cloud services in different VPCs can communicate with each other through [Cloud Connect Network](#) that is faster and steadier.

Note:

Shenzhen/Shanghai Finance Zones are tailor-made AZs for compliance with regulatory requirements in the finance industry and feature high security and isolation. Currently, they are available for CVM, finance database, distributed cache, and face recognition services. They can be activated by verified financial customers by [contacting customer service](#).

AZ

Overview

An availability zone (AZ) is a physical IDC of Tencent Cloud with independent power supply and network in the same region. It can ensure business stability, as failures (except for major disasters or power failures) in one AZ are isolated without affecting other AZs in the same region. By starting an instance in an independent AZ, users can protect their applications from being affected by a single point of failure.

Characteristics

Tencent Cloud services in the same VPC are interconnected via the private network, which means they can communicate using [private IPs](#), even if they are in different AZs of the same region.

Note:

Private network interconnection refers to the interconnection of resources under the same account. Resources under different accounts are completely isolated on the private network.

List of Regions and AZs

Note:

- Resources available in different regions and AZs may be sold out and become unavailable, and previously sold-out resources may be replenished. The resource availability will be assessed and adjusted based on the actual business usage as displayed on the purchase page in the console.
- The Memcached Edition is currently available in the following regions: Qingyuan, Chengdu, Beijing, Shanghai, Guangzhou, Shenzhen Finance, Shanghai Finance, Singapore, and Hong Kong (China).

China

Region	AZ	ZoneId
South China (Guangzhou) ap-guangzhou	Guangzhou Zone 1 ap-guangzhou-1	100001
	Guangzhou Zone 2 ap-guangzhou-2	100002
	Guangzhou Zone 3 ap-guangzhou-3	100003
	Guangzhou Zone 4 ap-guangzhou-4	100004
	Guangzhou Zone 6 ap-guangzhou-6	100006
	Guangzhou Zone 7 ap-guangzhou-7	100007
South China (Shenzhen Finance) ap-shenzhen-fsi	Shenzhen Finance Zone 1 (only financial institutions and enterprises can contact us to apply for activation) ap-shenzhen-fsi-1	110001
	Shenzhen Finance Zone 2 (only financial institutions and enterprises can contact us to apply for activation) ap-shenzhen-fsi-2	110002
	Shenzhen Finance Zone 3 (only financial institutions and enterprises can contact us to apply for activation) ap-shenzhen-fsi-3	110003

East China (Shanghai) ap-shanghai	Shanghai Zone 2 ap-shanghai-2	200002
	Shanghai Zone 3 ap-shanghai-3	200003
	Shanghai Zone 4 ap-shanghai-4	200004
	Shanghai Zone 5 ap-shanghai-5	200005
	Shanghai Zone 7 ap-shanghai-7	200007
	Shanghai Zone 8 ap-shanghai-8	200008
East China (Shanghai Finance) ap-shanghai-fsi	Shanghai Finance Zone 1 (only financial institutions and enterprises can contact us to apply for activation) ap-shanghai-fsi-1	700001
	Shanghai Finance Zone 2 (only financial institutions and enterprises can contact us to apply for activation) ap-shanghai-fsi-2	700002
	Shanghai Finance Zone 3 (only financial institutions and enterprises can contact us to apply for activation) ap-shanghai-fsi-3	700003
East China (Nanjing) ap-nanjing	Nanjing Zone 1 ap-nanjing-1	330001
	Nanjing Zone 2 ap-nanjing-2	330002
	Nanjing Zone 3 ap-nanjing-3	330003
North China (Beijing) ap-beijing	Beijing Zone 1 ap-beijing-1	800001
	Beijing Zone 3 ap-beijing-3	800003
	Beijing Zone 4 ap-beijing-4	800004

	Beijing Zone 5 ap-beijing-5	800005
	Beijing Zone 6 ap-beijing-6	800006
	Beijing Zone 7 ap-beijing-7	800007
	Beijing Zone 8 ap-beijing-8	800008
North China (Beijing Finance) ap-beijing-fsi	Beijing Finance Zone 1 (only financial institutions and enterprises can contact us to apply for activation) ap-beijing-fsi-1	460001
Southwest (Chengdu) ap-chengdu	Chengdu Zone 1 ap-chengdu-1	160001
	Chengdu Zone 2 ap-chengdu-2	160002
Southwest (Chongqing) ap-chongqing	Chongqing Zone 1 ap-chongqing-1	190001
Hong Kong/Macao/Taiwan (China Region) (Hong Kong, China) ap-hongkong	Hong Kong Zone 1 (Hong Kong nodes cover services in the China regions of Hong Kong, Macao, and Taiwan) ap-hongkong-1	300001
	Hong Kong Zone 2 (Hong Kong nodes cover services in the China regions of Hong Kong, Macao, and Taiwan) ap-hongkong-2	300002
	Hong Kong Zone 3 (Hong Kong nodes cover services in the China regions of Hong Kong, Macao, and Taiwan) ap-hongkong-3	300003

Other countries and regions

Region	AZ	Zoneld
Southeast Asia Pacific (Singapore) ap-singapore	Singapore Zone 1 (Singapore nodes cover services in Southeast Asia) ap-singapore-1	900001

	Singapore Zone 2 (Singapore nodes cover services in Southeast Asia) ap-singapore-2	900002
	Singapore Zone 3 (Singapore nodes cover services in Southeast Asia) ap-singapore-3	900003
	Singapore Zone 4 (Singapore nodes cover services in Southeast Asia) ap-singapore-4	900004
Southeast Asia (Jakarta) ap-jakarta	Jakarta Zone 1 (Jakarta nodes cover services in Southeast Asia) ap-jakarta-1	720001
	Jakarta Zone 2 (Jakarta nodes cover services in Southeast Asia) ap-jakarta-2	720002
Southeast Asia Pacific (Bangkok) ap-bangkok	Bangkok Zone 1 (Bangkok nodes cover services in Southeast Asia) ap-bangkok-1	230001
	Bangkok Zone 2 (Bangkok nodes cover services in Southeast Asia) ap-bangkok-2	230002
Northeast Asia (Seoul) ap-seoul	Seoul Zone 1 (Seoul nodes cover services in Northeast Asia) ap-seoul-1	180001
	Seoul Zone 2 (Seoul nodes cover services in Northeast Asia) ap-seoul-2	180002
Northeast Asia (Tokyo) ap-tokyo	Tokyo Zone 1 (Tokyo nodes cover services in Northeast Asia) ap-tokyo-1	250001
	Tokyo Zone 2 (Tokyo nodes cover services in Northeast Asia) ap-tokyo-2	250002
West US (Silicon Valley) na-siliconvalley	Silicon Valley Zone 1 (Silicon Valley nodes cover services in West US) na-siliconvalley-1	150001

	Silicon Valley Zone 2 (Silicon Valley nodes cover services in West US) na-siliconvalley-2	150002
East US (Virginia) na-ashburn	Virginia Zone 1 (Virginia nodes cover services in East US) na-ashburn-1	220001
	Virginia Zone 2 (Virginia nodes cover services in East US) na-ashburn-2	220002
Europe (Frankfurt) eu-frankfurt	Frankfurt Zone 1 (Frankfurt nodes cover services in Europe) eu-frankfurt-1	170001
	Frankfurt Zone 2 (Frankfurt nodes cover services in Europe) eu-frankfurt-2	170002
South America (São Paulo) sa-saopaulo	São Paulo Zone 1 (São Paulo nodes cover services in South America) sa-saopaulo-1	740001
Middle East(Riyadh) me-saudi-arabia	Saudi Arabia Zone 1 me-saudi-arabia-1	101001
	Saudi Arabia Zone 2 me-saudi-arabia-2	101002

Selection of Regions and AZs

When you purchase Tencent Cloud services, we recommend that you select the region closest to your end users to minimize access latency and improve download speed.

Terms

Last updated: 2026-03-17 16:36:35

This article introduces the relevant terms and concepts in the Distributed Cache help documentation to help you better understand the product features.

Instance

An instance is an independently running database environment in Tencent Cloud, serving as the basic unit for users to purchase Distributed Cache, existing as a separate process. A database instance can contain multiple databases created by users. You can create, modify, and delete instances in the console. Instances are independent of each other with resource isolation, and there is no contention for resources such as CPU, memory, or IO between them.

Region

Region (Region) refers to the geographic location of the servers where users purchase Tencent Cloud Distributed Cache instances, and each region is completely independent. When purchasing a Distributed Cache instance, you need to specify the region closest to you to reduce access latency, and changing the region is not supported after purchase. Additionally, Distributed Cache must be used with CVM, and you need to ensure that the cache database and CVM are in the same region.

AZ

An availability zone refers to a physical data center of Tencent Cloud that is independent of other data centers in the same region in terms of power supply and network. Availability zones are interconnected via private network. Network delay is lower within an availability zone. Failures within availability zones are ensured to be isolated from each other (except for major disasters or power failures), so as to prevent fault propagation and ensure that users' businesses continue to provide online services.

Proxy

Proxy server (Proxy) is an independent component in the Tencent Cloud Distributed Cache architecture and does not consume the resources of data shards. By deploying multiple proxy nodes, it can achieve CLB for traffic and failover of the system, thereby enhancing the availability and scalability of the Tencent Cloud Distributed Cache instance.

Replica (Replication)

The replica of Distributed Cache is a data redundancy and high availability solution implemented through the master-replica architecture (Master-Replica). A master node (Master) handles all write operations and asynchronously replicates data to one or more replica nodes (Replica), ensuring that multiple copies of data

are stored. When the master node fails, the system automatically triggers failover within seconds, promoting a replica node to the new master node to ensure service continuity.

Shard

The shard of Distributed Cache is a horizontal scaling solution implemented through a distributed cluster (Cluster). In the Tencent Cloud Memory Edition (cluster architecture), a shard is the smallest data unit composed of a group of primary/secondary nodes. The cluster defaults to distributing 16384 hash slots evenly across all shards; after a client Key is calculated by an algorithm, the shard where the corresponding slot resides is responsible for storage and response. By horizontally scaling the number of shards (1–128 shards), slot rebalancing and data migration can be completed online, achieving linear scaling of capacity and QPS.

Tencent Cloud Console

The Tencent Cloud console is the unified web operation panel for Tencent Cloud resources. By logging in to the console, you can visually manage and monitor the full lifecycle of database instances.

Basic Network

A basic network is a public network space used by multiple users. It cannot be segmented. IP addresses are unique, randomly assigned, and cannot be modified.

Virtual Private Cloud

VPC (Virtual Private Cloud, VPC) is a logically isolated network space customized by you on Tencent Cloud. You can build a logically isolated, user-configured network space for resources such as CVM and cloud database to enhance the security of your cloud resources and meet different use case requirements.

Security Group

A security group is a virtual firewall that features stateful data packet filtering. It is used to configure the network access control of instances for services such as CVM, CLB, and cloud databases while controlling their outbound and inbound traffic. It is an important means of network security isolation.

Number of Connections

The number of TCP session channels established between the client and the Distributed Cache instance directly reflects the concurrent access pressure. Its upper limit is strictly constrained by the instance's memory specification, and connection pool reuse can significantly improve resource utilization.

Slow query

Slow query refers to a query request that exceeds the specified time range, and the corresponding statement is called a slow query statement. Distributed Cache provides slow query statistics for both data cache nodes

and proxy nodes. The information includes the command of the slow query, its occurrence time, request duration, and specific details.

Tag

Tags are used to categorize resources and aggregate them by classification. If your account has multiple cloud database resources with various interconnections, you can use tags to group related or functionally similar resources. During daily Ops or troubleshooting, this enables quick resource search, batch operations, and efficient issue resolution.

Project

A project is the core logical unit of cloud resource management. By aggregating cloud products (such as databases, storage, and computing) into the same space, it enables resource grouping and isolation, unified permission management, and cost allocation accounting, providing infrastructure for team collaboration.

CVM

Tencent Cloud CVM (Cloud Virtual Machine, CVM) is a scalable computing service provided by Tencent Cloud. Only by connecting to the private network address automatically assigned to Distributed Cache via CVM can you access the database.

CVM

Cloud Access Management (CAM) is a user and permission management system provided by Tencent Cloud, which is designed to help customers securely and finely manage access to Tencent Cloud products and resources.

DTS

Data Transfer Service (DTS) is a comprehensive database data transfer service that integrates data migration, synchronization, and subscription. It enables users to complete database migration seamlessly without business interruption.

DBbrain

Distributed Cache instances are integrated with DBbrain's (TencentDB for DBbrain, DBbrain) diagnosis and optimization feature, which monitors and diagnoses database instance exceptions in real time. It automatically generates health reports and provides expert optimization recommendations, helping you comprehensively grasp database performance, quickly troubleshoot issues, and optimize database efficiency promptly.

Service Regions and Service Providers

Last updated: 2026-03-25 18:59:13

Service Regions

For this Feature, the available Service Regions are as follows:

- Sao Paulo
- Virginia
- Tokyo
- Hong Kong SAR
- Riyadh
- Frankfurt
- Silicon Valley
- Bangkok
- Jakarta
- Singapore

In addition to the above, this Feature is also available in any other service regions as shown on the product configuration page.

For further details on the entities that provide the data center and infrastructure for Service Regions and on which the Tencent Cloud services run, please refer to Tencent Cloud's general [Sub-processors and Other Third Parties](#) page.

Sub-Processors and service providers that provide services to and process Content for this Feature

The entities listed below are Sub-Processors that we engage to process your Content for this Feature.

Entity / Party / Recipient Category	Purpose(s) of Entity's Data Processing Activity	Processing Location
–	–	–

For Sub-Processors that we engage to process your Content (as defined in the [Data Processing and Security Agreement](#)) generally across the board for Tencent Cloud Features and/or services, please refer to the page for more information.

Service Providers or Partners that provide processing activities to Administrative Information for this Feature

The entities listed below are the service providers or partners that we engage to process your Administrative Information in connection with this Feature.

Entity / Party / Recipient Category	Purpose(s) of Entity's Data Processing Activity	Processing Location
-	-	-

For service providers or partners that we engage to process your Administrative Information (as defined in the [Data Processing and Security Agreement](#)) more generally (not specific to this Feature), please refer to the [Sub-processors and Other Third Parties](#) page for more information.