Alibaba Cloud Common Solutions

Gaming Solutions

Issue: 20181123

Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.
2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company, or individual in any form or by any means without the prior written consent of Alibaba Cloud.
3. The content of this document may be changed due to product version upgrades, adjustments, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and the updated versions of this document will be occasionally released through Alibaba Cloud-authorized channels. You shall pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.
4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides the document in the context that Alibaba Cloud products and services are provided on an "as is", "with all faults" and "as available" basis. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not bear any liability for any errors or financial losses incurred by any organizations, companies, or individuals arising from their download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, bear responsibility for any indirect, consequential, exemplary, incidental, special, or punitive damages, including lost profits arising from the use or trust in this document, even if Alibaba Cloud has been notified of the possibility of such a loss.
5. By law, all the content of the Alibaba Cloud website, including but not limited to works, products, images, archives, information, materials, website architecture, website graphic layout, and webpage design, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade

secrets. No part of the Alibaba Cloud website, product programs, or content shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates).

6. Please contact Alibaba Cloud directly if you discover any errors in this document.

Generic conventions

Table -1: Style conventions

Style	Description	Example
	This warning information indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	 **Danger:** Resetting will result in the loss of user configuration data.
	This warning information indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	 **Warning:** Restarting will cause business interruption. About 10 minutes are required to restore business.
	This indicates warning information, supplementary instructions, and other content that the user must understand.	 **Note:** Take the necessary precautions to save exported data containing sensitive information.
	This indicates supplemental instructions, best practices, tips, and other content that is good to know for the user.	 **Note:** You can use **Ctrl** + **A** to select all files.
>	Multi-level menu cascade.	**Settings** > **Network** > **Set network type**
Bold	It is used for buttons, menus, page names, and other UI elements.	Click **OK**.
`Courier font`	It is used for commands.	Run the `cd /d C:/windows` command to enter the Windows system folder.
Italics	It is used for parameters and variables.	`bae log list --instanceid` *`Instance_ID`*
[] or [a|b]	It indicates that it is a optional value, and only one item can be selected.	`ipconfig` *`[-all|-t]`*
{} or {a|b}	It indicates that it is a required value, and only one item can be selected.	`swich` *`{stand | slave}`*

Contents

1 Global Game Servers

Industry Overview

The game industry is evolving each day at a magnificent speed. Along with which, the internatio nal 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.

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.

Why Alibaba Cloud

- 1) unified management of resources

  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.

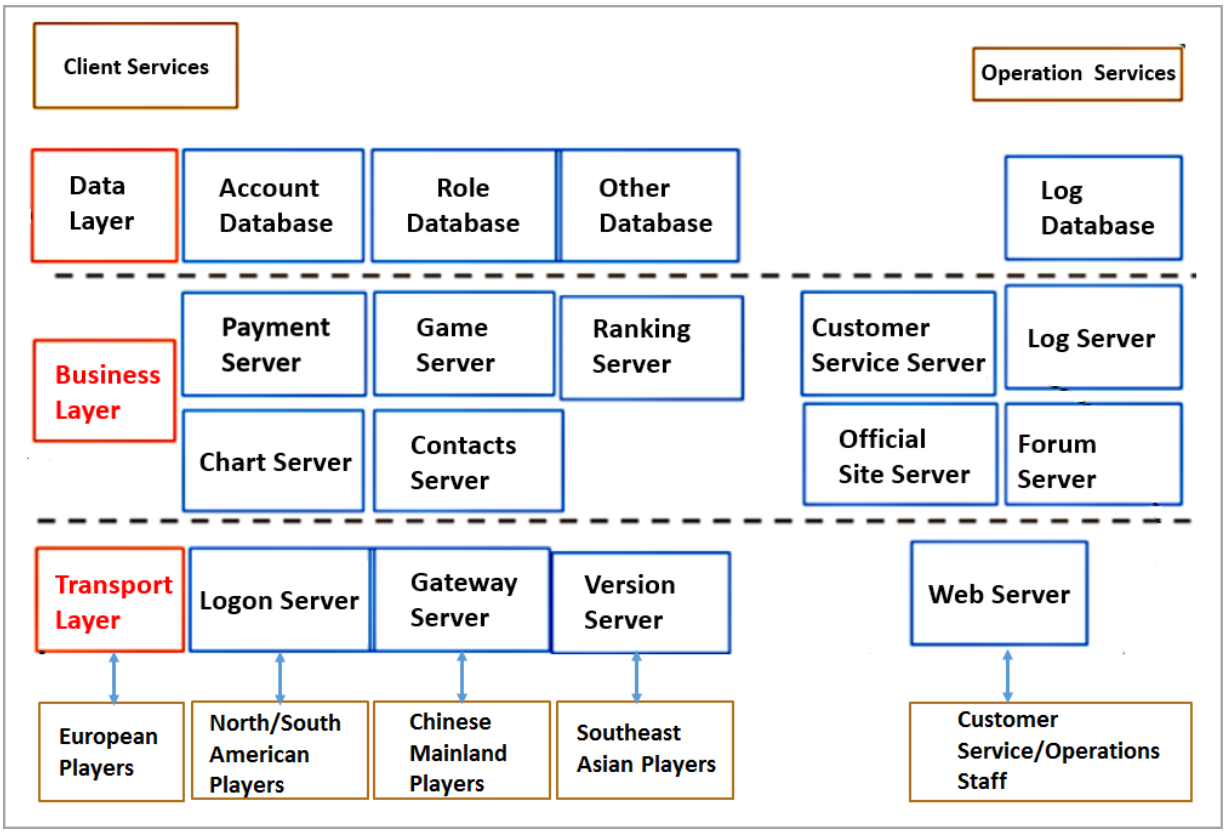
- Comprehensive suite of products

  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.

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.



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.

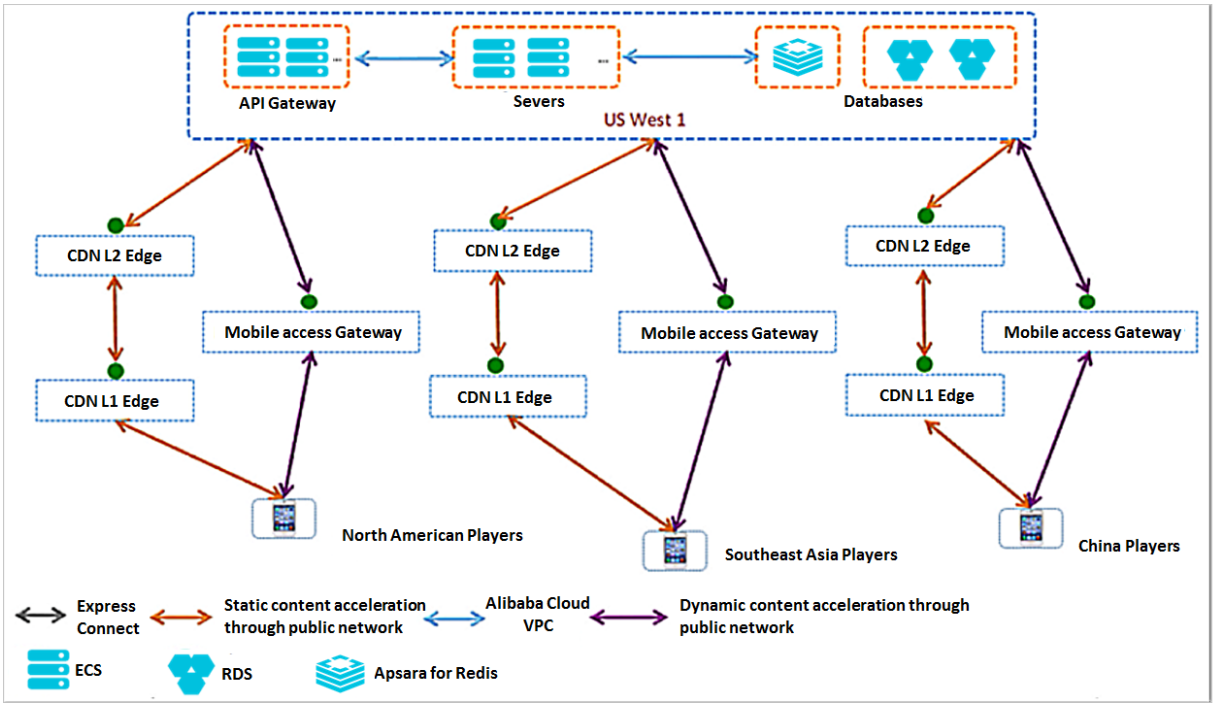
Mobile Accelerator

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.

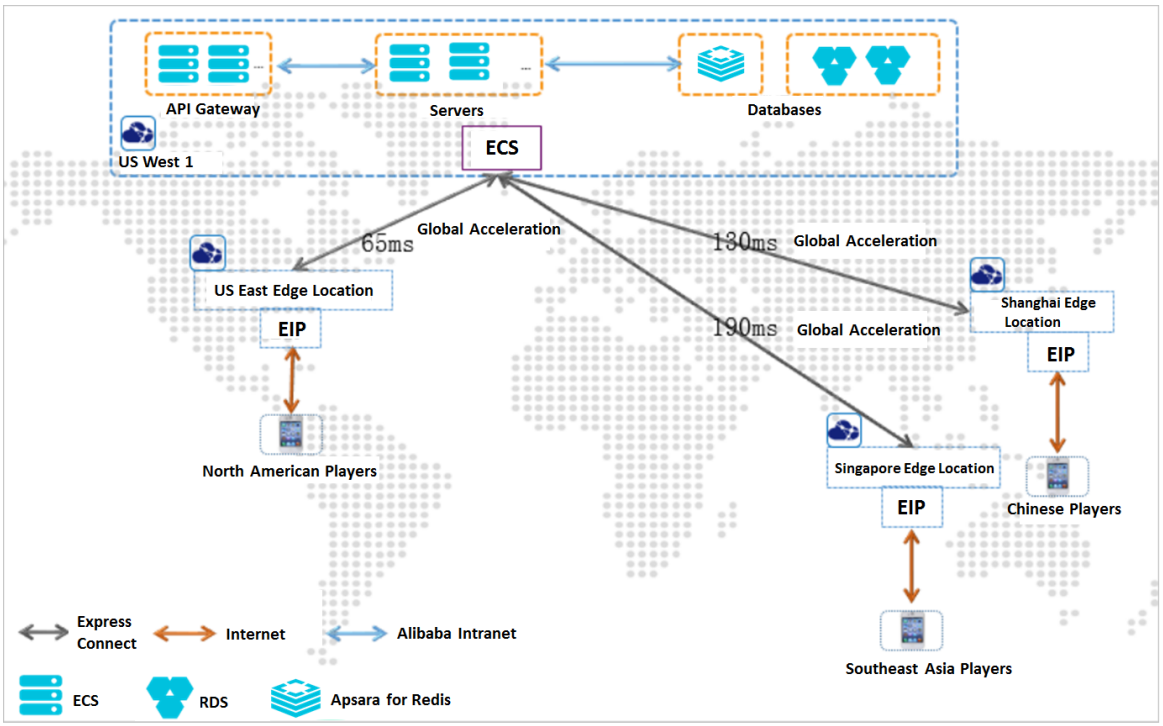
These problems make Alibaba Cloud Mobile Accelerator the most appropriate fit to resolve the “ last mile” acceleration between clients and edge locations. See the following figure.



Global Acceleration

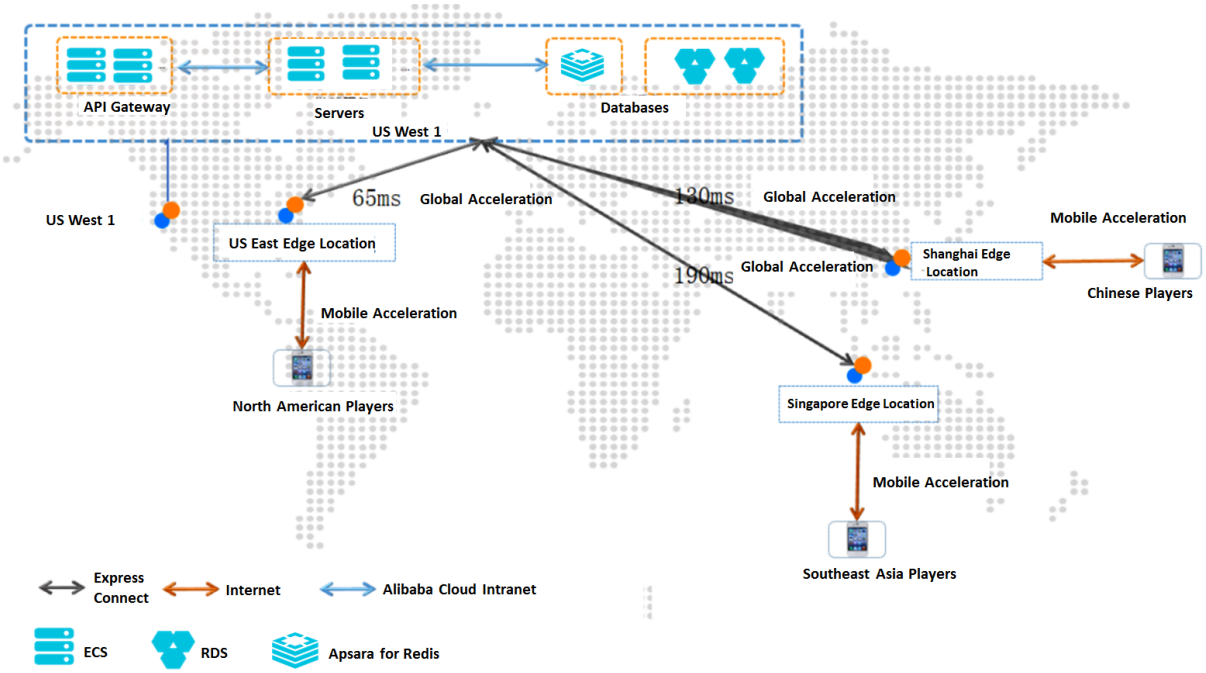
With Alibaba Cloud Global Acceleration, you do not have to manually configure content delivery acceleration to edge locations, which can be complicated, and requires longtime debugging. Global Acceleration ensures you the high availability, scalability, performance, and flexible routing.

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:



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.

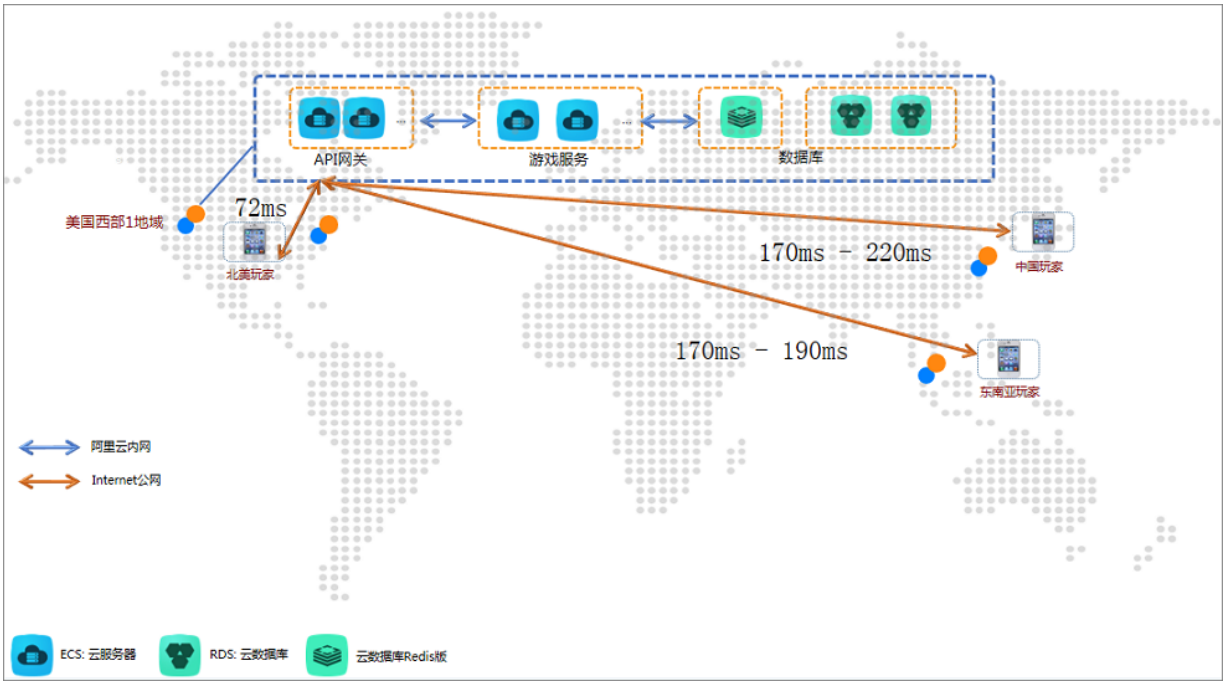


Gaming Architecture

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

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 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.

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.

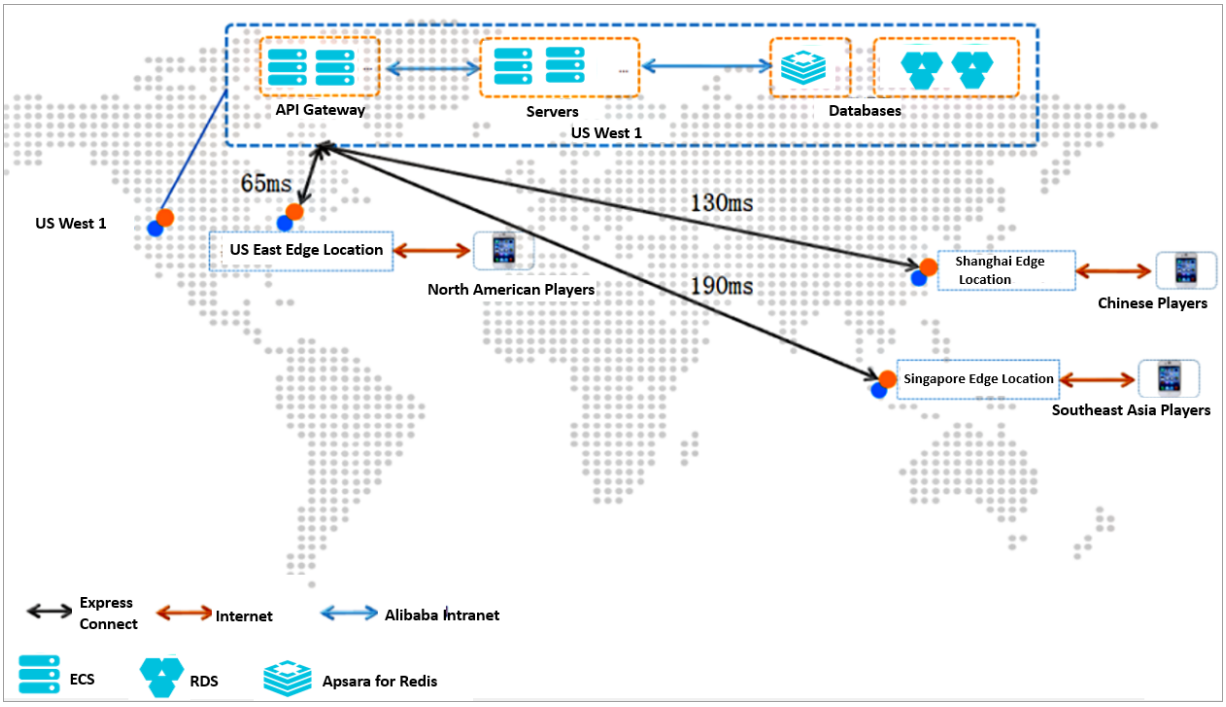
Architecture advantages and disadvantages

- 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.

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:

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.

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.

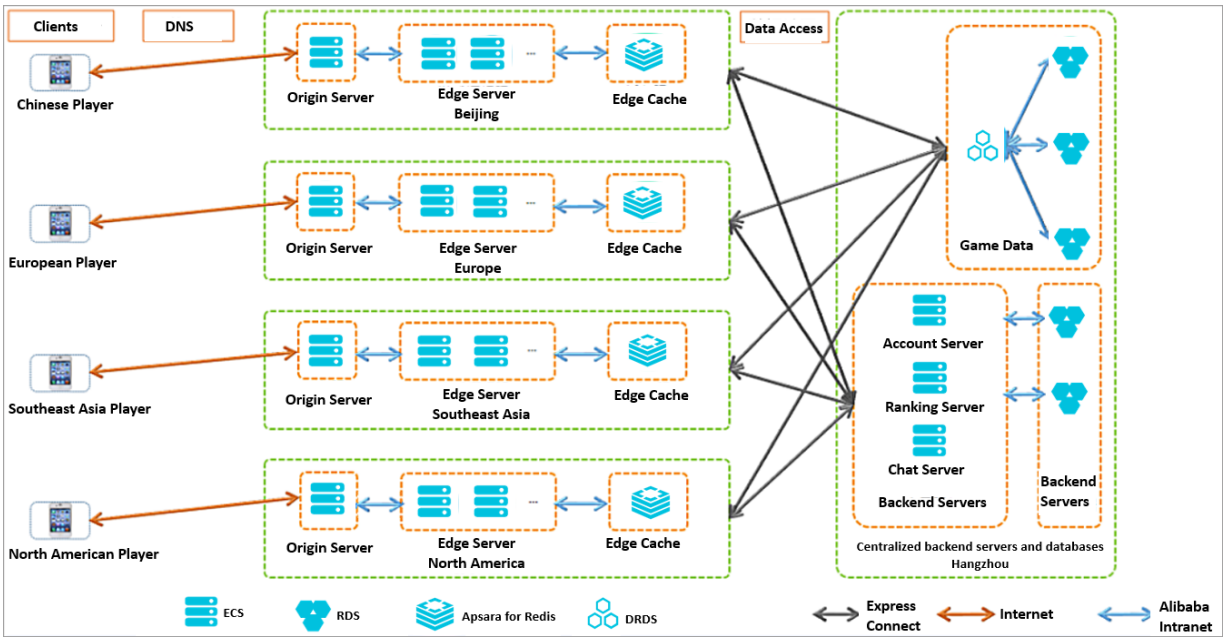
Architecture advantages and disadvantages

- 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).

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:

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 *hfc5 family of high-frequency computing instances* is suitable for MMO games and provides high-frequency specifications.

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.

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

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 intelligen t 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

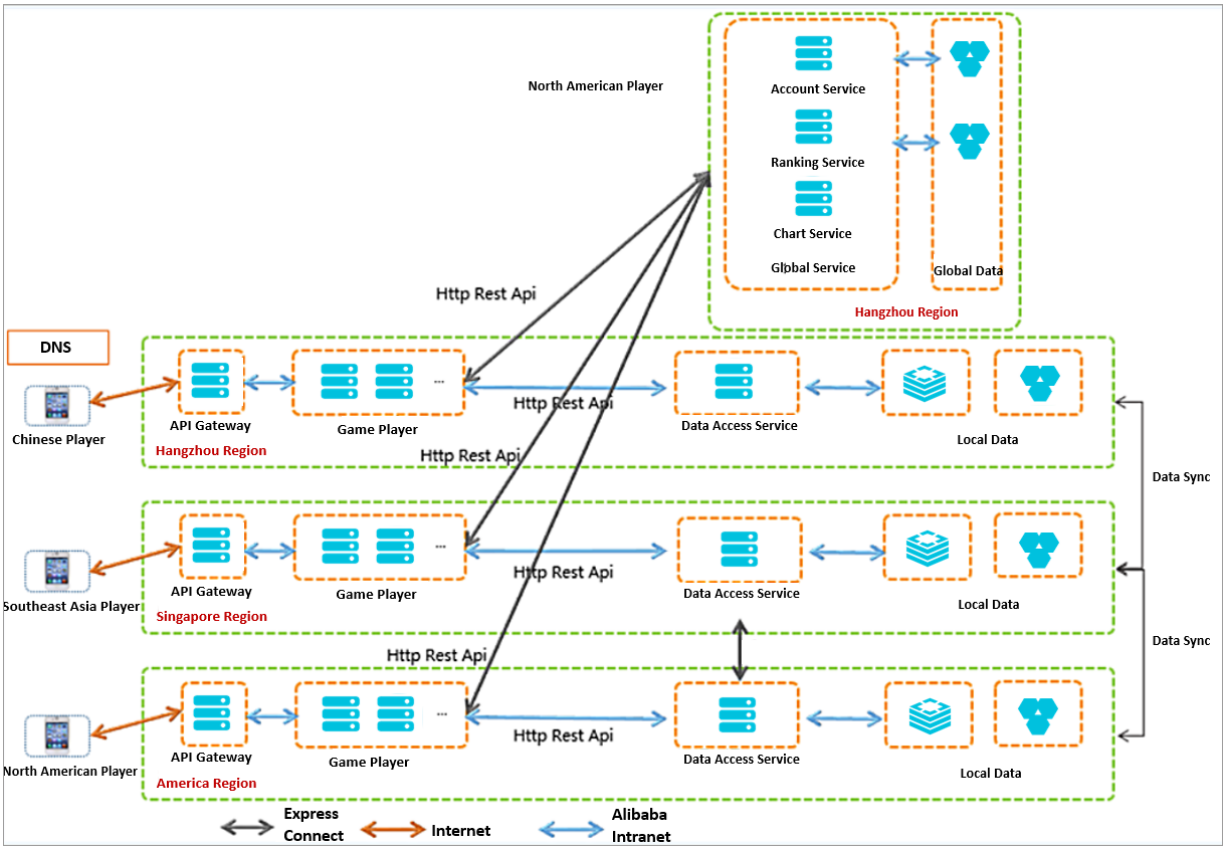
Because rankings and other global data is generated by summarizing data from all regions, this data must be centrally stored. Then, each region can regularly pull the necessary global data (the data pull interval must be set according to the ranking generation cycle). However, the previous data version 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.

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:

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.

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.

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

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.

Cloud product introduction

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. For more information, refer to *ECS*.

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. For more information, refer to *ApsaraDB for RDS*.

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. For more information, refer to *ApsaraDB for Redis*.

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. For more information, refer to *Express Connect*.

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. For more information, refer to *Global Acceleration*.

Cloud resolution

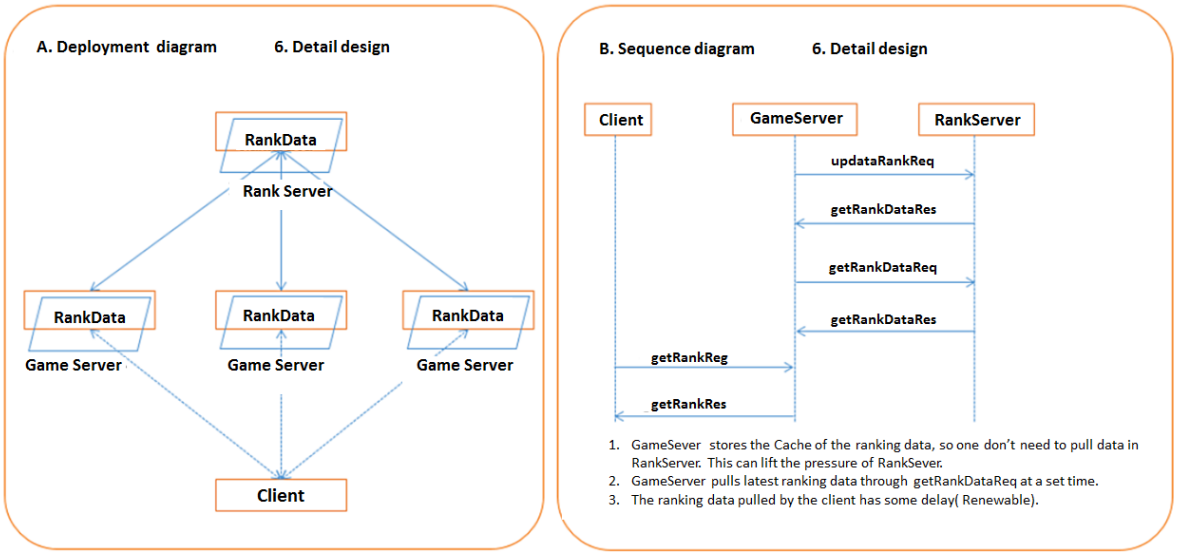
Alibaba Cloud DNS is a cloud computing service portal. It gradually integrates existing Alibaba Cloud products, forming an indispensable element in the cloud product family. ECS, ApsaraDB for RDS, CDN, Server Load Balancer, and other products provide users with efficient and reliable computing, storage, 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.

For more information, refer to *Alibaba Cloud DNS*.

Typical system design

Global rankings design

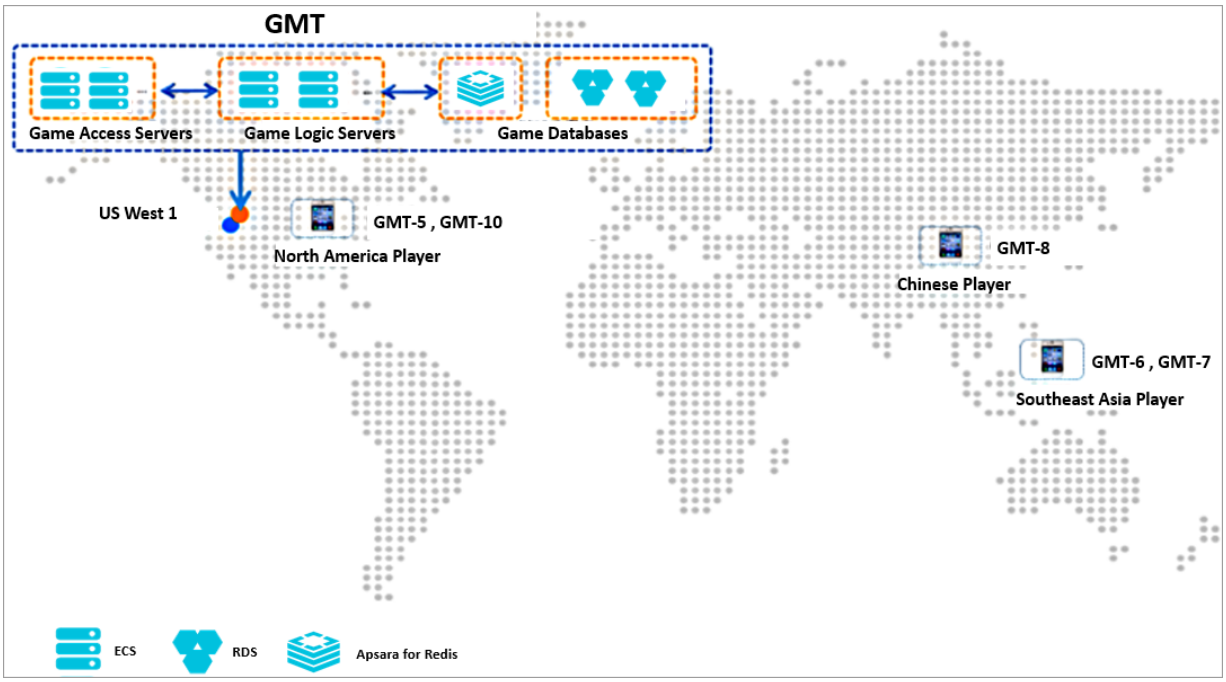
When designing a global rankings service, you must consider demand, analysis, data structure, rank data persistence, rank server SPOF issues, and other issues. A ranking service architecture is shown in the following figure:



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.

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.



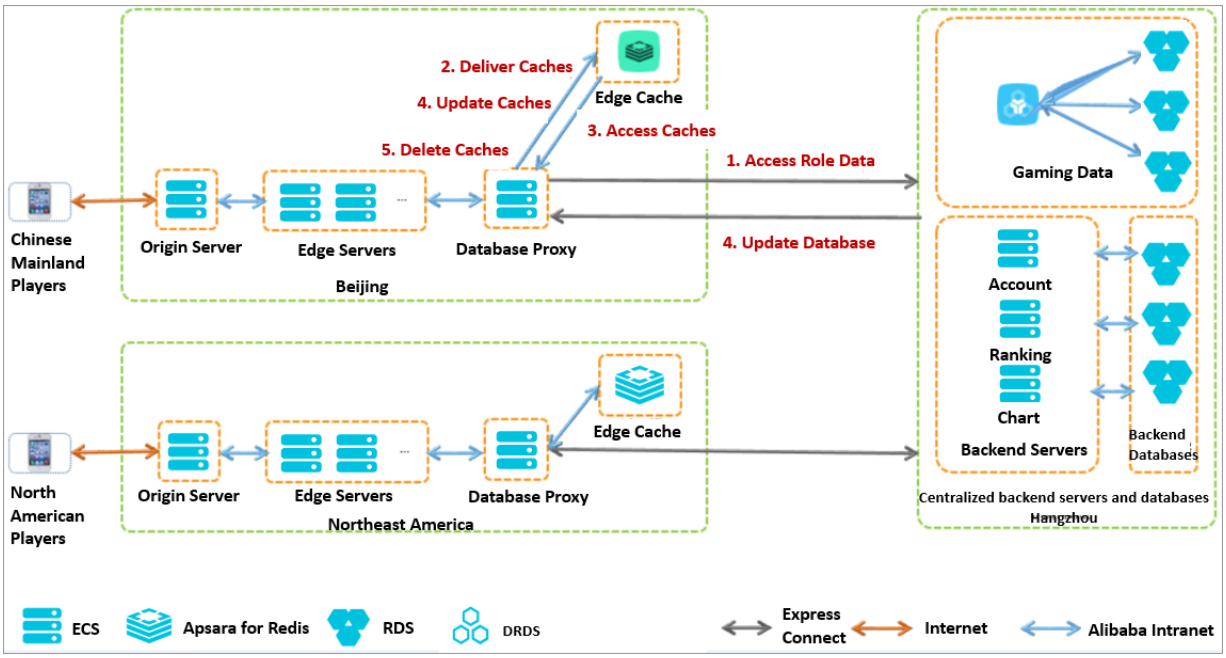
Why is the game time synchronization necessary?

- 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.

Game data synchronization

Cite LeftSolution 1 Use a cache for instant data writebackCite Right

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.



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).

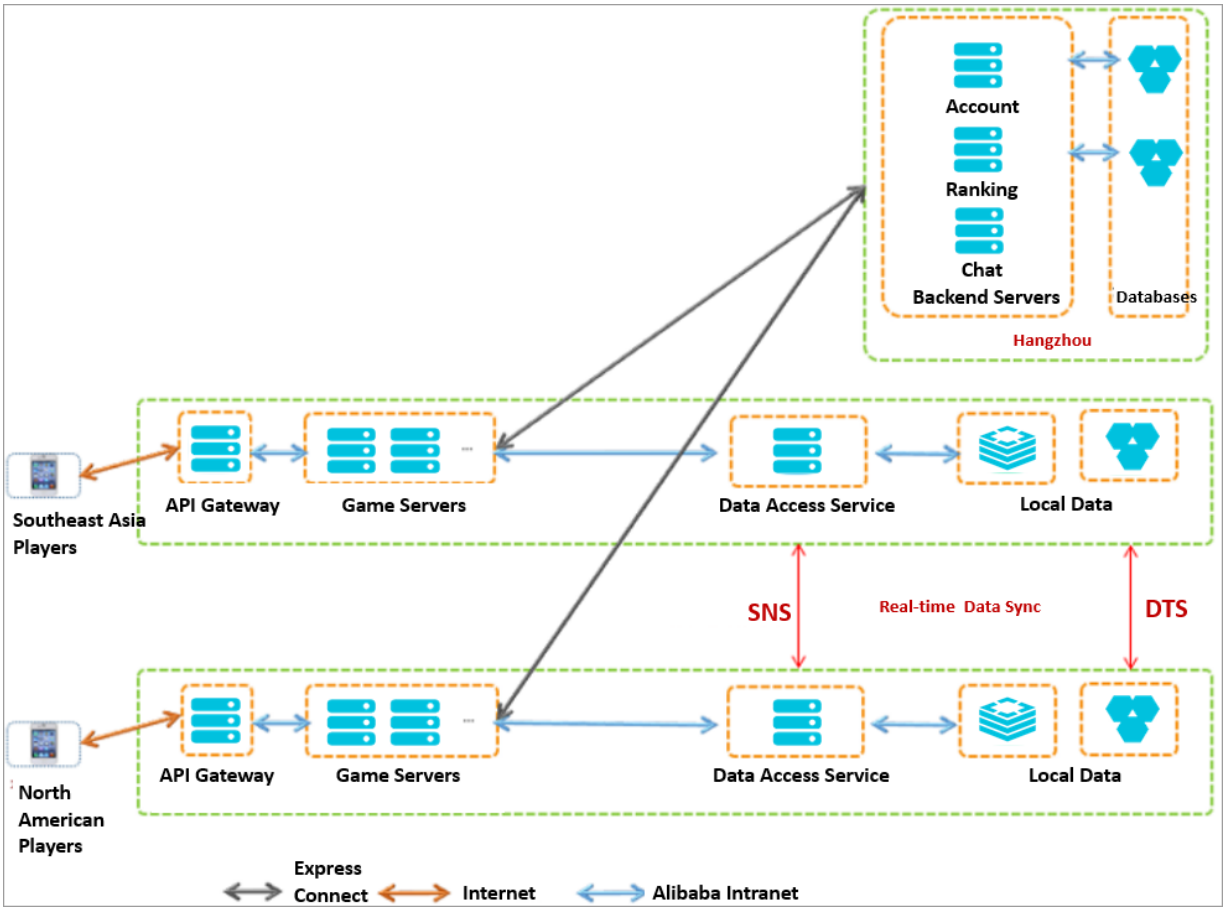
1. The database proxy reads role data from the database.
2. The database proxy inserts the role data into the cache.
3. The role data read from the cache is used in computing.
4. When the data changes, the cache data and database data are updated simultaneously.
5. When players exit the game, the data 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.

Cite LeftSolution 2 Real-time regional database synchronizationCite Right

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.



- Advantages:

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

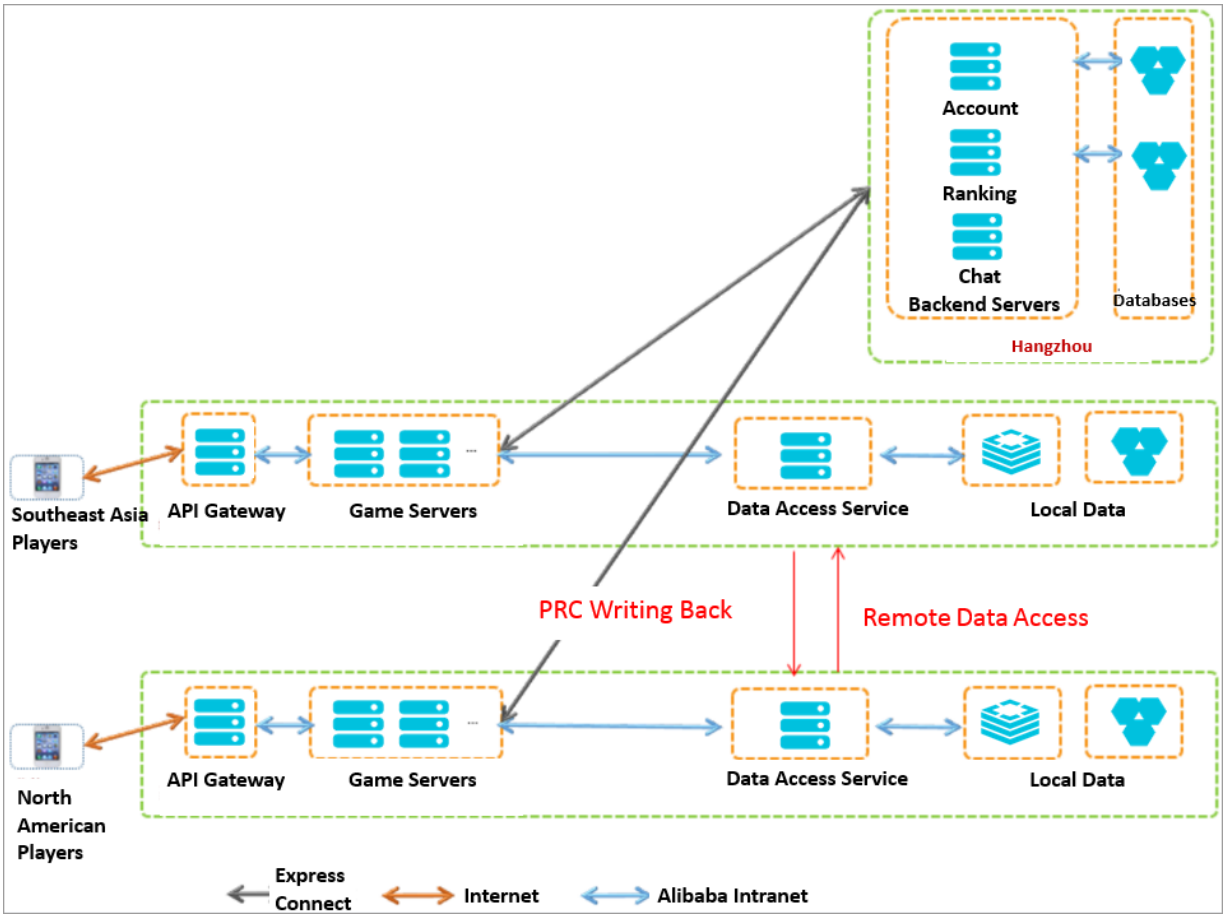
- Disadvantages:

  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.

Cite LeftSolution 3 Cross-region data updateCite Right

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.



- Advantages:

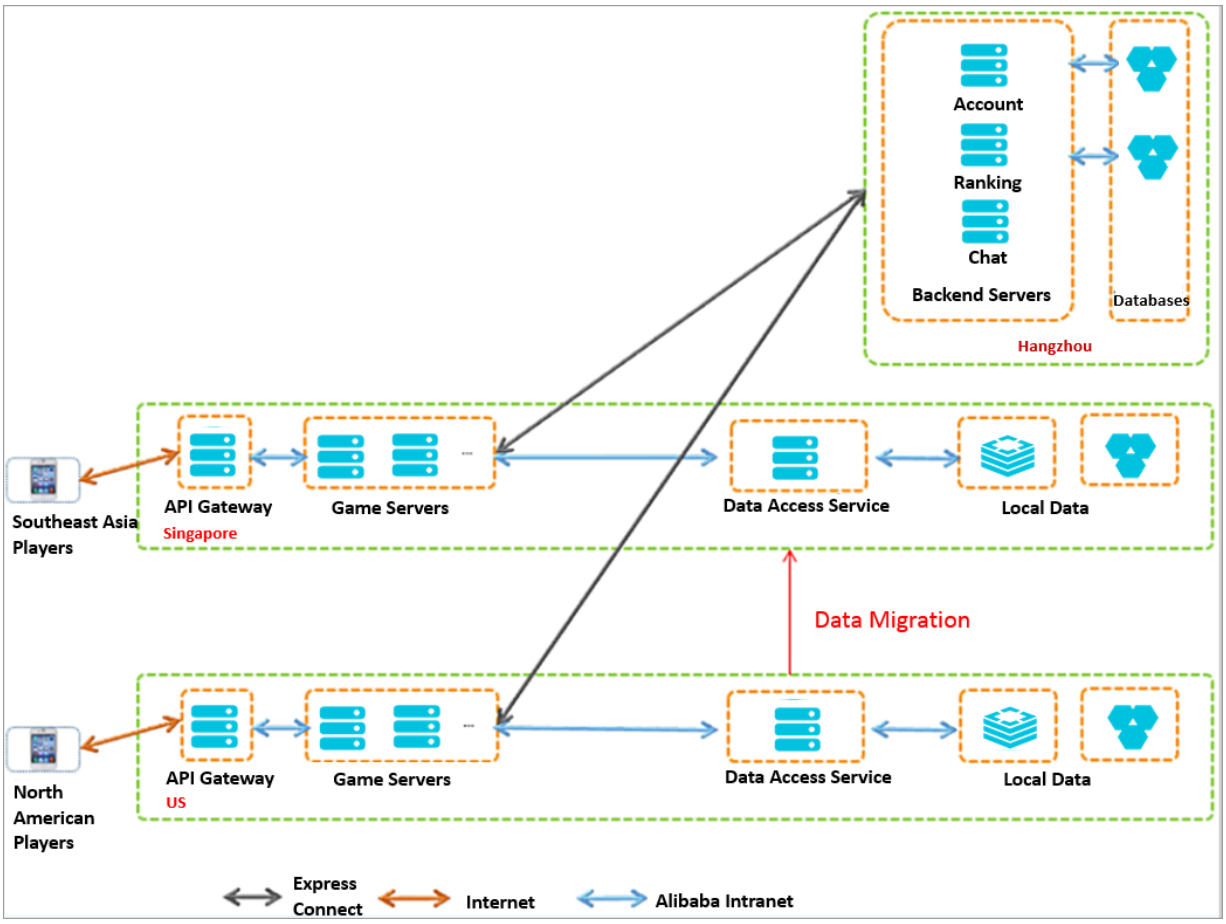
  For players in the same region, role data is read/written to/from the nearest local node.

- 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.

Cite LeftSolution 4 Remote data migrationCite Right

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.



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:

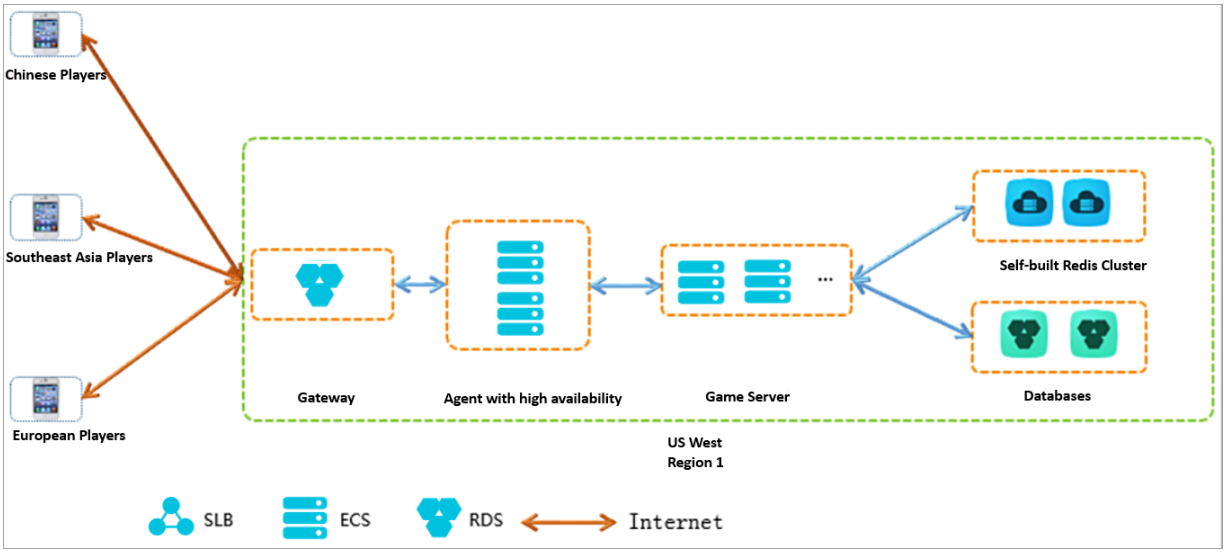
- After the client installation package is downloaded, the user manually sets the language version. Such large packages contain various language packs and support one-click language selection.
- When the game is installed on the client, it detects the language used by the mobile device to dynamically 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.

Case studies

Case study 1 Game A

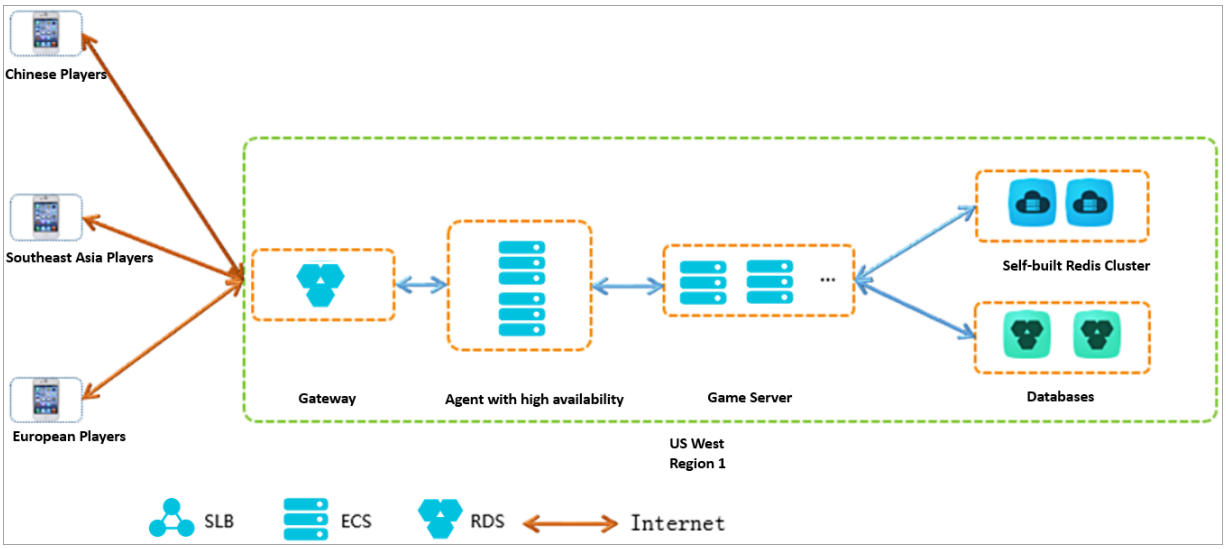
Game A is a global server card game. Currently, the game server is deployed in the Alibaba Cloud US West 1 region. The gameplay is not very sensitive to network latency and a latency below 300 ms does not affect the gaming experience. Therefore, the customer did not plan for any network access optimization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



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.



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.

2 MMO game solution

Industry overview

As the gaming industry expands, casual games and card and board games cannot keep up with gamers' demand for high-quality games. Meanwhile, the rapid increase in mobile terminal hardware configurations allows them to run MMO RPG (Massively Multiplayer Online Role Playing Game), MOBA (Multiplayer Online Battle Arena), and other medium-to-heavyweight games. In this context, more and more MMO games are released.

MMO client products are now entering the mobile game product market. This means that cloud services are tasked with more and more support responsibilities as the backend servers for the games. A reasonable platform architecture is the foundation that ensures stable system operation and business development.

Currently, the homogenization of game products is a serious problem. Game vendors are gradually coming back down to earth after an era of explosive growth. In the future, they must attempt to create higher-quality and globalized games. Only the top vendors and the most powerful newcomers will survive the current upheaval in the industry. Going forward, MMO games will continue to be an extremely important part of the high-quality game market.

Technical challenges

- Large bandwidth and high package volume

  MMO games generally wish to provide the largest field of vision possible, and movement and combat are the core gameplay aspects, which requires mutual real-time visibility between gamers on the same screen. Therefore, a large volume of movement and combat packets must be broadcast within a field of vision.

  In this case, MMO game servers produce a massive amount of communication packets when many gamers are playing simultaneously. Therefore, the access layers of MMO game servers require ample network bandwidth and high network packet throughput.
- Resource auto scaling

  Mobile and webpage-based MMO games are generally light games with time fragmentation. As a result, the industry requires maximum conservation and utilization of game server resources in order to efficiently achieve server activation and combination for MMO game servers. This is especially true for webpage-based games.
- High computing power

For MMO web games, game planners hope to use strong interaction between gamers to attract new gamers. Therefore, they must increase the number of concurrent gamers in individual zones as much as possible. The maximum number of concurrent gamers in individual zones is generally required to be in the thousands. Therefore, MMO web games require strong interactio n and validation as well as high game server computing capability.

- Nearby access

  MMO games generally adopt nearby deployment models based on zone and server division and in multiple centers across different regions. This gives gamers nearby game servers, ensuring smooth gameplay and enhancing the gaming experience.

Why Alibaba Cloud?

- Top-level infrastructure
  - Alibaba Cloud has self-built 5A data centers and 16 major IDCs around the world, with BGP exclusive bandwidth, 1000 gigabit optical fiber access, and over 1200 CDN nodes worldwide.
  - Our solutions leverage Alibaba Cloud’s global network of data centers and Express Connect service to easily implement global server sharing architectures. This forms a global network , with one console able to deploy resources in China and abroad in mere seconds.
- Comprehensive products for support
  - Alibaba Cloud's comprehensive product lines meet the needs of different business scenarios.
  - Our wide range of instance types provide the instance specifications and performance metrics required in various technical scenarios.
- High frequency instances

  Alibaba Cloud provides instances with various frequencies that are able to satisfy the different CPU computing power needs of various scenarios.
  - Our solutions meet the computing speed requirements of complex game logics.
  - Our solutions meet the single-core processing speed requirements of old client game architectures.
- High network throughput capacity

  Alibaba Cloud provides high network throughput ECS instances with a throughput capacity of over one million PPS to meet the needs of various high packet throughput scenarios.

- Our solutions provide the PPS capacity required for message broadcast scenarios for multiple gamers on a single screen.
- Our solutions provide the intranet throughput capacity required for public data reading.

- Stable computing power assurance

  Alibaba Cloud provides dedicated instances to ensure the continuous and stable output of computing power.

  - Our solutions provide the CPU computing power stability required for high load scenarios.
  - Our solutions provide the computing speed assurance required by CPU-intensive scenarios.
- Nearby access worldwide

  Alibaba Cloud has data centers around the world that provide nearby access for gamers and meet the needs of global server sharing infrastructures.

  - Our solutions ensure a good access experience for gamers sharing servers around the world.
  - Our solutions solve the last mile problem for gamer access.
- High quality BGP network

  Alibaba Cloud's multi-line BGP networks ensure gamers experience high-quality network access, solving network access problems for gamers on different carriers.
- Professional security protection capabilities

  Alibaba Cloud draws on the security protection experience accumulated by the Alibaba Group over more than a decade to provide comprehensive security protection solutions that extend from clients, the network layer, and the application layer to the infrastructure layer.

  - We provide defenses against DDoS, CC, and various other types of cyberattacks.
  - We prevent application layer attacks and penetration problems that affect login, payment, and other core business platforms.
  - We use big data analysis to predict security risks.

    Alibaba Cloud's Anti-DDoS Service Pro is the most powerful game security protection system in the industry. It can easily defend against DDoS attacks of hundreds of gigabytes and various application layer attacks.
- Efficient O&M support system

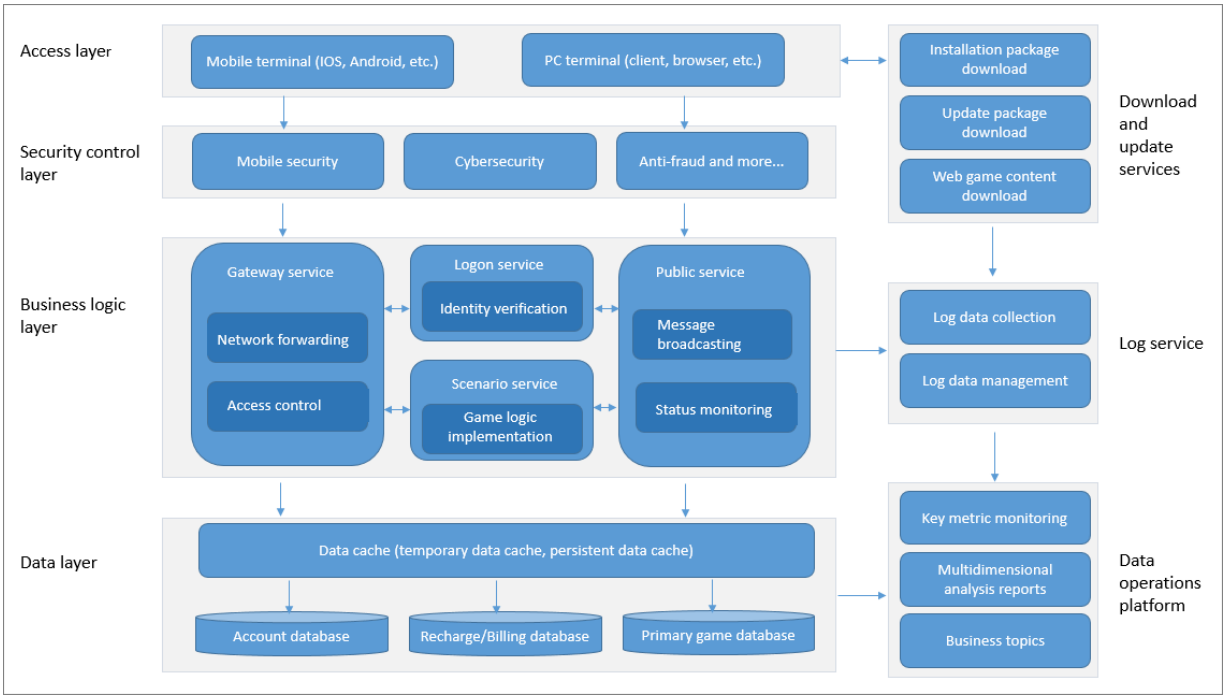
  - Alibaba Cloud's wide range of product APIs greatly increase customer's O&M flexibility and efficiency.

- Our comprehensive monitoring and alert system effectively reduces the difficulty of O&M monitoring.

- High-quality service experience
  - We provide a unified presales and aftersales service system worldwide along with local service support.
  - Our professional technical support service ensures rapid response and efficient problem resolution.

Business function architecture

The overall business process and specific business modules of MMO games are shown in the following figure:

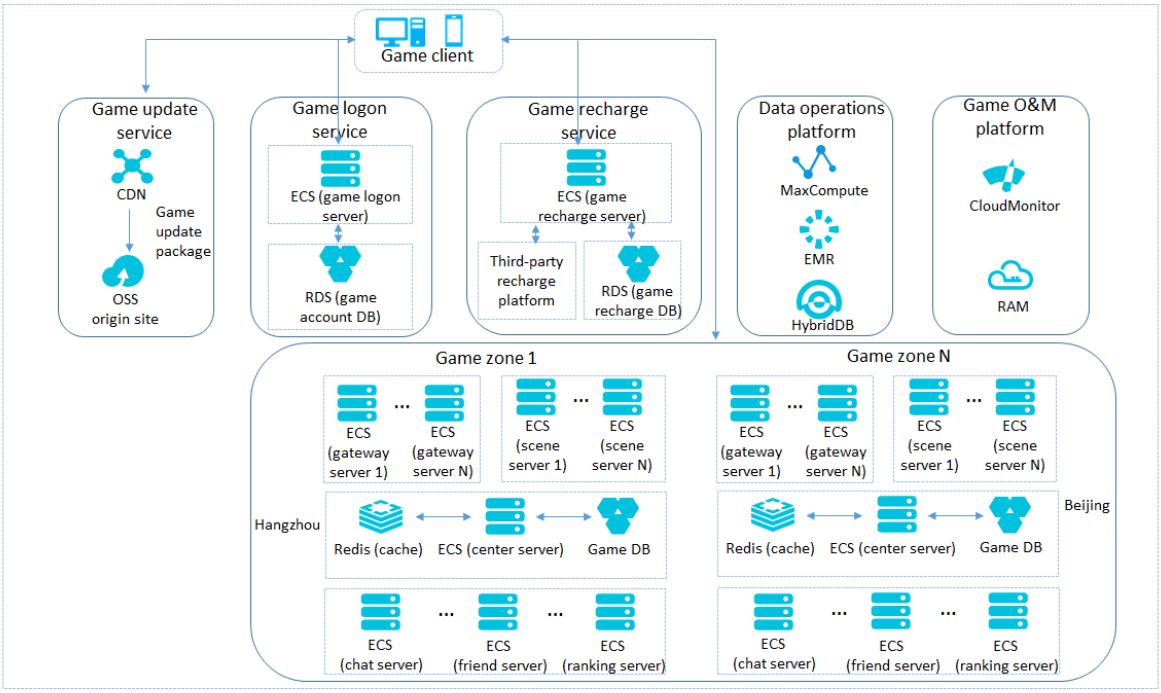


Product architecture

General reference architecture solution of MMO games

Alibaba Cloud's complete product lines provide comprehensive solutions for all scenarios, from game downloads and updates, game business servers, game logic servers, game database servers, and game data operations platforms to game O&M monitoring platforms.

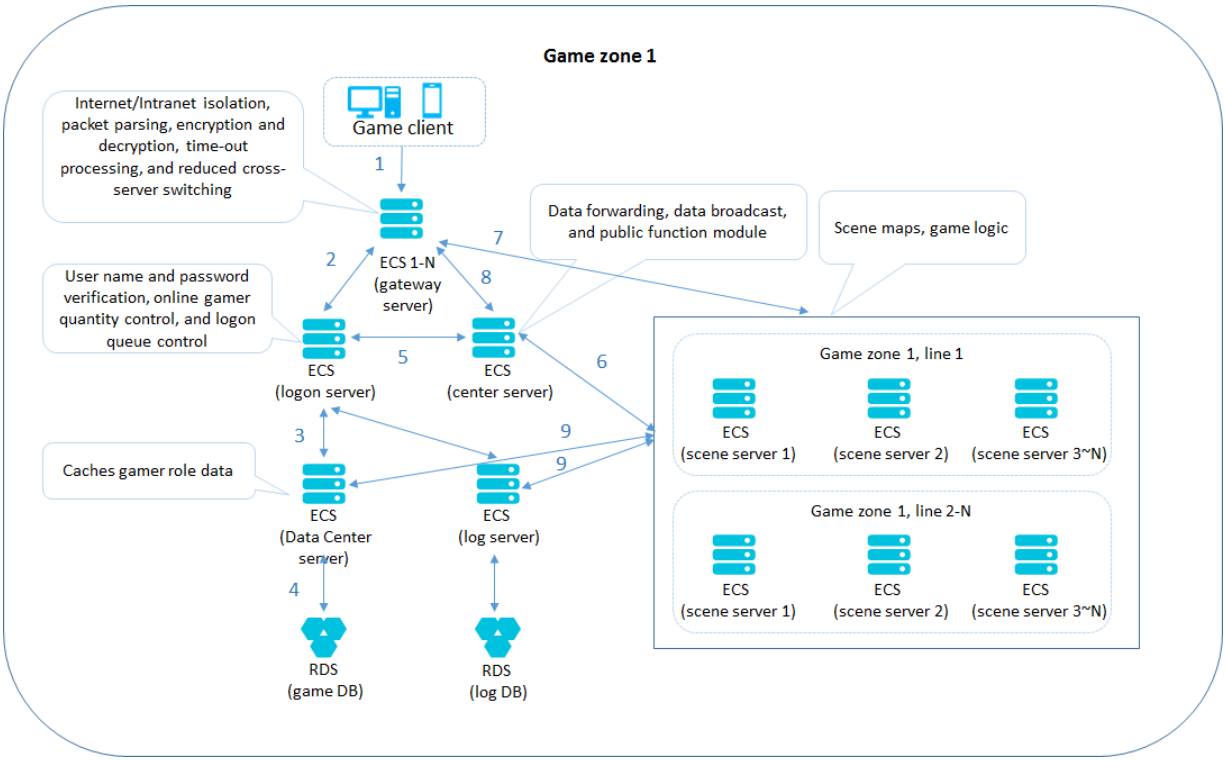
Using rational service combinations and resource configurations, you can effectively improve O&M efficiency, enhance service experiences, and reduce total operating costs.



NOTE: The instances shown in this figure do not represent the actual quantity. Each role server can use Server Load Balancer to implement HA and horizontal scaling.

MMO game architecture - Client games

You can use Alibaba Cloud's different products to implement all business processes or only specific processes. An MMO client game architecture is shown in the following figure:



Data interaction process

See the figure shown in the MMO game architecture - Client games part. The following list describes the specific data interactions numbered in the figure:

1. The client connects to the gate server to initiate a login request.
2. The gate server forwards the login request to the login server.
3. The login server initiates an ID data verification query to the DC server.
4. The DC server accesses the gameDB server to perform the data query and return the results.
5. If ID verification is successful, the login server continues to query data and returns account status data (role, level, attributes, last login scene server, geolocation, and other information). It also synchronizes login status and information to the center server.
6. The center server is responsible for distributing information to the corresponding scene server. At the same time, a gamer online notification is broadcast to the gamer's friends and the online gamer status monitor (to control reconnection after disconnection and disconnection time-out).
7. After the gate server receives authentication information, it establishes a connection with the scene server, and the gamer successfully logs on to the scene server.
8. If the gamer has public information to broadcast, a request is sent to the center server. The center server performs message packet broadcast.
9. The scene server starts writing all user behavior to the log. At the same time, the relevant gamer data storage or query request is sent to the DC server.

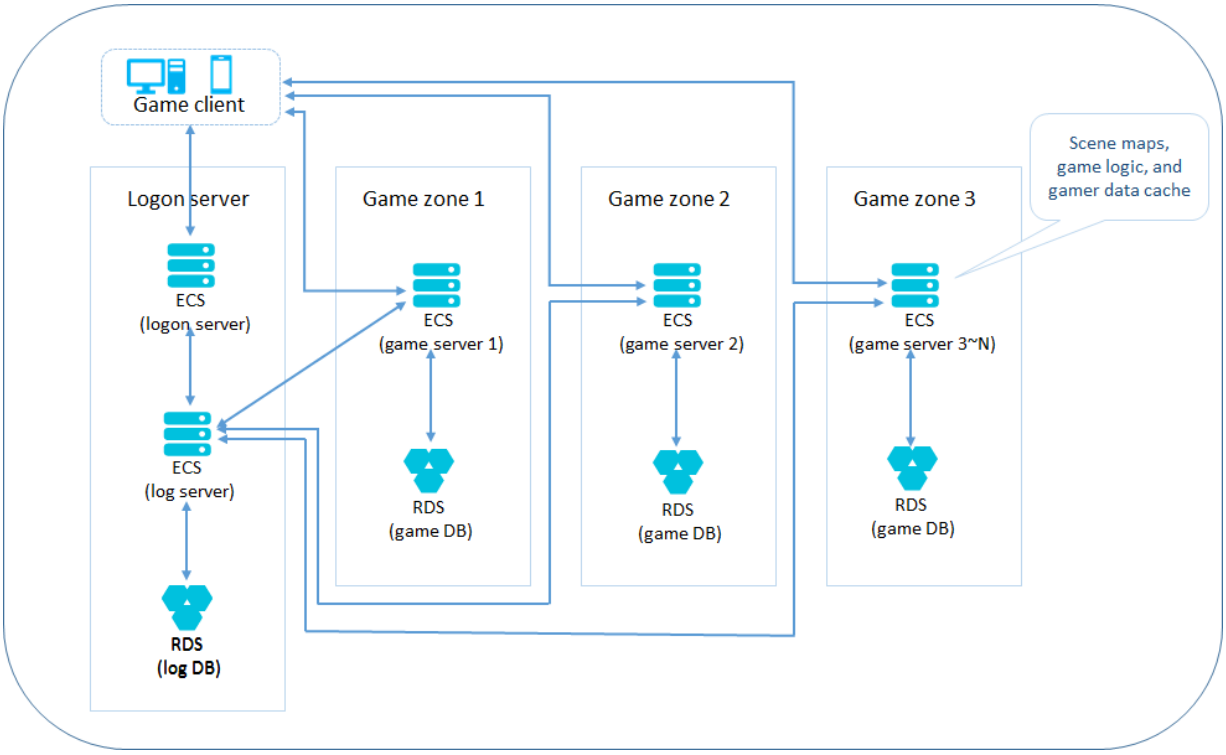
Architecture features

MMO client game architectures have the following main features:

- The gateway server is responsible for all network packet forwarding. Generally, the network load is concentrated here, so it has high network throughput requirements.
- The scene server contains the game logic and is relatively dependent on CPU power and requires a certain level of network packet forwarding capabilities.
- A single game zone serves over 10,000 gamers. The logic server generally divides gamers according to scene maps.Larger scales can be achieved by line division.
- The DC server caches gamer data and writes it to the database asynchronously. This ensures gamer clients can rapidly read and write data. It has high availability requirements, so it must use the application layer to implement data error tolerance mechanisms.
- The log server collects and processes all service behavior logs for a region. It has high disk write performance requirements. Generally, its function is implemented by grouping multiple servers.

MMO game architecture - Mobile games

You can use Alibaba Cloud's different products to implement all business processes or only specific processes. An MMO mobile game architecture is shown in the following figure:



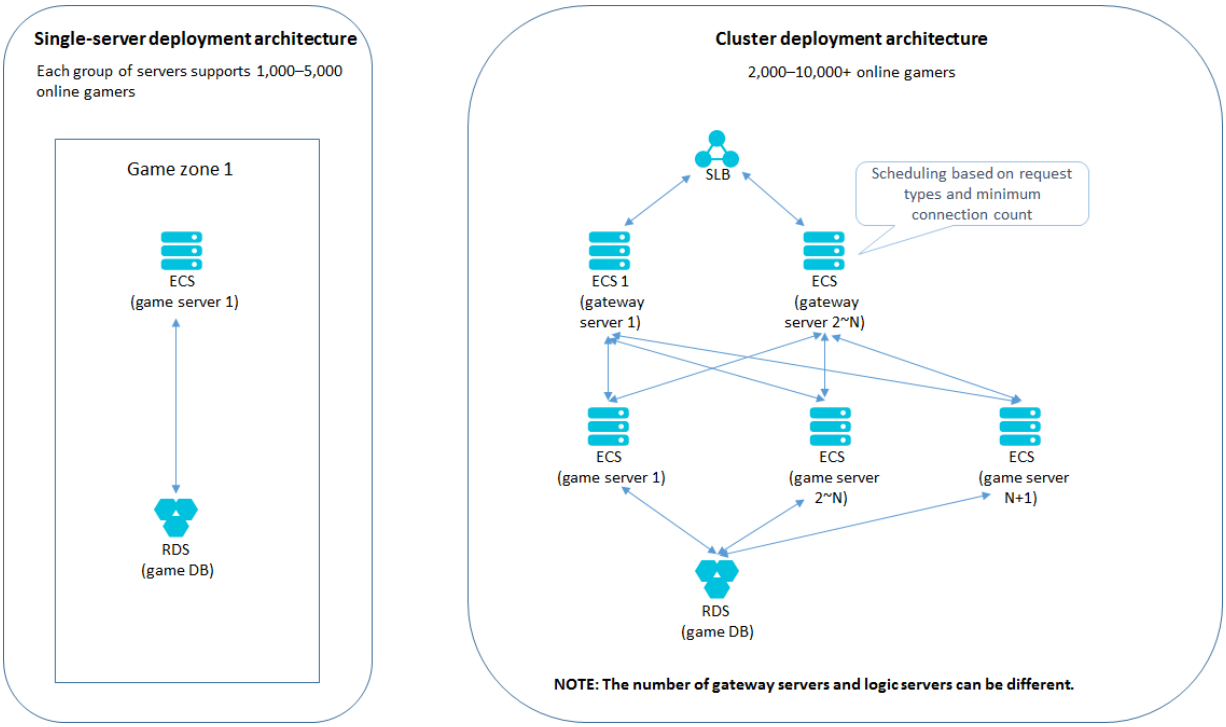
Architecture features

MMO mobile game architectures have the following main features:

- Compared to client games, mobile games have less complex gameplay and shorter lifecycles . Considering this feature together with operator policies and the resource economics model, mobile games usually adopt relatively simple deployment architectures. However, a minority of MMO mobile games use the client game deployment architecture.
- Clients generally connect directly to the game server. For a small number of games, a gateway is configured at the frontend of the game server or the gateway and game server are deployed on the same machine. Mobile games are relatively dependent on the CPU power and network packet forwarding capabilities of individual servers. A single game zone generally supports 1, 000 - 5,000 online gamers.
- A dedicated game database server can be deployed for each game server, or a single game database server can be deployed for multiple zones.

Derivative architecture of MMO mobile games

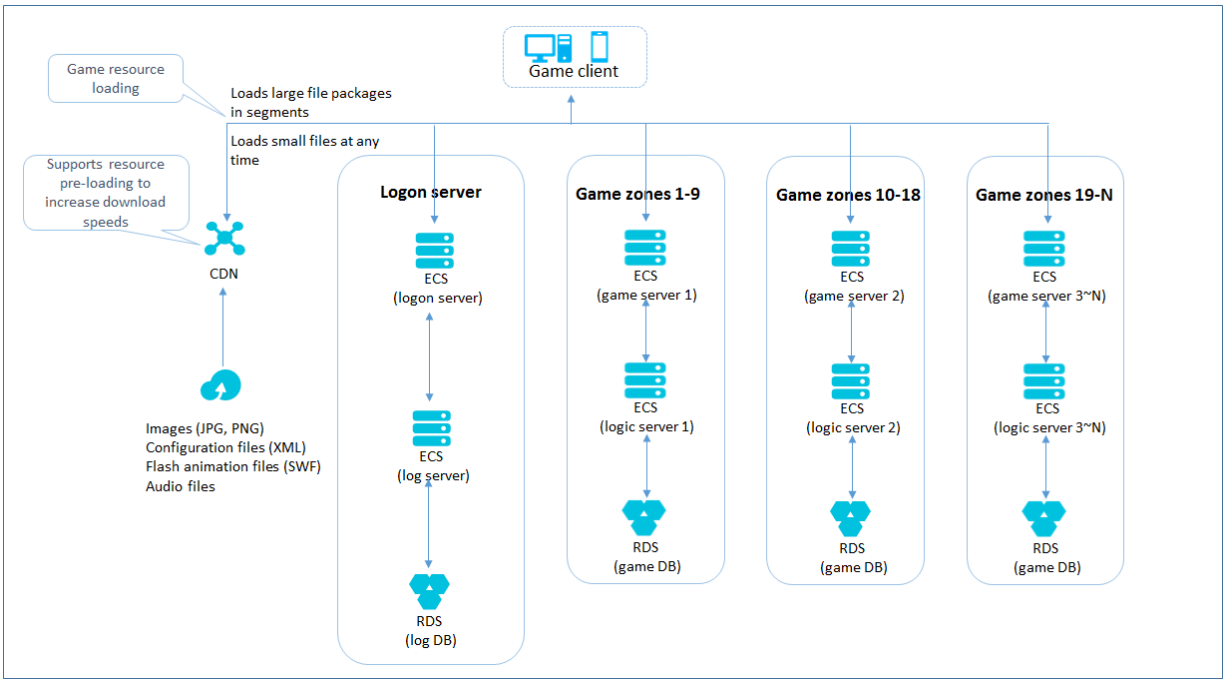
A detailed derivative architecture of MMO mobile games is shown in the following figure:



- In the derivative architecture of MMO mobile games, Server Load Balancer and gateway servers are used to organize the multiple logic servers into a large server. This server can support more online gamers to help build a gamer ecosystem.
- The gateway server implements gamer request scheduling and connection status monitoring.
- Battle servers (or PVP battlefield servers) require greater computing power and load capabiliti es. These servers can be constructed with high-configuration or dedicated instances, and controlled centrally by the gateway server.

Centralized deployment architecture of MMO web games

When the logic server layer and game database layer adopt a centralized deployment method, the specific architecture is shown in the following figure:



- Through the web game process, content loading relies on CDN. Therefore, the game requires a certain level of CDN stability and download speed. In resource loading, large file packages are generally loaded in segments, while small files can be loaded at any time.
- Web games generally adopt the operating policy of rapid server activation and combination , so that they can activate thousands of game zones in a short period of time and gradually combine servers as the number of active gamers in each zone falls. Using custom images and automatic server activation scripts, they can quickly create game regions and logic servers.
- Compared to client and mobile games, web games have low computing complexity. A single game zone does not require any special levels of server computing power or network throughput. Generally, a single server is used to activate multiple zones. The computing load determines how many game zones can run on each server.

MMO page-tour partition deployment Schema

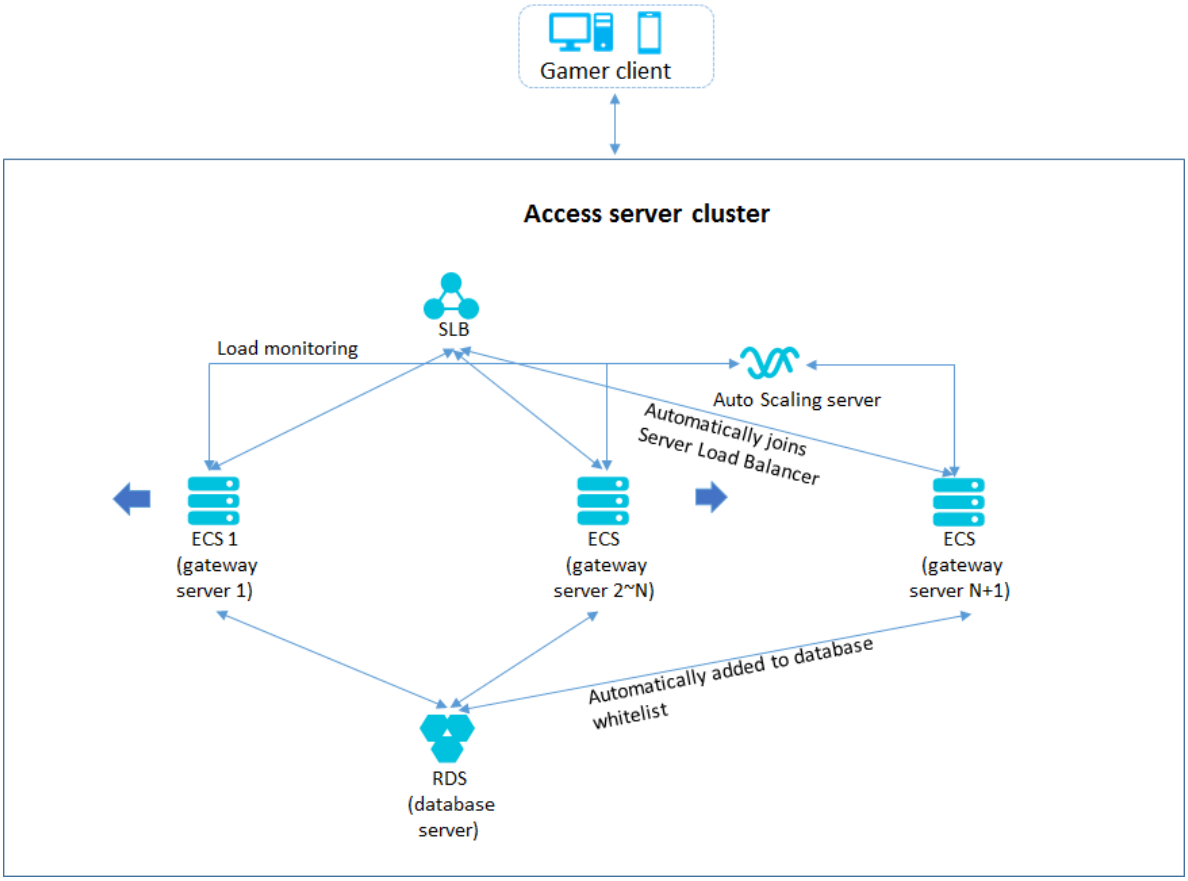
When the logic server layer and game database layer adopt a divisional deployment method, the specific architecture is shown in the following figure:

General technology solution of MMO games

Elastic scaling up & down for the access layer server cluster

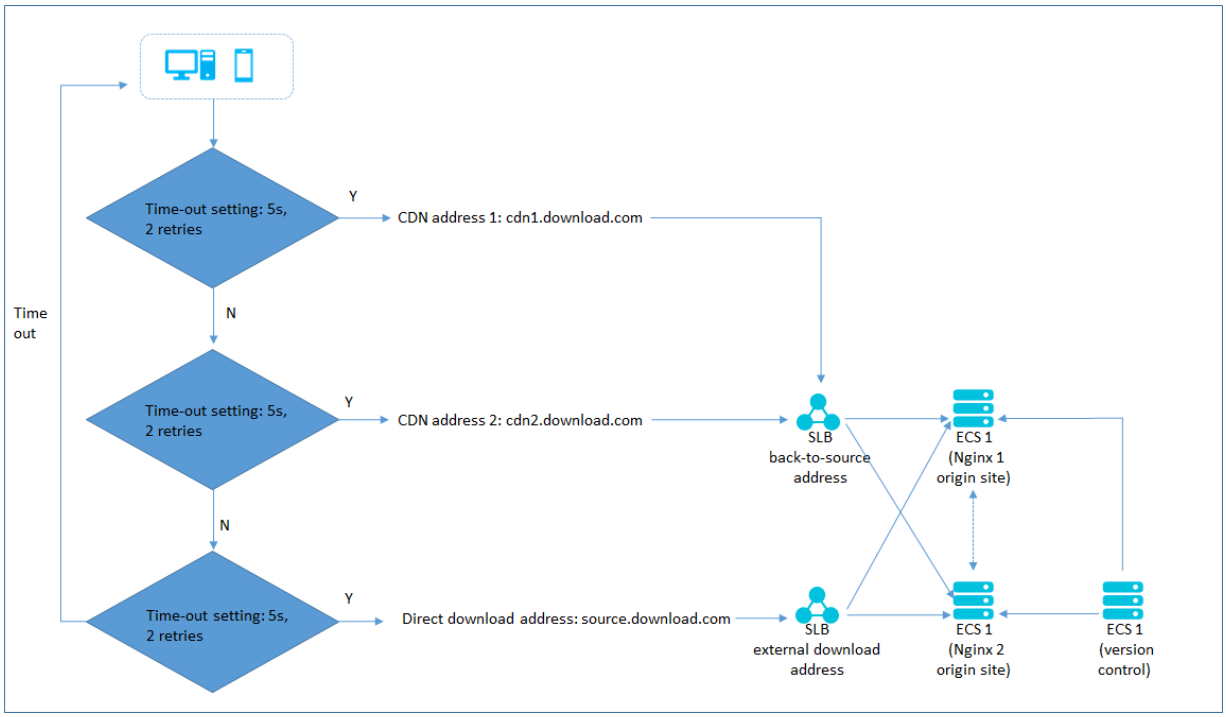
- Using Auto Scaling, you can automatically scale up and scale down your access layer server cluster. This allows you to effectively cope with boot storms, battle event traffic peaks, and other scenarios, and ensure your server cluster resources have sufficient load capabilities.

- This solution also applies to login servers and other servers that require elastic scaling.



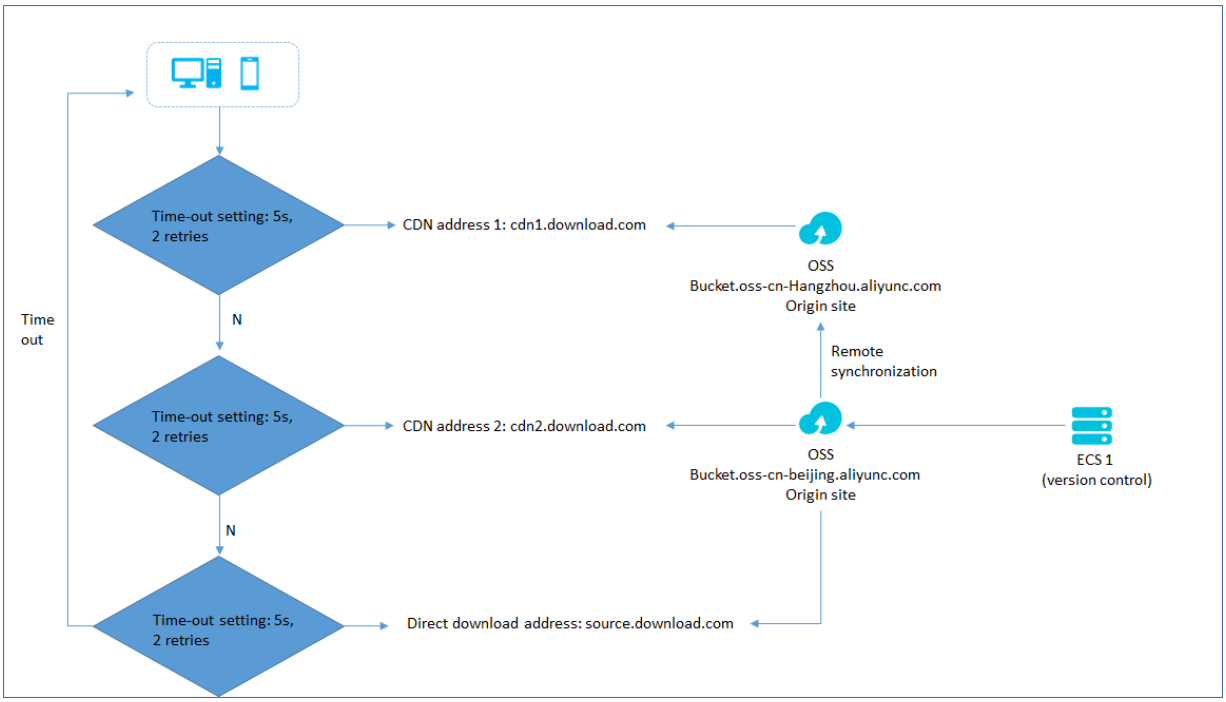
High availability of game downloads and updates - Self-built ECS origin site

- A multi-level download retry mechanism ensures the high availability of downloads and updates, reducing the proportion of gamers lost during this stage.
- By splitting the back-to-source address and direct external download address, this solution avoids potential security risks from exposed addresses and Server Load Balancer unavailability .
- You can use rsync+inotify to synchronize files across multiple origin site servers in real time.



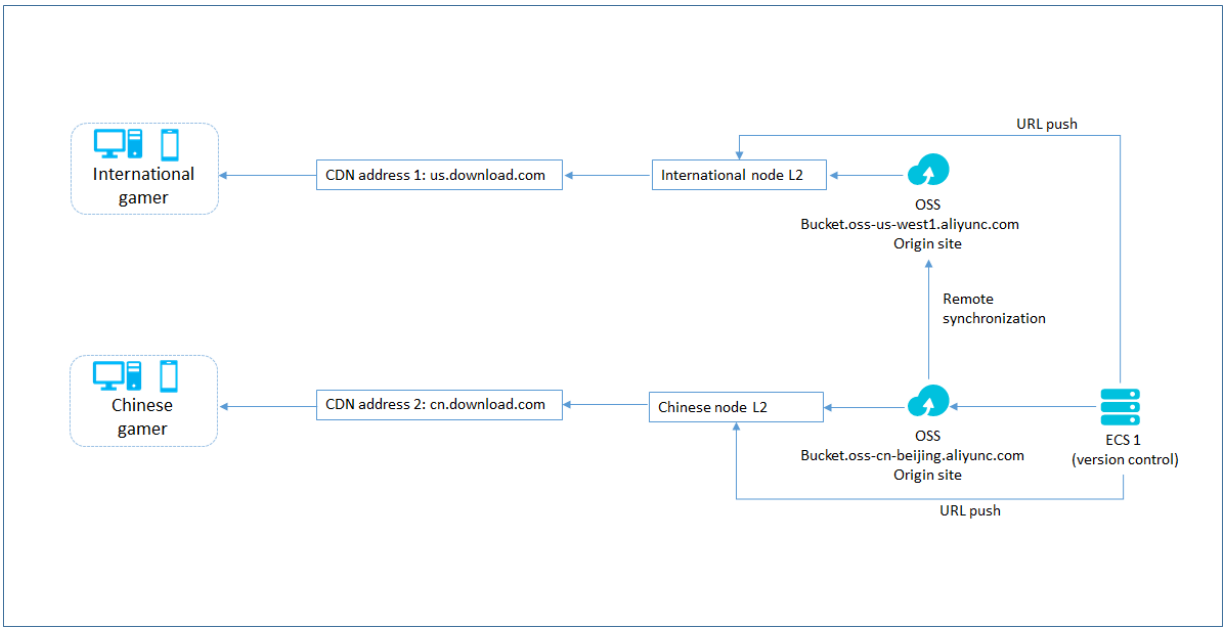
High availability of game downloads and updates - OSS origin site

- A multi-level download retry mechanism ensures the high availability of downloads and updates, reducing the proportion of gamers lost during this stage.
- By splitting the back-to-source address and direct external download address, this solution avoids potential security risks from exposed addresses.
- By setting OSS as the CDN origin site and taking advantage of OSS's automatic remote replication function, it further improves origin site availability and throughput.



Large game file downloads & foreign back-to-source - OSS origin site

- By deploying origin sites in China and abroad, this solution ensures high-speed and stable back-to-source operations.
- OSS's cross-region replication function automatically synchronizes origin site files.
- The URL push function pushes large files to L2 nodes, accelerating the speed of the first download and reducing the number of back-to-source requests.



Cloud product introduction

ECS

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. You can release the resources when you do not need them, so that you can save your cost.

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 operating entity presented to each user.

For more information, refer to *ECS*.

RDS

Alibaba Cloud Relational Database Service (RDS) is a stable and reliable online database service providing the auto scaling capacity.

Based on Alibaba Cloud's distributed file system and high-performance storage, 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.

For more information, refer to *RDS*.

Redis

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

For more information, refer to *Redis*.

MaxCompute

MaxCompute is a big data processing platform that processes and stores massive batch structural data to provide effective data warehousing solutions and big data modeling. MaxCompute provides comprehensive data import solutions and a variety of typical distributed computing

models, enabling you to solve the massive data computing problem in a fast way, effectively reduce costs, and ensure data security.

MaxCompute is mainly used to store and compute batches of structured data. It provides data warehouse solutions for massive volumes of data, as well as big data analysis and modeling services.

MaxCompute aims to provide a convenient means of analyzing and processing massive data volumes. You no longer need to concern yourself with the details of distributed computing. Rather , you can directly and conveniently achieve your data analysis goals.

MaxCompute is already widely used within the Alibaba Group, in applications such as data warehousing and BI analysis for large-scale Internet companies, log analysis for websites, transaction analysis for e-commerce websites, and mining of user features and interests.

For more information, refer to *MaxCompute*.

CloudMonitor

CloudMonitor is a service that monitors Alibaba Cloud resources and Internet applications. The service can be used to collect monitoring metrics for Alibaba Cloud resources, to test the availability of Internet services, and to set alarms for these metrics.

For more information, refer to *CloudMonitor*.

Use cases

Dawn of War, a mobile fantasy MMORPG independently developed by ZPLAY, is set in a world based on European mythology. The game is rich in content, with 3D mounts, costumes and accessories, and other distinctive gameplay systems, as well as a 360° view.

This game used Alibaba Cloud's cloudification solution to form a game server architecture that can effectively avoid single point of failure, support smooth resizing, and allow for flexible deployment . The high-performance ECS delivers the high computing power required by intensive game logic. The ApsaraDB for RDS read-only instances effectively support read/write splitting for MMO games .

Summary and outlook

MMO games have always received a high degree of attention from the industry. Since the days of physical IDCs to the current cloud platform era, the MMO client games have evolved into mobile and web games, presenting developers with a host of technical challenges and opportunities for innovation.

This document provides a comprehensive introduction to the MMO product architecture, discussing the general reference architecture of MMO games along with MMO client game architectures, MMO mobile game architectures, and MMO web game deployment architectures. It also describes certain general technical solutions for MMO games, and briefly discusses the cloud products used in these solutions. Although this document does not discuss every scenario and issue, we hope that it will inspire customers to learn more.

As the industry develops and technology advances, Alibaba Cloud will continue to refine its various basic services, network services, security services, and big data services, providing customers with simpler, more stable, and more complete solutions.