



电子科技大学
University of Electronic Science and Technology of China

大数据计算技术

Big Data Computing Technology

Lecture 24 阿里云

目录

24.1 阿里巴巴阿里云服务

- ▶ 24.1.1 阿里云计算体系架构
- 24.1.2 弹性计算服务（ECS）
- 24.1.3 开放存储服务OSS
- 24.1.4 开放结构化数据服务OTS
- 24.1.5 开放数据处理服务ODPS
- 24.1.6 关系型数据库RDS



- 阿里云计算体系架构

阿里云专注于云计算领域的研究，依托云计算的架构做一个可扩展、高可靠、低成本的基础设施服务，支撑包括电子商务在内的互联网应用的发展，从而降低进入电子商务生态圈的门槛、成本，并提高效率

阿里云发展历程

2009

- 阿里云计算有限公司成立。

2014

- 阿里云海外数据中心开服，成为中国第一家提供全球云计算服务的公司。

2017

- 成为奥运会全球指定云服务商。
- 阿里巴巴达摩院成立，进行核心基础技术研究。

2013

- 成为世界上第一个对外提供5K云计算能力的公司。

2015

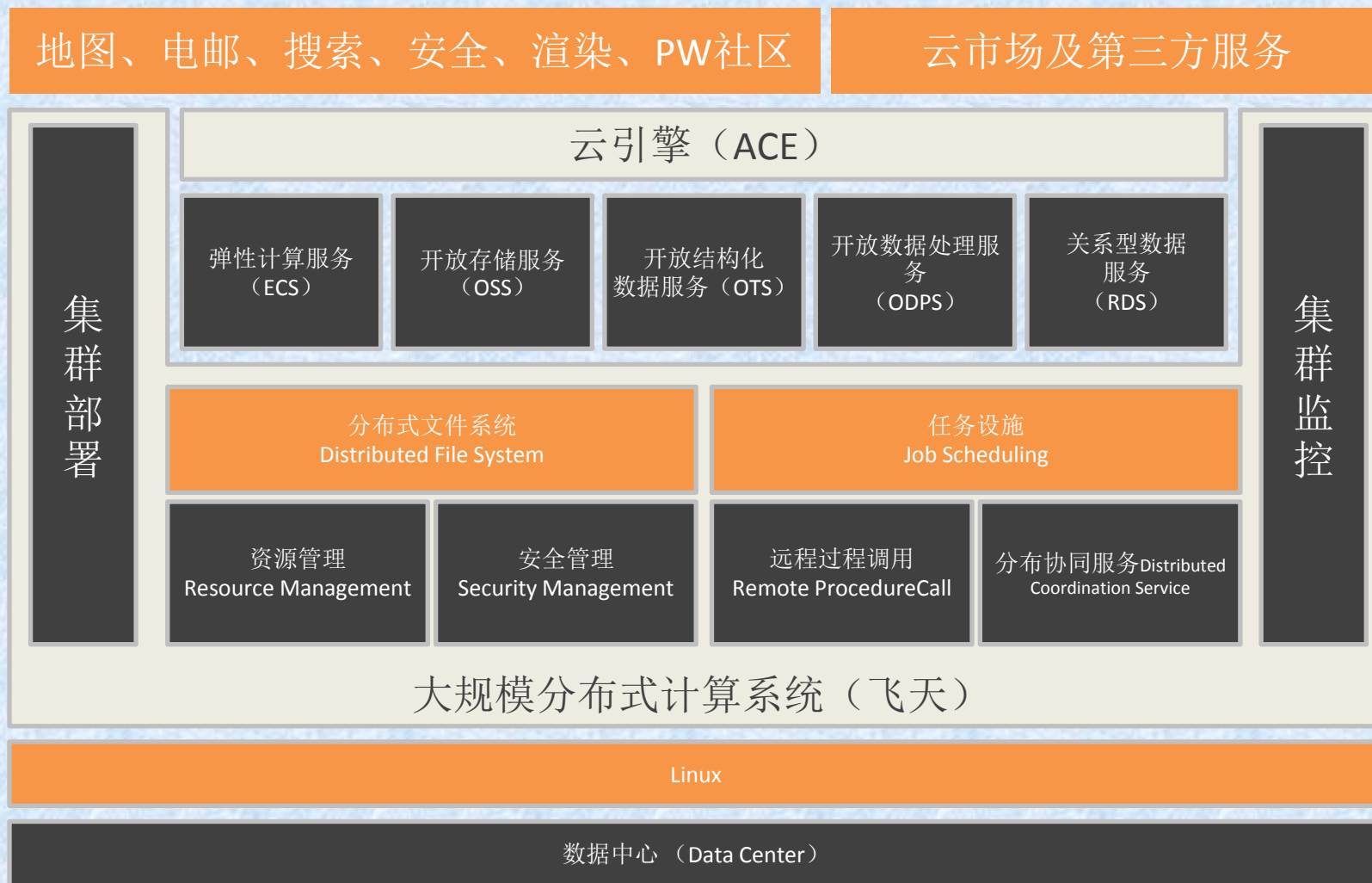
- 阿里巴巴启动“中台战略”，构建“大中台、小前台”。

2019

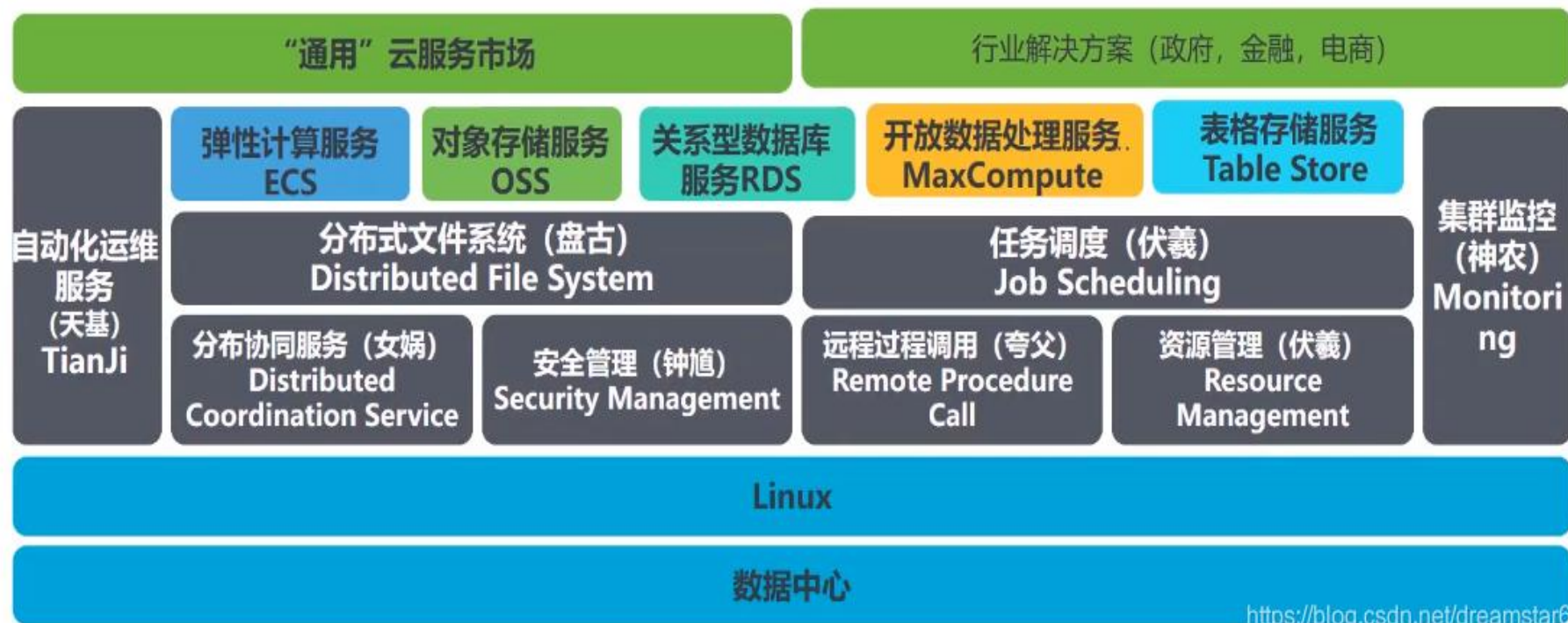
- 阿里云战略升级为阿里云智能。
- 阿里巴巴核心系统100%迁移到阿里云。
- 平头哥发布首款AI芯片含光800。

[dn.net/dreamstar613](https://www.aliyun.com/industry/dn.net/dreamstar613)

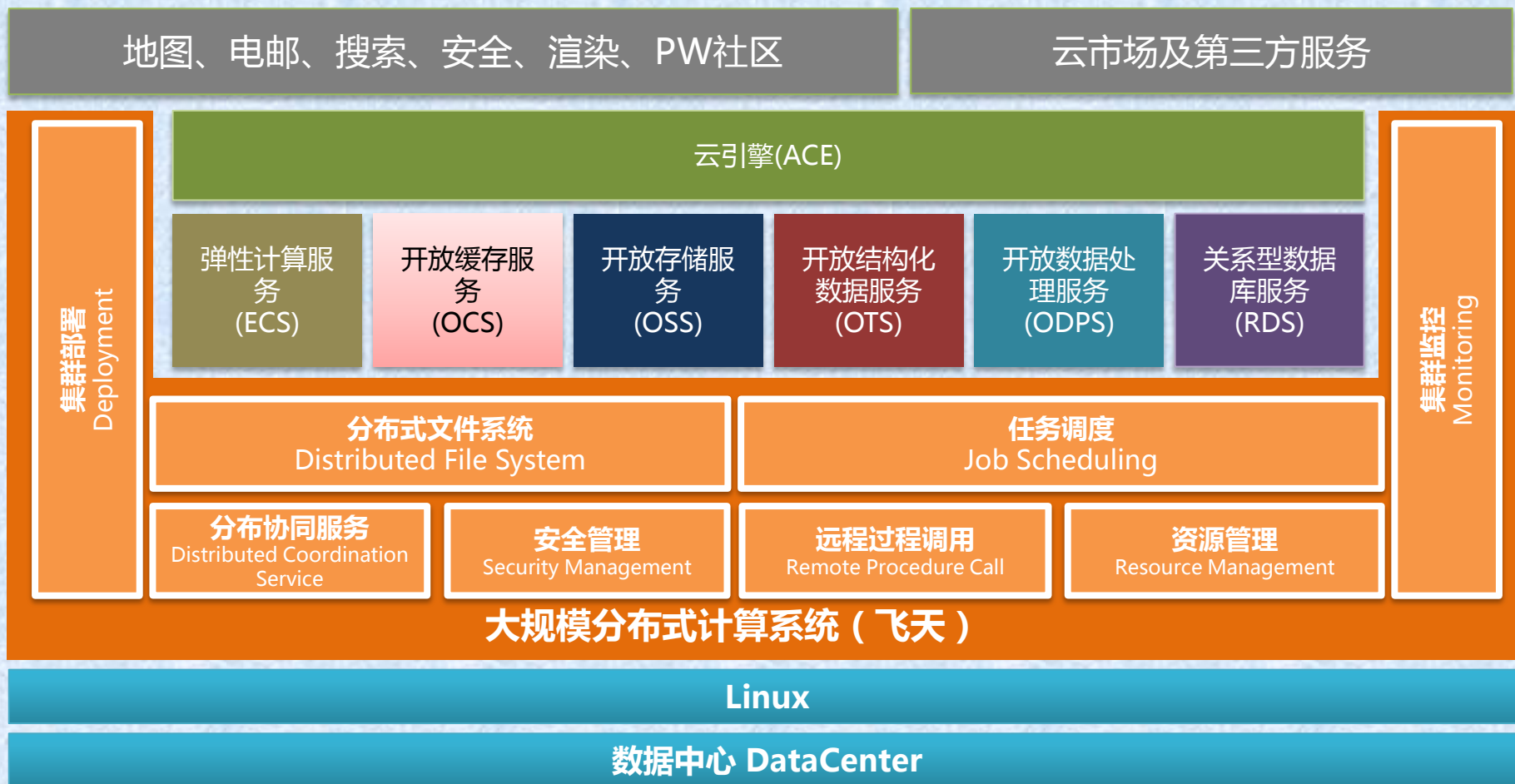
阿里云计算体系



阿里云技术架构



阿里云体系架构图



阿里云系统架构

地图、电邮、搜索、安全、渲染、PW社区

云市场及第三方服务

云引擎(ACE)

弹性计算服务
(ECS)

开放缓存服务
(OCS)

开放存储服务
(OSS)

开放结构化
数据服务
(OTS)

开放数据处
理服务
(ODPS)

关系型数据
库服务
(RDS)

集群
部署
Depl
oym
ent

集群
监控
Mon
itori
ng

分布式文件系统
Distributed File System

任务调度
Job Scheduling

分布协同服务
Distributed
Coordination Service

安全管理
Security
Management

远程过程调用
Remote Procedure Call

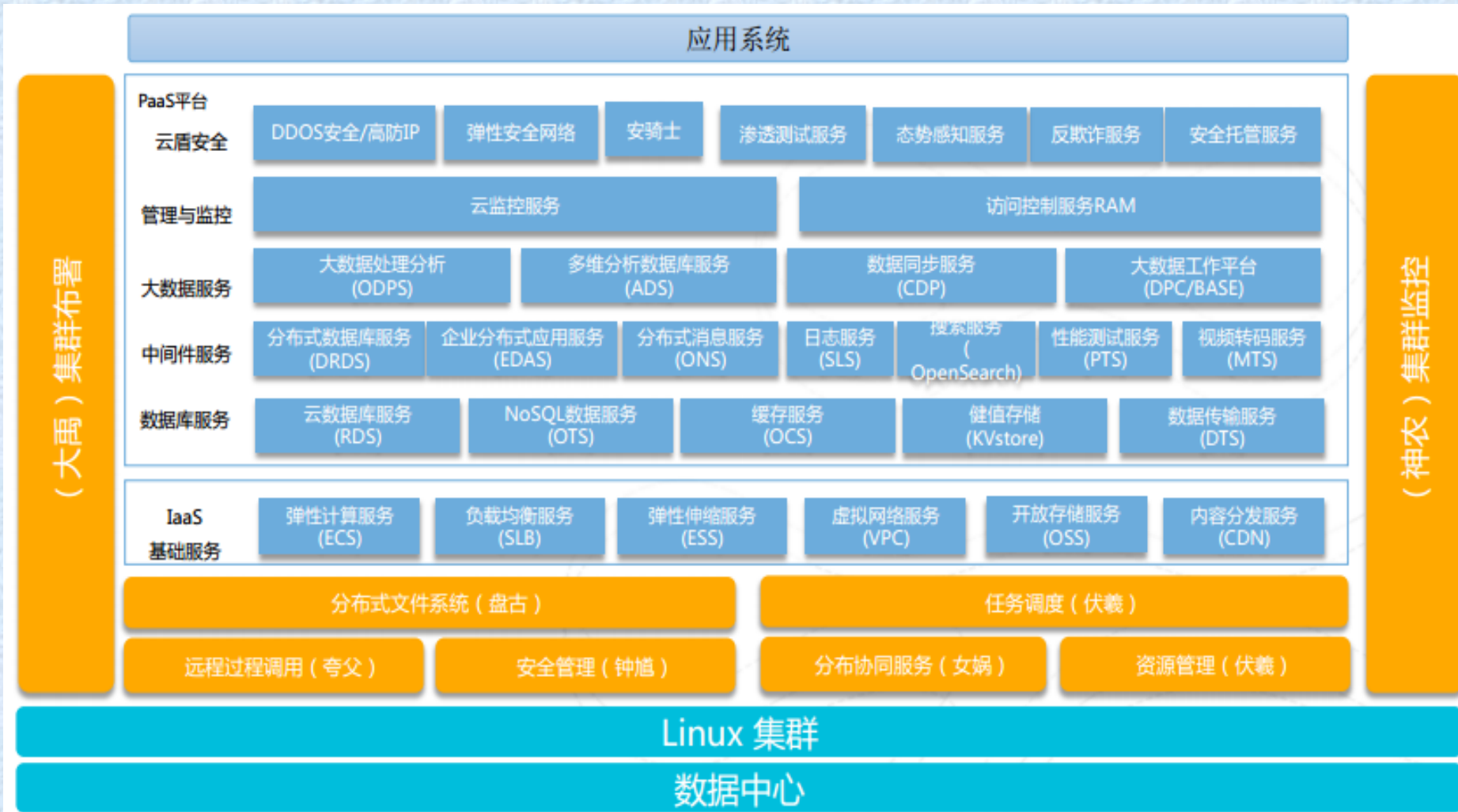
资源管理
Resource Management

大规模分布式计算系统 (飞天)

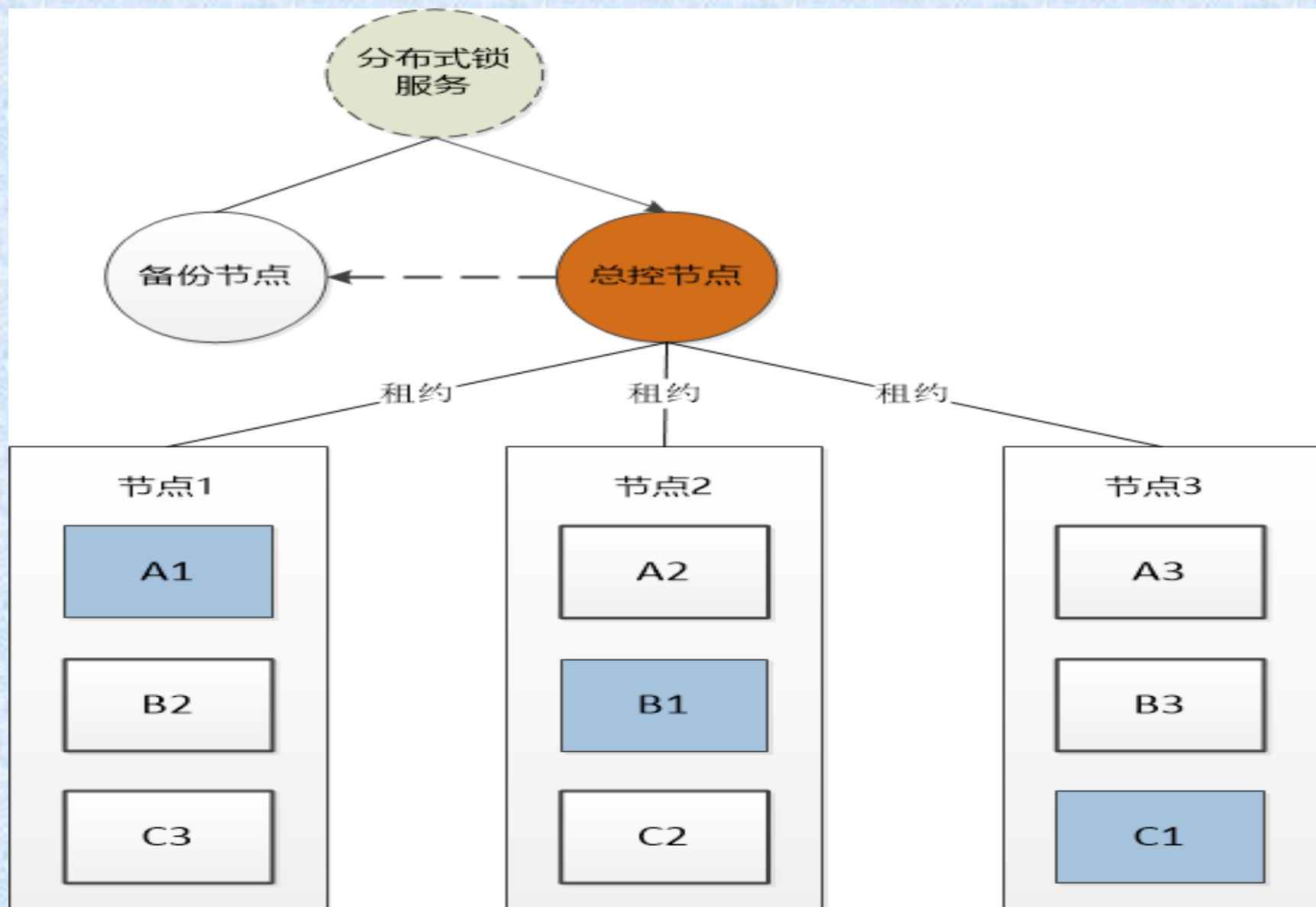
Linux

数据中心 DataCenter

阿里云的产品体系



分布式文件系统（一层结构）



分布式文件系统-盘古

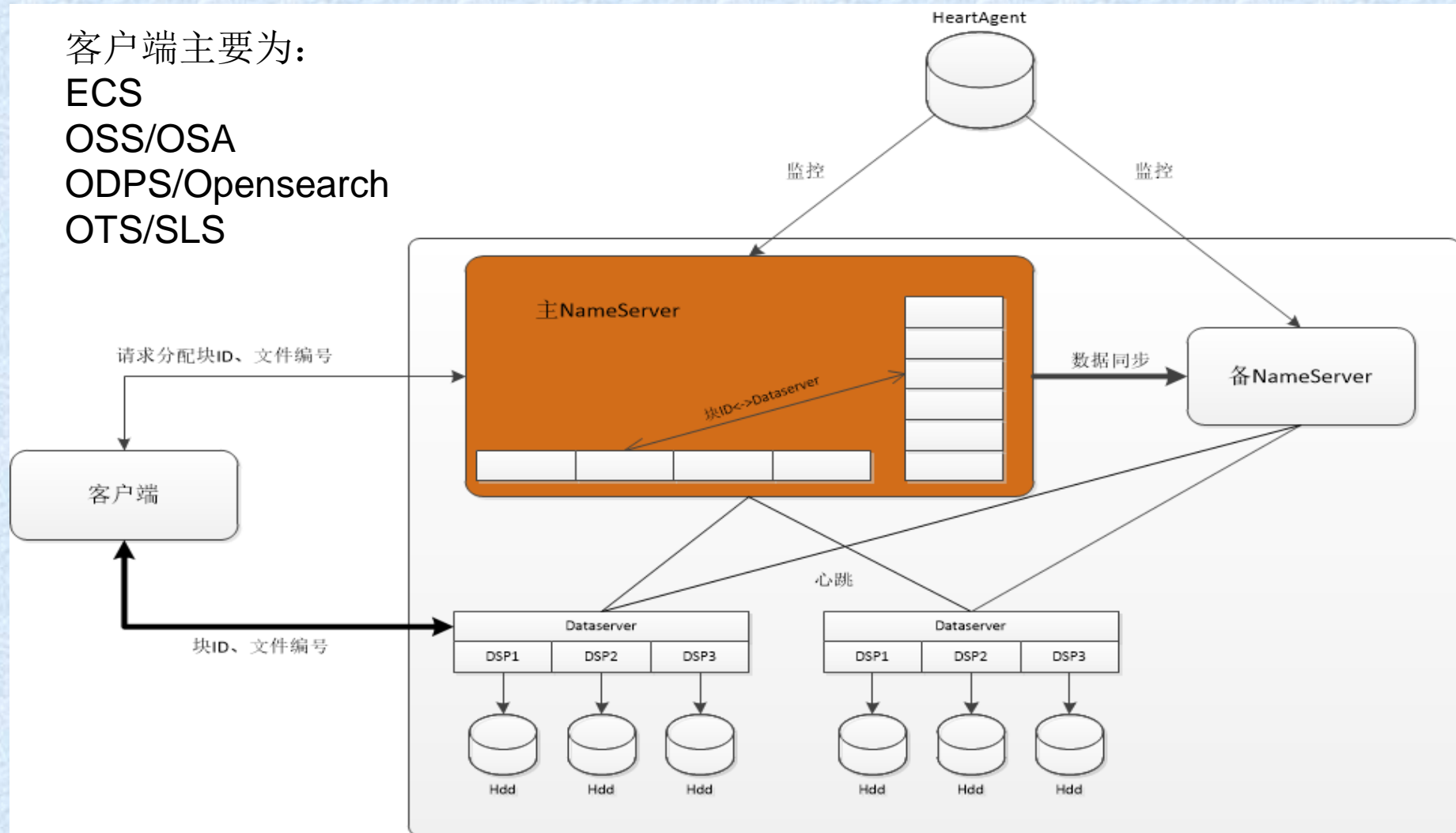
客户端主要为:

ECS

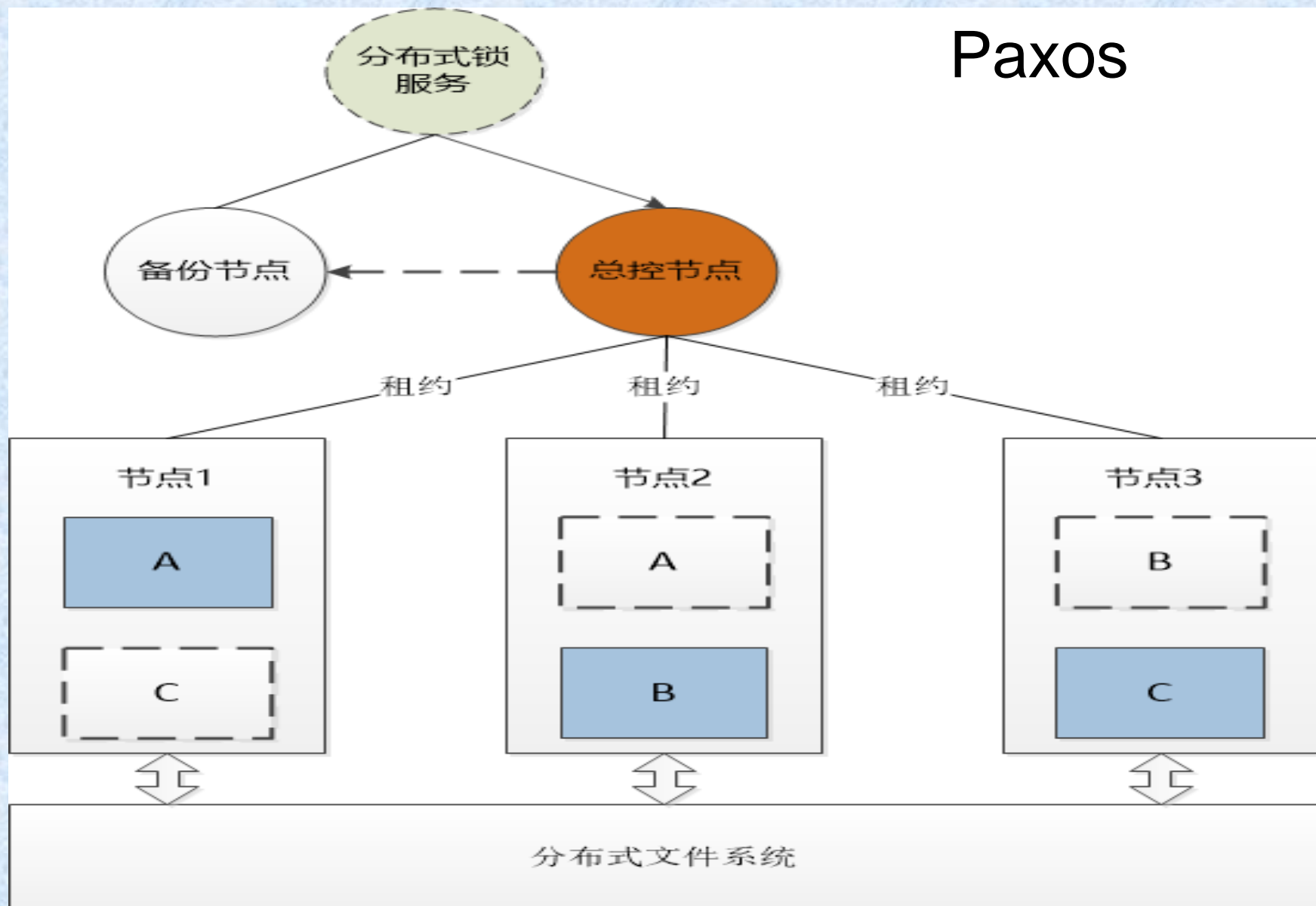
OSS/OSA

ODPS/Opensearch

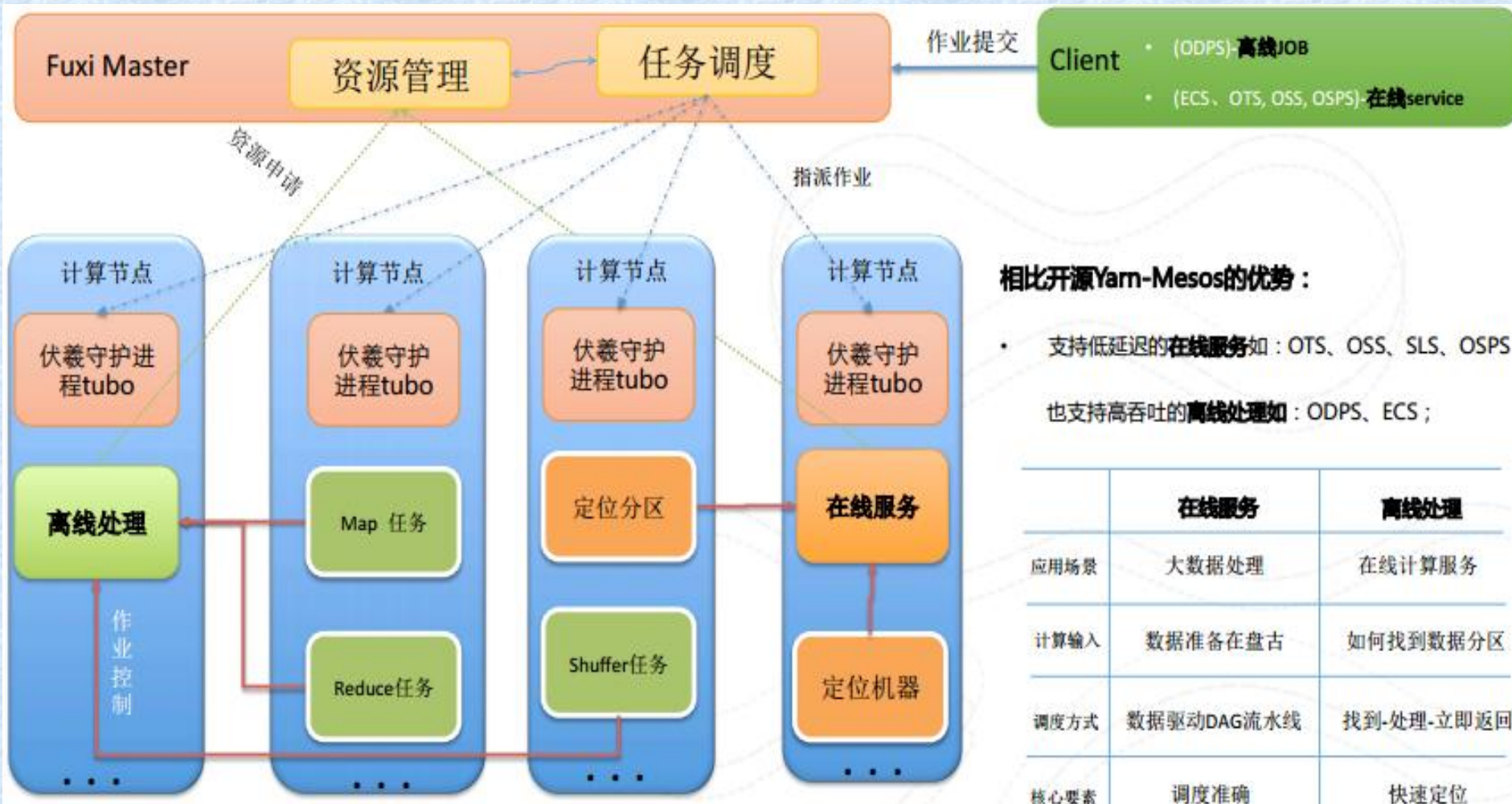
OTS/SLS



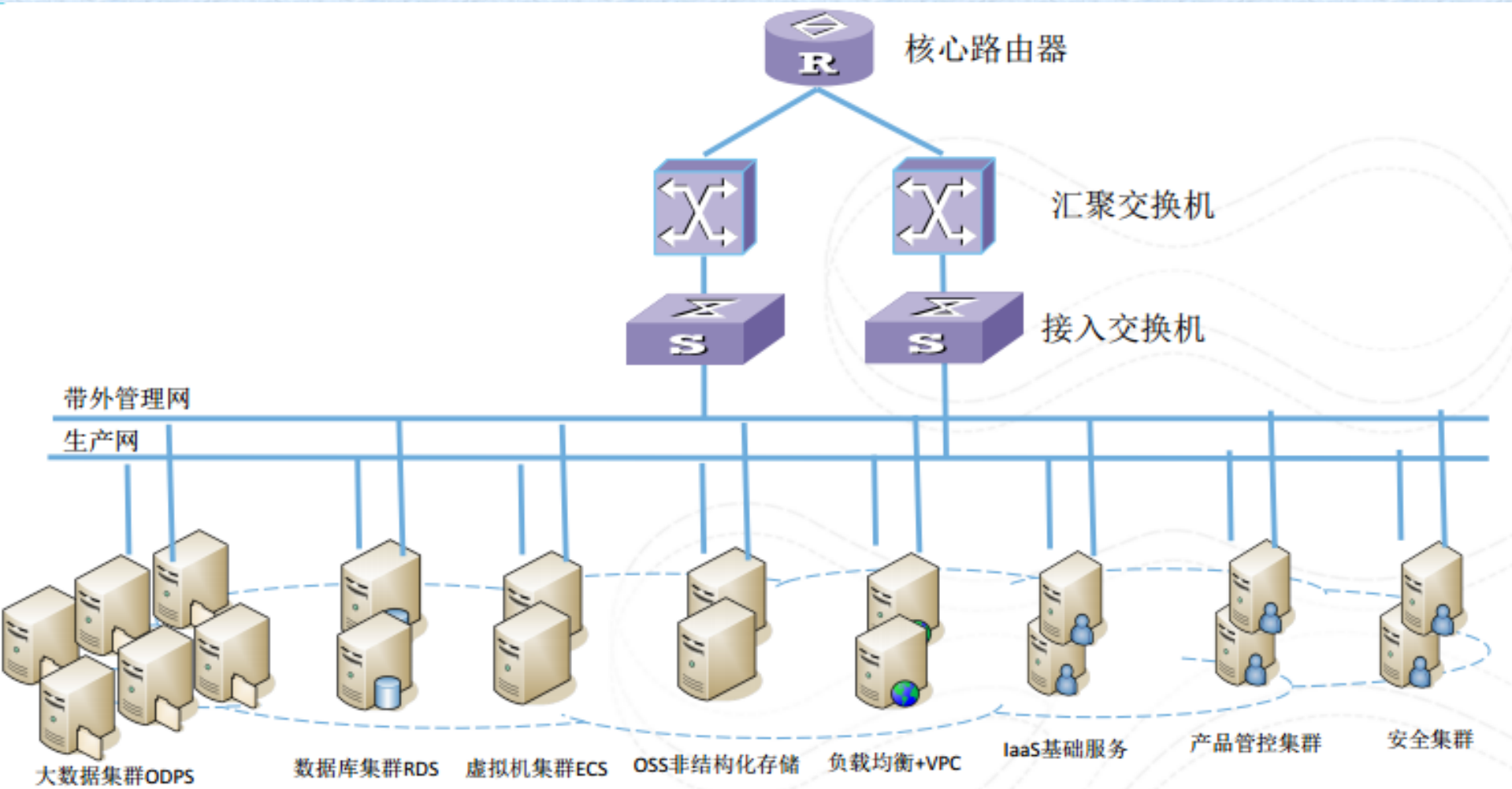
分布式文件系统（二层结构）



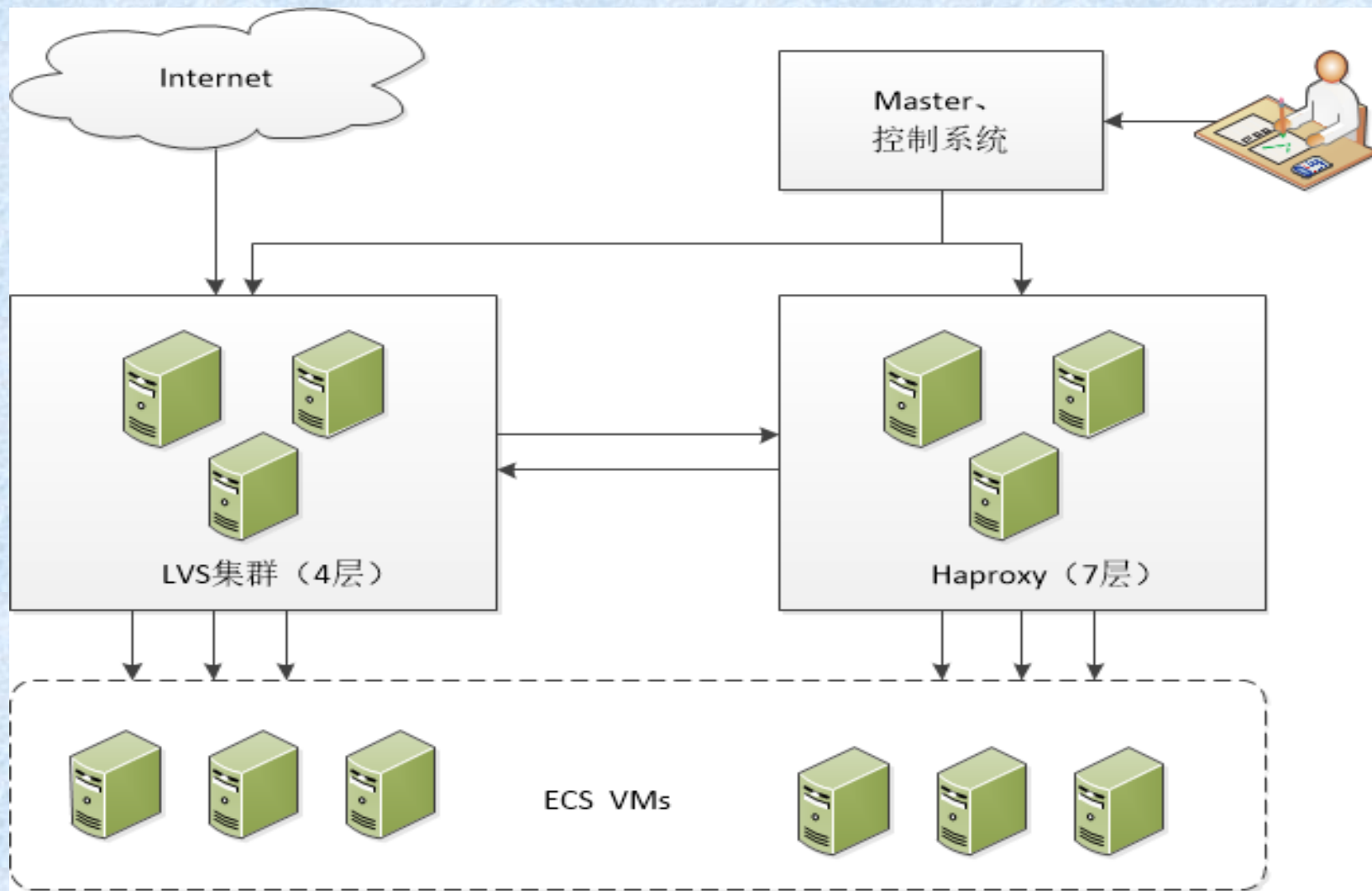
任务调度系统-伏羲



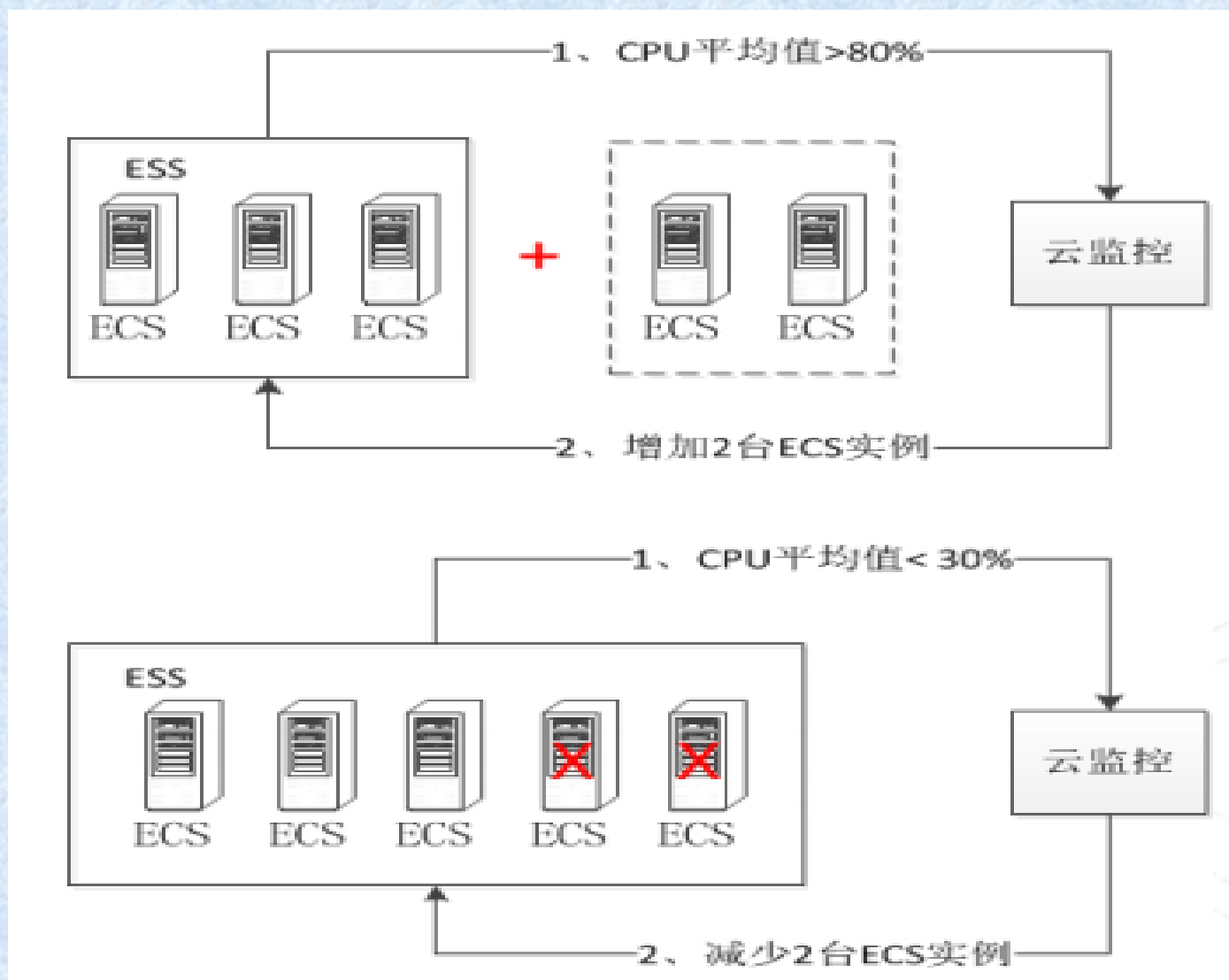
物理部署示意图



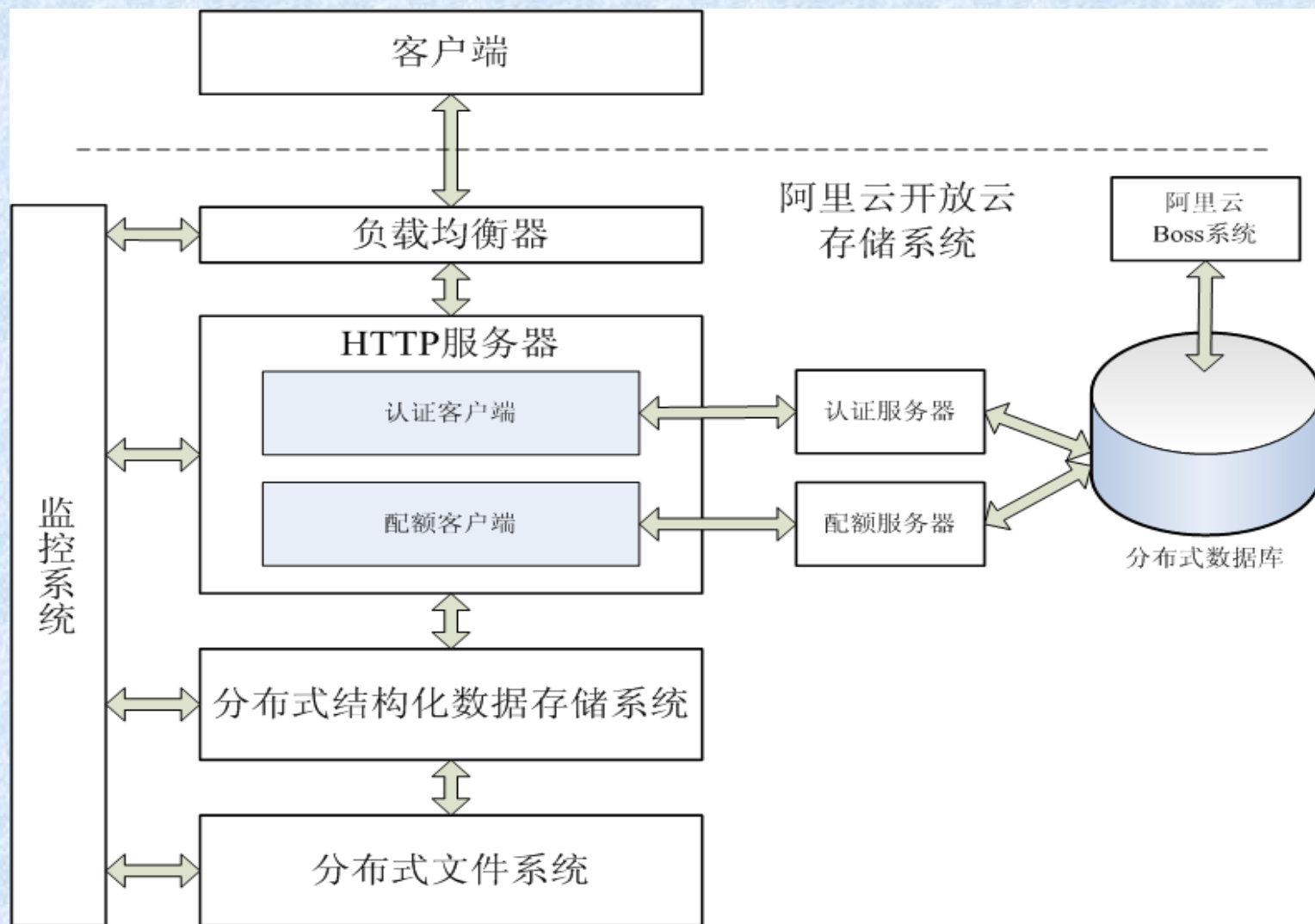
I AAS基础服务-SLB负载均衡



I AAS基础服务-ESS弹性伸缩服务

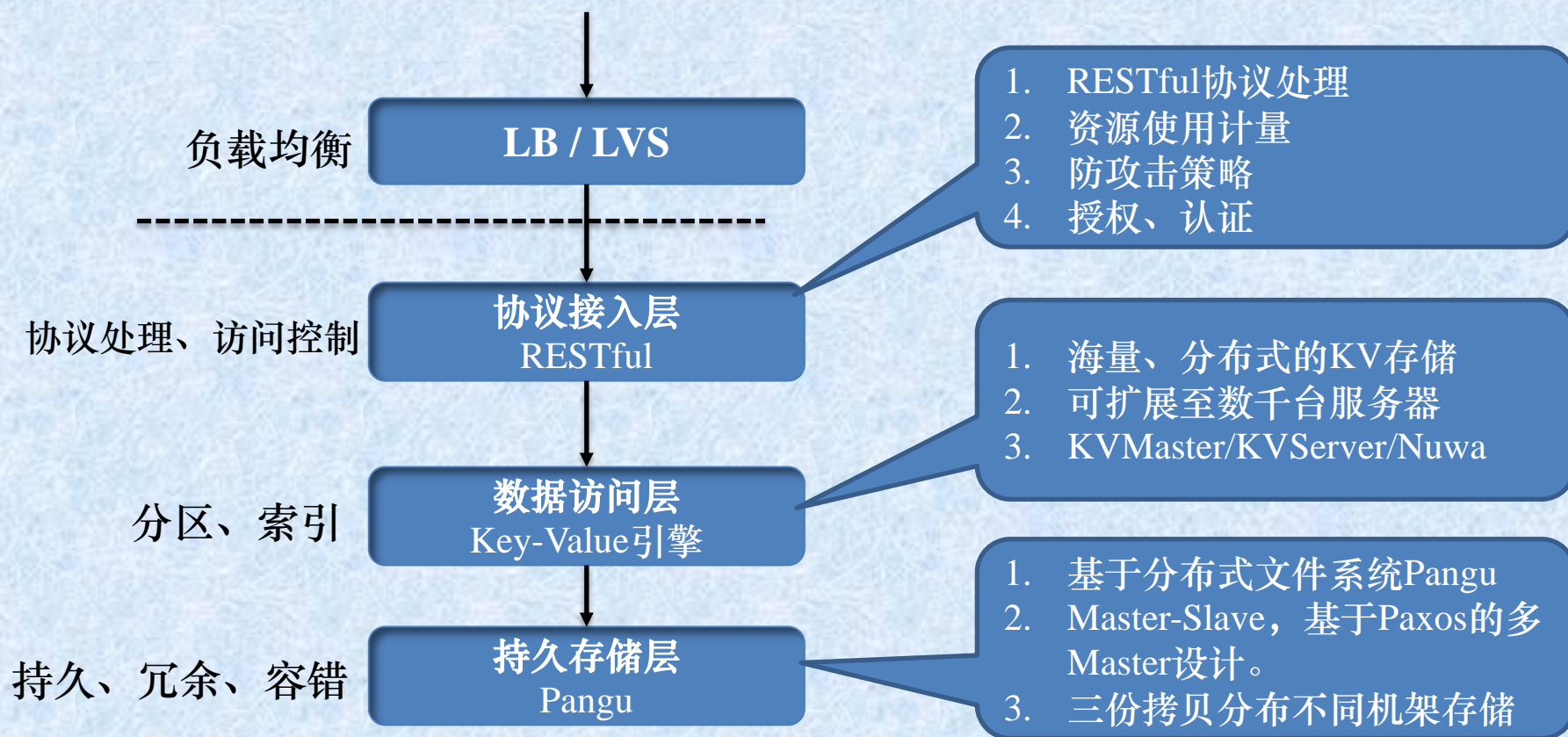


IAAS基础服务-OSS基础架构

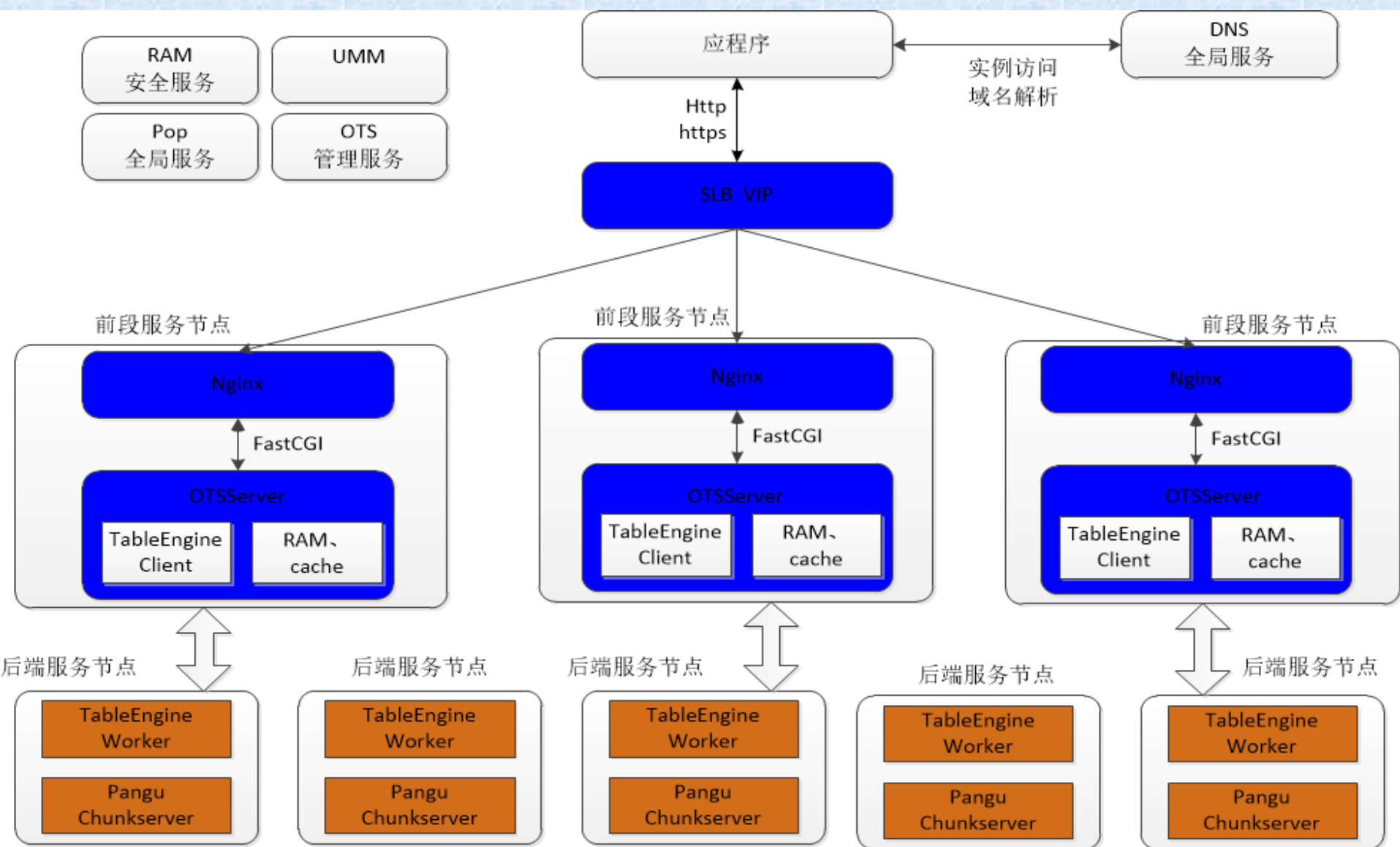


I AAS基础服务-OSS基础架构

`http://<bucket>.oss.aliyuncs.com/<object>`



阿里云体系架构PAAS-开放表格服务



PAAS-开放表结构服务(OTS)

大规模：基于 阿里云云操作系统

通用性：支持 表，视图，事务，增删更新查 等功能，支持

partition, primay key

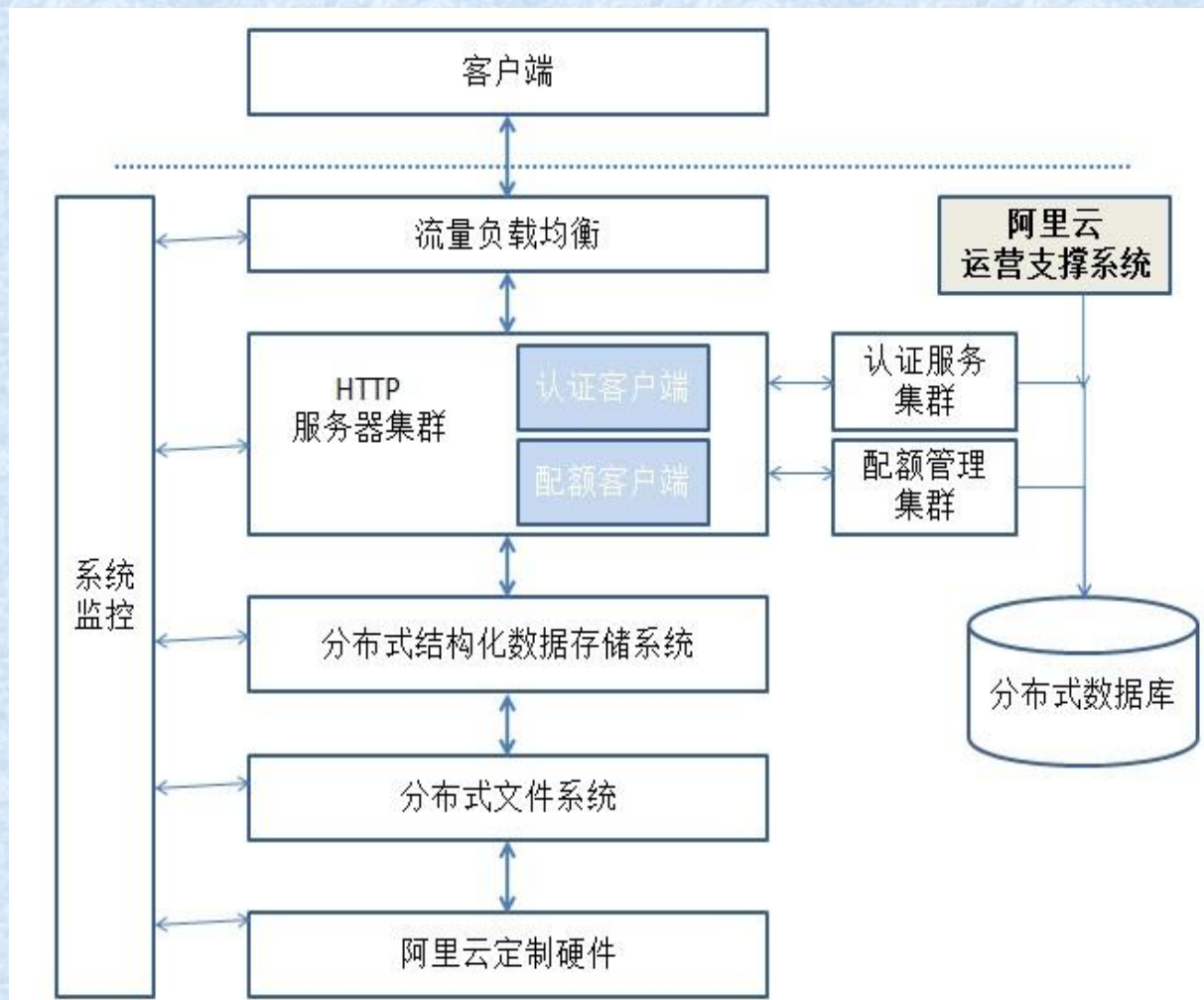
高可用：通过自动故障迁移来提供高可用性；通过存储的冗余备份来确保数据的安全。

实时性：支持实时的数据插入、删除和更新；支持实时的随机查询、范围查询和偏移查询，并可通过视图和分页的方式让查询更高效。

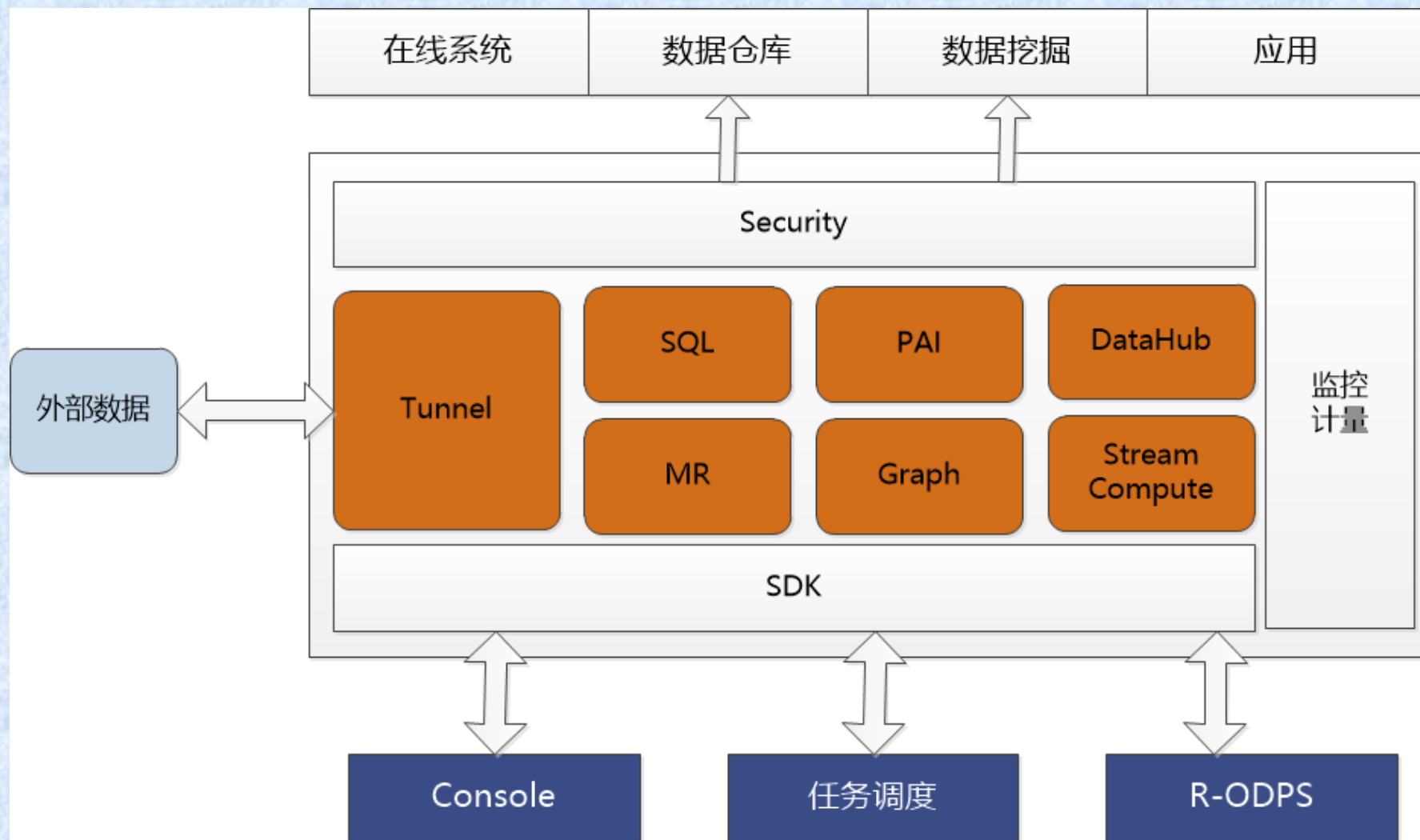
一致性：通过数据分区内的事务，使得数据的更改和查询保持强一致性

宜用性：通过提供多语言SDK（包括C++、Python、C#等）让用户的开发更便捷。

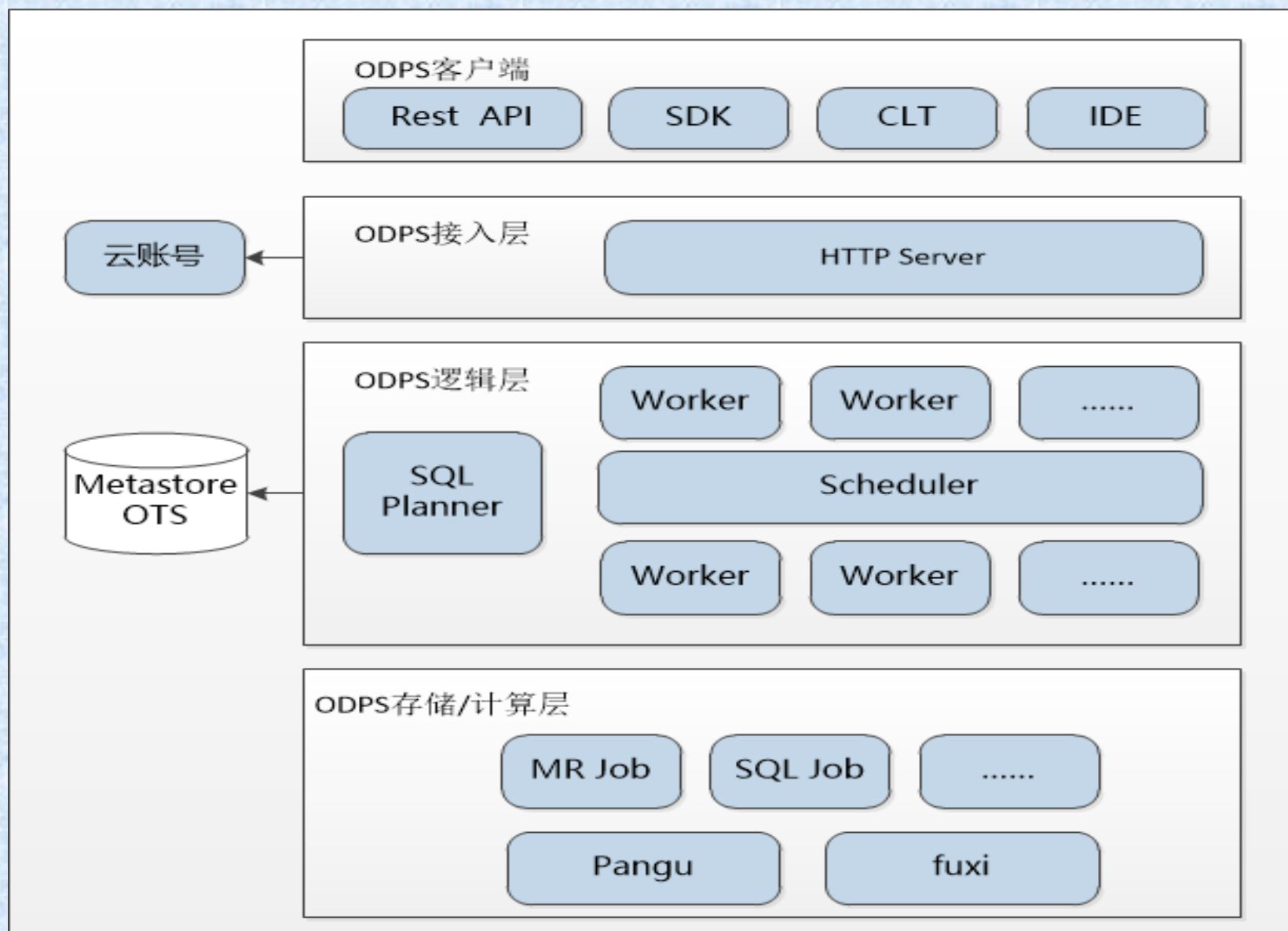
经济性：仅按用户的实际资源使用进行计费。



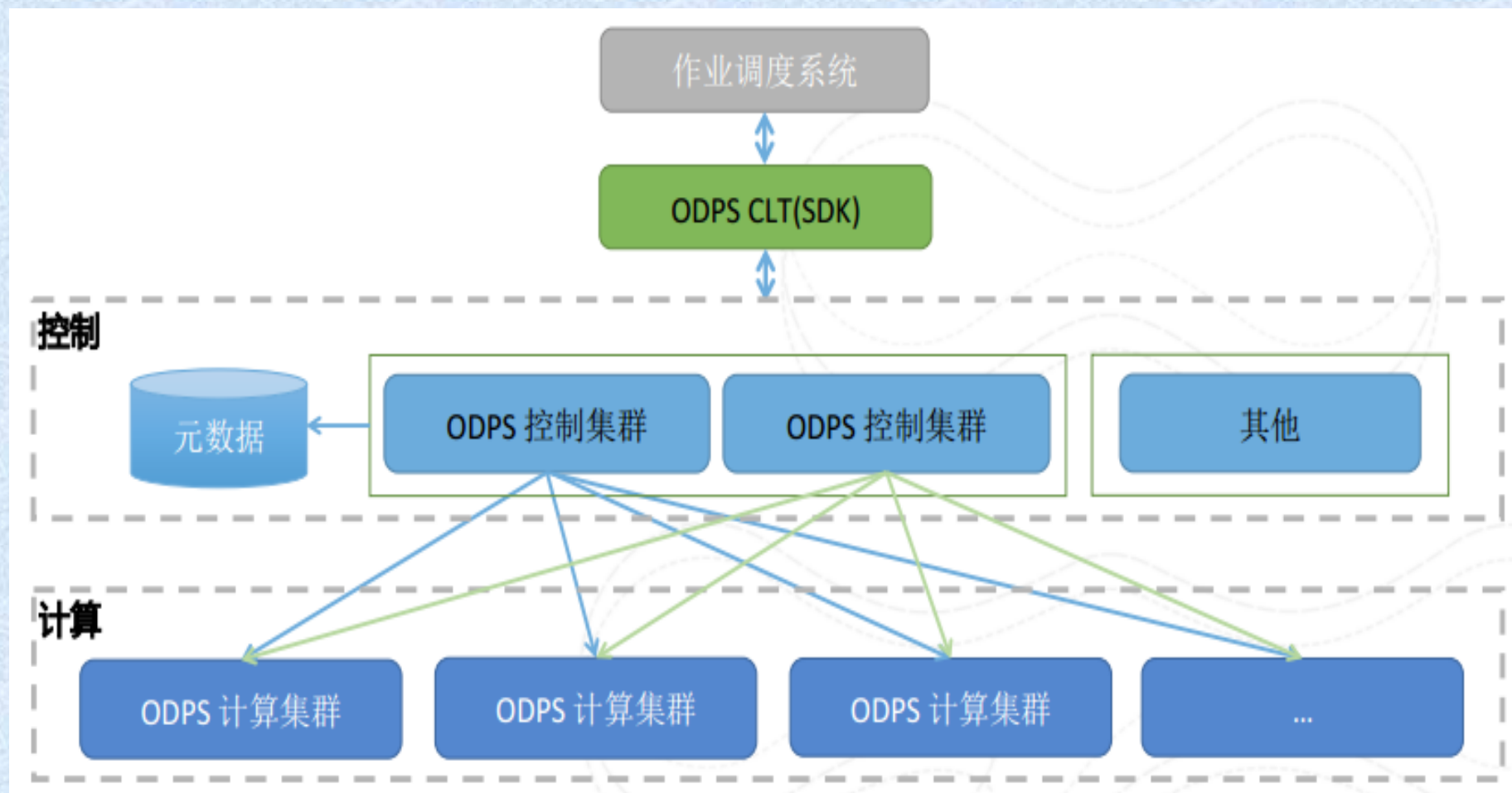
PAAS-大数据计算服务ODPS功能组成



PAAS-大数据计算服务ODPS架构图



PAAS-大数据计算服务ODPS多数据中心



PAAS-云数据库服务-RDS

云数据库-关系型数据库服务，是一种稳定、可靠、可弹性伸缩的在线服务数据库。

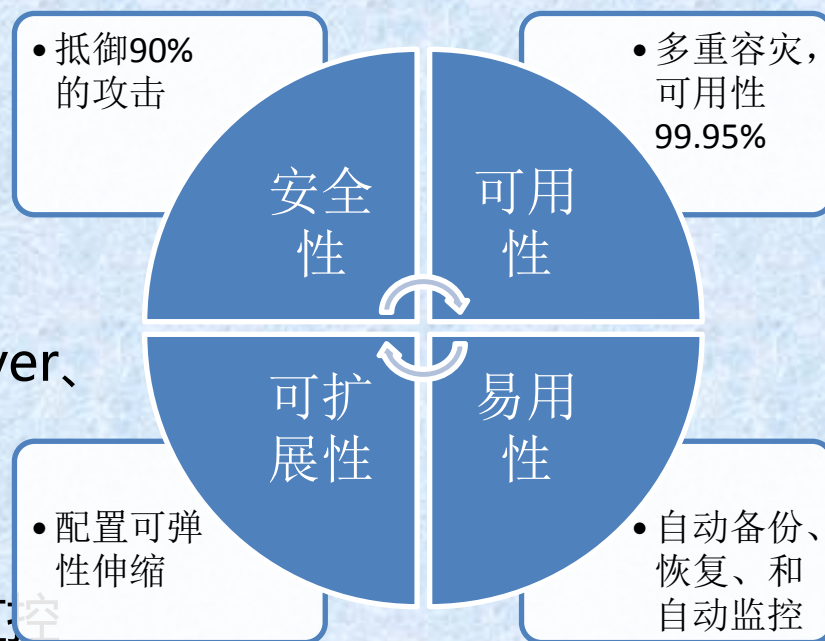
□ 在线的关系数据库服务

□ 即开即用

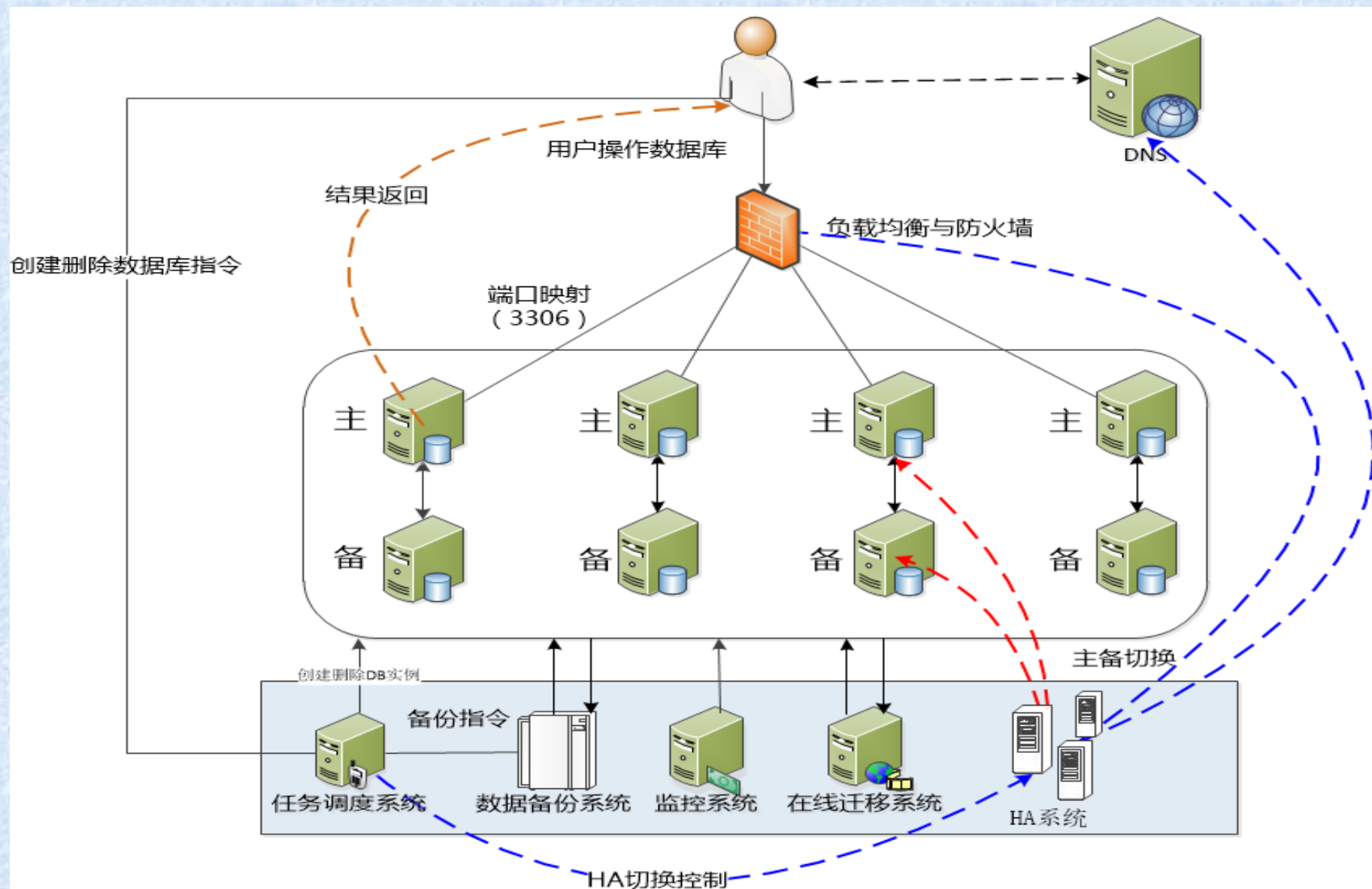
□ 支持四种DB引擎;MySQL、SQLServer、
postgresql , PPAS

□ 支持在线的扩容、备份回滚、性能监控
和分析等功能

RDS采用主从备份架构，具有高可用和数据可靠性

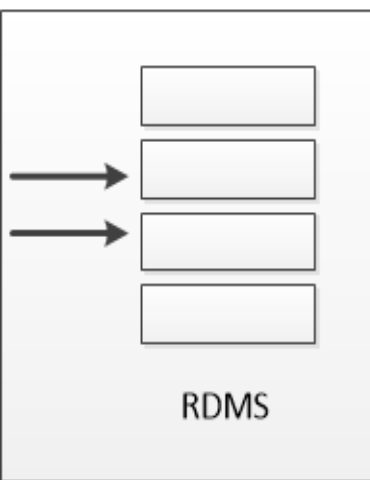
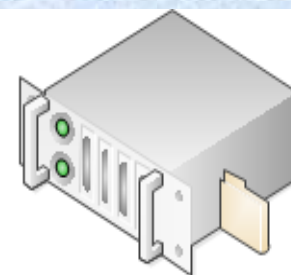
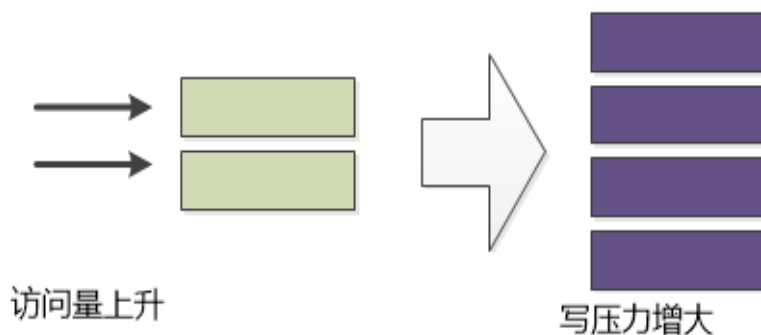


PAAS-云数据库服务-RDS总体架构

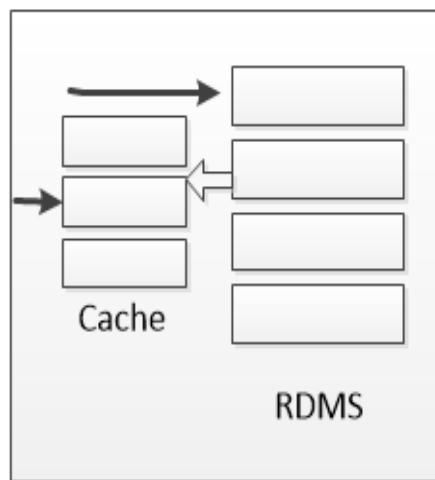


PAAS-云数据库服务-RDS

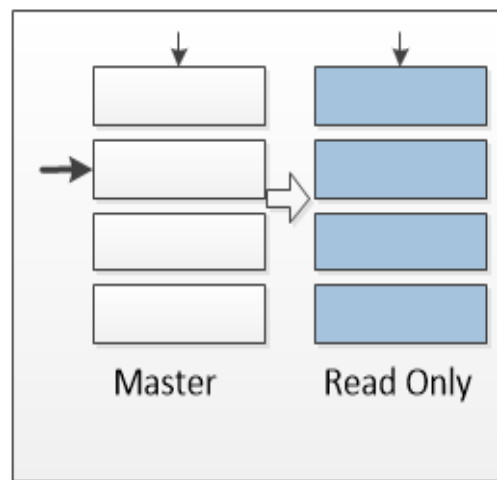
数据量性能瓶颈，单实例最大14000IOPS、25000OPS，承载2T



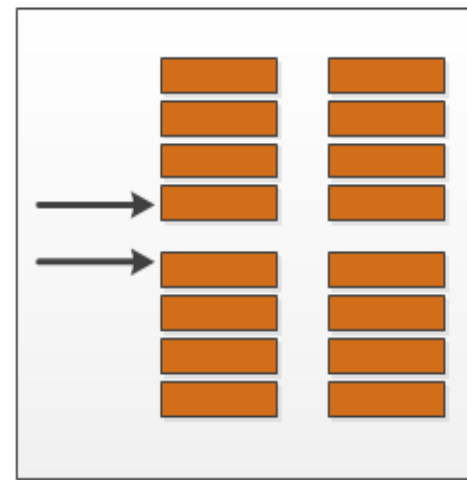
经典架构



缓存架构



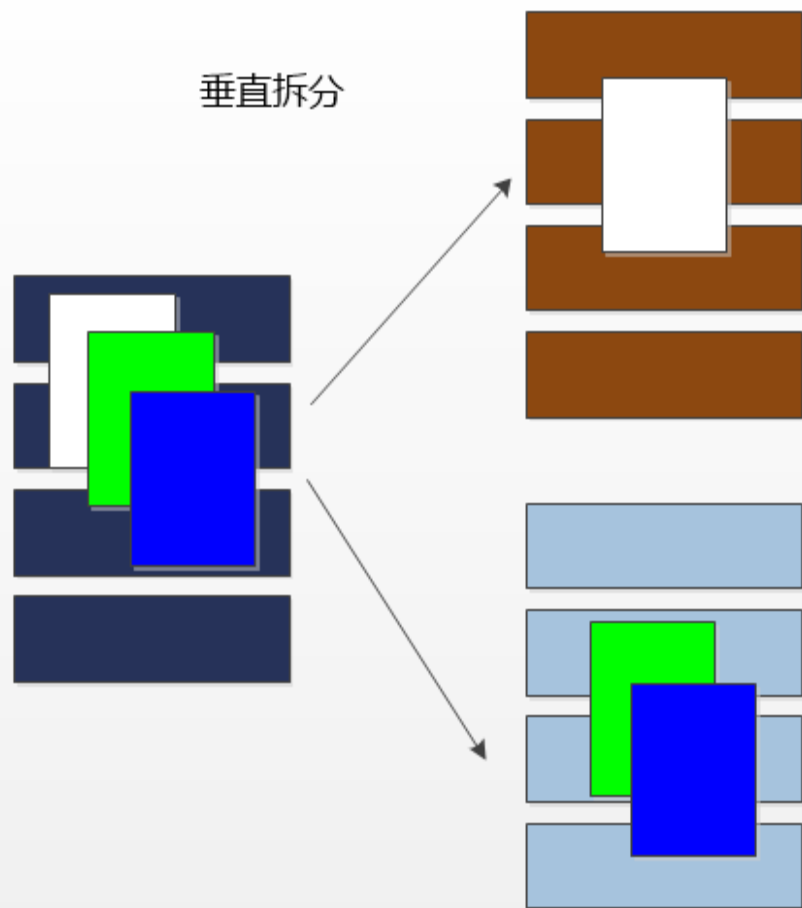
读写分离



分库分表

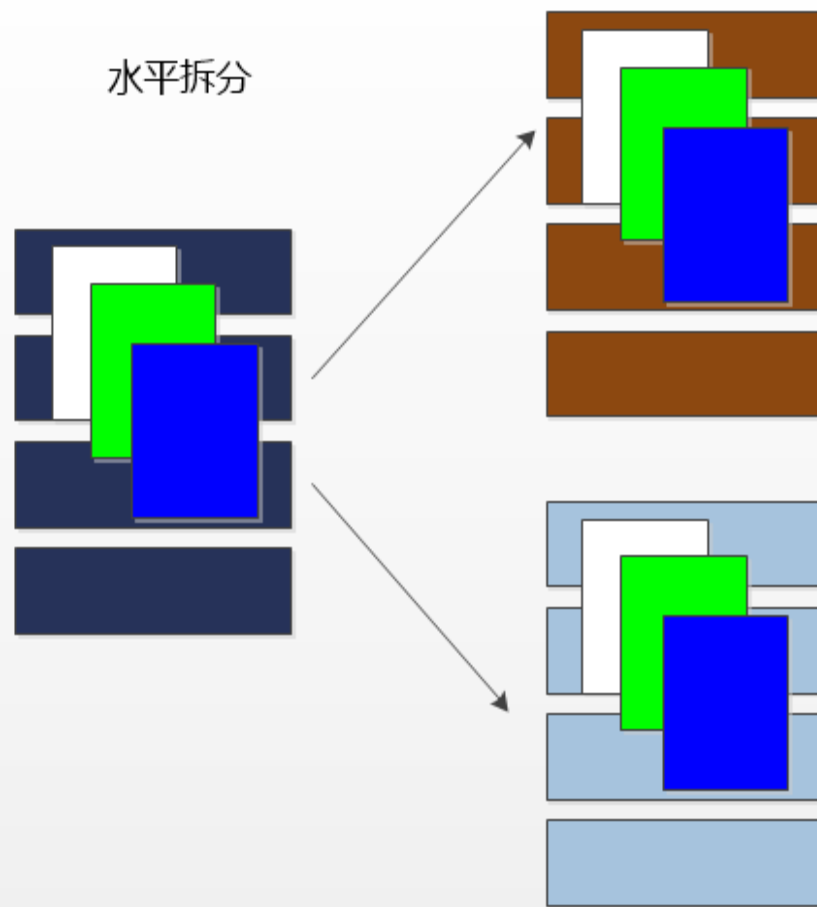
PAAS-云数据库服务-RDS

垂直拆分



解决了表与表之间的IO竞争

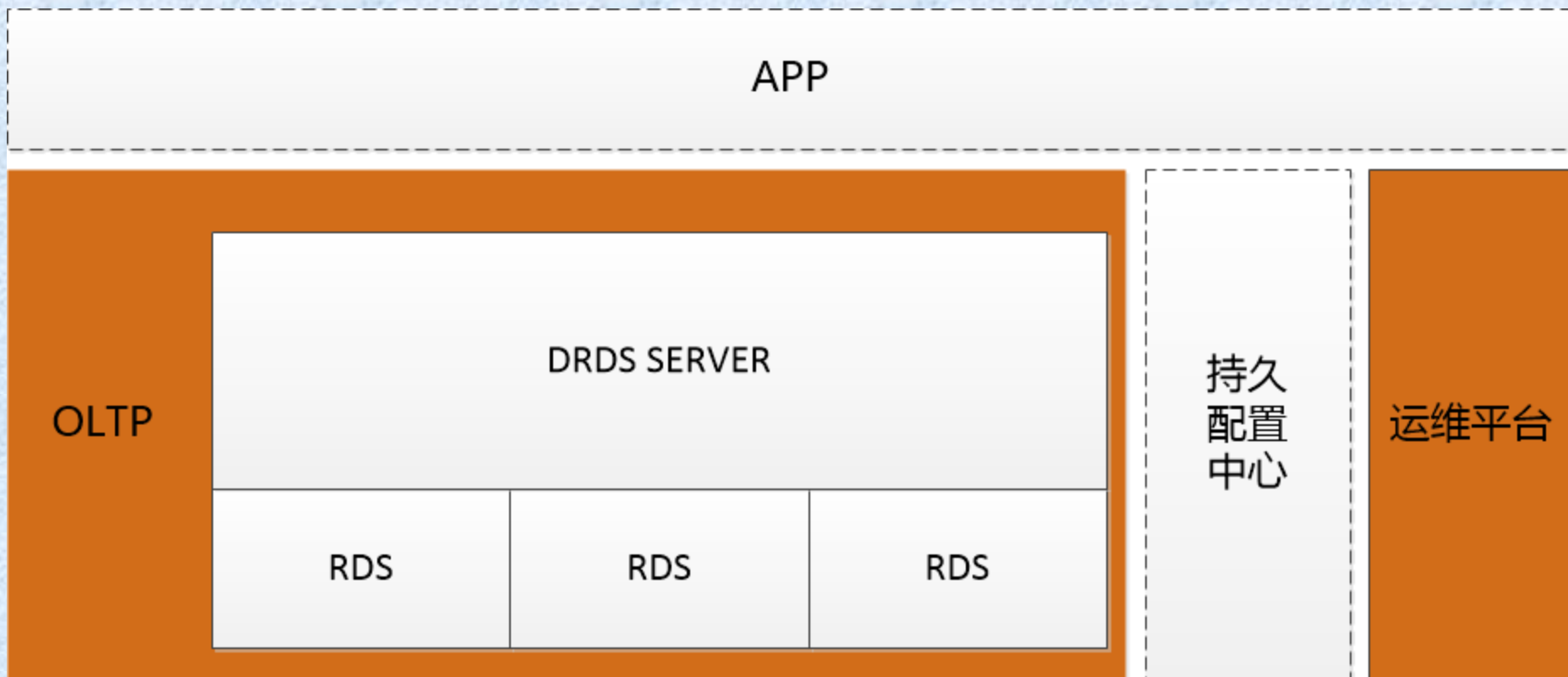
水平拆分



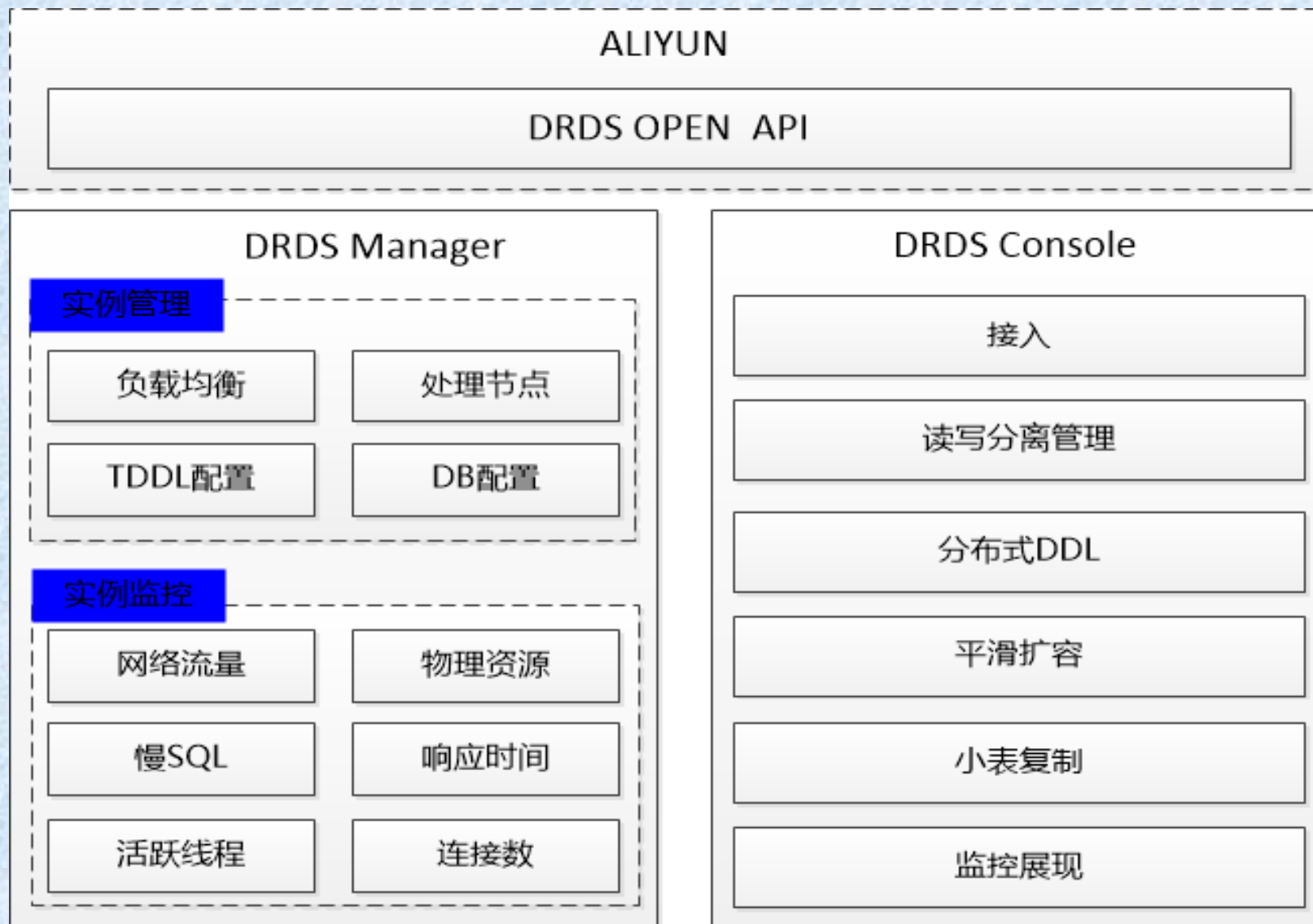
解决了单表数据量过大的问题

PAAS中间件-分布式关系数据库DRDS

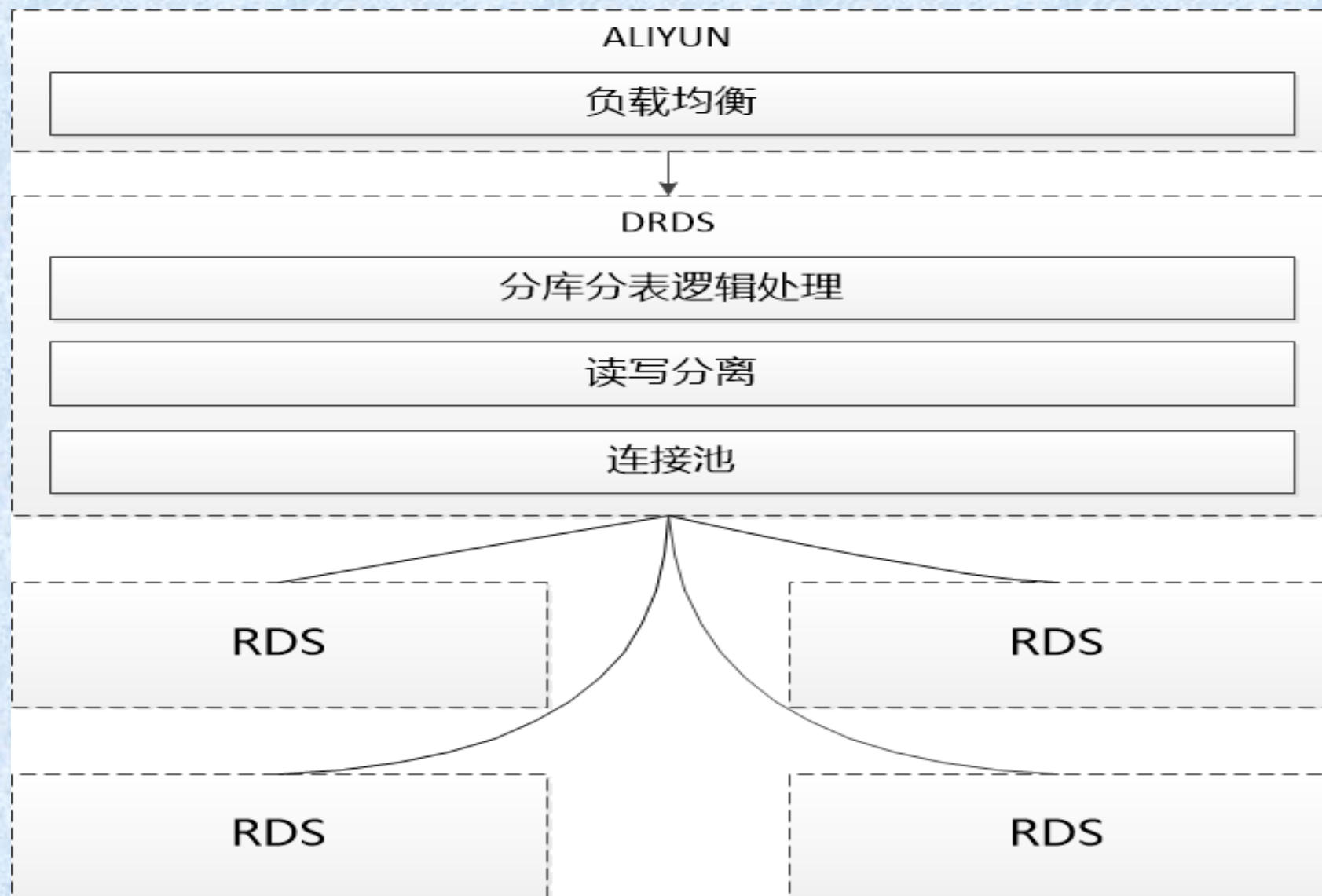
分布式关系数据库DRDS，是一种水平拆分、可平滑扩缩容，读写分离的在线分布式数据库服务。DRDS使用中间件产品。



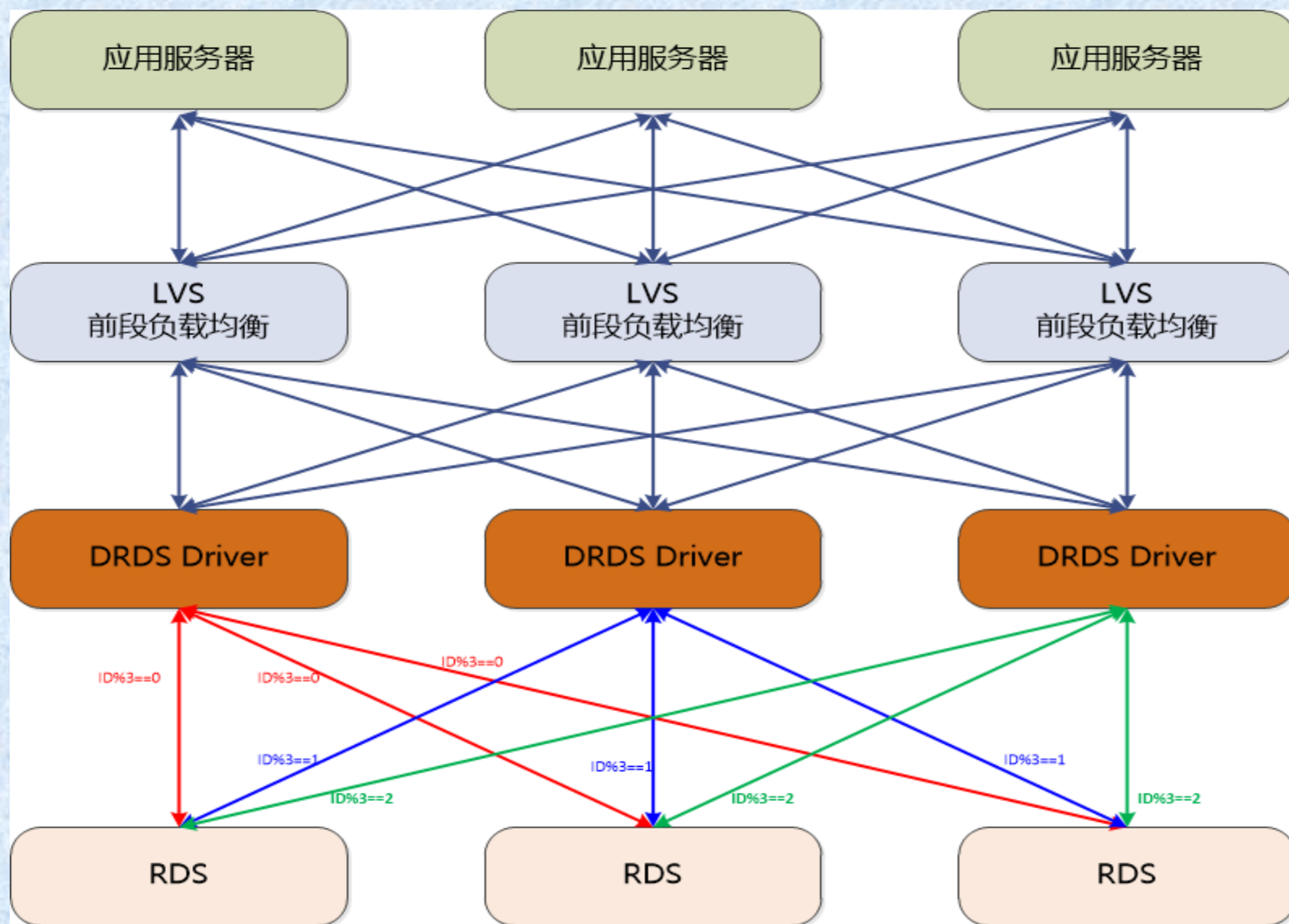
PAAS中间件-分布式关系数据库DRDS



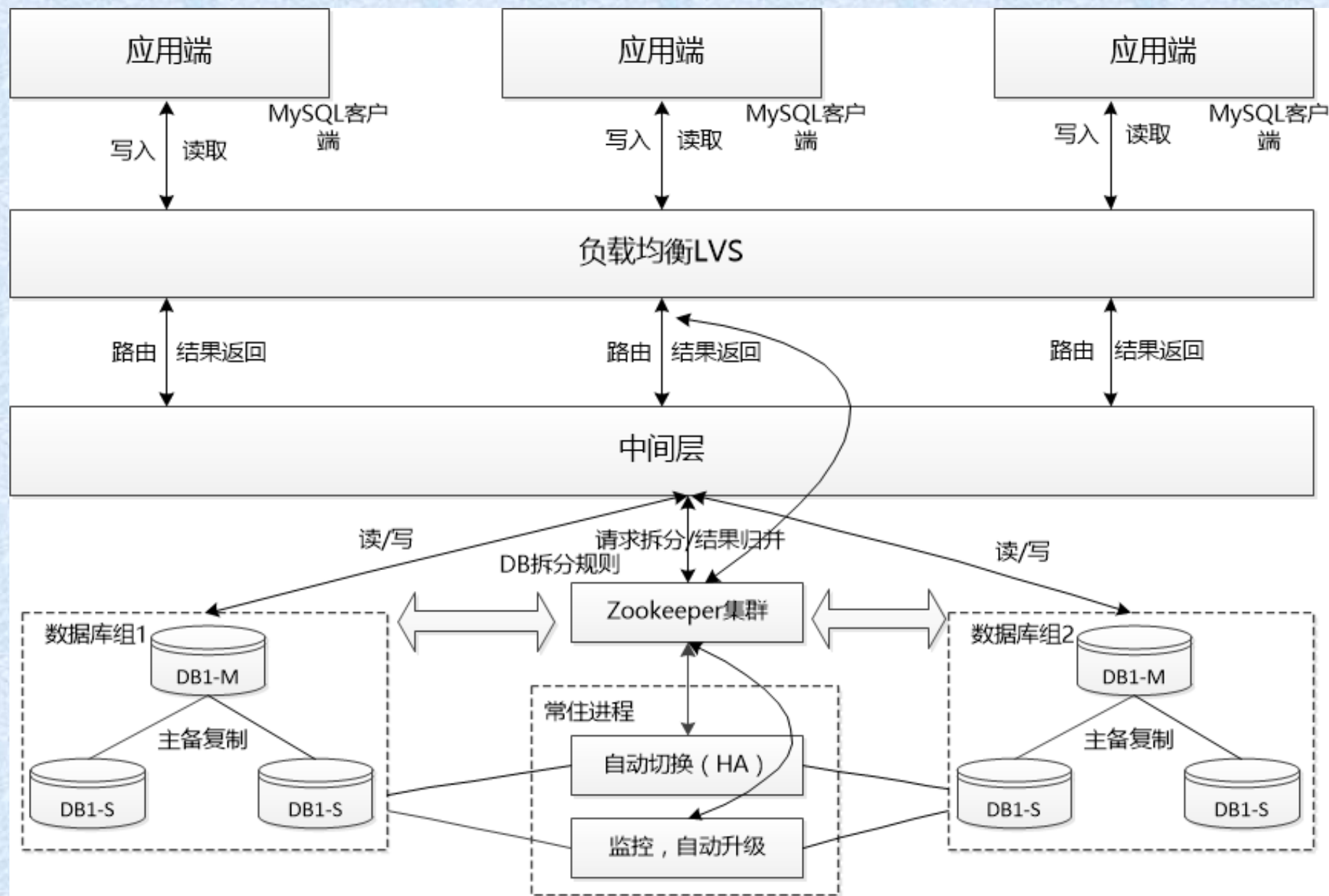
PAAS中间件-分布式关系数据库DRDS



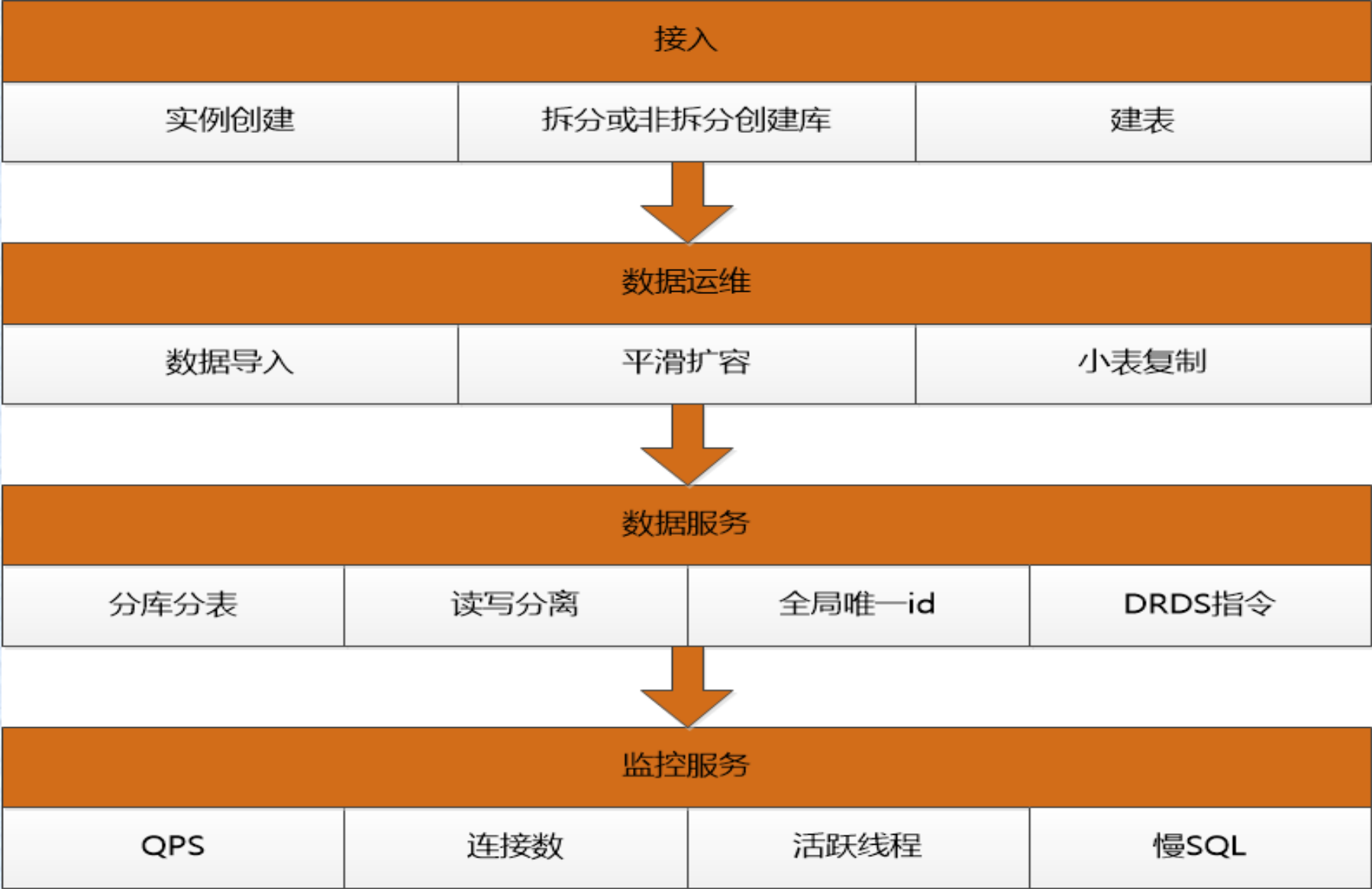
PAAS中间件-分布式关系数据库DRDS



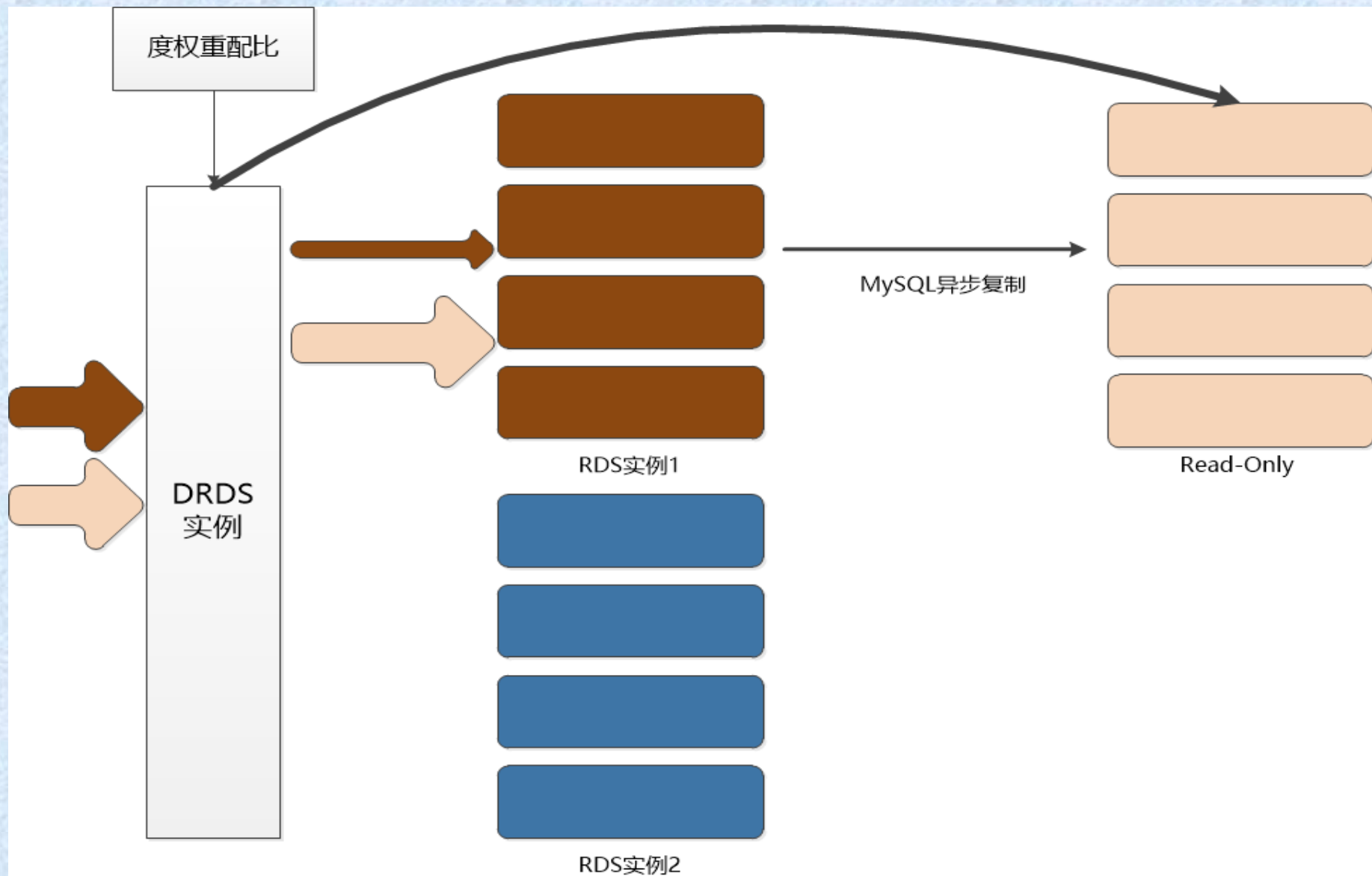
PAAS中间件-分布式关系数据库DRDS



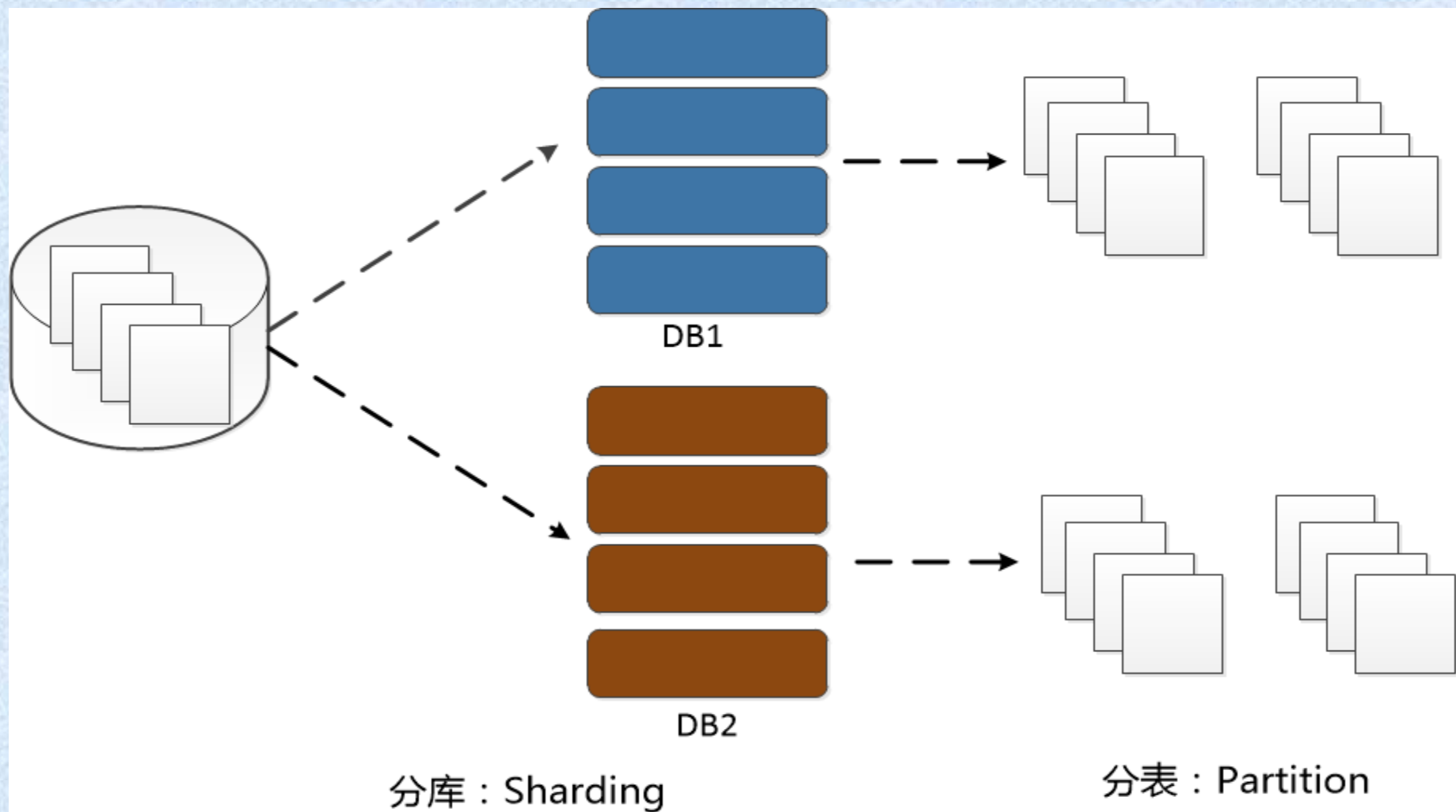
Paas中间件-分布式关系数据库DRDS-功能栈



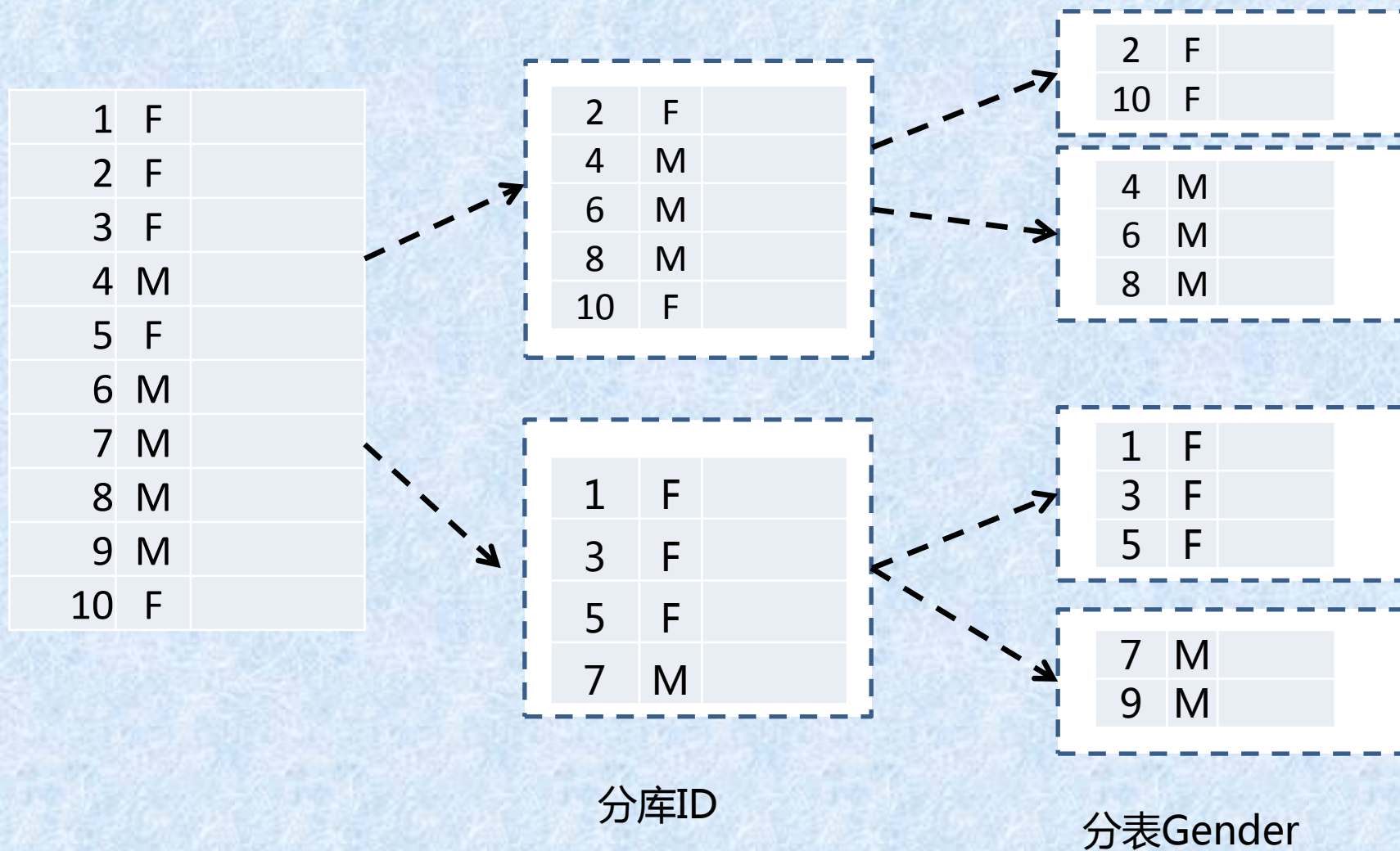
PAAS中间件-分布式关系数据库DRDS-读写分离



PAAS中间件-DRDS-分库分表



PAAS中间件-DRDS-分库分表



PAAS中间件-DRDS-SQL路由

按照查询条件做路由:

- 1: 定位数据库
- 2: 定位表分片

Select * from customer
Where id=7
And gender='M'

1	F	
2	F	
3	F	
4	M	
5	F	
6	M	
7	M	
8	M	
9	M	
10	F	

DRDS
实例

T1P1

T1P2

DB1

T1P3

T1P4

DB2

2	F	
10	F	

4	M	
6	M	
8	M	

1	F	
3	F	
5	F	

7	M	
9	M	

阿里云分析性数据库为什么快

- 底层存储使用SSD盘，更强的磁盘IO能力
- 表关联的时候，使用快速hash join
- 智能索引自动结合数据状况创建合理的索引
- 数据做就近处理、没有传输的开销
- 尽量在内存中做各种表的关联和计算

目录

24.1 阿里巴巴阿里云服务

24.1.1 阿里云计算体系架构

► 24.1.2 弹性计算服务（ECS）

24.1.3 开放存储服务OSS

24.1.4 开放结构化数据服务OTS

24.1.5 开放数据处理服务ODPS

24.1.6 关系型数据库RDS

24.1 阿里巴巴阿里云服务

- 弹性计算服务（ECS）

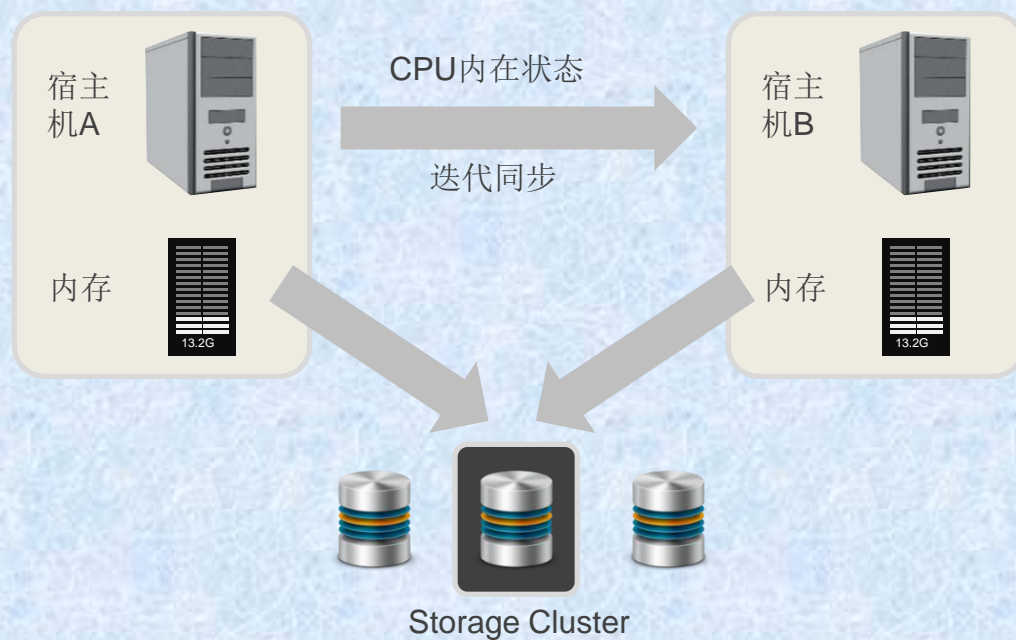
ECS底层基于分布式计算平台飞天，飞天平台

- 负责管理实际的硬件资源
- 向用户提供安全可靠的云服务器
- 任何硬件的故障都可以自动恢复
- 同时提供防网络攻击等高级功能
- 能够简化开发部署过程
- 降低运维成本
- 构建按需扩展的网络架构



24.1 阿里巴巴阿里云服务

- 弹性计算服务（ECS）



目录

24.1 阿里巴巴阿里云服务

24.1.1 阿里云计算体系架构

24.1.2 弹性计算服务（ECS）

► 24.1.3 开放存储服务OSS

24.1.4 开放结构化数据服务OTS

24.1.5 开放数据处理服务ODPS

24.1.6 关系型数据库RDS

24.1 阿里巴巴阿里云服务

- 开放存储服务OSS

开放存储服务（Open Storage Service，OSS），是阿里云对外提供的海量、安全和高可靠的云存储服务。它的主要特点如下。

弹性扩展

海量的存储空间，随着用户使用的增加，存储空间弹性增长，无需担心存储容量的限制。

大规模并发读写

数据并发读写，在短时间内可以进行大量数据的读/写操作。

图片处理优化

对存储在OSS上的图片，支持缩略、裁剪、水印、压缩和格式转换等图片处理功能。

目录

24.1 阿里巴巴阿里云服务

24.1.1 阿里云计算体系架构

24.1.2 弹性计算服务（ECS）

24.1.3 开放存储服务OSS

► 24.1.4 开放结构化数据服务OTS

24.1.5 开放数据处理服务ODPS

24.1.6 关系型数据库RDS

24.1 阿里巴巴阿里云服务

- 开放结构化数据服务OTS

开放结构化数据服务（Open Table Service，OTS）是构建在阿里云飞天分布式系统之上的NoSQL数据库服务，提供海量结构化数据的存储和实时访问。

数据的海量存储

支持互联网应用用于服务海量的终端用户

简单易用的表管理

用户根据业务需求创建多个实例进行管理

数据的管理

数据的单行读写、多行读写以及范围读取

目录

24.1 阿里巴巴阿里云服务

- 24.1.1 阿里云计算体系架构
- 24.1.2 弹性计算服务（ECS）
- 24.1.3 开放存储服务OSS
- 24.1.4 开放结构化数据服务OTS
- 24.1.5 开放数据处理服务ODPS
- 24.1.6 关系型数据库RDS

24.1 阿里巴巴阿里云服务

- 开放数据处理服务ODPS

开放数据处理服务（Open Data Processing Service，ODPS）[9]由阿里云自主研发，提供针对TB/PB级数据、实时性要求不高的分布式处理能力，应用于数据分析、挖掘、商业智能等领域。

海量运算

彻底无极限解决大数据存储与运算瓶颈，可以专心于数据分析和挖掘，最大化发挥数据价值。

数据安全

多层次数据存储和访问安全机制，保护数据不丢失、不泄露、不被窃取。

开箱即用

无需关心集群的搭建和运维，仅需简单的几步操作，即可开始数据的分析和挖掘任务。

目录

24.1 阿里巴巴阿里云服务

- 24.1.1 阿里云计算体系架构
- 24.1.2 弹性计算服务（ECS）
- 24.1.3 开放存储服务OSS
- 24.1.4 开放结构化数据服务OTS
- 24.1.5 开放数据处理服务ODPS
- 24.1.6 关系型数据库RDS

24.1 阿里巴巴阿里云服务

- 关系型数据库RDS

关系型数据库云服务（Relational Database Service, RDS）[9]，通过云服务的方式让关系型数据库设置、操作和扩展变得更加简单。

安全稳定，数据可靠

RDS集群处于多层防火墙的保护之下，可以有力地抗击各种恶意攻击，保证数据的安全

自动备份，管理透明

RDS根据自定义的备份策略自动备份数据库，防止数据丢失和误删除，保证数据安全可靠

性能卓越，灵活扩容

采用高端高性能服务器配置，为高性能提供了有效的硬件平台。

目录

24.2 云创存储万物云服务

- ▶ 24.2.1 平台简介
- 24.2.2 系统架构
- 24.2.3 功能服务

24.2 云创存储万物云服务

- 平台简介



一个数据服务逻辑层和一套面向应用的编程接口，满足物联网应用各个层次的数据存储、查询、处理需求，保障用户数据安全和 服务稳定，并提供一系列协助用户开发调试、监控性能和优化性能的工具。

目录

24.2 云创存储万物云服务

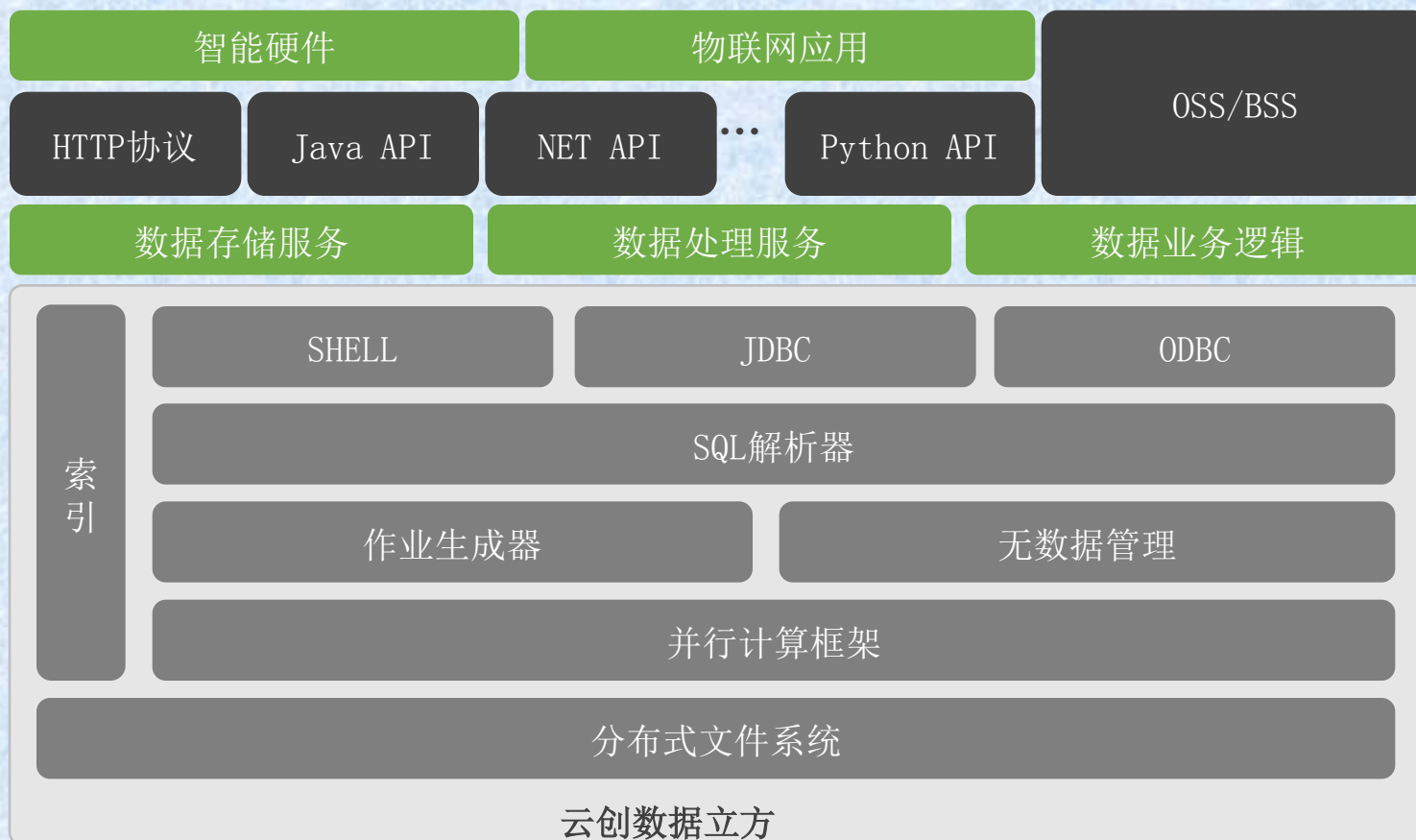
24.2.1 平台简介

► 24.2.2 系统架构

24.2.3 功能服务

24.2 云创存储万物云服务

- 系统架构



目录

24.2 云创存储万物云服务

24.2.1 平台简介

24.2.2 系统架构

► 24.2.3 功能服务

24.2 云创存储万物云服务

- 功能服务



数据存储服务

提供海量、弹性、安全、高可用和高可靠的云存储

数据处理应用服务

提供针对TB/PB级数据、实时性要求不高的处理服务，主要应用于数据挖掘和数据智能分析等领域

24.2 云创存储万物云服务

- 功能服务

编程接口

平台提供丰富的编程接口，实现了大数据平台所遵循的设计理念。

基于HTTP协议的数据服务调用接口

构建基于HTTP的 RESTful协议接口的目的主要是使智能硬件便捷地通过发送HTTP POST请求直接访问平台大数据服务。

应用程序编程接口（JavaAPI等）

大数据平台通过提供软件应用开发包的形式为物联网应用提供具有针对性的数据应用服务，目的在于减少物联网应用端的开发量。

24.2 云创存储万物云服务

- 功能服务

操作名称	操作目标
AddTableRow	插入一行数据
AddTableRows	插入多行数据
GetTableRow	读取一个表中单行数据
GetTableRows	批量读取一个表中若干行数据
GetTableRowsByRange	读取指定主键范围内的数据
DeleteTableRow	删除一个表中一行数据
DeleteTableRows	删除一个表中若干行数据
CreateTable	根据给定的表结构信息创建相应的表
DeleteTable	删除指定的表
GetAllTableInfo	获取当前用户所有表的结构信息
GetTableInfo	获取指定表的结构信息

24.2 云创存储万物云服务

- 功能服务

编程接口方法调用示例

1. 智能硬件数据提交示例

调用地址	http://cdsserver: 8080/api/addTableRow
方法	POST
参数示例	<pre>user={"AccessID":"user", "AccessKey": "xxxxxxxx"} row={"values": {"location": "123245", "temp": "21"}} tableName={"tableName"}</pre>
返回	1

2. 物联网应用数据读取示例

- 创建CDS对象，在构造函数中指定供服务安全认证所需的AccessID和AccessKey;
- 构造请求对象;
- 调用CDSClient对象相关接口发送请求。

4.2 云创存储万物云服务

- 功能服务

数据安全机制

访问许可验证

面对数据处理服务的请求，物联网大数据平台通过使用 **AccessID/AccessKey** 对称加密的方法来验证发送请求的用户身份

大数据平台对用户数据的建表操作采用用户名+实例名+表名

攻击防范机制

大数据平台内建了基本的攻击监测及防范措施

24.2 云创存储万物云服务

- 功能服务

智能硬件直通接入方案

- 用户可以使用**HTTP、TCP或MQTT**等协议接入智能硬件。
- 物联网智能设备基于常用**嵌入式MCU**，如ARM、Intel等，可通过**SPI/RS2485/RS232/I2C**等接口外接各种传感器。
- 物联网智能设备搭载的**标准Linux、mbed OS**等操作系统均支持CoAP、HTTP、MQTT、LWM2M等协议，支持多种通信手段，包括3G/LTE/Bluetooth Smart/Wi-Fi及6toWPAN。
- 智能设备通过**无线网络**将数据可靠传输到平台服务器，为用户提供一个涵盖**数据采集、可靠传输、大数据存储和处理**的完整解决方案。

24.2 云创存储万物云服务

- 功能服务

- 规模和性能

- 数据立方存储系统支持**弹性扩展**，用户无需担心存储空间不足。
- 分布式存储系统中**各存储节点副本数据实时同步**，读写性能不会因数据量增加而受影响。
- 现有平台提供**毫秒级单行数据读/写延迟**，数据入库可在**每秒十万条以上量级**，查询在**百万QPS级别**。
- 根据物联网应用的特点，物联网大数据平台还提供一系列有针对性的工具，如**数据迁移同步、性能监控、辅助调试工具**等。