Global Game Servers

Contents

1. Industry Overview

The game industry is evolving each day at a magnificent speed. Along with which, the international consumer base is growing manifold. Hence, to enhance players' experience becomes the prime objective of all game industries. A player with a high latency internet connection may show slow responses, and a disconnection in a flash can render all players offline. Therefore, reliable servers, real-time data analysis and low-latency are essential for high-quality contents and outstanding player experience, which makes gaming a natural fit to run on cloud. Moreover, gaming companies no longer need to estimate the game servers they will need or even to make additional purchases in advance, for cloud computing offers flexible and scalable infrastructure following the online player number.

Mobile gaming is now the largest segment. With mobile technologies developed enough to handle medium-to-heavy workload games, many console gaming companies have entered the mobile segment. The trends of mobile gaming toward innovation, specification, heavy workload, VR/AR, return of classic copyright, going international with global servers make cloud computing the best choice for mobile gaming companies to obtain reliable backend and suitable architecture supports for their business.

In 2016, Supercell launched a new game - Clash Royale, which laid the benchmark of the global server architecture in the gaming industry. In China, Clash of Kings, a game developed by Elex Technology, also achieved remarkable success. Following the success ladder of these games, many mobile gaming companies choose global server architecture as a solution to go global.

2. Technical Challenges

- **Architecture design**

The key challenge to the Global Server architecture for mobile games is how to design and deploy the transport layer, business layer, and data layer to fit different game genres.

- **Network latency**

Latency is inevitable during data transmission. But how to reduce latency and minimize its effect on players' visual experience is critical to guarantee the seamless access and fair competition of players across the world.

- **Data read and write**

Efficient data read and write, with data consistency.

- **Resource management**

Centralized and efficient game O&M and resource management.

3. Why Alibaba Cloud

Centralized resource management

Alibaba Cloud provides features that match games with global players and mass data with ease.

- Data centers with high compute power across multiple regions worldwide

- One account for maintenance and resource management

- Centralized presale and aftersales service system along with the localized support

Sustained and secure availability

Alibaba Cloud provides stable and low-latency network to guarantee global players seamless access and fair competition.

- Dedicated and consistent network connection by using Express Connect to link data centers across the world

- High network quality and low latency level specified in SLAs

Comprehensive suite of products

Alibaba Cloud offers properly designed architectures and flexible deployment plans to fulfill the business logic of different genres of games.

- Reliable global serverless architecture and customizable deployment plans

- Customer-based solutions for network latency, data consistency, and other technical difficulties.

4. Business Logic Architecture

The following architecture is applicable to browser games, mobile games, and console games. It is constructed by two main modules: client services and operation services.
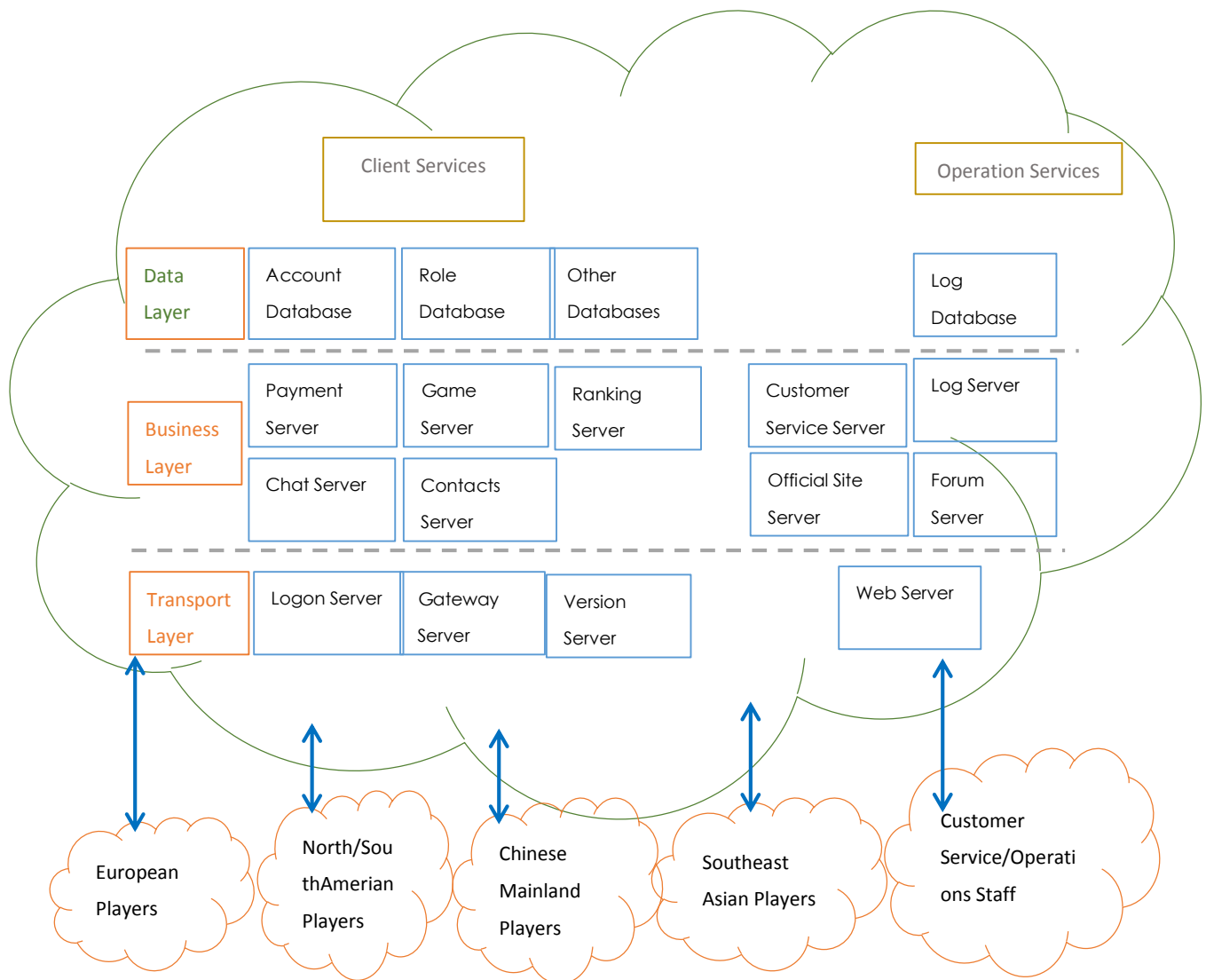
5. Global Server Game Acceleration

The architecture of Global Server Game Acceleration (or Global Acceleration) is built with the centralized game servers deployed in a single region, and edge locations deployed worldwide with the accelerated public network access. Players are geographically routed to the closest edge locations through Express Connect. This aids to minimize the latency, ensures high-quality performance and availability and to realize the global acceleration.

Currently, the prime pain area of global server games is the unfair gaming experience caused by different network latencies of players located in different regions. The technical complications such as price and quality variations of dedicated network connection from the third-parties, the high O&M costs involved in the worldwide edge location deployment, complex proxies are to name a few.

5.1. Mobile Accelerator

Game developers may encounter the following problems:

- Slow app installation and launch

- Slow game loading

- High latency

- Slow interaction among the carriers like China Mobile, Unicom, and Telecom

- Chaotic carrier IP libraries

- High failure rate of server access

- Low availability with non-WiFi connection

- High packet loss rate and domain name hijacking

- Poor interactive experience

Differences in gaming experience of the users spread across the world

These problems make Alibaba Cloud Mobile Accelerator the most appropriate fit to resolve the "last mile" acceleration between clients and edge locations. See the following figure.

Note: For a better view, please zoom in by 150%.
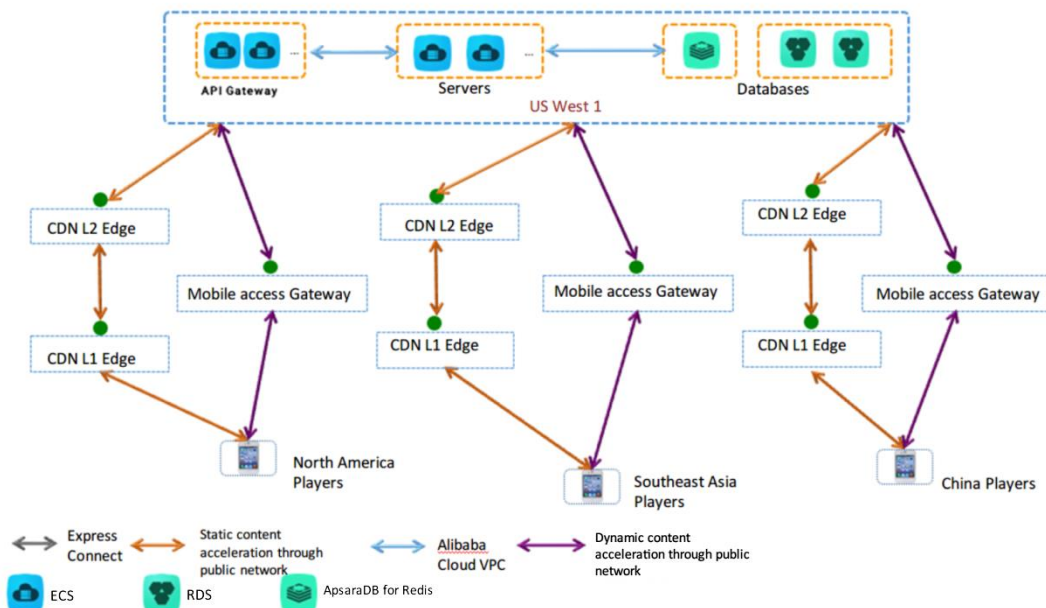
5.2. Global Acceleration

With Alibaba Cloud Global Acceleration, you do not have to manually configure content delivery acceleration to edge locations, which can be complicated, and requires longtime debugging. Global Acceleration ensures you the high availability, scalability, performance, and flexible routing.

Basically, Global Acceleration provides the point-to-point acceleration by using EIPs (elastic IP addresses) to map the ECS instances or VPC Server Load Balancer instances on your centralized servers to the public network. Global Acceleration speeds up the cross-region and cross-country connection to the servers, as shown in the following figure.

Note: For a better view, please zoom in by 150%.
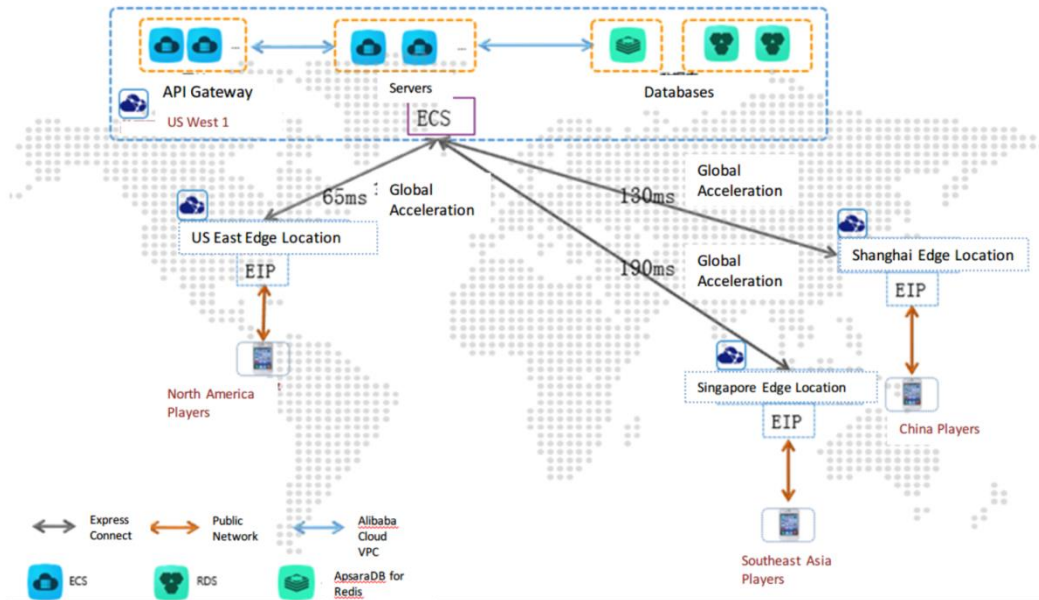
5.3. Whole-path Acceleration

By combining Mobile Acceleration with Global Acceleration, we can accelerate the entire data path from the server to the client. In this solution, Mobile Acceleration speeds up the connection between clients and edge locations based on dynamic routing, while Global Acceleration speeds up the connection between the centralized servers and edge locations through Express Connect. See the following figure.

Note: For a better view, please zoom in by 150%.
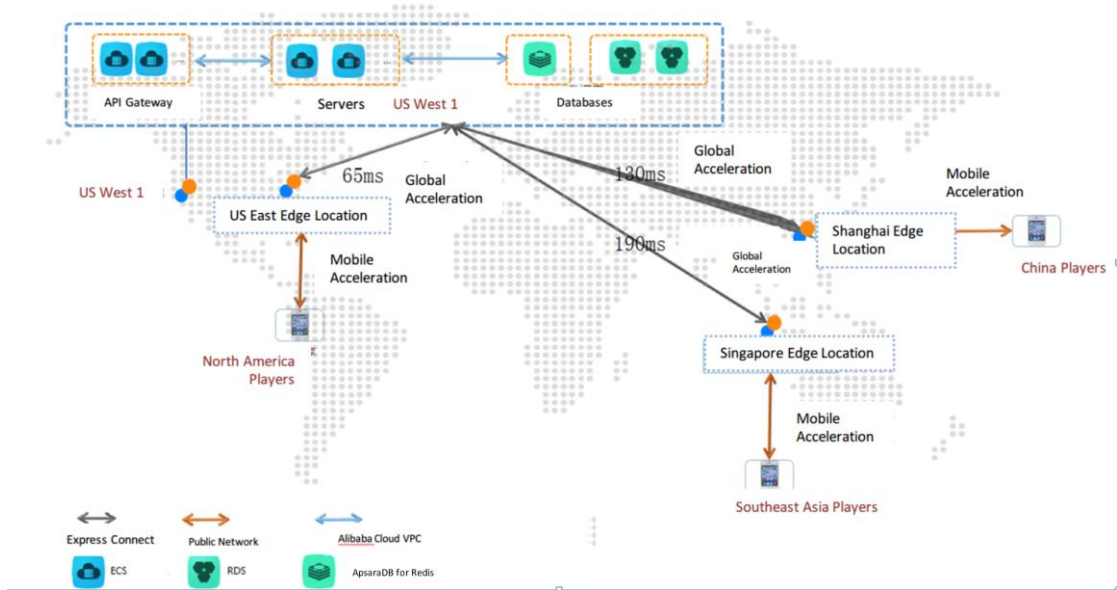
6. Gaming Architecture

Based on the market research, the in-depth discussion with our clients, and our own research on game architectures, we have designed and developed the following four architectures for global server games.

• Fully-centralized deployment

• Centralized deployment and network optimization

• Centralized data and distributed logic

• Fully-distributed deployment

6.1. Fully-centralized deployment

For global server games, fully-centralized deployment is the preferred choice of architecture for games that are not sensitive to network latency. In this architecture, the game access layer, business logic layer, and data layer are all centrally deployed in the same region. Players across the world, access the game over the Internet. The following figure shows the deployment architecture.

Note: For a better view, please zoom in by 150%.
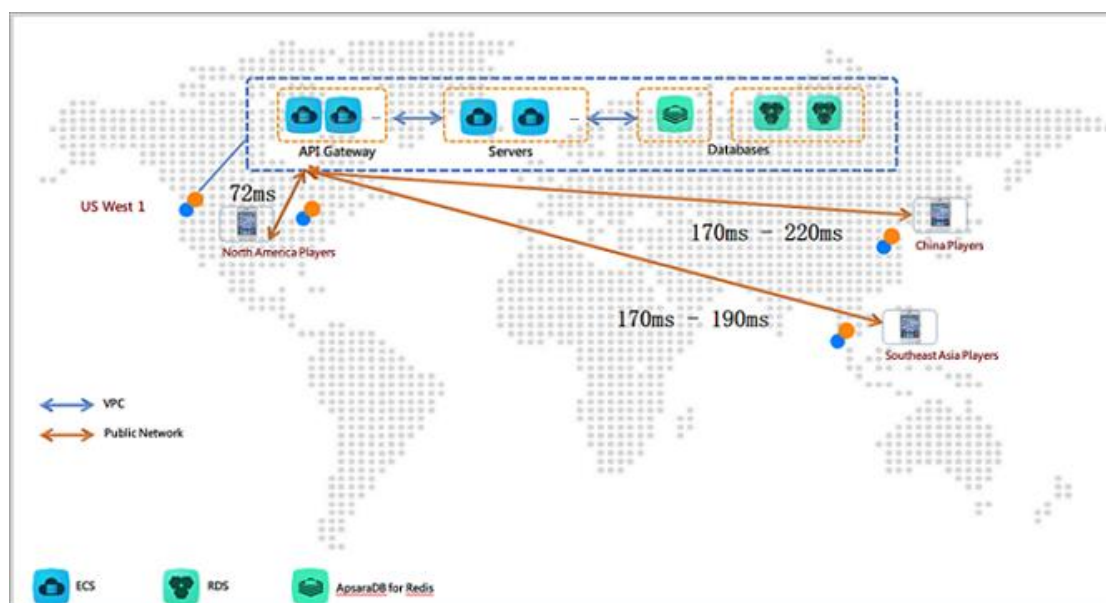
For this architecture, we recommend the c5 family of computing-type instances. This instance type can support most game services that are not sensitive to network latency.

6.1.1. Applicable scenarios

This architecture is suitable for global server games where players are concentrated in a certain region and the gameplay method is designed to be insensitive to network latency. If your preliminary game server architecture design is not suitable for the distributed deployment (for example, if no data synchronization mechanism is set in the logic architecture), you must select fully-centralized deployment when you launch your game.

6.1.2. Architecture advantages and disadvantages

Architecture advantages:

• Easy deployment

• Better gaming experience in the primary coverage region

• No data consistency issues

Architecture disadvantages:

In this architecture, not all players can access through the nearest node

6.2. Centralized deployment and network optimization

In this architecture, the game access layer, business logic layer, and data layer are centrally deployed in a single region. Then, Global Acceleration is deployed on the Alibaba Cloud nodes for the regions you need to cover. After using intelligent DNS for scheduling, players in the various regions automatically access the game from the nearest node. Alibaba Cloud Express Connect provides an intranet connection

between the game service and the various access points. The following figure shows the deployment architecture.

Note: For a better view, please zoom in by 150%.
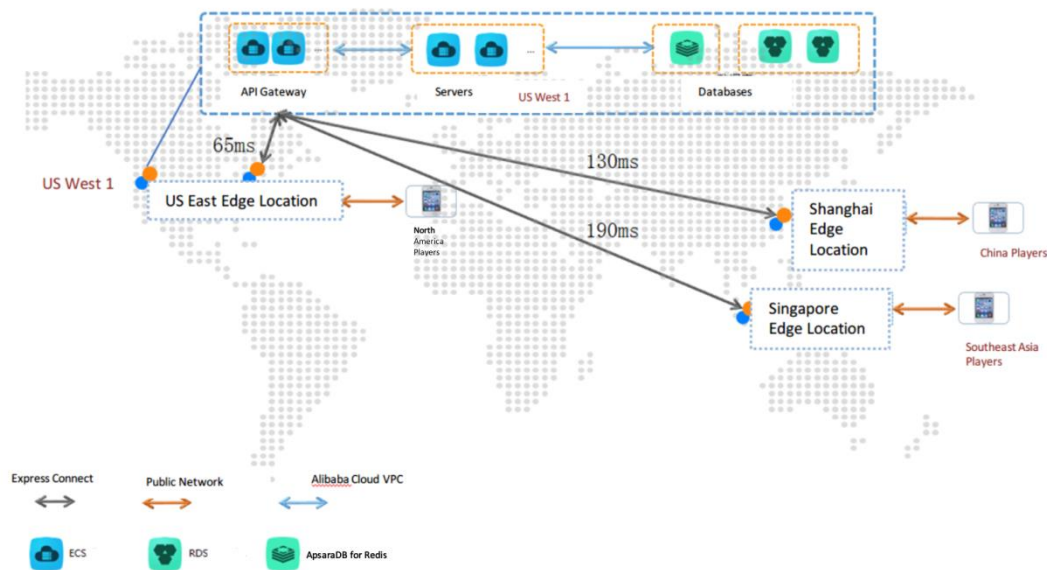
We recommend using the sn1ne family of computing network enhanced instances. This instance type is an enhanced network model, which can meet the architecture's low-latency needs.

6.2.1. Applicable scenarios

This solution is appropriate for game server architectures that are unsuitable for the distributed deployment. It is a better choice for operators who want to cover as many regions as possible, while keeping the game's network latency below 200ms.

6.2.2. Architecture advantages and disadvantages

Architecture advantages:

- Easy deployment

- Network acceleration

- No data consistency issues

Architecture disadvantages:

Fixed latency (For some games, differences in fixed latency can lead to an unfair gameplay. In this case, you must use frame synchronization or other methods to eliminate the latency difference)

6.3. Centralized data and distributed logic

In this architecture, the data layer is centrally deployed in a single data center. Then, the game access layer, business logic layer, and cache layer are deployed in each of the regions that must be covered. The distributed architecture is shown in the following figure.

Note: For a better view, please zoom in by 150%.
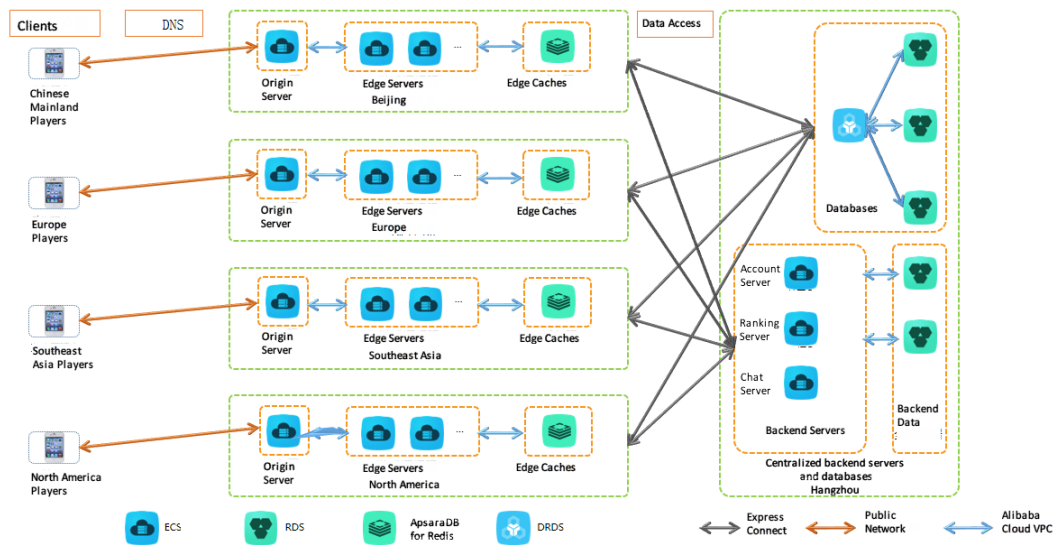
We recommend using one of the following two instance type families:

- The se1ne family of memory network enhanced instances provides high network packet sending and receiving capabilities. As this is an enhanced network model, it also provides high PPS.

- The c4 family of high-frequency computing instances is suitable for MMO games and provides high-frequency specifications.

6.3.1. Applicable scenarios

This architecture is suitable for games where players mostly interact with others in the same region and that have high requirements for network latency (for example, below 120ms, the minimum latency noticeable by human eyes). This architecture is a good choice for action games that want to equally cover all regions of the world.

6.3.2. Architecture advantages and disadvantages

Architecture advantages:

- Players can access through the nearest node

- The game logic is computed on the nearest node (local cache server data sync function: After a player exits, data is written back to the database in real time, 100 rows of dirty data is written every 5 seconds)

- Almost no data consistency issues

- Flexible distributed node adjustment

Architecture disadvantages:

- Must be deployed in multiple regions

- When players interact across regions, the latency of one party increases (a special solution can resolve the problem of latency differences)

- A complete dirty data writeback function is required to ensure data consistency

6.3.3. Key design aspects

This section gives a detailed description of several key design aspects.

- Key aspect 1: Centralized game database deployment

In global server games, gaming rule interaction may occur between any players. Therefore, gaming data, player account data, and global game data (such as rankings) between, must be centrally deployed in a single IDC.

The player data read/write frequency is high and a large proportion of records have a single line. Therefore, it is best to use distributed storage architecture. For example, you can use the Alibaba Cloud DRDS and ApsaraDB for RDS products for a database and table-based splitting. This avoids the performance bottleneck of a single database instance.

- Key aspect 2: Regional player access

As this type of game is a service provided to players around the world and access to

Chinese networks from other countries may be poor, you need to provide the nearest access to the players in various regions around the world.

For example, based on the distribution of Alibaba Cloud data centers, you could deploy access nodes in South China, North China, Southeast Asia, Europe, and North America. Specifically, you can deploy access services in the China East 2, China North 2, Singapore, Germany, and the East US regions.

- Key aspect 3: Player data is regionally cached and regularly written back to the central data center

The players in various regions play together. To avoid the network latency of the remote data reading to affect the overall gaming experience, player data must be regionally cached and then regularly written back to the central data center in batches. This way, the regional game logic servers only need to remotely read data once, during player log on. Then, they can quickly read player data from the cache server.

For example, you can use the Alibaba Redis product for caching, allowing you to implement data caching and persistence. In this way, you do not lose data even if the leased line connection is unavailable.

- Key aspect 4: Intelligent DNS allows nearby access

When players from around the world access the game, the best option is to use the intelligent DNS service for auto scheduling. You can also create your own scheduling service. During scheduling, players' locations must be used to schedule players in the same region to the same access point. If the gameplay involves player matching, the matching algorithm must consider the player's location.

- Key aspect 5: Players playing against each other should be controlled within the same region as much as possible

Because a game logic server is deployed in each region and player data is cached regionally, the backend should limit direct combat to players in the same region as far

as possible.

- Key aspect 6: Centralized storage of global data

Because rankings and other global data is generated by summarizing data from all regions, this data must be centrally stored. Then, each region can regularly pull the necessary global data (the data pull interval must be set according to the ranking generation cycle). However, the previous data version in the local cache cannot be deleted before the latest data is pulled, and the services used to generate global data must also be centrally deployed.

- Key aspect 7: Cross-region player access

It is possible that a player may log on from different regions, during each logon, the system must check if the current logon access point is the same as the last access point used. If they are different, the player data in the cache of the previously used access point must be written back to the database. Then, the player is permitted to log on from the new access node. This prevents data inconsistencies.

6.4. Fully-distributed deployment

In this architecture, the game logic and game data are deployed in a distributed manner. Only global game services and data are centrally deployed. This architecture is suitable for games with low read/write frequencies and less-strict network latency requirements. The following figure shows the deployment architecture.

Note: For a better view, please zoom in by 150%.
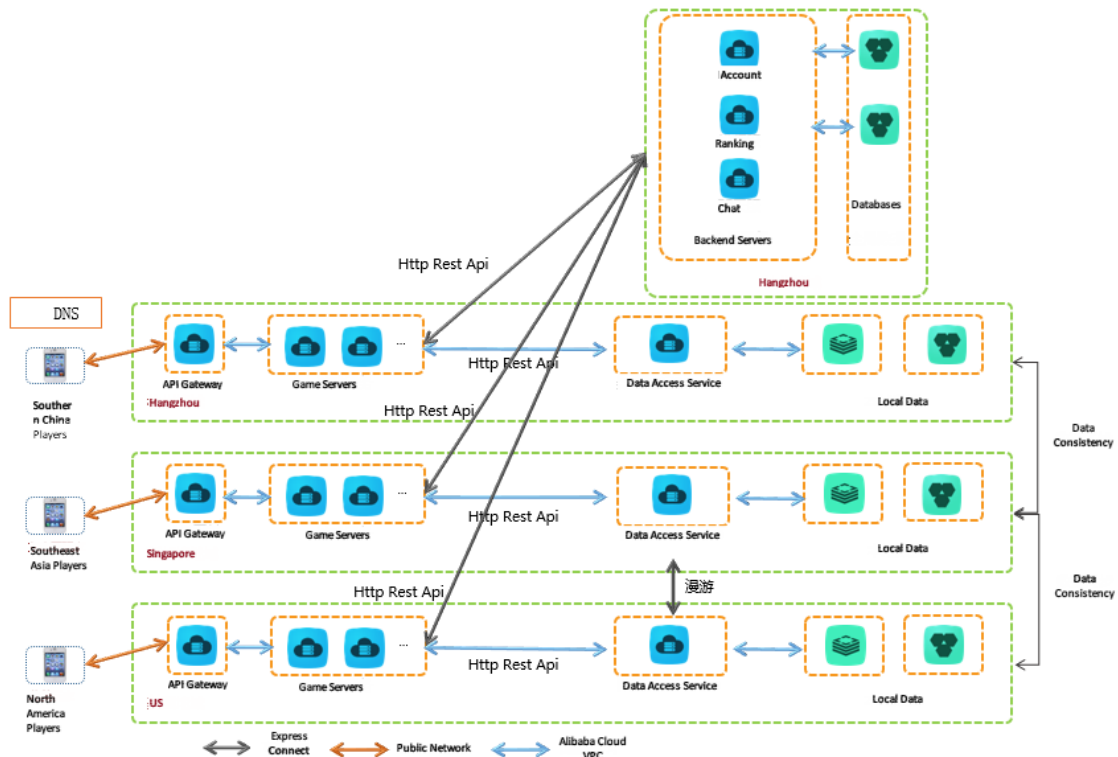
We recommend using one of the following three instance type families:

- The cm4 family of high-frequency general instances is suitable for fully-distributed architectures with high requirements for network, computing, and storage resources.

- The se1ne family of memory network enhanced instances provides high network packet sending and receiving capabilities.

- The g5 family of general instances provides a distributed cache function while balancing the ratios of various resources.

6.4.1. Applicable scenarios

This architecture is suitable for games with high network latency requirements, a great deal of interaction between players in different regions, equivalent coverage in all regions, and a complete data synchronization mechanism in the game architecture.

6.4.2. Architecture advantages and disadvantages

Architecture advantages:

- Players can access through the nearest node

- Game logic is computed on the nearest node

- The game business logic layer is completely stateless

- Fast data reading and writing speed

Architecture disadvantages:

- Must be deployed in multiple regions

- Large amounts of data must be synchronized across different regions

6.4.3.   Key design aspects

This section gives a detailed description of several key design aspects.

- Key aspect 1: Centralized storage of global data

As the rankings and other global data are generated by summarizing data from all regions, this data must be centrally stored. Then, each region can regularly pull the necessary global data (the data pull interval must be set according to the ranking generation cycle). However, the previous data version in the local cache cannot be deleted before the latest data is obtained, and the services used to generate global data must also be centrally deployed.

- Key aspect 2: Regional player access

As this type of game is a service provided to players around the world and access to Chinese networks from other countries may be poor, you need to provide nearby access for players in multiple regions around the world. For example, based on the

distribution of Alibaba Cloud data centers, you could deploy access nodes in South China, North China, Southeast Asia, Europe, and North America. Specifically, you can deploy access services in the China East 2, China North 2, Singapore, Germany, and US East regions.

- Key aspect 3: Intelligent DNS allows nearby access

When players from around the world access the game, it is best to use the intelligent DNS service for automatic scheduling. You can also create your own scheduling service. During scheduling, players' locations must be used to schedule players in the same region to the same access point. If the gameplay involves player matching, the matching algorithm must consider the player's location.

- Key aspect 4: Real-time game database synchronization

This architecture allows players in different regions to play across servers. Therefore, the game databases of the different regions must be synchronized. You can use Alibaba Cloud DTS for real-time data synchronization, or migrate data when players interact across servers. The data synchronization method used in this solution is described as a special solution later in this article.

7. Cloud product introduction

7.1. ECS product introduction

Elastic Compute Service (ECS) is a basic cloud computing service provided by Alibaba Cloud. You can create any number of ECS instances at any time according to your business needs, without having to purchase hardware in advance. As your business grows, you can resize the disks and increase the bandwidth of your ECS instances. When you no longer need an ECS instance, you can release it to reduce your fees.

An ECS instance is a virtual computing environment which includes a CPU, memory, operating system, disks, bandwidth and other basic server components. It is the actual operating entity presented to each user.

7.1.1. Product advantages

Compared to normal IDCs and server vendors, Alibaba Cloud's ECS instances provide high availability, security, and flexibility.

7.1.1.1. High availability

Compared with Internet Data Centers (IDCs) and server vendors, Alibaba Cloud adopts more stringent IDC standards, server access standards, and O&M standards to ensure data reliability and the high availability of cloud computing infrastructure and ECS.

In addition, each region of Alibaba Cloud has multiple zones. For higher availability, you can build active/standby or active/active services in multiple zones.

For a finance-oriented solution with three centers in two locations, you can build higher-availability services in multiple regions and zones. These services include disaster tolerance and backup, which are supported by Alibaba Cloud's mature

solutions.

Services can be switched smoothly within Alibaba Cloud's framework. Alibaba Cloud's industry solutions support a variety of services, such as finance, ecommerce, and video services.

Alibaba Cloud provides the following support services:

- Products and services for increased availability, including ECS, Server Load Balancer, multi-backup database, and Data Transmission.

- Industry partners and ecosystem partners that help you build a more advanced and stable architecture and ensure service continuity.

- Diverse training services that enable you to achieve high availability from the business end to the underlying basic service infrastructure.

7.1.1.2. Security

Users of cloud computing are most concerned about security and stability. Alibaba Cloud has recently passed a host of international information security certifications, including ISO27001 and MTCS, which demands strict confidentiality of user data and user information, as well as user's privacy.

Alibaba Cloud VPC offers more business possibilities. You only need to perform a simple configuration to connect your business environment to global IDCs, and then your business becomes more flexible, stable, and scalable.

- **This solution does not cause any problems to your self-built IDC.** Alibaba Cloud VPC can connect your IDC through a leased line to build a hybrid cloud architecture. With Alibaba Cloud's ecosystem, you can create a business ecosystem beyond imagination.

- **Alibaba Cloud VPC provides greater stability and security:**

Stability: After building your business on VPC, you can update your network architecture and network functions on a daily basis as the network infrastructure

constantly evolves, and allows your business to run steadily. VPC allows you to divide, configure, and manage your network as needed.

Security: VPC features traffic isolation and attack isolation to protect your services against the constant attack traffic on the Internet. By building your business on VPC, you have already established your first line of defense.

7.1.1.3. Elasticity

Currently, Alibaba Cloud is capable of providing the IT resources required by a medium-sized Internet enterprise within just a few minutes. In this way, most enterprises that build business on cloud can process huge business volumes.

- **Computing elasticity**

Vertical elasticity refers to configuration changes to a single server. With Alibaba Cloud, however, you can change the configuration of the ECS or storage capacity you have purchased based on the increase or decrease of business volume.

- **Transverse elasticity**

In a traditional IDC, it is basically impossible to prepare resources during peak hours of a game application or live broadcast platform. However, ECS gives you the elasticity to cope with peak hours. In off-peak hours, you can release redundant resources to reduce costs.

- **Storage elasticity**

In traditional IDC solutions, you have to add more servers to increase your storage capacity, but only a limited number of servers can be added due to the expensive costs. Cloud computing can provide you with mass storage. Storage assurance enables you to purchase the storage you need at any time, which provides storage assurance.

- **Network elasticity**

If you purchase Alibaba Cloud VPC, all network configurations and offline IDC

configurations can remain the same, but more possibilities are allowed. This solution provides interconnection and security domain isolation between data centers, and all VPC network configurations and plans are flexible.

7.1.2. Product type families

An instance is the minimal unit that can provide computing services for your business. It uses a certain specification to provide you with the associated computing capabilities.

Note: The instance types available for sale vary by region. The specific information is shown on the Instance Purchase Page.

ECS instances can be divided into multiple type families based on their business and usage scenarios. In the same type family, there are also multiple sub-types based on the CPU and memory configuration. ECS instances only take on their unique and definite service formats when equipped with disks, images, and network types.

7.2. ApsaraDB for RDS product introduction

Alibaba Cloud ApsaraDB for RDS (Relational Database Service) is a stable, reliable, and elastically scalable online database service.

Based on Alibaba Cloud's distributed file system and high-performance storage, ApsaraDB for RDS supports the MySQL, SQL Server, PostgreSQL, and PPAS (Postgre Plus Advanced Server, a database highly compatible with Oracle) engines. It provides a complete set of solutions for disaster tolerance, backup, recovery, monitoring, migration, and other functions.

7.2.1. Product advantages

- The improvements to the underlying source code increase database performance three times over. ApsaraDB for RDS provides performance monitoring, remote disaster tolerance, primary/standby database failover, data

backup and recovery, and other functions. It ensures the security, stability, and reliability of cloud databases.

- ApsaraDB is easy to use and elastically scalable. You do not need to purchase any hardware or install any software. You simply need to purchase instances based on your business needs. At the same time, you do not have to worry about hardware planning problems, as you can adjust your ApsaraDB for RDS instance specifications and capacities at any time to cope with the pressure on your business.

- ApsaraDB for RDS supports full-data hot migration. This allows you to smoothly migrate your business to the cloud, without affecting its operation.

- ApsaraDB for RDS provides automated O&M management to simplify complicated O&M work and reduce manpower costs, allowing you to focus on business development.

7.2.2. Functions and features

This section introduces the various features of ApsaraDB for RDS.

7.2.2.1. Read/Write splitting

The implementation of read/write splitting is shown in the following figure.
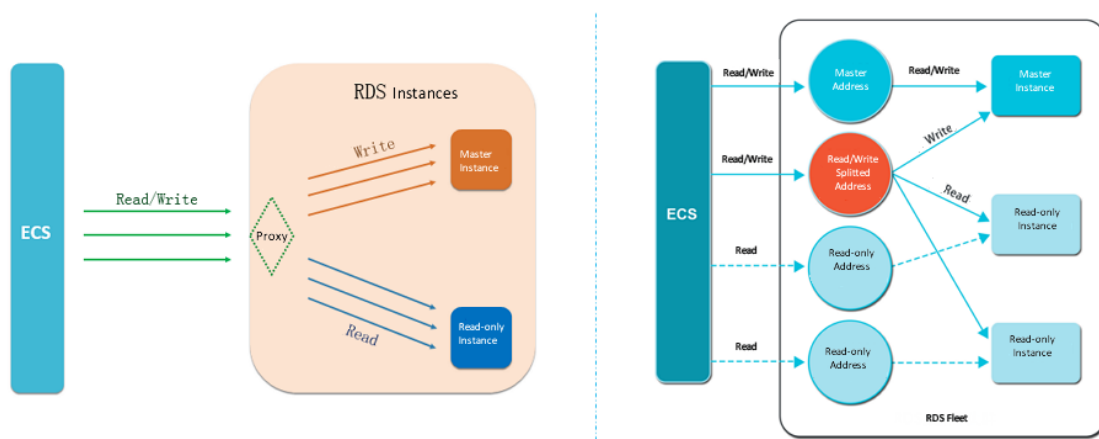
- After activating a read-only instance, you can activate the read/write splitting

function for free.

- Read/Write splitting provides real-time health checks, custom read weighting, and other functions to achieve 99.95% of availability.

- With the read/write splitting function activated, the instance has three connection addresses: master instance connection address, read-only connection address, and read/write splitting address.

7.2.2.2. Remote disaster tolerance

ApsaraDB for RDS allows you to create remote disaster recovery instances to ensure availability in the case of faults affecting multiple zones.

In addition, you can use DTS to synchronize databases in self-built IDCs to Alibaba Cloud databases on any ApsaraDB for RDS instance in real time. Even if the IDC is damaged, a data backup always exists on ApsaraDB for RDS.
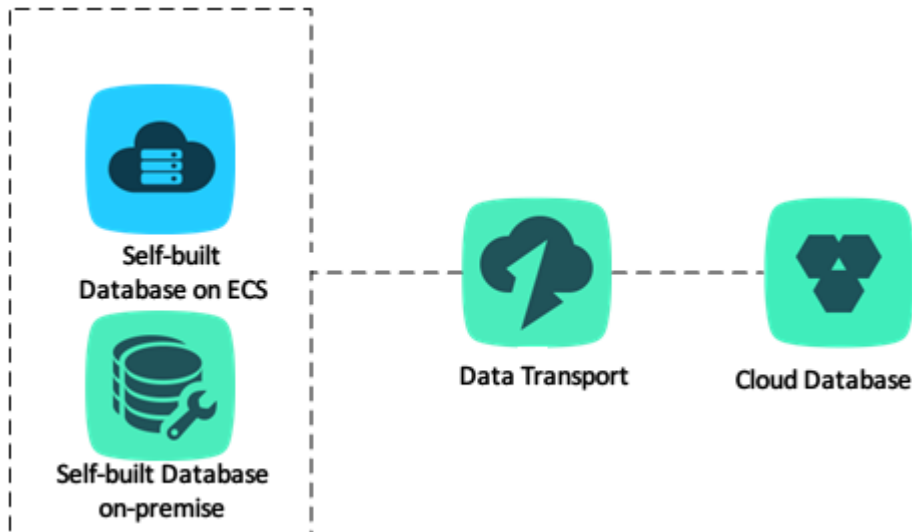
7.2.2.3. Big data analysis

Alibaba Cloud provides MaxCompute for the storage and processing of massive amounts of structured data. The service offers mass data warehouse solutions and

analytical modeling services for big data.

You can import data from an ApsaraDB for RDS instance to MaxCompute using Data IDE, to achieve large-scale data computing, as shown in the following figure.
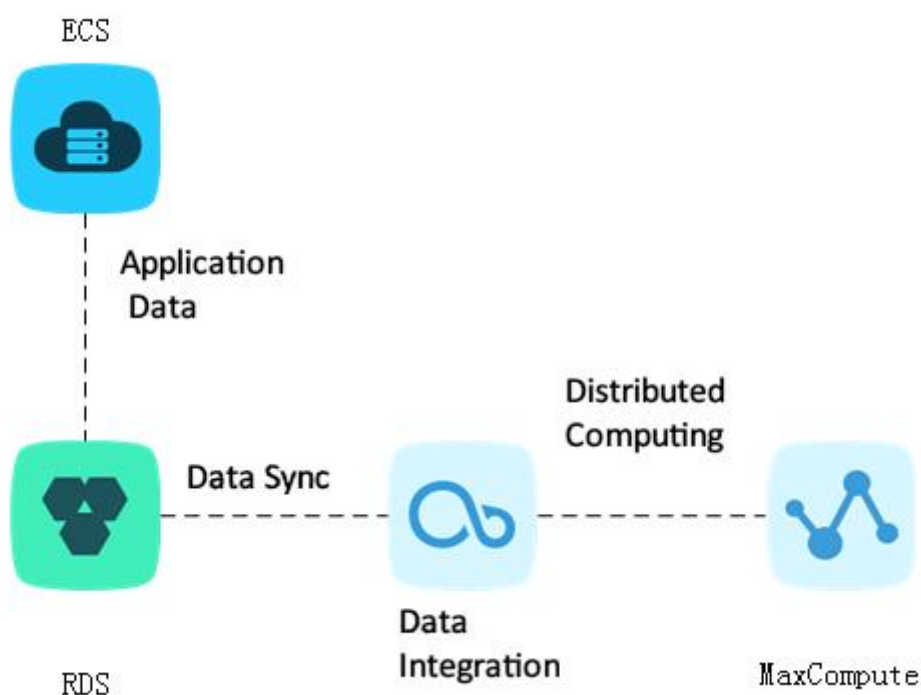
7.2.2.4. OpenSearch

OpenSearch is a structured data search hosting service, which provides mobile application developers and website owners with simple, efficient, stable, low-cost, and scalable search solutions. OpenSearch provides a function to auto synchronize data in ApsaraDB for RDS to OpenSearch for various complex search types.

ECS

Application
Data

structured Data
Search

RDS   OpenSearch
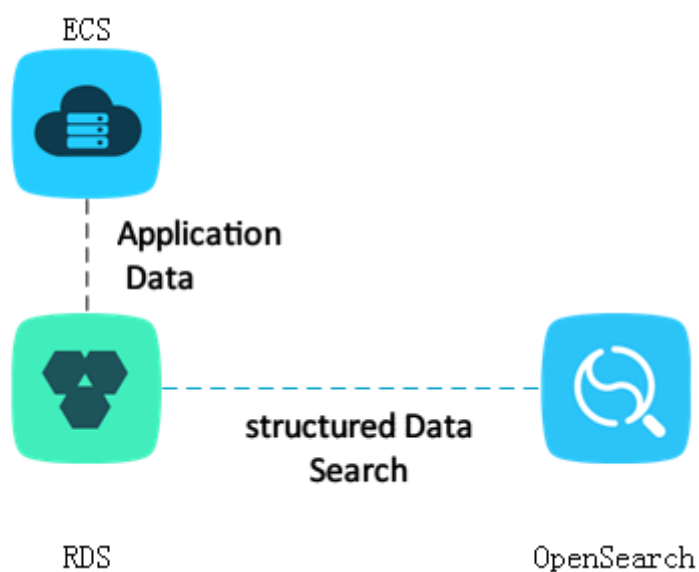
7.2.2.5. Diversified data storage

ApsaraDB for RDS supports diversified storage extensions when used in combination with ApsaraDB for Memcache, ApsaraDB for Redis, OSS, and other storage products.
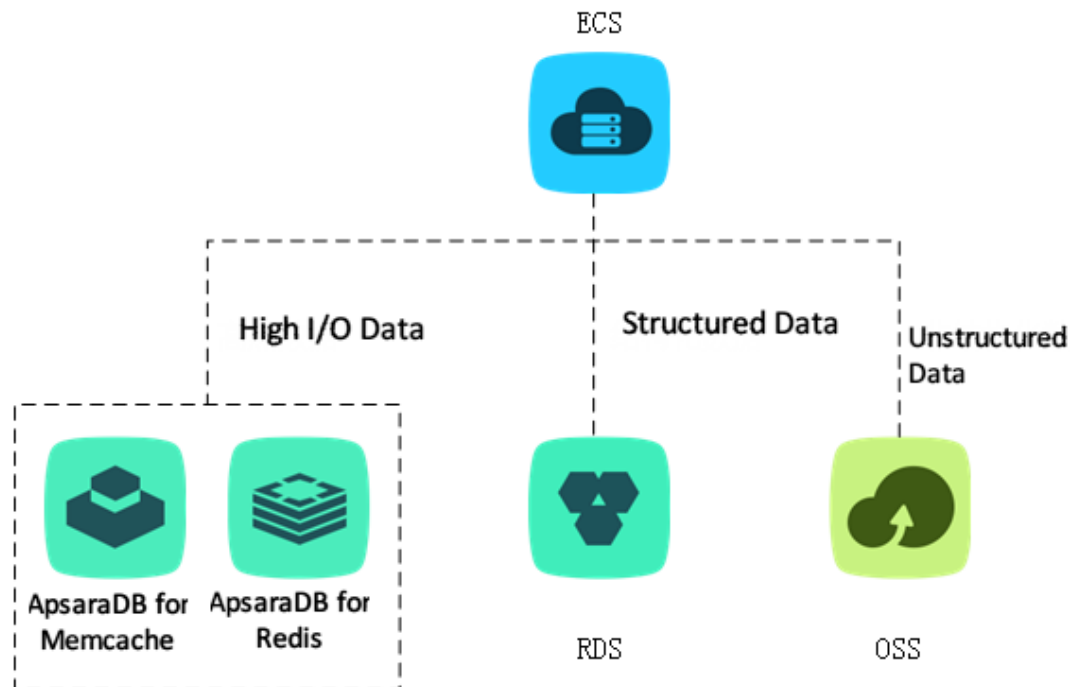
- **Cache data persistence**

ApsaraDB for RDS can be used together with ApsaraDB for Memcache and ApsaraDB for Redis to form a high-throughput and low-latency storage solution.

Compared with ApsaraDB for RDS, ApsaraDB for Memcache has two advantages:

- High response speed: The request latency of ApsaraDB for Memcache and ApsaraDB for Redis is usually within several milliseconds.

- High response to queries: The cache area supports a higher number of Queries per Second (QPS) than ApsaraDB for RDS.

- **Heterogeneous data storage**

OSS is an Alibaba Cloud storage service that features massive capacity, robust security, low cost, and high reliability. ApsaraDB for RDS and OSS can work together to form heterogeneous data storage solutions.

7.3. Redis product introduction

ApsaraDB for Redis is compatible with open-source Redis protocol standards and provides persistent memory database services. Based on its high-reliability, dual-machine hot standby architecture, and seamlessly scalable cluster architecture, this service can meet the needs of businesses that require high read/write performance and flexible capacity adjustment.

Using a memory + hard disk storage layout, ApsaraDB for Redis can meet your persistence requirements, while providing high-speed data read/write capability.

7.3.1. Product advantages

Superior performance

The cluster function supports ultra-high capacity and performance. This service supports cluster functions and provides cluster instances of 128 GB or higher to meet large capacity and high performance needs. It provides master-slave dual-node instances of 64 GB or smaller capacity, meeting the average user's requirements for capacity and performance.

Elastic resizing

- One-key storage capacity resizing: You can use the console to adjust the storage capacity of your instances as needed.

- Online resizing with no service interruption: You can adjust the instance capacity online without suspending your services or affecting your business.

Data security

- Persistent data storage: With its memory + hard disk storage layout, ApsaraDB for Redis provides high-speed data read/write capability, while satisfying data persistence requirements.

- Master-slave dual-backup for data: All data on the master node has a backup copy on the slave node. Password authentication is required for secure and reliable access.

High availability

Dual-copy and cluster version instances have a master node and a slave node. This prevents service interruptions caused by SPOF. Automatic hardware fault detection and recovery: This feature auto detects hardware failures and fail over to the slave node, and also restores the service in a matter of seconds. Instance-level resource isolation provides enhanced stability for individual services.

Seconds-level monitoring

Real-time seconds-level monitoring and minute-level historical monitoring. ApsaraDB for Redis provides monitoring information for data structures and interfaces. This allows you to easily monitor access and operations performed on your ApsaraDB for Redis instances.

Ease of use

- Out-of-the-box service: This product requires no setup or installation and can be used right after purchase for quick and convenient business deployment.

- Compatible with open-source Redis: This product is compatible with Redis commands, and any Redis client can easily establish a connection with ApsaraDB for Redis to perform data operations.

- Visual management and monitoring panel: The console provides monitoring statistics for multiple metrics and allows you to conveniently manage ApsaraDB for Redis instances.

7.3.2. Gaming industry applications

Game companies can use ApsaraDB for Redis as an important part of their deployment architectures.

- **Scenario 1: Using ApsaraDB for Redis for data storage**

Game deployment architecture is relatively simple. With the main program deployed on ECS, all business data is stored in Redis as a persistent database. ApsaraDB for Redis supports persistence functions, with a primary-standby dual-machine redundant data storage.

- **Scenario 2: Using ApsaraDB for Redis as a cache to accelerate application access**

Using Redis as a cache layer can accelerate application access. Data are stored in a backend database (ApsaraDB for RDS).

ApsaraDB for Redis provides a hot standby high-availability architecture that ensures extremely high service reliability. The master node provides external services. If this node fails, the system automatically switches over to the slave node. The entire failover process is completely transparent to users.

7.4. Express Connect

Alibaba Cloud Express Connect helps you build private network communication channels between VPC instances and between a VPC instance and your data center. This increases the flexibility of the network topology and enhances the quality and security of cross-network communication.

Express Connect helps you avoid unstable network quality caused by public network bypass and the risk of data theft during transmission.

- Intranet communication between VPC instances

Express Connect supports intranet communication between VPC instances in the same region or different regions, of the same account or different accounts. By creating virtual router interfaces on the VPC instances at both ends to build an Express Connect channel on Alibaba's backbone transmission network, you can implement fast, secure, reliable, and convenient communication between VPC instances.

- Intranet communication between your physical data centers and Alibaba Cloud VPC

You can use a physical leased line to connect your data centers to Alibaba Cloud at the physical layer, and then establish a virtual border router and a virtual router interface to connect your data centers with an Alibaba Cloud VPC.

7.4.1. Infrastructure

Based on a three-layer overlay and switch virtualization technology in the SDN architecture, Alibaba Cloud isolates customers' physical connection access ports and abstracts them into virtual border routers.

Using mainstream tunneling technology, Alibaba Cloud encapsulates customers' packets in a switch. In addition, tunnel encapsulation is used between your physical leased lines and the VPC router. Then, the data is transmitted to the VPC.
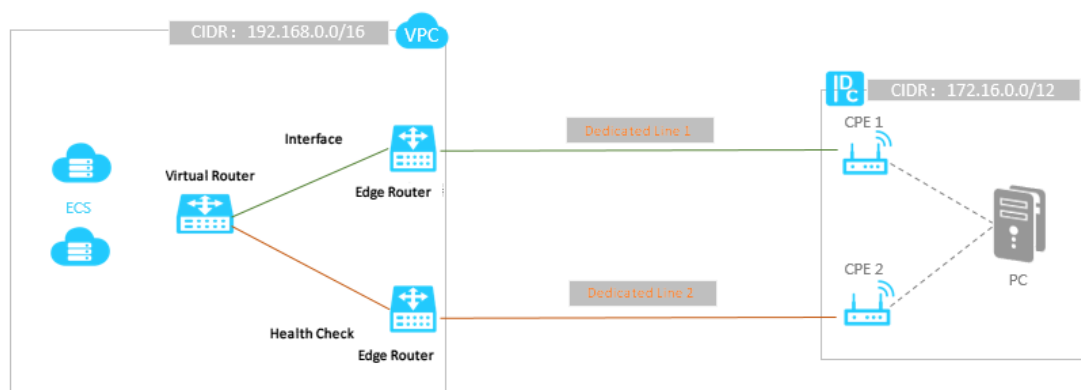
Difference between the Public network and Express Connect access

Comparison point	Public network	Express Connect
Communication quality and availability	Long-distance public network communication is affected by a variety of factors that make it hard to ensure latency stability and low packet loss rates.	Alibaba Cloud's high-quality infrastructure delivers enhanced link quality and availability: • Delay jitter $\leqslant$ 20% • Packet success rate $\geqslant$ 99.8% • Availability $\geqslant$ 99.95% • Packet loss rate < 0.2%
Cost	You must pay traffic and bandwidth fees for using the public network.	Alibaba Cloud provides free bandwidth, and virtual devices can be purchased as needed at affordable prices.
Security	Communications data is at a risk of being monitored or stolen when transmitted on the public network.	Based on Alibaba Cloud's virtual network technology, Express Connect isolates different communication links to enhance data security.

7.4.2. Product advantages

• **High-speed intercommunication**

Powered by Alibaba Cloud's network virtualization technology, this product can

connect different network environments, so that both sides can communicate directly by intranet and bypass public networks. Even over a long distance, you enjoy low latency and high bandwidth comparable to intranet communication.

- **Stable and reliable**

Powered by the Alibaba Group's outstanding infrastructure, Alibaba Cloud Express Connect ensures that communication between networks is stable and reliable.

- **Security**

Express Connect performs inter-network communication at the network virtualization layer, so all data is transmitted through Alibaba's own facilities. This eliminates the need to use public networks and achieves multi-tenant isolation, allowing you to avoid the risk of data theft during transmission.

- **Flexible topology**

Whether for a multi-center deployment, merged networks created from physical machine rooms and cloud resources, or the use of multiple leased lines for disaster tolerance, Express Connect can deliver a flexible network topology.

- **Ease of use**

Express Connect is easy to use. With just a few simple operations you can replace complex physical network configurations. At the same time, you have complete control over Express Connect's communication bandwidth, operation statuses, and route configurations on both sides. What you see is what you get.

- **Purchase only what you need**

A diverse range of specifications are available. Different specifications provide different data transmission speeds. You can purchase only what you need for your business to control your costs.

7.5. Global Acceleration

Global Acceleration is a web acceleration product. Supported by Alibaba's global backbone network, Global Acceleration enables nearest possible access globally. This helps minimize the impact of network problems such as latency, jitter, and packet loss on the quality of service, and brings a better experience to the global users of your services.

At its far end, Global Acceleration only requires a GA public network IP address portal. The backend is the same game server that helps to truly achieving a global server architecture.

7.6. Alibaba Cloud DNS

Alibaba Cloud DNS is a cloud computing service portal. It gradually integrates existing Alibaba Cloud products, forming an indispensable element in the cloud product family. ECS, ApsaraDB for RDS, CDN, Server Load Balancer, and other products provide users with efficient and reliable computing, storage, website acceleration, and load balancing services. Alibaba Cloud DNS provides a powerful and stable resolution scheduling portal. This ensures that users have a smooth access experience and provides them with an all-in-one service experience.

7.6.1. Product advantages

Exclusive DNS server cluster

- Our self-developed high-performance DNS kernel significantly improves resolution speed and processing capabilities by a factor of over 1,000 compared to traditional DNS resolution.

- We offer exclusive DNS servers for paying users.

- Alibaba Cloud VIP users enjoy exclusive maintenance services from the O&M and security protection teams.

Maximum 100% SLA for stable operation assurance

Our maximum DNS server SLA promises zero downtime throughout the year.

Alibaba Cloud DNS cyber-attack protection

This product offers high-level protection capabilities against ultra-large DDoS attacks and DNS Query attacks. It monitors and defends against DNS attacks in real time.

Rapid, real-time resolution effectiveness

You can intelligently customize minimum TTL values. Our rapid first resolution technique syncs resolution records to all Alibaba Cloud DNS nodes within one second.

Smooth imperceptible switchover

After you add a domain name, you can preset a resolution record. Once the change takes effect in DNS, the resolution service smoothly switches over without affecting the access experience.

Load balancing, URL forwarding, search engine line, and plan domain name changes

Our complete set of resolution management tools allow you to conveniently add, modify, and delete various domain name records after logging on to Alibaba Cloud DNS. This includes A, CNAME, MX, AAAA, SRV, TXT, and NS records.

In addition, we also provide URL forwarding services, TTL customization, and support for the resolution of Chinese domain names, wildcard domain names, and subdomain names.

7.6.2. Application scenarios

Alibaba Cloud DNS can also work with other Alibaba Cloud products, so that the resolution record settings for these products can quickly be effective.

CDN/OSS: Alibaba Cloud DNS uses APIs to interact with CDN and OSS. It supports synchronization settings for CNAME records generated during domain name acceleration.

- Alibaba Cloud Security - Web Vulnerability Detection: DNS works with Web Vulnerability Detection to scan for resolution record-related website vulnerabilities.

- Alibaba Cloud Security - Anti-DDoS Pro: DNS works with Anti-DDoS Pro to synchronize anti-DDoS resolution record settings.

- Alibaba Cloud Security - Server Guard: Alibaba Cloud DNS works with Server Guard to inspect resolution records related to the Server Guard statuses.

- Alibaba Mail: Alibaba Cloud DNS works with Alibaba Mail to help users easily set common Alibaba Mail resolution records.

- ECS: Alibaba Cloud DNS works with ECS to help users easily set host resolution records.

8. Typical system design

8.1. Global rankings design

When designing a global rankings service, you must consider demand, analysis, data structure, rank data persistence, rank server SPOF issues, and other issues. A ranking service architecture is shown in the following figure:

**Note:** For a better view, please zoom in by 150%.
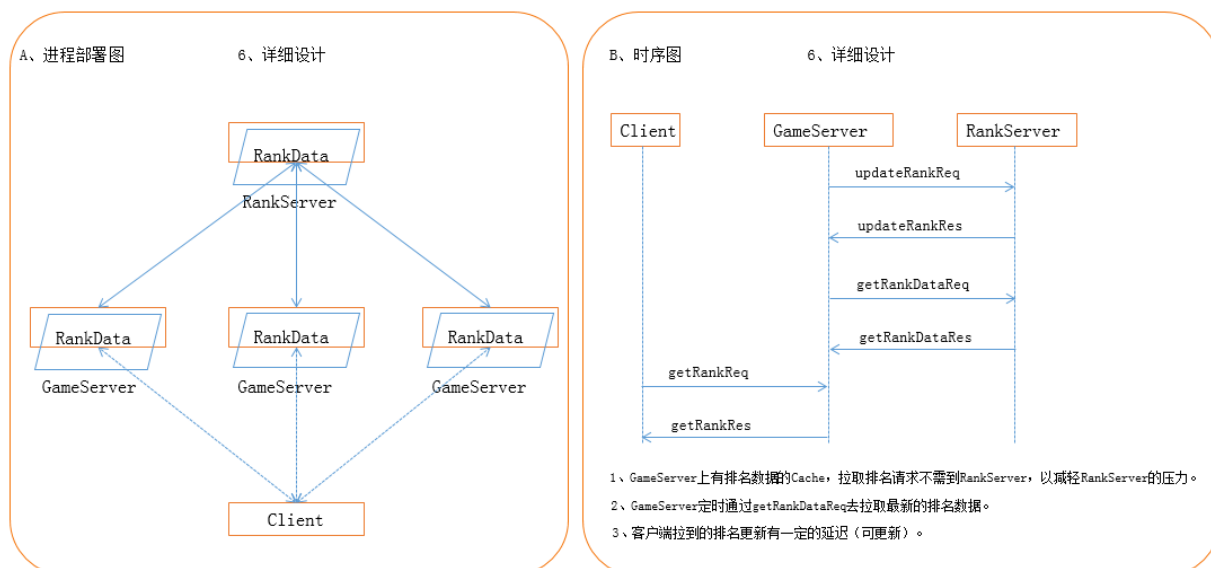
This architecture uses Redis to implement data at regular intervals. The game server reports rank data to the rank server. Clients pull rank data from the game server and the game server pulls rank data from the rank server.

8.2. Game time design

All the game servers use GMT Jan 1, 1970 00:00:00 offset (generally an absolute value of 1 for the second count) to express the in-game time. This time is synced to game clients, who use the time zone set on the cell phone to compute the game time

to be displayed in the client. The specifications are shown in the figure below:

Note: For a better view, please zoom in by 150%.
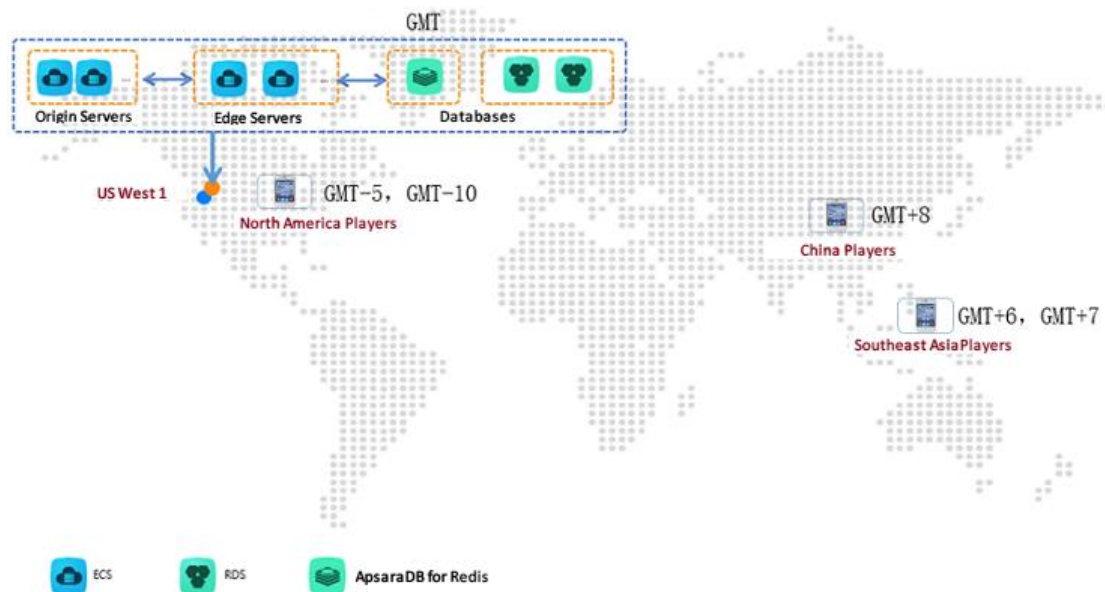
**Why is the game time synchronization necessary?**

- To prevent client modifications to the local time from confusing the game logic, the client must use the server time.

- Basically all techniques to solve game status synchronization problems, such as predictive pull or server verification sync, require time synchronization.

- In games, some timed events or time-related gameplay features require a standard and uniform game server time to ensure fair play.

- During the client and server communication, a more secured method is to add a timestamp to each packet so that the server can verify the validity of the packets.

8.3. Game data synchronization

8.3.1. Solution 1 Use a cache for instant data writeback

In this method, data is stored centrally and the local caches instantly write data back to the database. The specific architecture is shown in the following figure.

Note: For a better view, please zoom in by 150%.
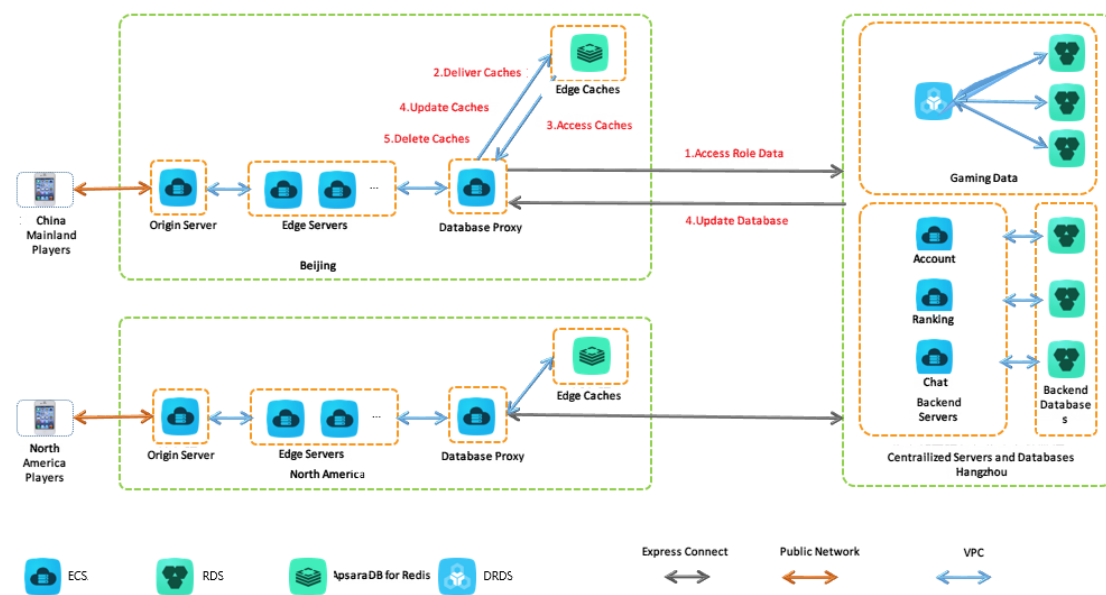
Currently, there are two main scenarios that involve data synchronization:

Scenario 1 Players log on in their local regions (explanation marked in red text in the previous figure).

1. The database proxy reads role data from the database.

2. The database proxy inserts the role data into the cache.

3. The role data read from the cache is used in computing.

4. When the data changes, the cache data and database data are updated simultaneously.

5. When players exit the game, the data in the cache is deleted.

Scenario 2 Players log on across regions.

1 First, the system checks if the player is logged on in another region. If yes, it goes to step 2. If no, it goes to step 3.

2 The player's role data is written back to the database and deleted from the cache when the player is logged out.

3 The database proxy reads role data from the database.

4 The database proxy inserts the role data into the cache.

5 The role data read from the cache is used in computing.

6 When the data changes, the cache data and database data are updated simultaneously.

7 When players exit the game, the data in the cache is deleted.

8.3.2. Real-time regional database synchronization

The various regional databases are synchronized in real time. You can accomplish this using Message Service or Alibaba Cloud DTS. Each of the regional databases store data for all players.

Note: For a better view, please zoom in by 150%.
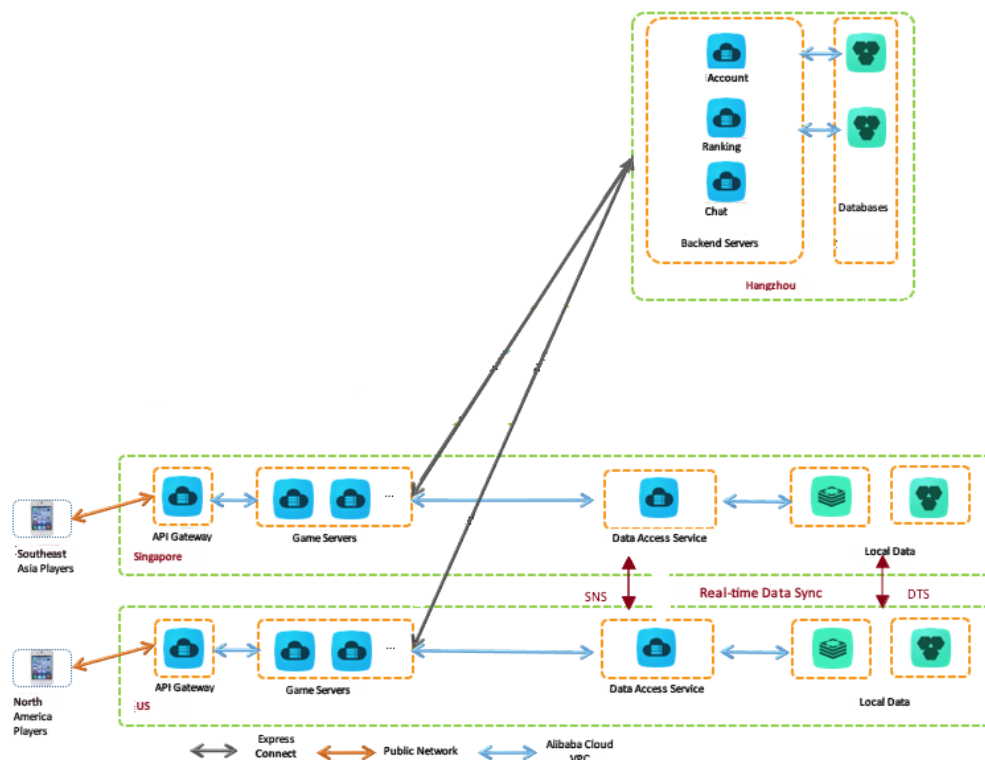
Advantages:

The players can play the game from different regions, their data need not be migrated. All regions support local data reading and writing.
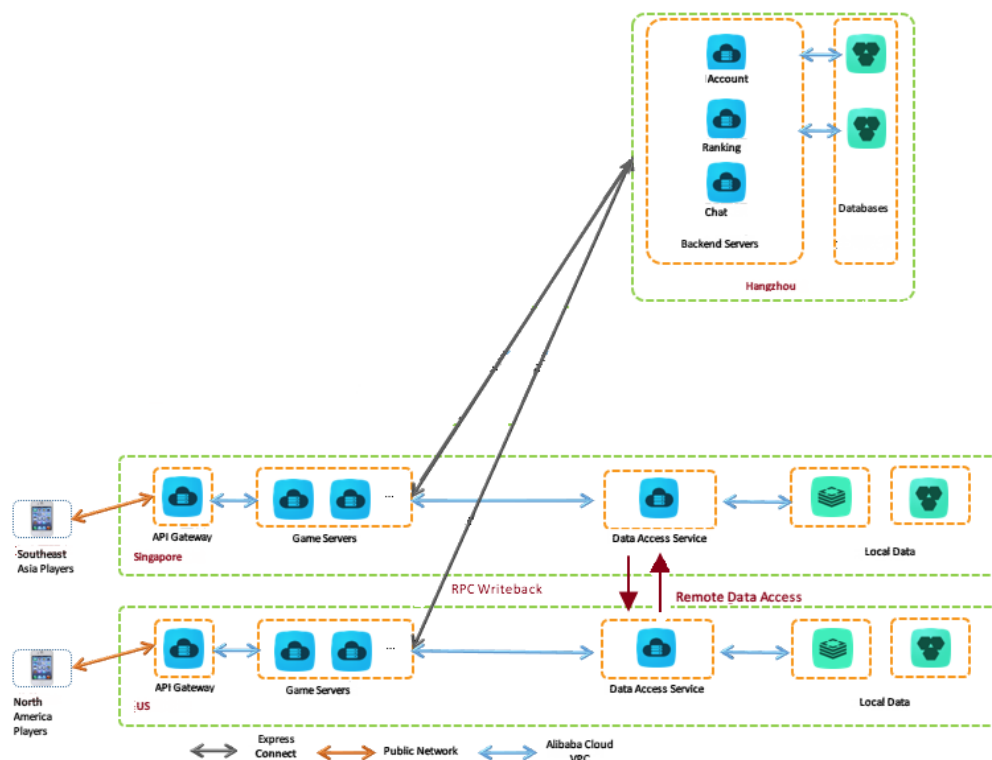
Disadvantages:

The asynchronous data synchronization can produce data inconsistencies. If the player uses a VPN to access the game, a transient VPN disconnection can cause the player to log on again in another region. If data synchronization messages are lost or delayed, the data read when a player logs in again may not be up to date.

The various regions contain all player data, so real-time synchronization may put a high level of pressure on the database. Real-time synchronization between regional databases demands a great deal of cross-region leased line bandwidth.

8.3.3. Solution 3 Cross-region data update

When a player logs on from a different region, the player's local server must remotely read combat data. The combat results call an API to trigger a data update on the local server.

Note: For a better view, please zoom in by 150%.

Advantages:

For players in the same region, role data is read/written to/from the nearest local node.

Disadvantages:

The implementation logic is complicated. All gameplay results must be abstracted to an interface and data is centrally changed by the original server interface.

8.3.4. Solution 4 Remote data migration

When a player logs on from a different region, the player data is remotely migrated from the player's previous server. Each time a player logs on, the system must check if data migration is required.

Note: We suggest that you zoom in by 150% for a better view.
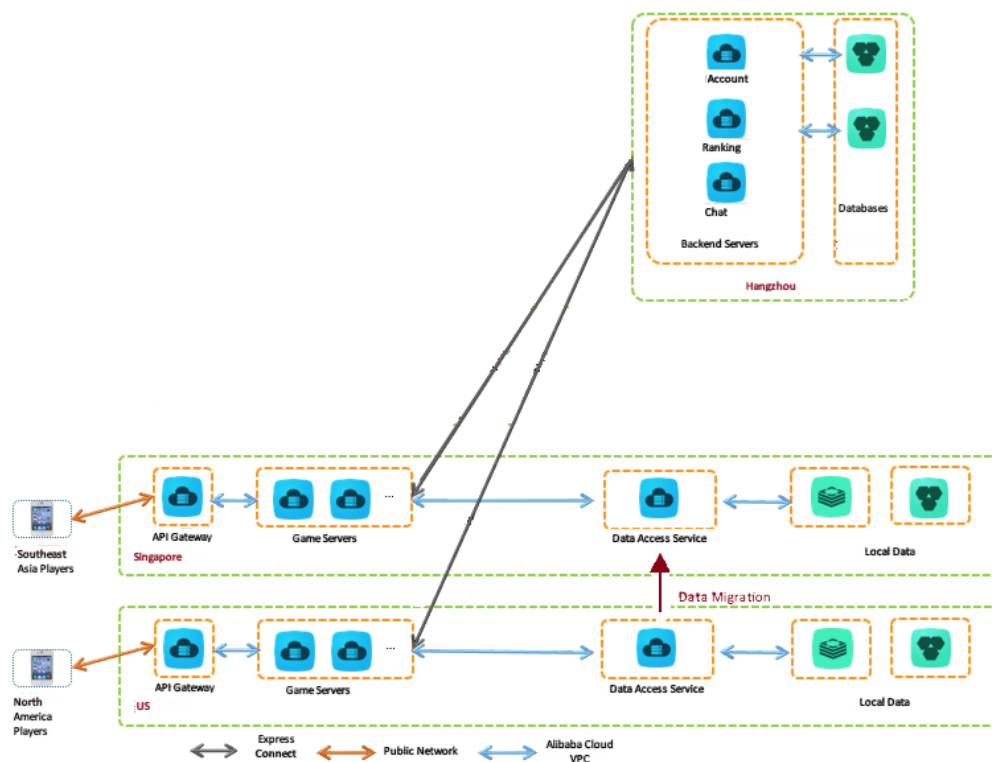
8.4. Game localization solutions

Game localization is an important factor for consideration of games seeking to enter overseas markets. Games, either being released overseas or looking to expand to new regions must be attentive to localization work.

The general approach is to build a standard client installation package, which contains several basic material packages, art material packages for different languages, and some program materials. This allows the game to be dynamically rendered based on the phone's language version.

**There are three common installation strategies:**

- After the client installation package is downloaded, the user manually sets the language version. Such large packages contain various language packs and support one-click language selection.

- When the game is installed on the client, it detects the language used by the mobile device to dynamically select the language version to install. These installation packages generally have a built-in default language and, if another language is needed, a language pack is downloaded from the Internet.

- Different language versions of the client installation package are submitted to the app store. Then, different installation package versions are served to different regions. Each installation package must be customized.

9. Case studies

9.1. Case study 1 Game A

Game A is a global server card game. Currently, the game server is deployed in the Alibaba Cloud US West 1 region. The gameplay is not very sensitive to network latency and a latency below 300 ms does not affect the gaming experience. Therefore, the customer did not plan for any network access optimization or distributed deployment, adopting the fully-centralized deployment architecture.

**Note:** For a better view, please zoom in by 150%.
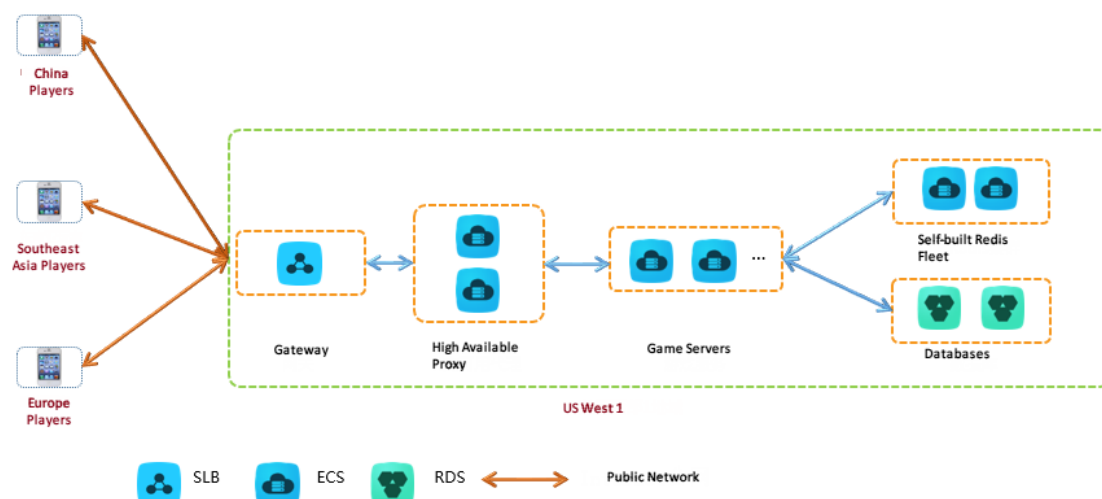
9.2. Case study 2 Game B

Game B is a global real-time multiplayer war game. The game adopted the global server game reference architecture with centralized deployment and network optimization.

The game's access layer, business logic layer, and data layer services are all deployed in the Alibaba Cloud US West 1 region. Chinese players use Alibaba Cloud public network BGP to access Express Connect and connect to the VPC for the US West 1 region. Global Acceleration is deployed in China North 2 and set as the access layer's public network portal, with intelligent DNS used for traffic scheduling. This layout improves the game access speed for players on the Chinese mainland.

Note: For a better view, please zoom in by 150%.
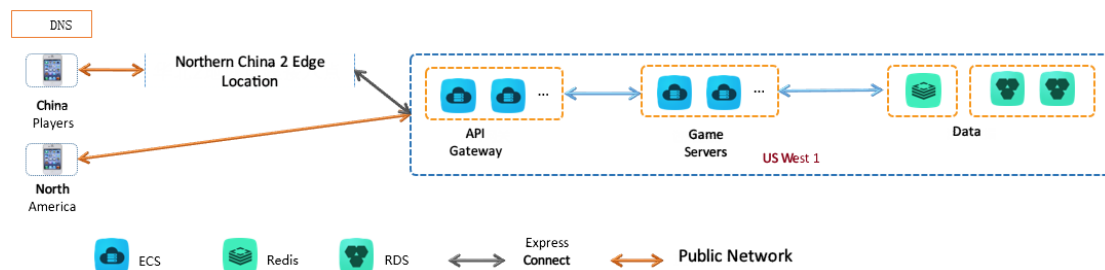
10. Conclusion and future prospects

By describing business needs, technical difficulties, cloud products, player acceleration methods, reference architectures, and typical designs, this article aims to provide complete solutions to more customers who want to develop global server games.

Alibaba Cloud already offers the following technical solutions to address a series of technical difficulties (such as time synchronization, localization, and latency) faced by global server games:

- Distributed deployment cross-region data synchronization

- Global serverless game time (such as GMT)

- Global serverless game localization (such as text, materials, and code)

- Global serverless latency elimination (such as server frame sync)

In future, Alibaba Cloud plans to perfect and provide general solutions and solutions for different global server game architectures:

- Distributed node traffic proxy construction best practices

- Global server SLG game architecture solutions

- Global server card/board game architecture solutions

In a nutshell, Alibaba Cloud's global data centers and Express Connect form a global network that assists the global deployment of games.

11. Presales consultation

Presales consultation contact number (seven days a week, during business hours):95187-1

- Our presales consultation service offers a team of professionals to assist you with the extensive service and support.

- You can also go to the Global Server Solution Website for more information and comprehensive consulting services.

- You can also seek consultation in terms of purchase consulting, configuration recommendations, pricing plans, and other one-on-one services.