阿里云
通用解决方案

数据库解决方案

文档版本：20181213

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的”现状“、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含”阿里云”、Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 禁止： 重置操作将丢失用户配置数据。
	该类警示信息可能导致系统重大变更甚至故障，或者导致人身伤害等结果。	 警告： 重启操作将导致业务中断，恢复业务所需时间约10分钟。
	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 说明： 您也可以通过按**Ctrl** + **A**选中全部文件。
>	多级菜单递进。	设置 > 网络 > 设置网络类型
粗体	表示按键、菜单、页面名称等UI元素。	单击 **确定**。
`courier` 字体	命令。	执行 `cd /d C:/windows` 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	`bae log list --instanceid` *`Instance_ID`*
[]或者[a|b]	表示可选项，至多选择一个。	`ipconfig` *`[-all|-t]`*
{}或者{a|b}	表示必选项，至多选择一个。	`swich` *`{stand | slave}`*

目录

1 数据传输解决方案

云计算凭借弹性、经济等优势，已经被越来越多的企业所认可，企业越来越多的将业务搬迁至云上。然而为应对企业发展的不同时期、不同的业务需求，企业不可避免的需要面对数据库间数据迁移传输的问题。

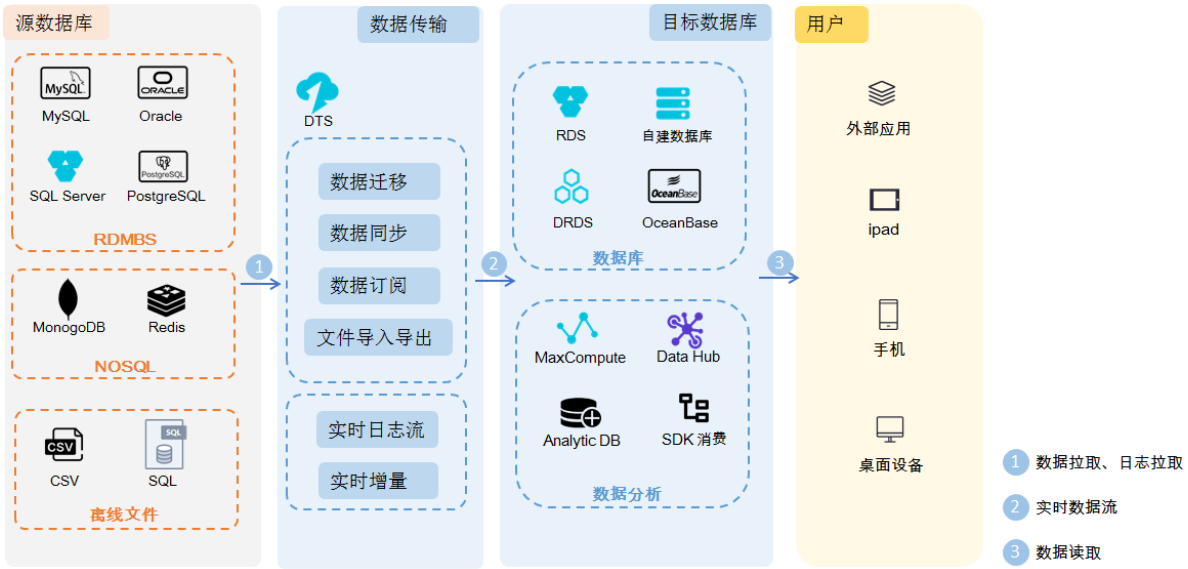
阿里云数据传输解决方案支持数据迁移、实时数据订阅及数据实时同步等多种数据传输方式，提供更丰富多样、高性能、高安全可靠的数据传输服务。

阿里云数据传输解决方案具有以下优势：

- **类型多样**：阿里云数据传输解决方案通过核心产品DTS（Data Transmission Service）可实现多种同异构数据库间的数据传输，且支持数据迁移、数据订阅、数据同步等多种单双向传输方式。
- **高性能**：DTS 使用高规格服务器来保证每条迁移或同步链路都能拥有良好的传输性能。对于数据迁移，DTS 底层采用了多种性能优化措施，全量数据迁移高峰期时性能可以达到 70MB/s，20w TPS。
- **安全可靠**：DTS 底层为服务集群，集群内任何一个节点宕机或发生故障，控制中心都能够将这个节点上的所有任务秒级切换到其他节点上，链路稳定性高达 99.95%。DTS 内部对部分传输链路提供 7×24 小时的数据准确性校验，快速发现并纠正传输数据，保证传输数据可靠性。
- **简单易用**：DTS 提供可视化管理界面，提供向导式的链路创建流程，用户可以在其控制台简单轻松得创建自己的传输链路。

解决方案架构及核心产品

阿里云数据传输解决方案架构如下图所示。



架构说明：

- 源数据库与目标数据库间数据传输时，通过阿里云DTS服务，可实现数据迁移、订阅、同步等多种传输方式。
- 阿里云数据传输解决方案支持多种数据源间的数据传输：

  — 数据迁移功能矩阵：

数据源	结构迁移	全量数据迁移	增量数据迁移
MySQL-->MySQL（RDS及自建）	支持	支持	支持（DML，部分DDL）
MySQL-->DRDS/PetaData/OceanBase	支持	支持	支持（DML）
MySQL-->Oracle（数据回流，需加白）	支持	支持	支持（DML）
Oracle-->MySQL（RDS及自建）	支持	支持	支持（DML，部分DDL）
Oracle-->DRDS	手工	支持	支持（DML）
Oracle-->RDS For PPAS	支持	支持	支持（DML）
Oracle-->ADS	支持	支持	支持（DML）
Oracle-->Oceanbase	支持	支持	支持
SQLServer-->SQLServer	支持	支持	支持（DML）
PostgreSQL-->PostgreSQL	支持	支持	支持（有主键表的DML）
MongoDB-->MongoDB	支持	支持	支持
Redis-->Redis	支持	支持	支持
DB2-->MySQL	支持	支持	支持（DML，部分DDL）

  — 数据同步功能矩阵：

数据源	说明
RDS For MySQL-->RDS For MySQL	1000 km+ 情况下，实现秒级同步延迟
RDS for MySQL<-->RDS for MySQL	双向数据同步
RDS for MySQL-->MaxCompute	数据同步到 MaxCompute 中，每张表对应两张表：全量基线表；增量日志表。通过全量基线表 + 增量日志表的 merge，可以获取任意时刻的全量数据
RDS for MySQL-->Datahub	通过这个功能支持 MySQL->流计算 的数据实时同步
RDS for MySQL-->AnalyticDB	支持实时报表分析、实时可视化大屏等实时数仓系统
MySQL-->DRDS	目前通过数据迁移 - 增量数据迁移实现
MongoDB-->MongoDB	高德异地容灾项目

- 数据订阅功能矩阵：

数据源	DML	DDL	备注
RDS for MySQL	支持	支持	-
DRDS	支持	支持	支持 DRDS 实例层面数据订阅

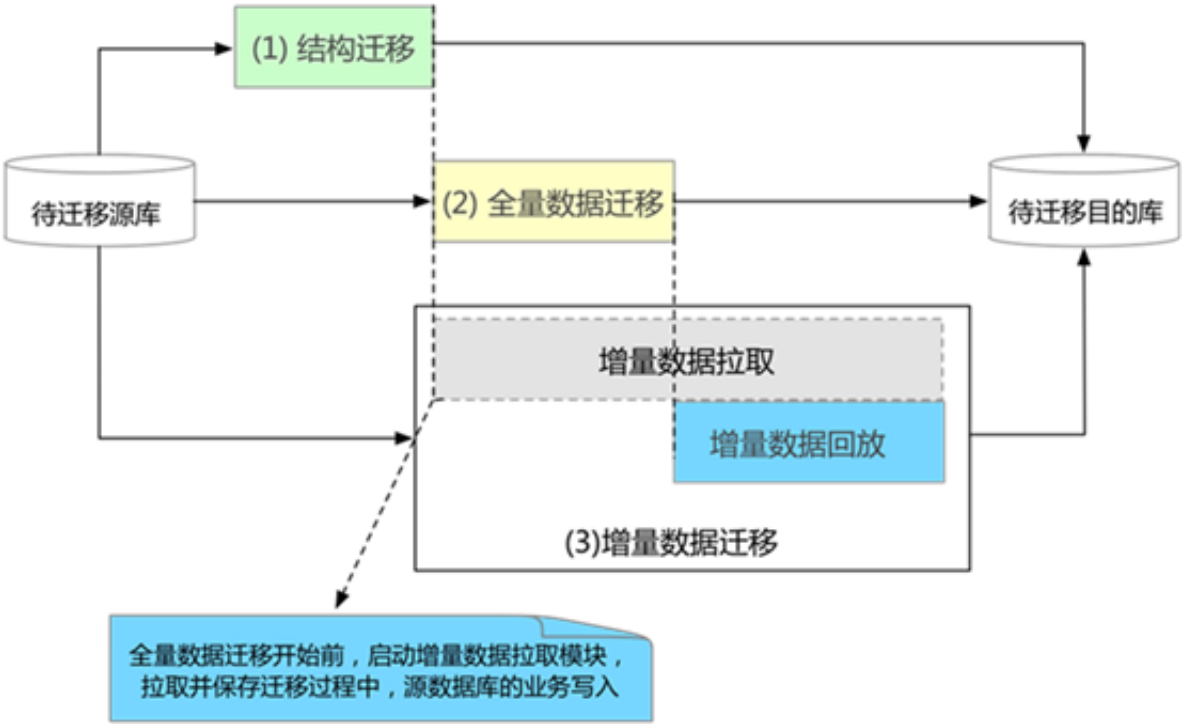
核心产品：

DTS（Data Transmission Service），是阿里云提供的一种支持多种数据源之间数据交互的数据流服务。它提供了数据迁移、实时数据订阅及数据实时同步等多种数据传输能力。更多DTS产品介绍请参考数据传输章节。

技术原理

- 数据迁移

  数据传输服务DTS在数据迁移的过程中，通过数据的全量迁移和增量迁移结合，迁移的源端数据库无需在迁移过程中停机，应用服务不会因为数据迁移出现中断。 数据迁移的技术原理如下图所示。



数据迁移过程：

1. 结构迁移：将源实例中的结构对象定义一键迁移至目标实例。
2. 全量迁移：源实例中的历史存量数据迁移至目标实例。
3. 增量数据迁移：全量迁移的同事进行增量数据拉取迁移，保障被迁移数据的完整性和一致性。

- **数据迁移**

  阿里云数据迁移支持：

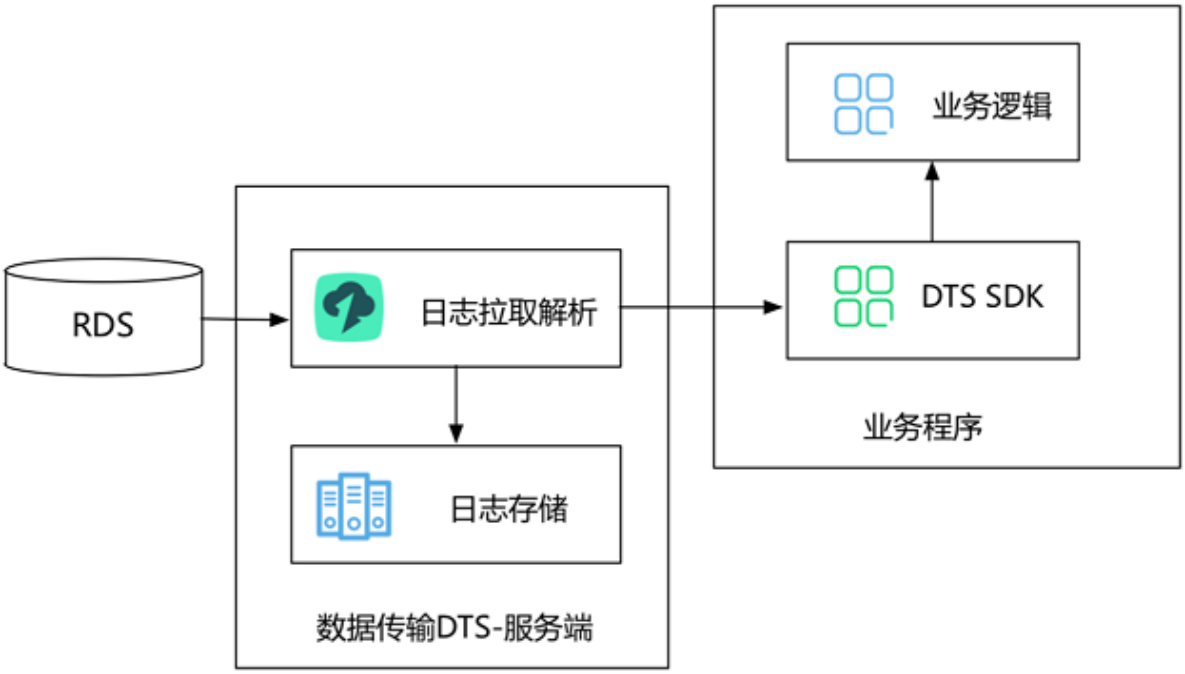
  - 多种迁移类型：结构对象迁移、全量数据迁移以及增量数据迁移。
  - 不停服迁移，迁移过程需要经历：

    1. 结构对象迁移
    2. 全量迁移
    3. 增量数据迁移

  通过有效的规划和演练，整个数据迁移的中断时间可以缩短至应用流量的切换时间，从而实现秒级切换。

- **数据订阅**

数据订阅支持实时拉取RDS实例的增量日志，用户可以通过DTS SDK来数据订阅服务端订阅增量日志，根据业务需求，实现数据定制化消费。

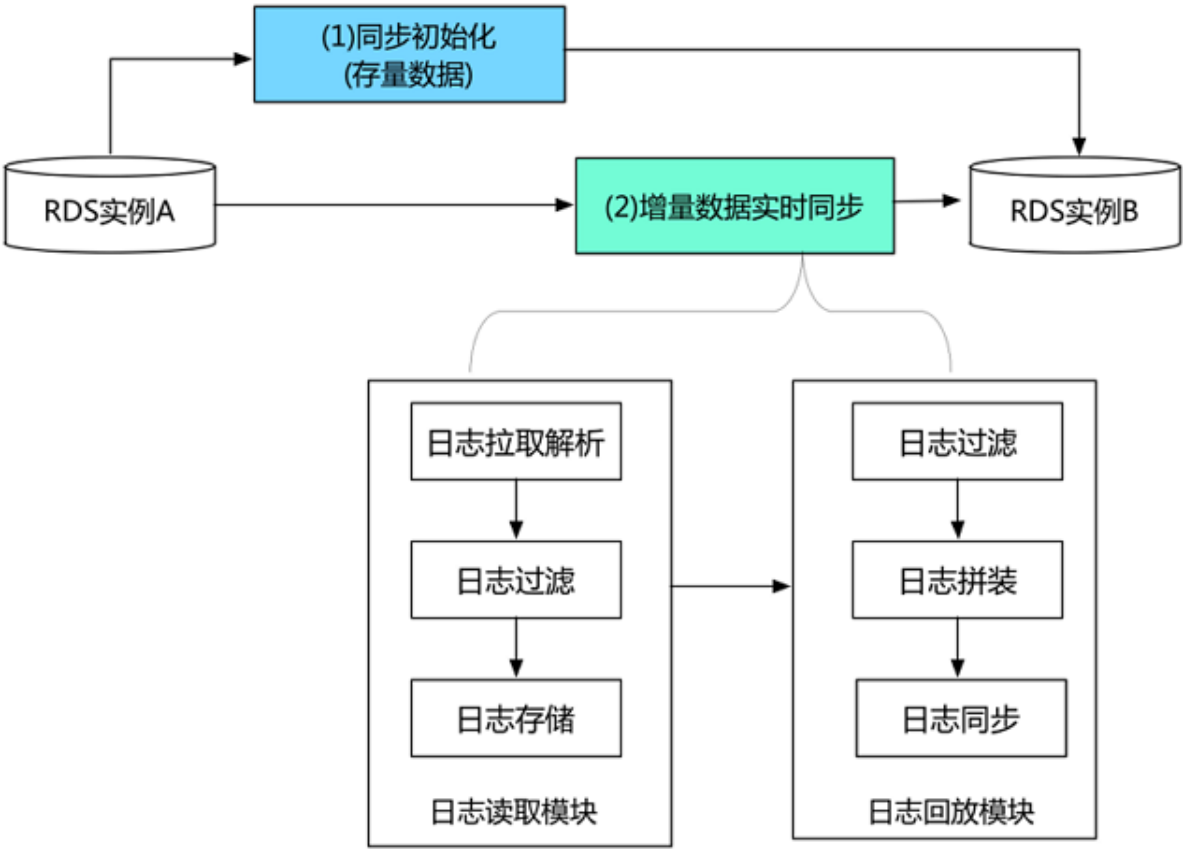


DTS服务端的日志拉取模块主要实现从数据源抓取原始数据，并通过解析、过滤、标准格式化等流程，最终将增量数据在本地持久化。

日志抓取模块通过数据库协议连接并实时拉取源实例的增量日志。例如源实例为RDS For MySQL，那么数据抓取模块通过Binlog dump协议连接源实例。

- 数据同步

  数据传输服务的实时同步功能能够实现任何两个RDS实例之间的增量数据实时同步，并支持RDS实例到ADS和ODPS等分析型数据库的数据实时同步。



同步链路的创建过程包括： 1. 同步初始化， 同步初始化主要将源实例的历史存量数据在目标实例初始化一份。 2. 增量数据实时同步， 当初始化完成后进入两边增量数据实时同步阶段，在这个阶段，DTS会实现源实例跟目标实例之间数据动态同步过程。

增量数据实时同步过程，DTS的底层实现模块主要包括： 1. 日志读取模块 2. 日志读取模块从源实例读取原始数据，经过解析、过滤及标准格式化，最终将数据在本地持久化。日志读取模块通过数据库协议连接并读取源实例的增量日志。如果源DB为RDS MySQL，那么数据抓取模块通过Binlog dump协议连接源库。 3. 日志回放模块 4. 日志回放模块从日志读取模块中请求增量数据，并根据用户配置的同步对象进行数据过滤，然后在保证事务时序性及事务一致性的前提下，将日志记录同步到目标实例。

DTS实现了日志读取模块、日志回放模块的高可用，DTS容灾系统一旦检测到链路异常，就会在健康服务节点上断点重启链路，从而有效保证同步链路的高可用。

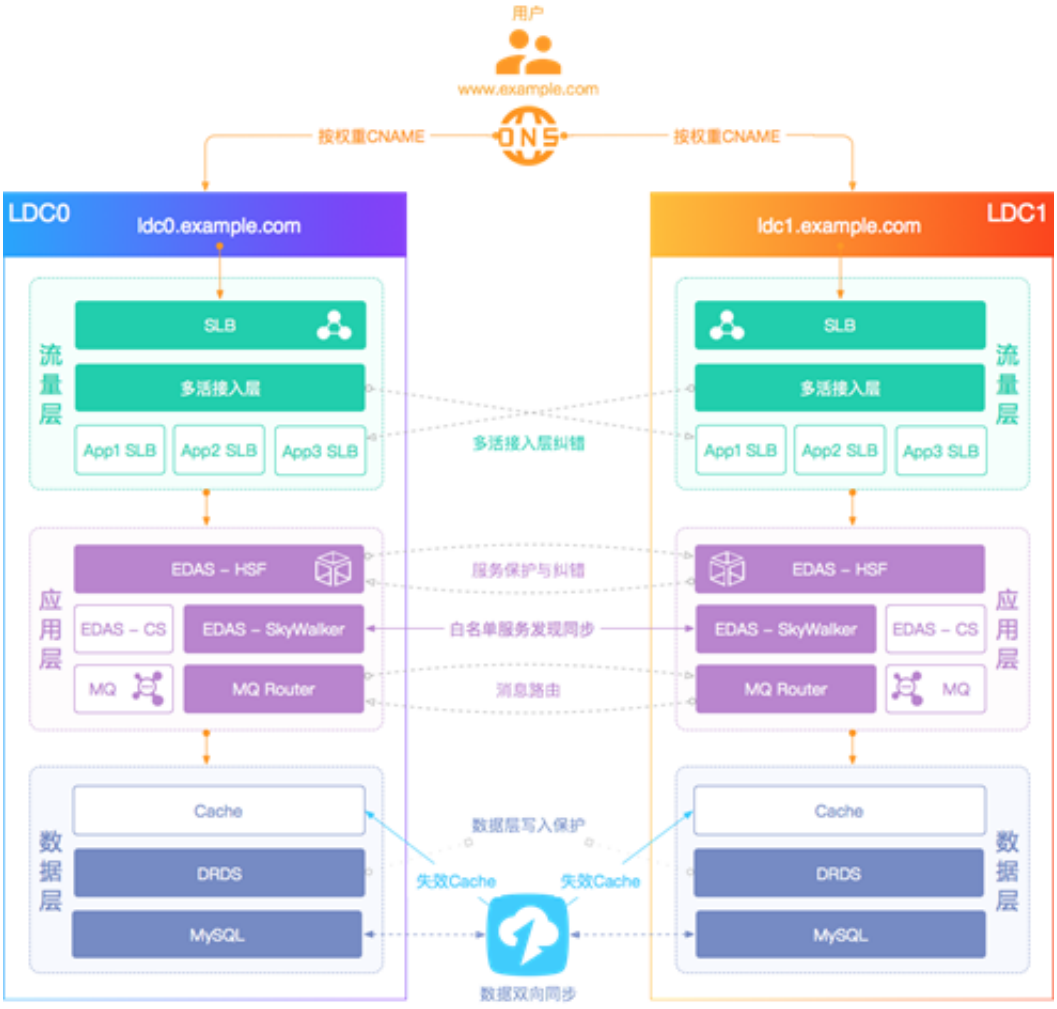
典型应用场景

数据传输几乎在各行各业各种场景均有涉及，尤其是异地灾备、异地多活、实时分析、精准营销等场景。更多场景介绍请参考使用场景章节。

2 数据库异地多活解决方案

异地多活指分布在异地的多个站点同时对外提供服务的业务场景。异地多活是高可用架构设计的一种，与传统的灾备设计的最主要区别在于“多活”，即所有站点都是同时在对外提供服务的。

以一个简单的业务单元的IT系统为例，整个IT系统的异地多活方案如下图所示。



整个方案将各站点分为：分流量层、应用层和数据层。

本文聚焦数据层的异地多活解决方案。目前针对数据层设计异地多活解决方案时，需遵循以下几个原则：

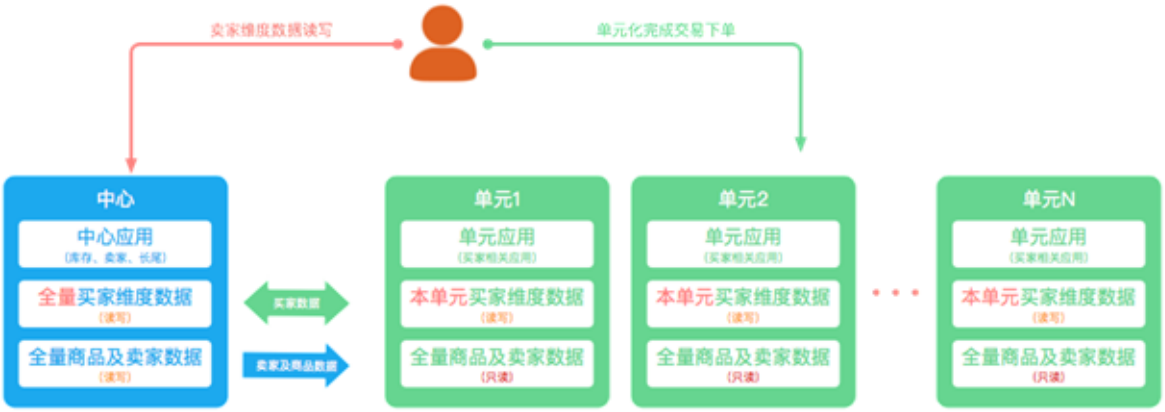
- 单元封闭：

  应用要走向异地，首先要面对的便是物理距离带来的延时。 如果某个应用请求需要在异地多个单元对同一行记录进行修改，为满足异地单元间数据库数据的一致性和完整性，我们需要付出高昂的时间成本。此外，如果某个应用请求需要多次访问异地单元，且各单元间服务再被服务调用，那物理延时将无法预估。 因此，数据库异地多活的问题转移到了如何避免跨单元的问

题，即要做到单元内数据读写封闭，不能出现不同单元对同一行数据进行修改，所以我们需要找到一个维度去划分单元，避免数据写冲突。

- **数据拆分**：

  选择什么维度来解决单点写的问题，要从业务本身入手去分析。例如电商的业务，最重要的流程即下单交易流程，从下单业务对数据划分单元时，改造成本最低、用户体验相对好的便是买家维度，即通过买家ID进行下单业务拆分，如下图所示。

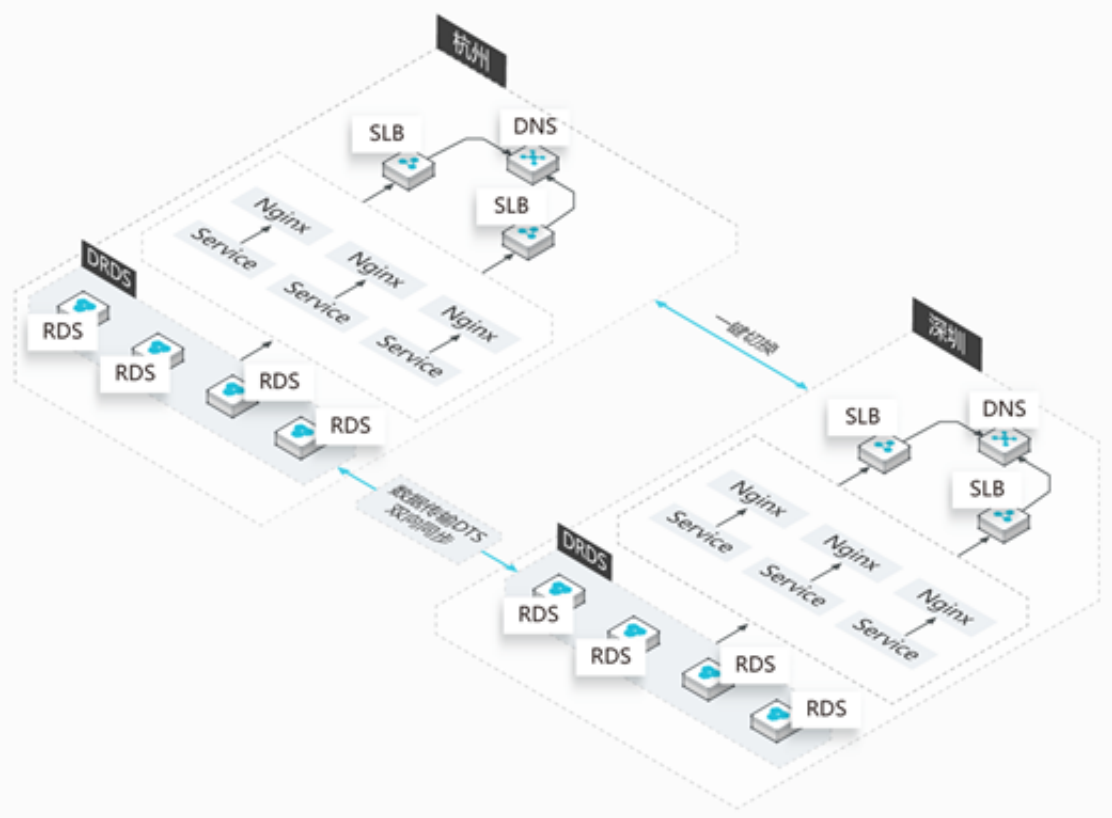


  买家的下单操作在买家所在的本单元内即可完成读写封闭。 按照上述示例划分业务，就意味着非买家维度的就需要做一定的妥协。对于非买家维度的操作，比如卖家操作（例如商品数据的修改）就可能会跨单元，对于买家与卖家数据读一致的问题， 可以接受“最终一致”的就通过读写分离的方案， 不能接受“最终一致”的则需要跨单元访问。

- **数据同步**：

  将业务划分多个单元以后，对于数据库来说，面临的最大挑战就是数据同步，因为对于单元封闭的买家维度的数据（unit类型）需要把单元的数据全部同步到中心，对于读写分离类型的(copy类型)，我们要把中心的数据同步到单元。

  原生的数据库复制无法满足单元化复杂的同步需求（例如，只同步部分的数据、库表的过滤、双向循环、API服务化等定制化的需求 ），在普通业务逻辑场景下，DTS的同步性能更高效、更稳定，DTS也成为多活的基础设施。具体的数据同步如下图所示。



- 缓存失效

  实现数据层异地多活后，业务希望实时获取数据库变更消息，单元的缓存失效实现可通过数据订阅实现。 通过数据订阅提供的消费 SDK，业务层可订阅 RDS 增量数据然后触发更新单元的缓存，通过这样的方式，应用无需实现缓存更新逻辑，架构更加简单。

核心产品

阿里云数据库异地多活解决方案使用以下阿里云核心产品，按照架构设计原则提供数据层多活解决方案。

DRDS

按照之前说的业务数据拆分的维度，阿里云DRDS有两种集群分别支持买家维度与卖家维度：

- unit 模式的DRDS集群：多地用户分别在本地域读写本地域的数据，且本地域的数据会和中心数据做双向同步。
- copy 模式的DRDS集群：此集群数据在中心数据库写，完成后全量同步到各个单元。 需要注意的是，DRDS层面需要增加对数据写入路由的判断：如果是跨单元的写，则判断为非法操作并抛出异常，确保数据不会跨单元写。

更多DRDS的介绍请参考分布式关系型数据库*DRDS*一文。

DTS

数据复制是数据库多活设计关键的一环，其中数据复制的正确性是第一位，同时效率也很关键。阿里云DTS支持多重的check ，避免循环复制（用事务表，或者thread_id方案），采用并行复制（串行的分发，冲突检测，并行的执行）、大事务切割来保证最终一致性。

数据校验也是关键的一环，阿里云DTS 通过全量校验工具（TCP）和增量校验工具（AMG）来保证实时/定时检查中心和单元的数据准确性，确保线上数据的万无一失。

更多的数据传输相关内容请参考数据传输服务一文。

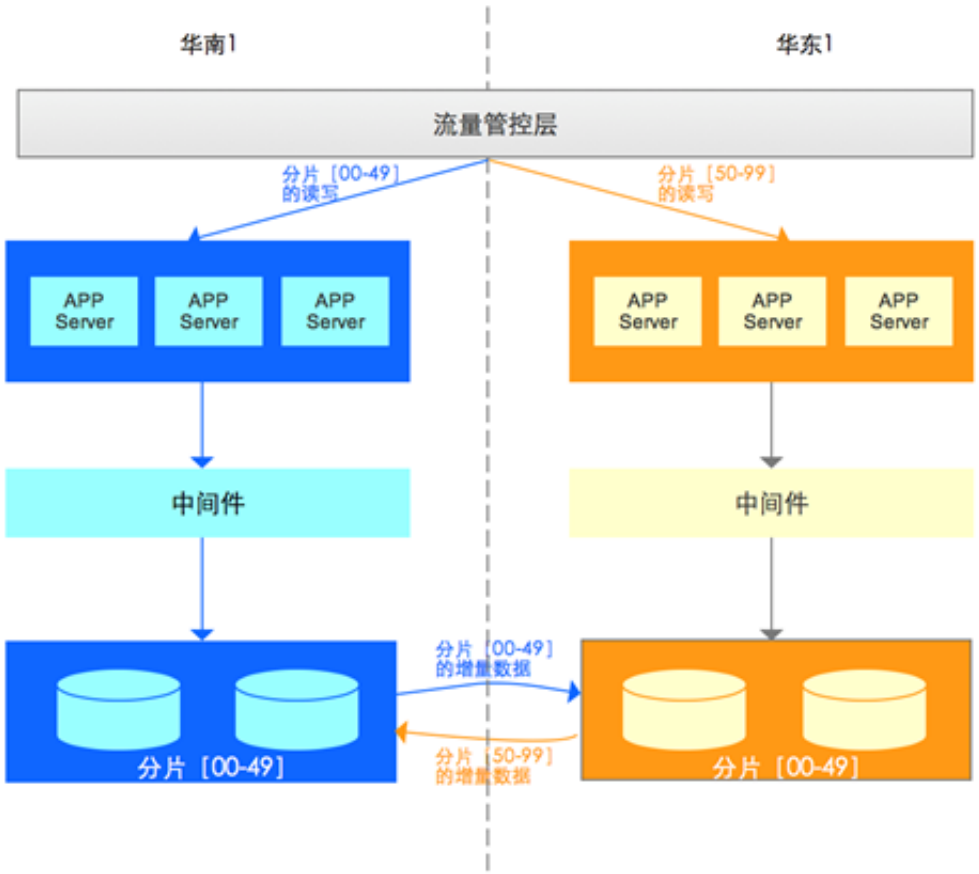
HDM

阿里云HDM提供了DRDS集群的搭建、同步链路的创建、多活的数据库监控、数据校验、集群扩缩容以及自动化的容灾等服务，都可通过HDM完成，通过HDM实现了异地多活场景下数据库的管理。

更多的数据管理请参考混合云数据库管理一文。

两地容灾切换方案

容灾是异地多活中最核心的一环， 以两个城市异地多活部署架构图为例：



- 在两个城市（城市1位于华南1地域、城市2位于华东1地域）均部署一套完整的业务系统。
- 下单业务按照“user_id”% 100 进行分片，在正常情况下：
  - [00~49]分片所有的读写都在城市1的数据库实例主库。
  - [50 ~ 99]分片所有的读写都在城市2的数据库实例主库。
- “城市1的数据库实例主库”和“城市2的数据库实例主库”建立DTS双向复制。

当出现异常时，需要进行容灾切换。可能出现的场景有以下4种：

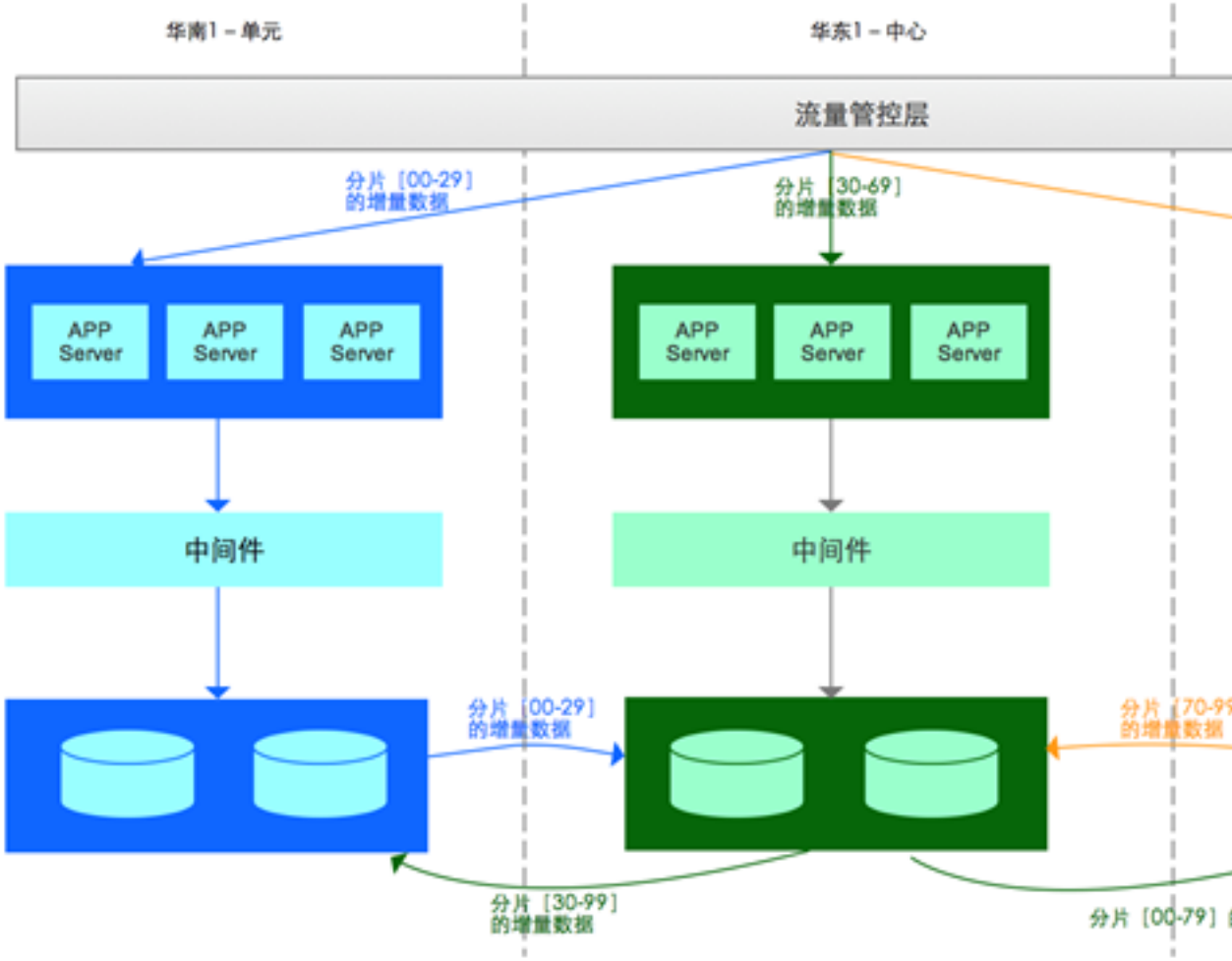
序号	异常情况	操作
1	城市1数据库主库故障	1. 数据库引擎完成主备切换 2. DTS自动切换到城市1新主库读取新的增量更新，然后同步到城市2的数据库实例
2	城市1所有APP Server故障	有两种处理方案： • 方案1：数据库层无任何操作，APP Server切换到城市2，并跨城市读写城市1的数据库 • 方案2：APP Server和数据库都切换到城市2

序号	异常情况	操作
3	城市1所有数据库故障	有两种处理方案： • 方案1：数据库层切换到城市2，APP Server跨城市读写城市2的数据库 • 方案2：APP Server和数据库都切换到城市2
4	城市1整体故障（包括所有APP Server + 数据库等）	1. 城市1的全部数库流量切换到城市2 2. 城市1数据库到城市2数据库的DTS数据同步链路停止 3. 在城市2中，DTS启动，保存 [00-49] 分片的变更 4. 城市1故障恢复后，[00-49] 的增量数据同步到城市1的数据库实例 5. 同步结束后，将 [00-49] 的数据库流量从城市2切回到城市1启动 [00-49] 分片从城市1到城市2的DTS同步

将第2种、第3种异常情况，全部采用第2种方案进行处理，那么不管是所有的APP Server异常、所有的数据库异常、整个城市异常，就直接按照城市级容灾方案处理，直接将APP Server、数据库切换到到另一个城市。

多城异地多活

多城市异地多活模式指的是3个或者3个以上城市间部署异地多活。该模式下存在中心节点和单元节点：

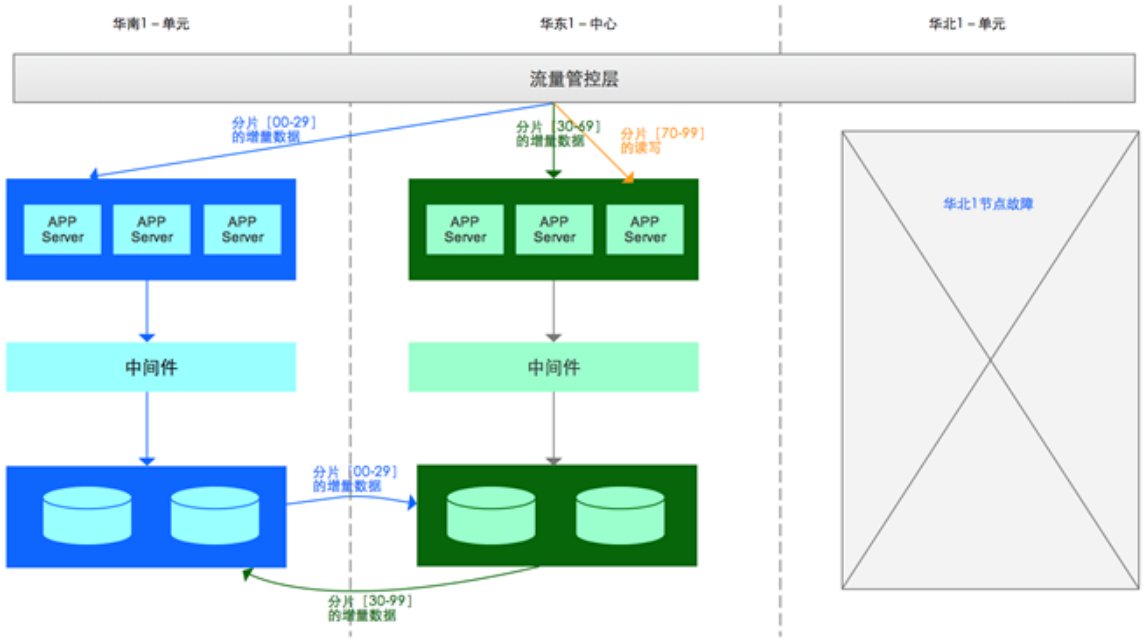


- 中心节点：指单元节点的增量数据都需要实时的同步到中心节点，同时中心节点将所有分片的增量数据同步到其他单元节点。
- 单元节点：即对应分片读写的节点，该节点需要将该分片的增量同步到中心节点，并且接收来自于中心节点的其他分片的增量数据。

下图是3城市异地多活架构图，其中华东1就是中心节点，华南1和华北1是单元节点。

单元城市级故障

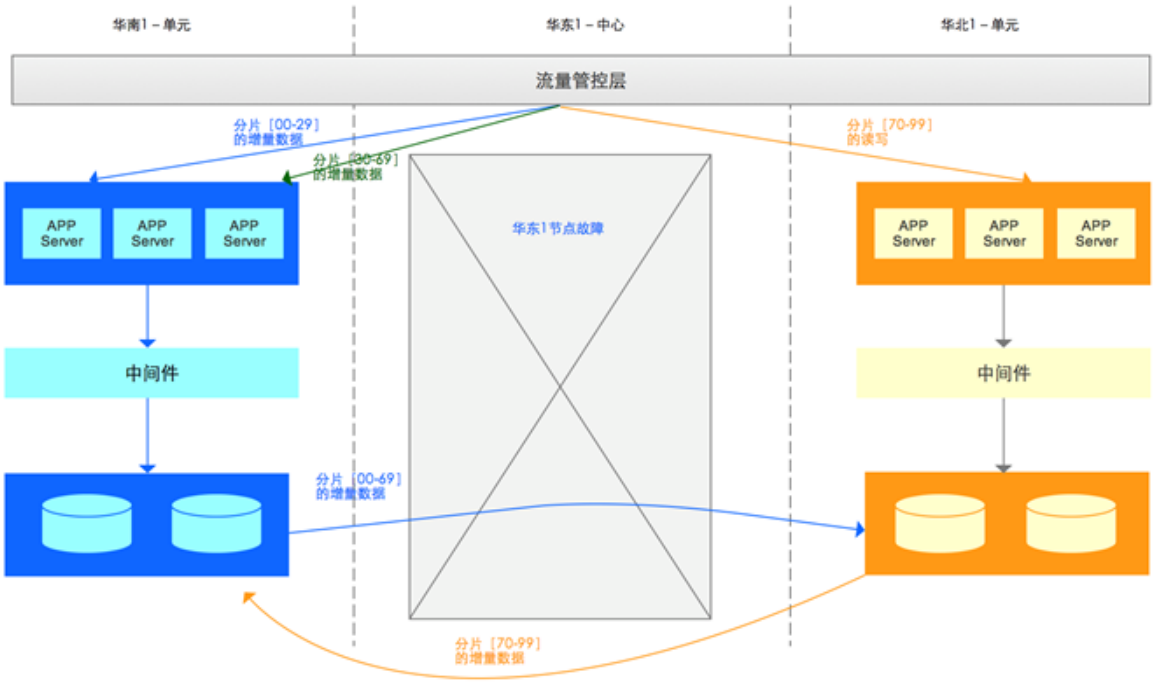
当单元城市出现故障，业务需要切换时，以华北1城市级故障为例：



1. 容灾
   a. 华北1（单元）的全部数库流量切换到华东1（中心）；
   b. 华北1（单元）数据库到华东1（中心）数据库的DTS数据同步链路停止，并记录同步位点
   c. 分片 [70-99] 的读写切换到华东1（中心）
2. 恢复
   a. 重建华北1（单元）；
   b. 华北1（单元）数据迁移和同步完成后，停止分片 [70-99] 在华东1（中心）的读写；
   c. 停止华东1（中心）到华北1（单元）分片 [70-99] 的数据同步；
   d. 创建华北1（单元）到华东1（中心）的数据同步；
   e. 将分片 [70-99] 的读写切换到华北1（单元）；
   f. 华北1（单元）的数据库主库开启写入；
   g. 检查；

中心城市级故障

当中心城市出现故障，业务需要切换时，以华东1城市级故障为例：



1. 容灾

   a. 华东1（中心）的全部数库流量切换到华南1（单元）；

   b. 华东1（中心）数据库到华南1（单元）数据库的DTS数据同步链路停止；

   c. 华东1（中心）数据库到华北1（单元）数据库的DTS数据同步链路停止；

   d. 华南1（单元）数据库到华东1（中心）数据库的DTS数据同步链路停止；

   e. 华北1（单元）数据库到华东1（中心）数据库的DTS数据同步链路停止；

   f. 新增华南1（单元）数据库到华北1（单元）分片 [30～99] 的DTS数据同步链路；

2. 恢复

   a. 重建华东1（中心）；

   b. 华东1（中心）数据迁移和同步完成后，停止分片 [30-69] 在华南1（单元）的读写；

   c. 停止华东1（中心）到华南1（单元）分片 [00-29] 的数据同步；

   d. 创建华东1（中心）到华南1（单元）的数据同步；

   e. 创建华东1（中心）到华北1（单元）的数据同步；

   f. 将分片 [00-29] 的读写切换到华南1（单元）；

   g. 华南1（单元）的数据库主库开启写入；

   h. 检查；