Alibaba Group 阿里巴巴集团 | 阿里技术保障 Alibaba Infrastructure Service

支付宝的高可用与容灾架构演进

曾欢 @阿里巴巴

shanheng.zh@alibaba-inc.com

Geekbang.
极客邦科技

整合全球最优质学习资源，帮助技术人和企业成长

Growing Technicians, Growing Companies

InfoQ

专注中高端技术人员的
技术媒体

EGO EXTRA GEEKS' ORGANIZATION NETWORKS

高端技术人员
学习型社交网络

StuQ

实践驱动的
IT职业学习和服务平台

GiT GEEKBANG INTERNATIONAL TRAINING 极客邦培训

一线专家驱动的
企业培训服务

地址：北京市朝阳区洛娃大厦C座6层1607室

网址：www.geekbang.org

个人经历

2012年

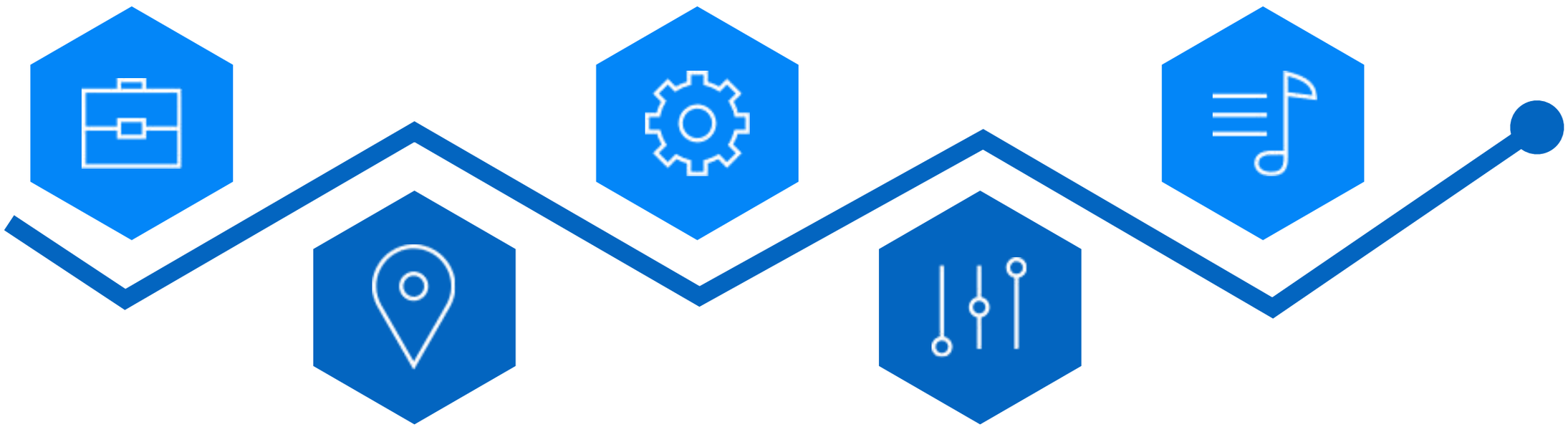
加入技术保障部应用运维团队

2013-2014年

负责天猫、淘宝等相关核心系统的运维工作

2015年

负责蚂蚁金服运维，稳定性相关



连续两年『天猫双十一卫士』业务稳定性平台

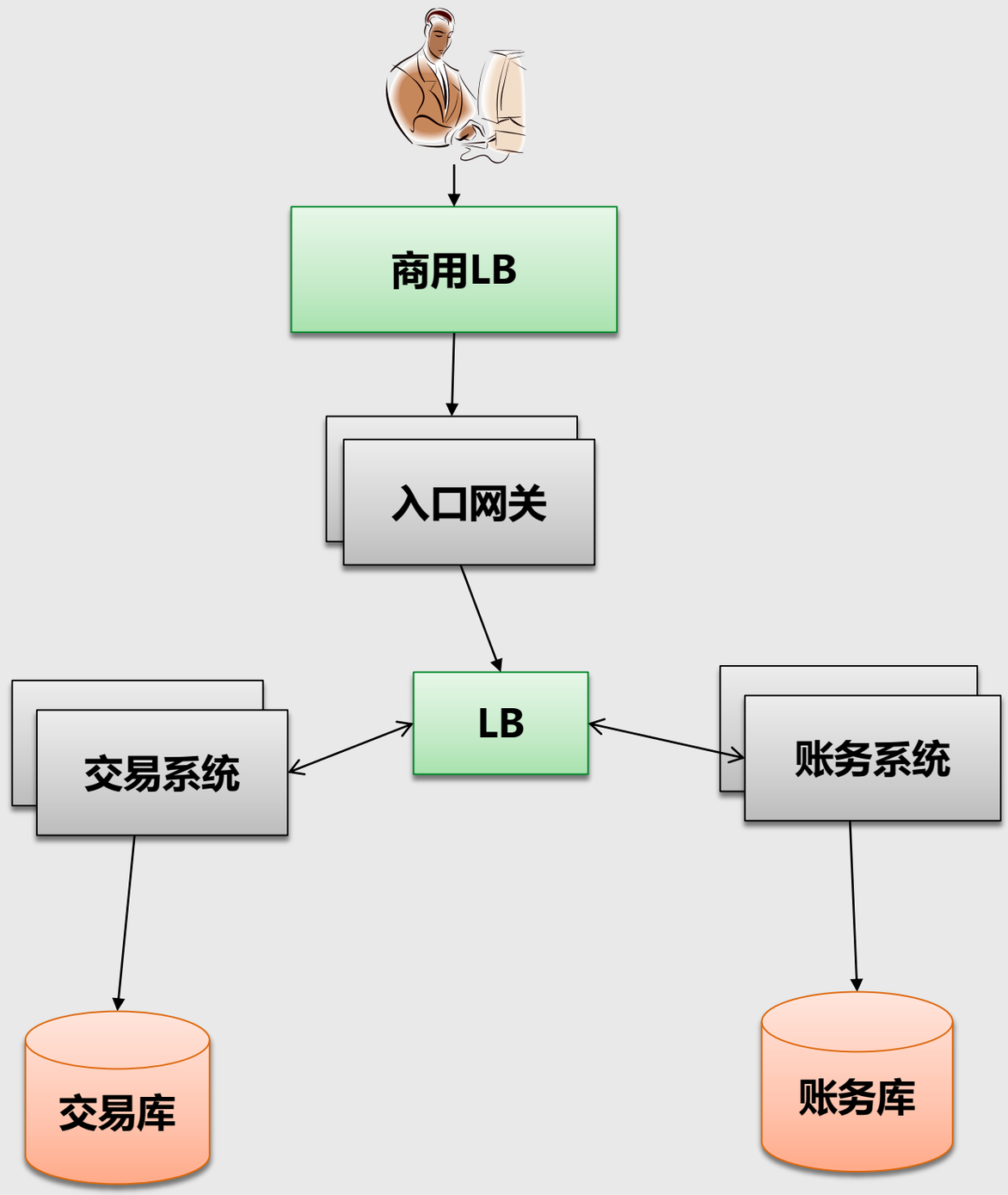
蚂蚁异地多活项目

目录

纯真-童年时期

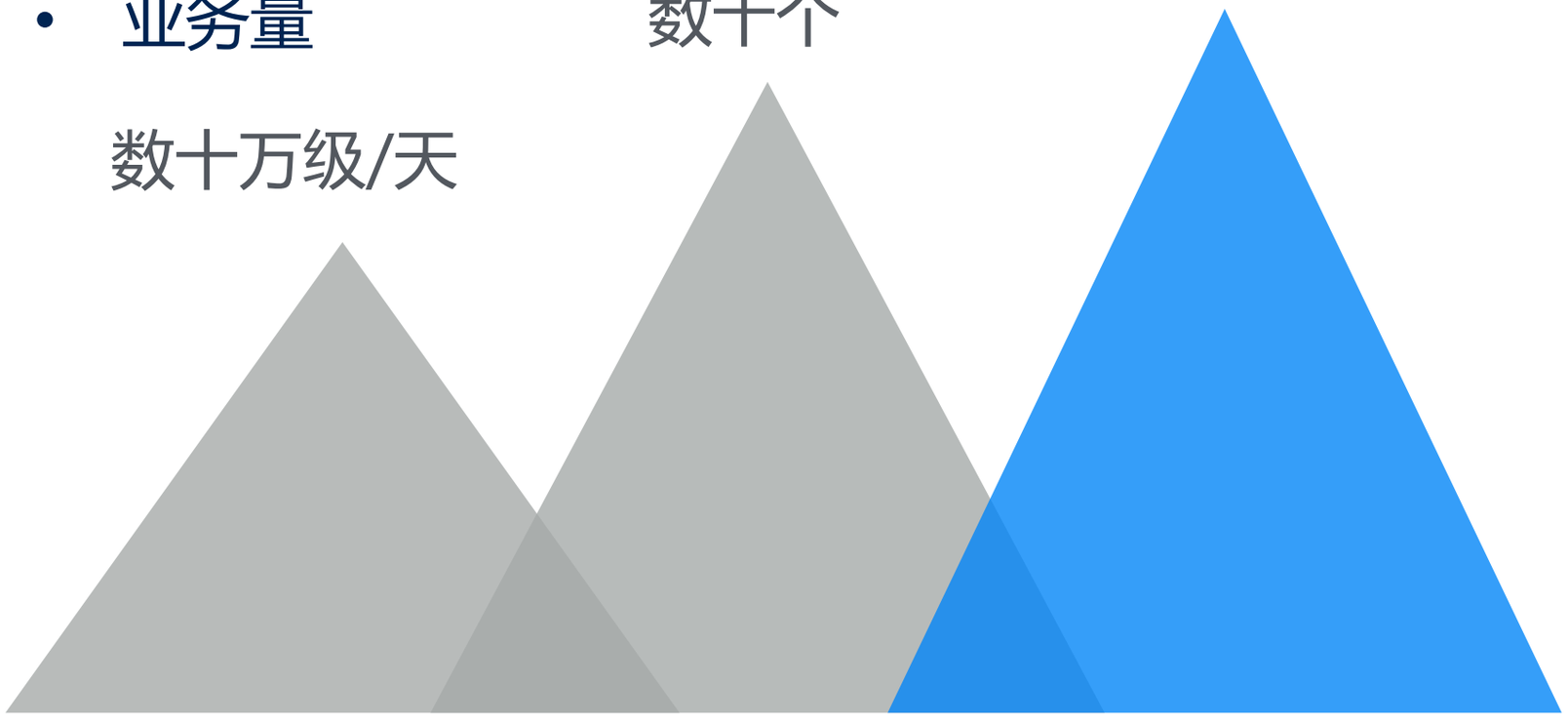
1

初期系统架构



『单机房』架构的背后

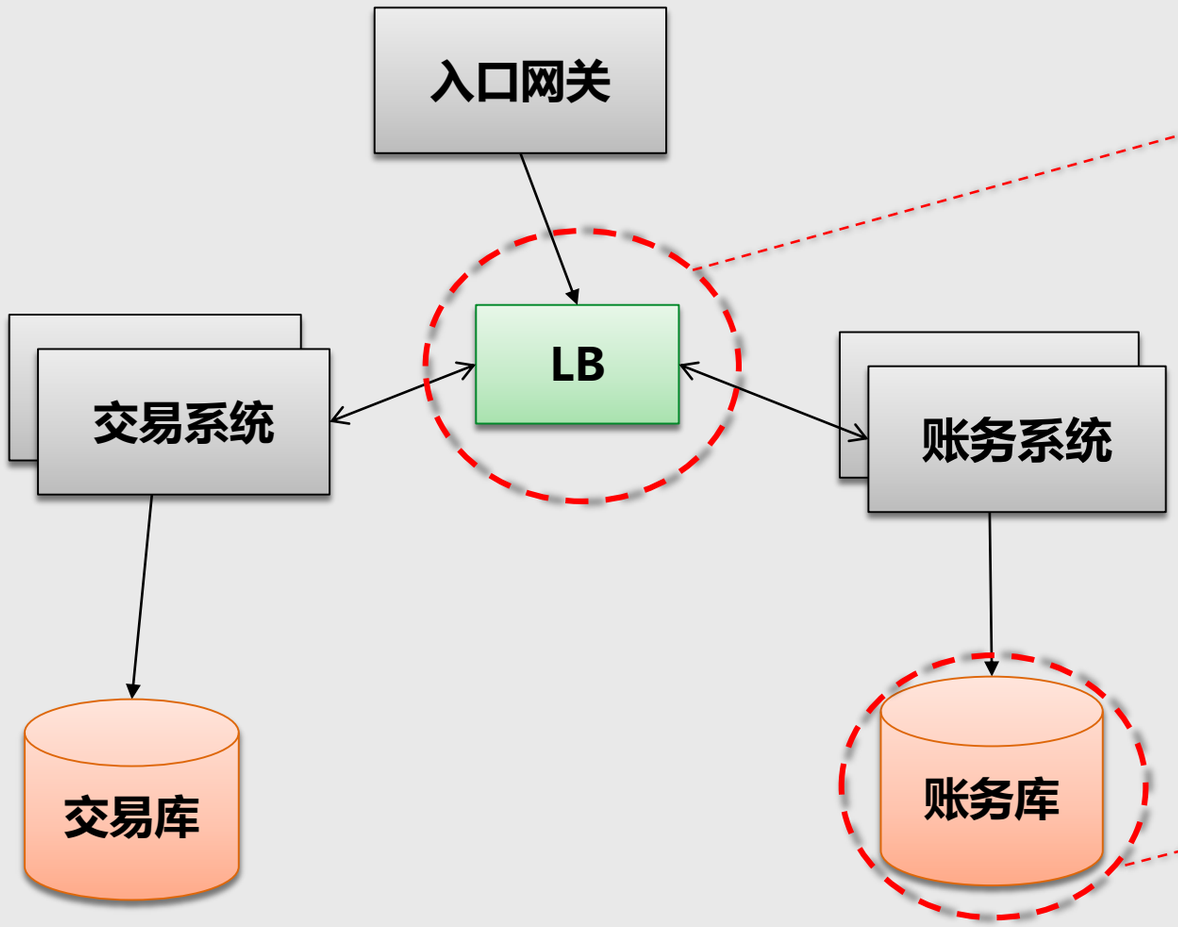
- 应用系统 数十个
- 业务量 数十万级/天



- 容灾能力低

  单应用出现问题无法切换
  主备切换业务中断
  停机维护

业务量增长带来的问题



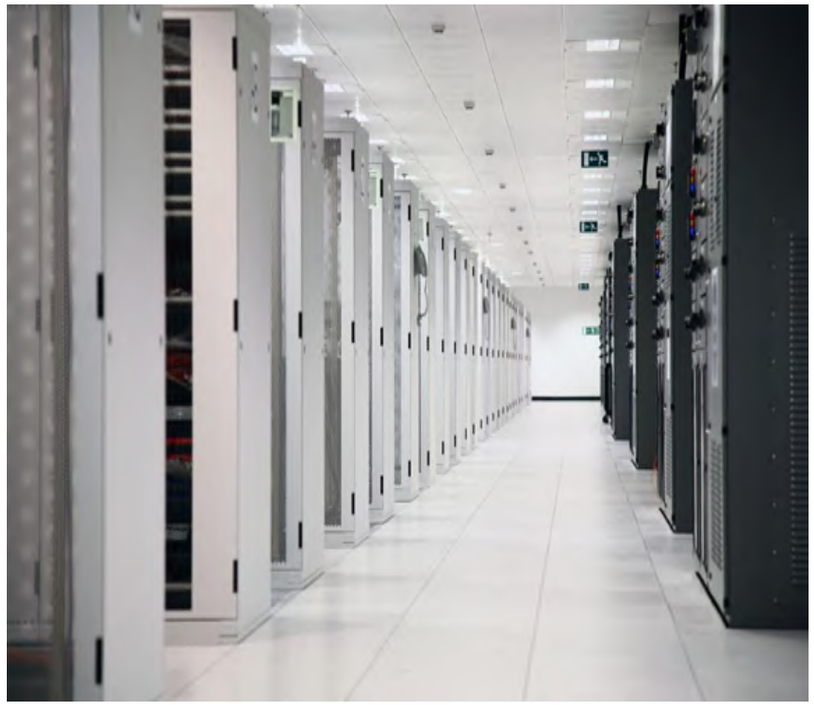
- 商用LB瓶颈：VIP数量只增不减，LB宕机影响巨大
- 核心DB单库瓶颈

新架构要解决的问题：

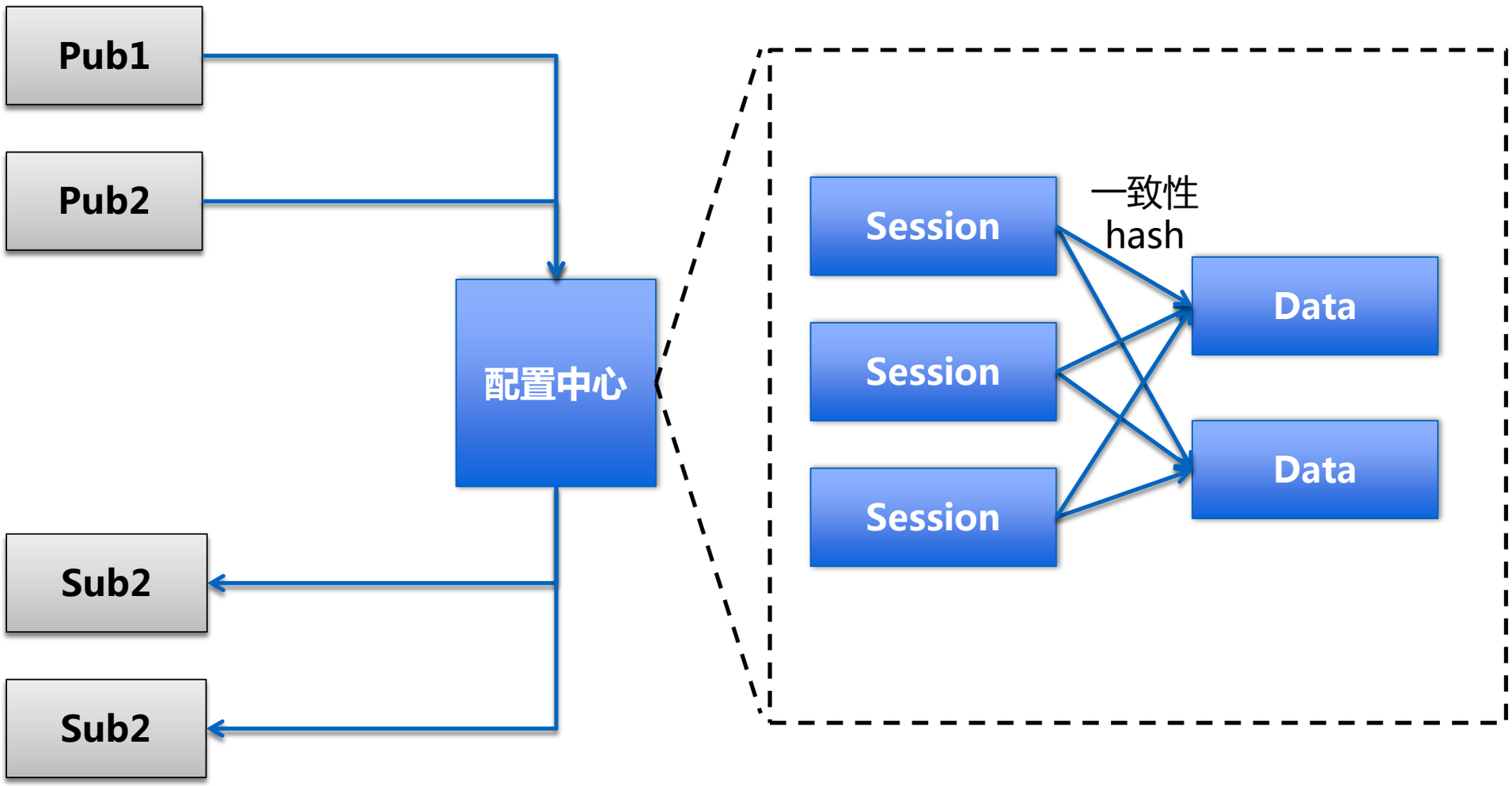
- 支撑日益增长的业务量：数百万级/天
- 解决容量问题：数百个系统
- 消除架构单点

懵懂-少年时期

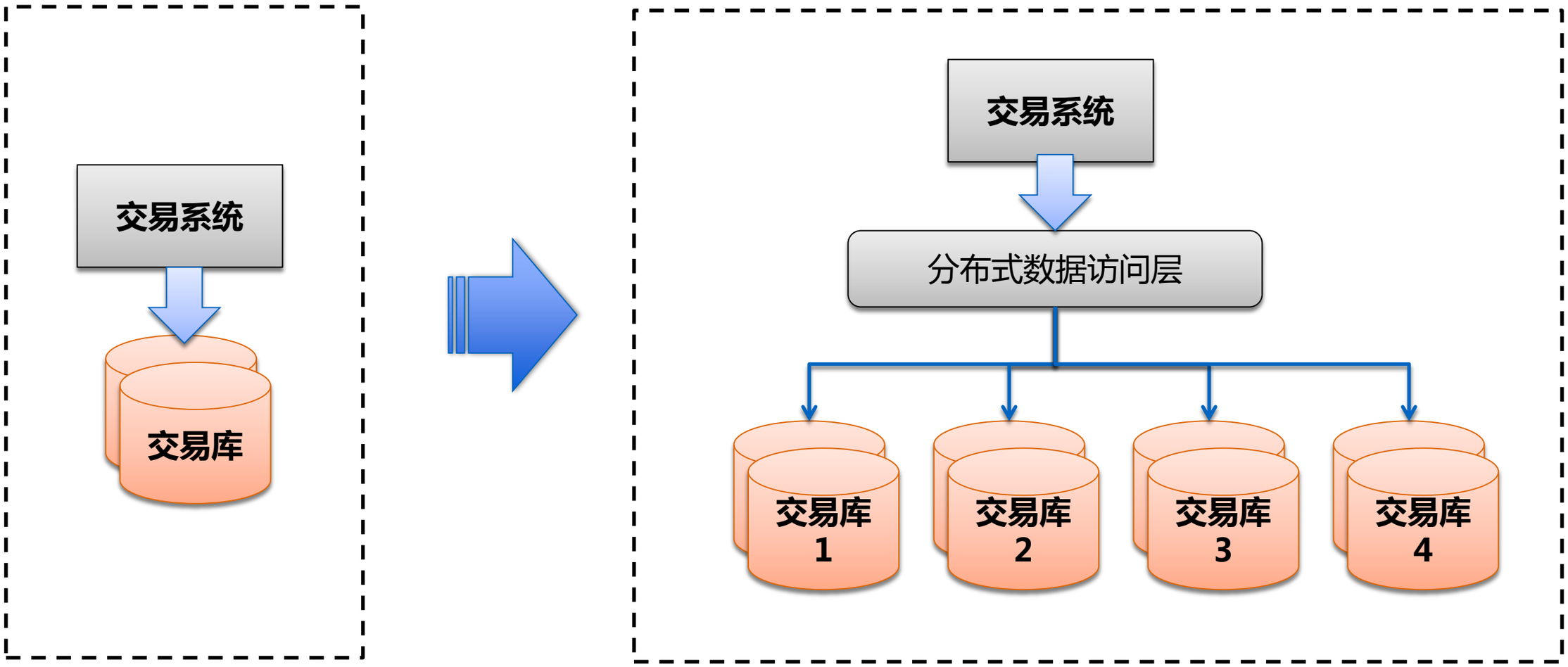
2



硬负载=>软负载：分布式服务调用



数据可扩展：UID水平拆分



GSLB

50%

50%

机房A

配置中心

交易系统

账务系统

交易库01

账务库01

交易备库2

账务备库2

机房B

配置中心

交易系统

账务系统

交易库02

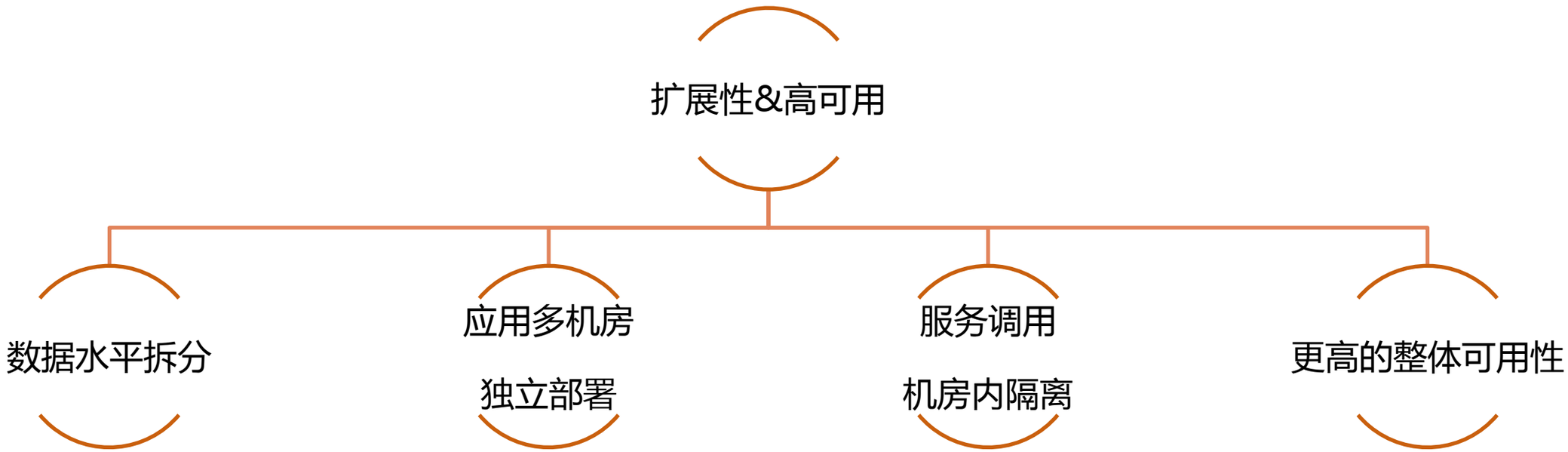
账务库02

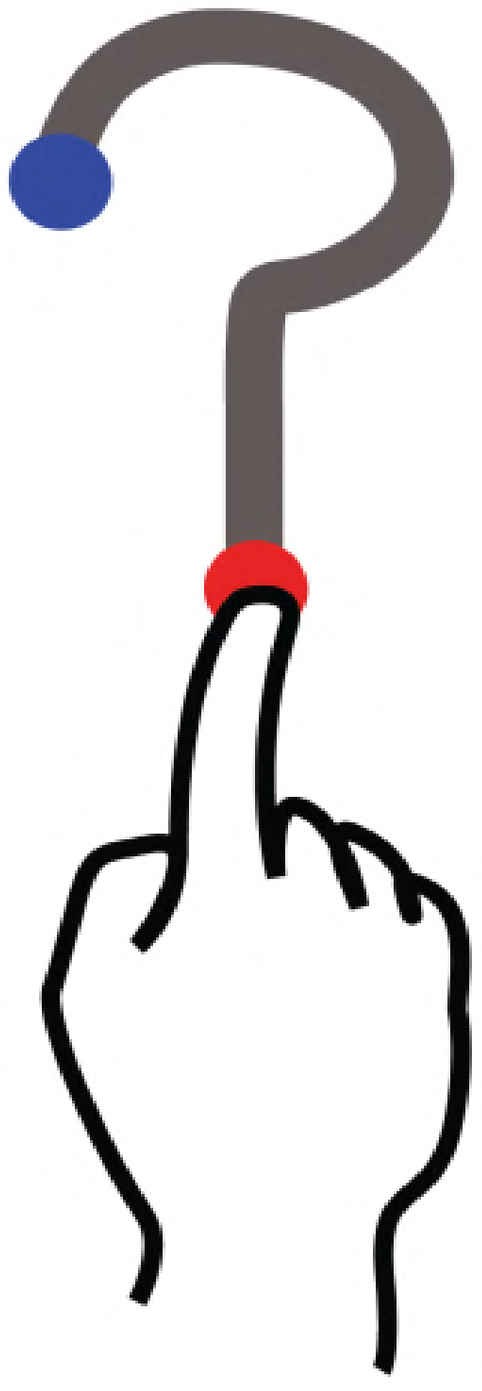
交易备库1

账务备库1

数据同步

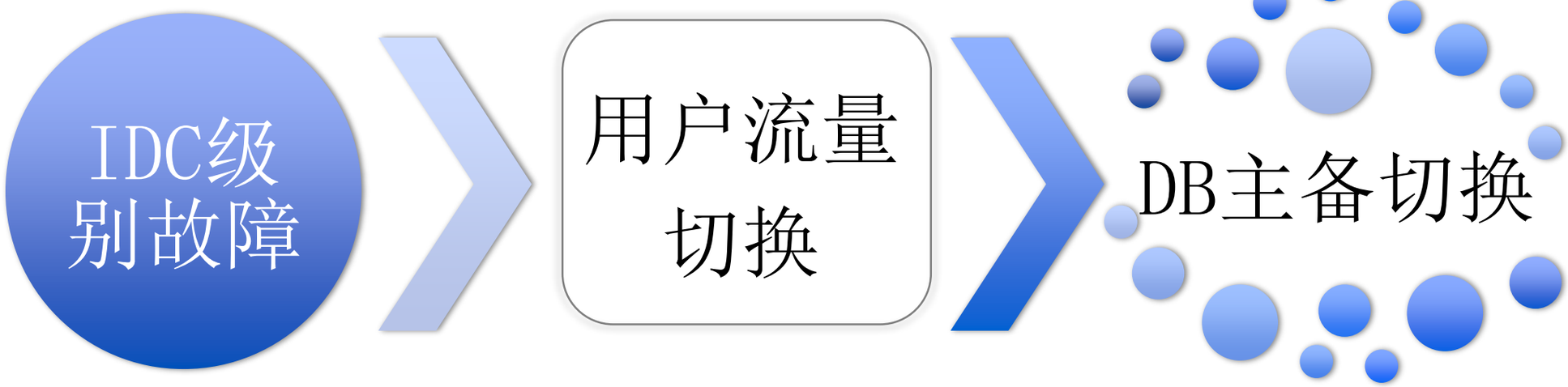
『多机房』架构优势





够用了吗？

容灾切换过程

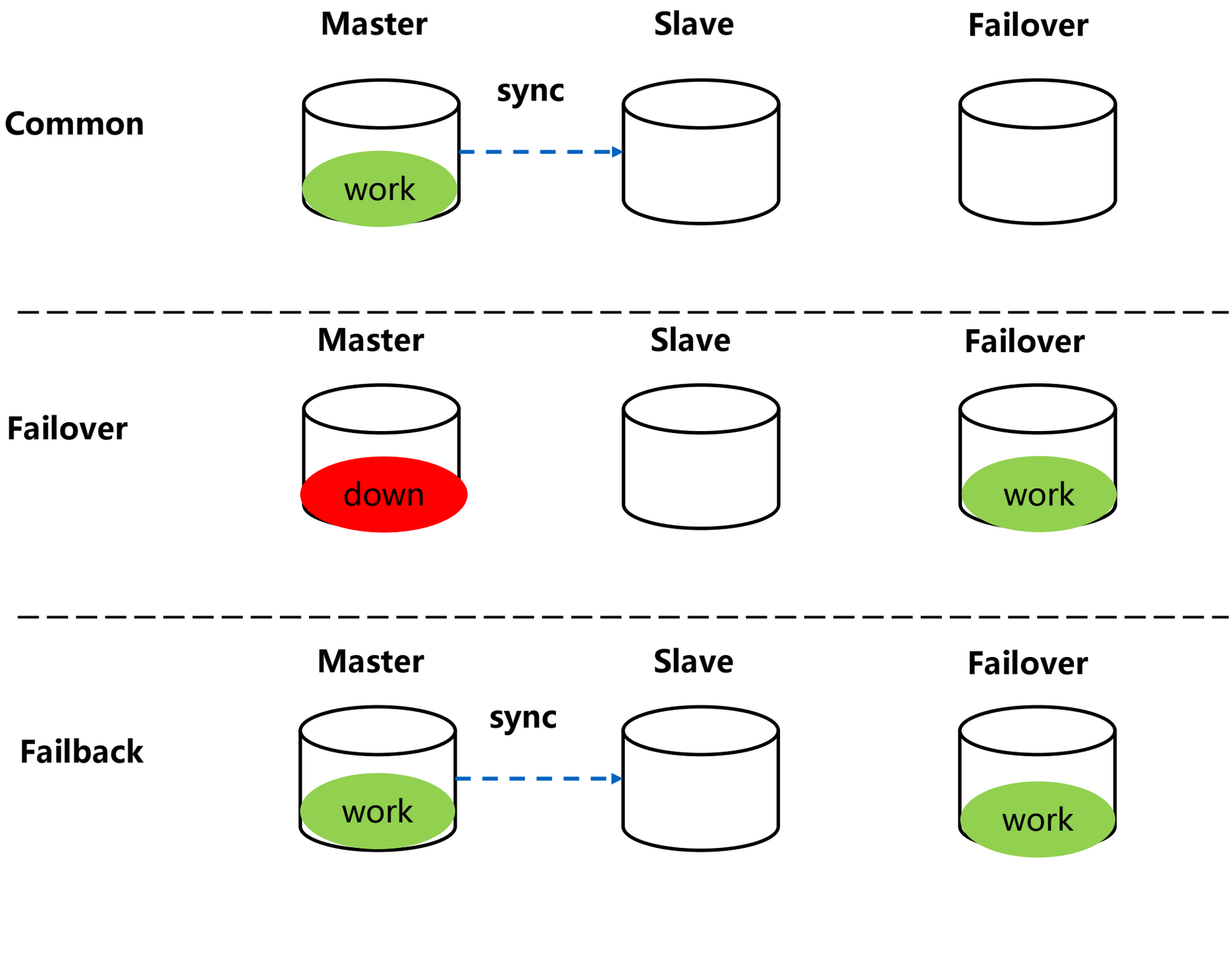


Oracle异常情况下的主备切换10min

数据高可用

FAILOVER方案

Failover方案


Master
Slave
Failover
Common
sync
work
Master
Slave
Failover
Failover
down
work
Master
Slave
Failover
Failback
sync
work
work

GSLB

50%

50%

机房A

配置中心

交易系统

账务系统

交易库01

账务库01

交易FO库2

交易备库2

账务备库2

机房B

配置中心

交易系统

账务系统

交易FO库1

交易库02

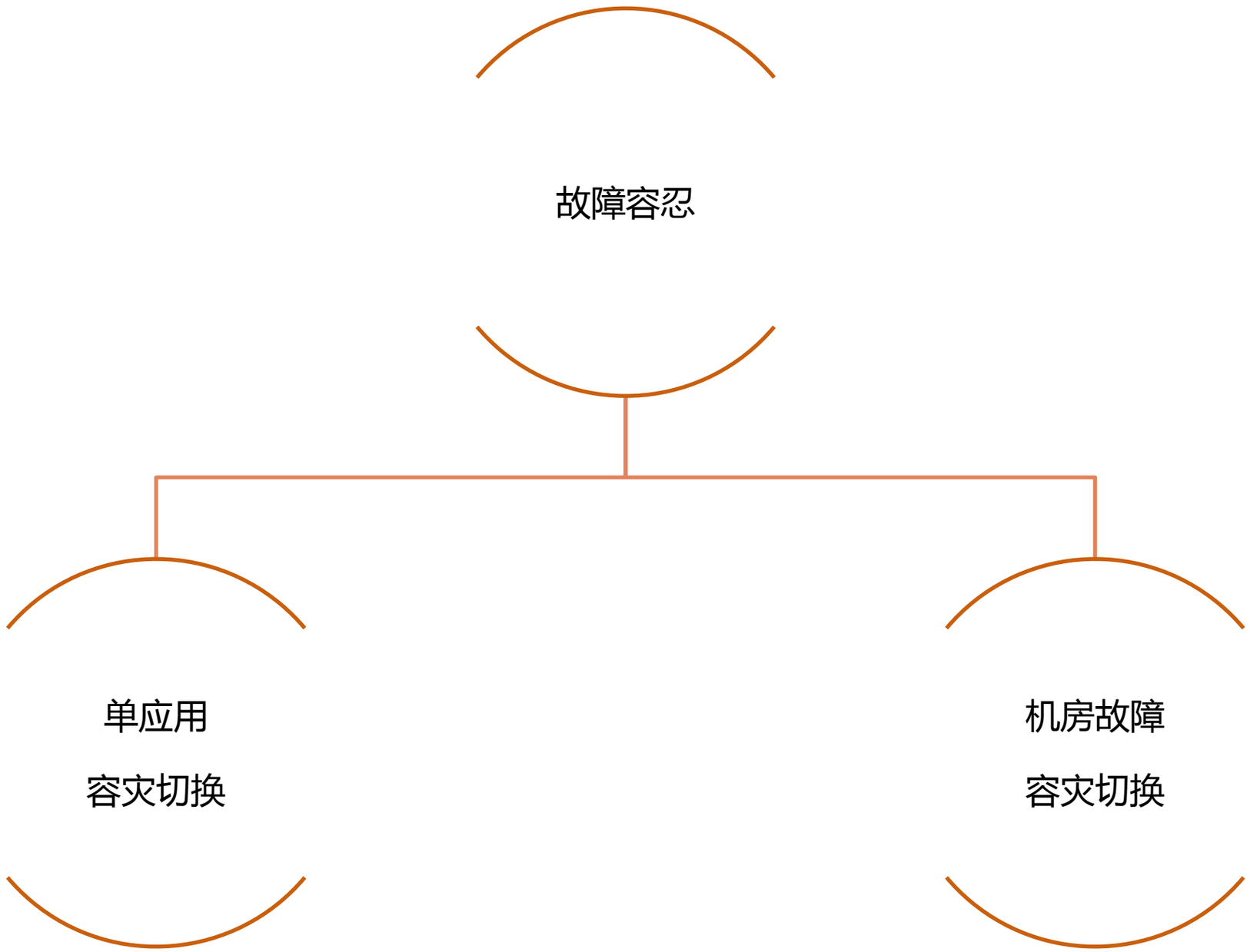
账务库02

交易备库1

账务备库1

数据同步

『多机房多活』架构的容灾能力

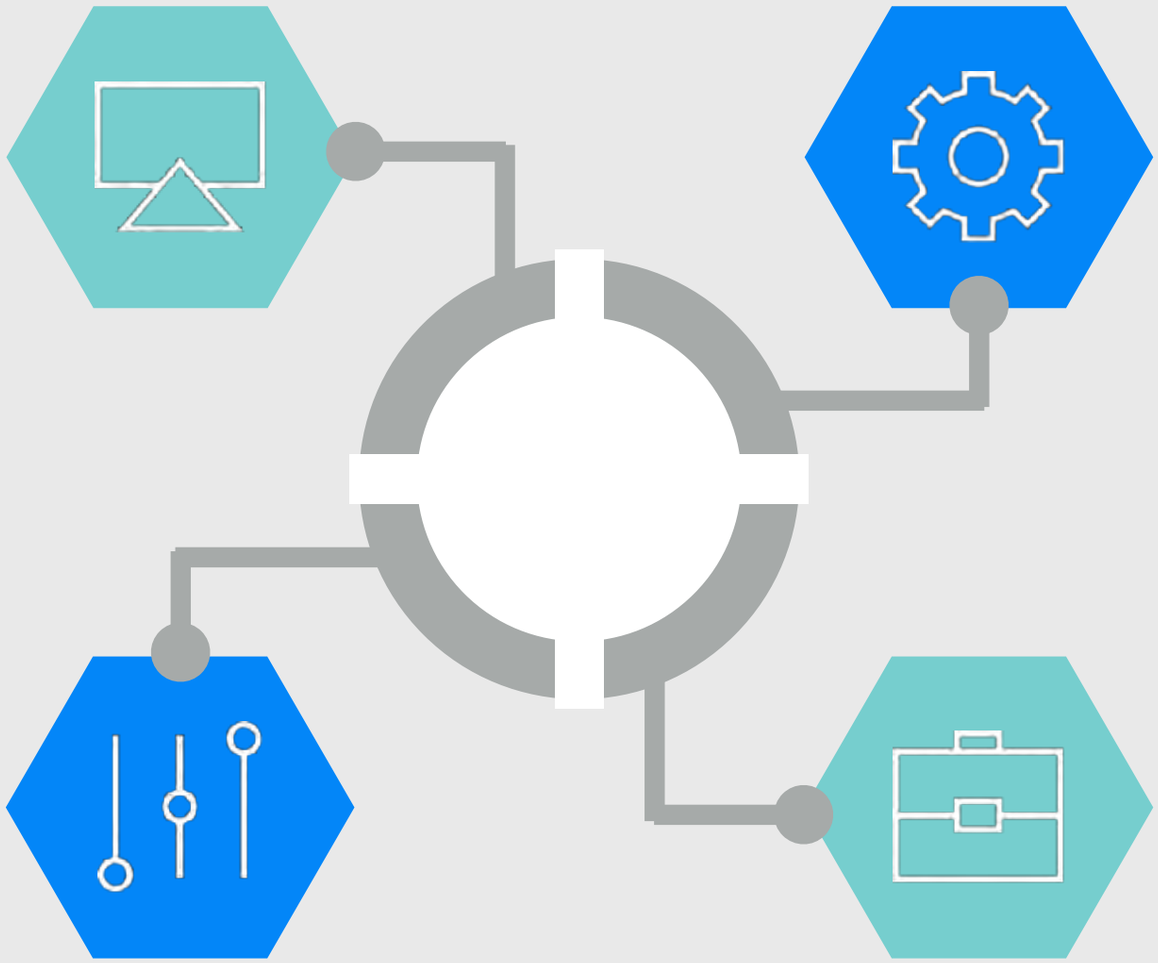


成熟-青年时期

3

严峻的新问题

DB连接数



城市IDC资源限制

跨IDC网络延时

机房间流量

对新一代架构的需求

01 彻底解决DB连接数问题

02 彻底解决IDC资源限制

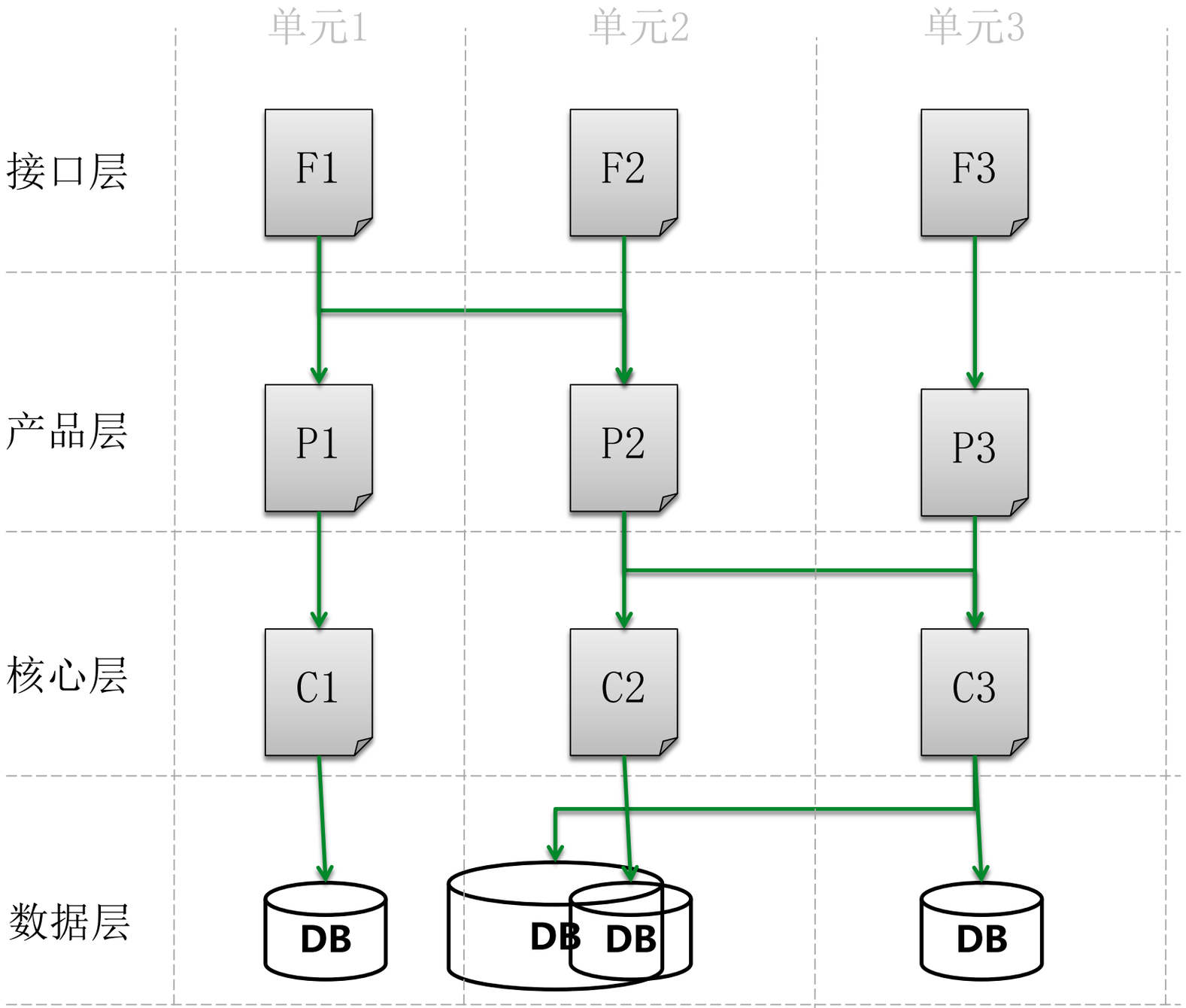
03 保证业务连续性

04 蓝绿发布

05 高可用-异地多活

06 单元化

单元化的实质

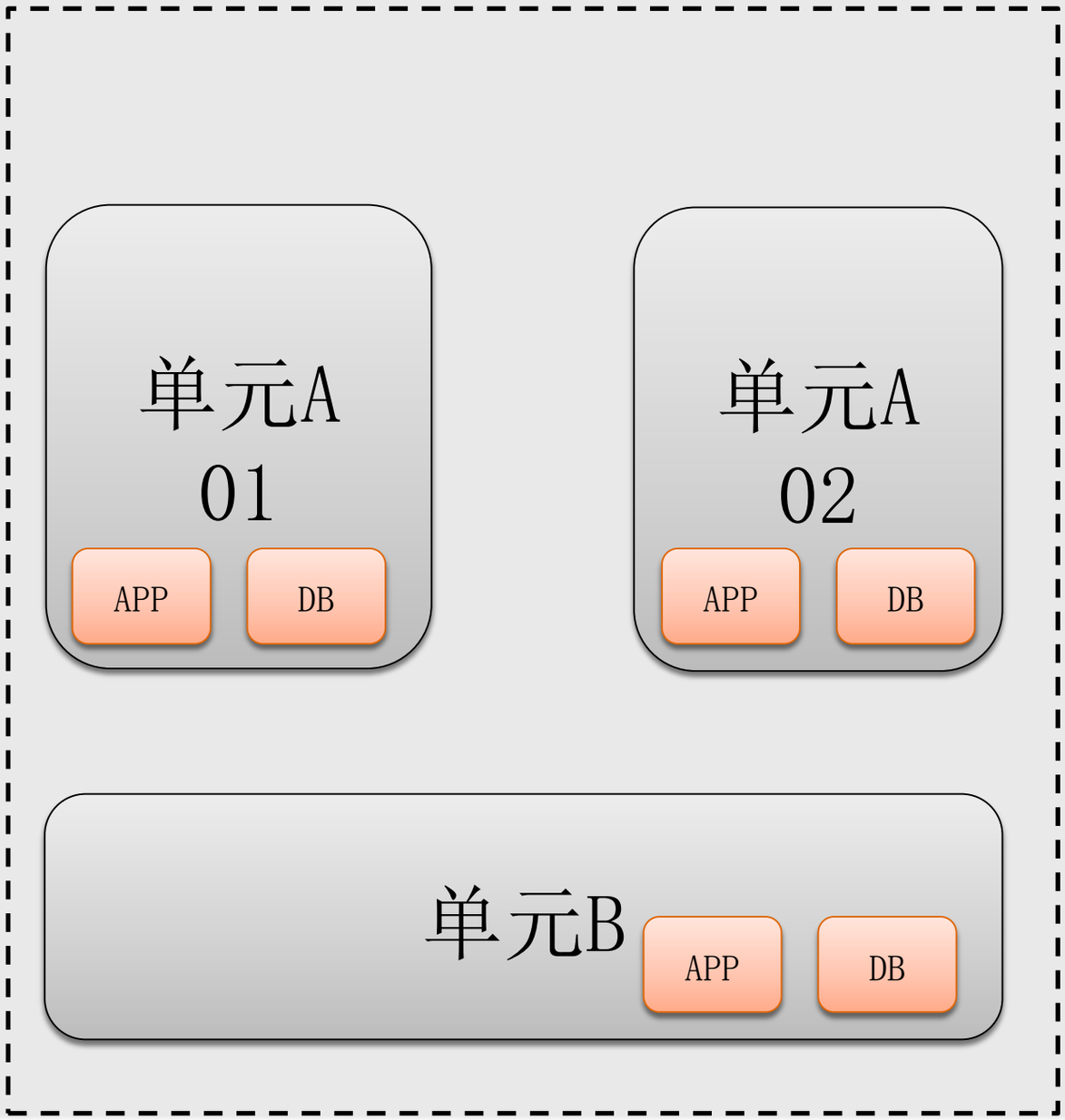


单元化核心思想

- 核心业务
- 数据按照UserID拆分，多机房部署
- 调用封闭
- 部分数据，不共享

- 非核心业务
- 不能按照UID拆分
- 核心不依赖长尾

支付宝单元化架构实现

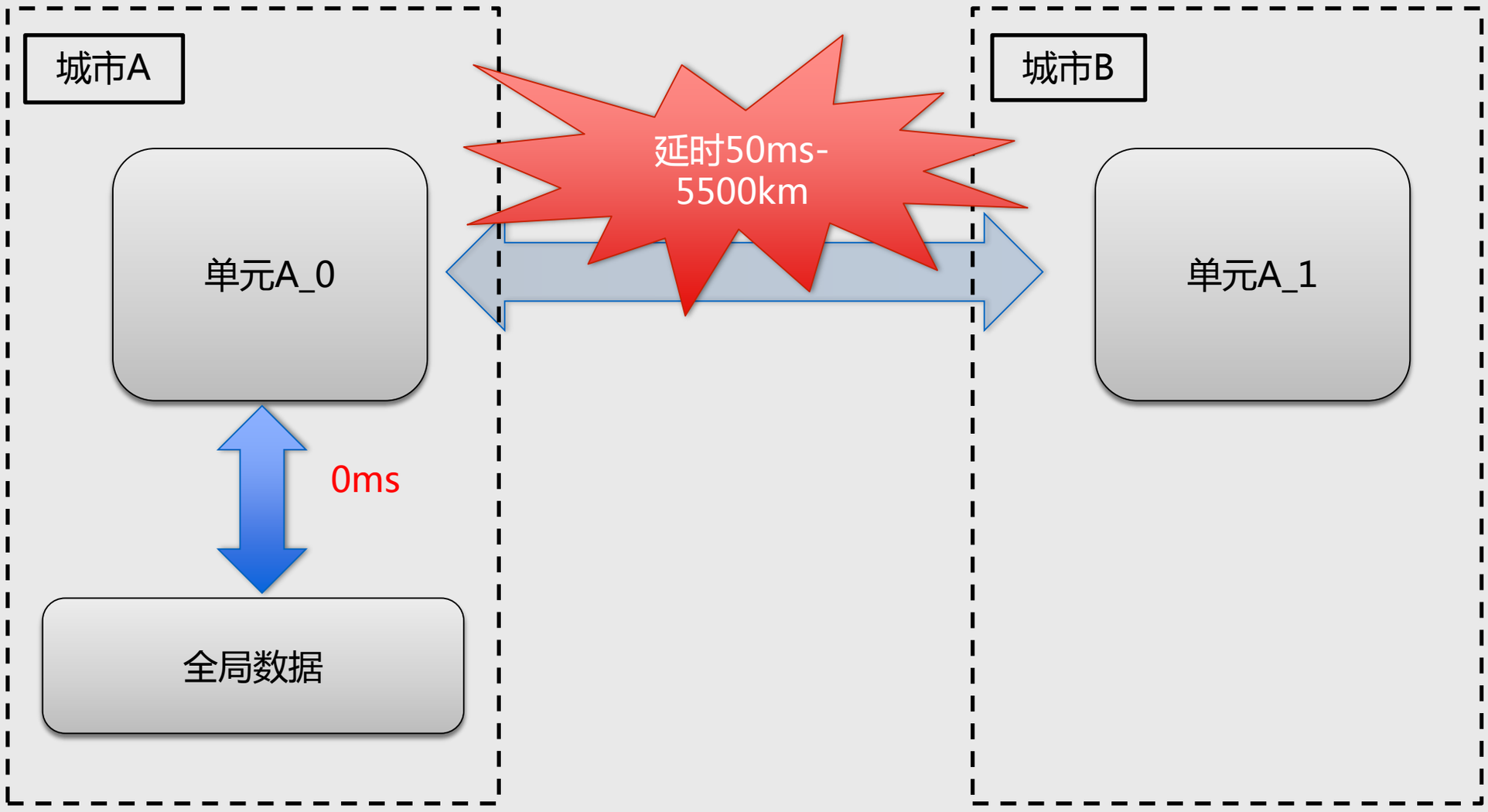


实现思路：

- 水平拆
  交易、支付、账务等，每个单元只有部分数据。

- 上层单元化改造
  从DB层往上延伸水平拆分概念，包括应用层到入口层。

走向异地的挑战



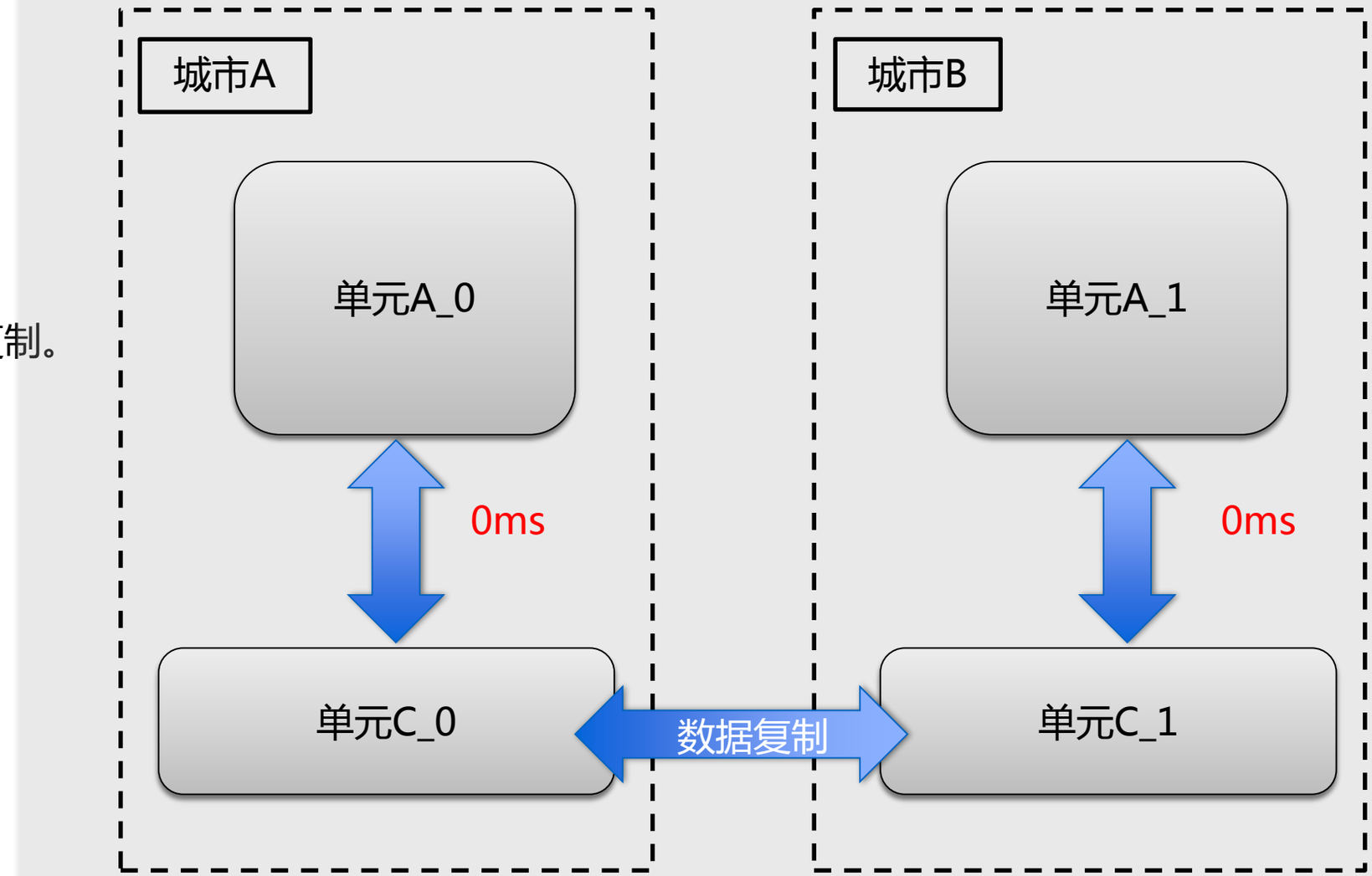
解决延时问题

引入单元C

无法拆分的全量数据，全局复制。

支撑核心业务本地调用

读写分离



解决数据一致性和时效性问题

- 基于DB同步的数据复制：

延时非敏感业务的异地复制方案；部分业务数据，可忍受3s时效性延迟（比如大部分的配置数据）。

- 基于消息系统的数据复制：

对于延时非常敏感的业务，更低延时的实现方案；上层基于应用进行复制，减少延时。底层DB主备同步同时进行。

回顾核心问题

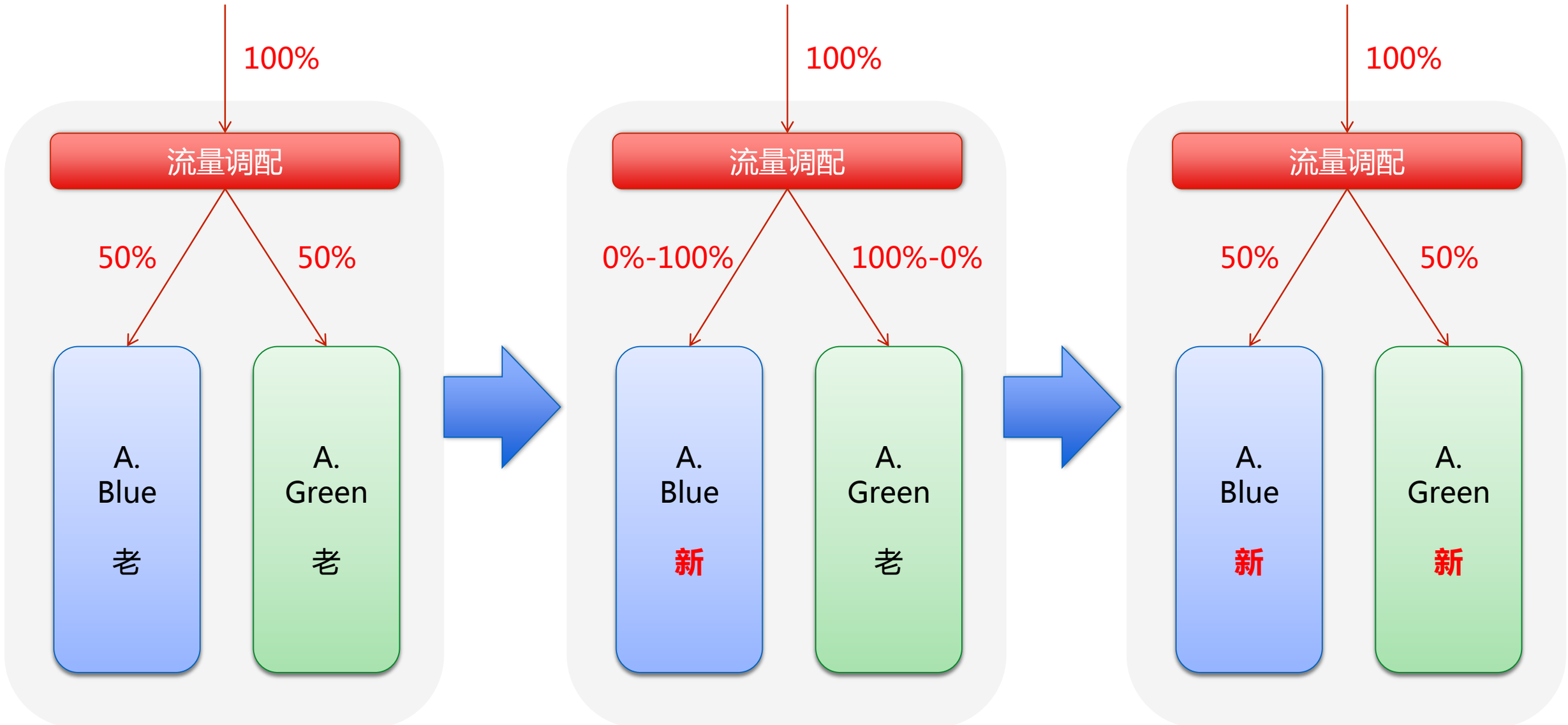
调用本地化
单元封闭

- DB连接数
- 机房间流量
- 跨IDC网络延时

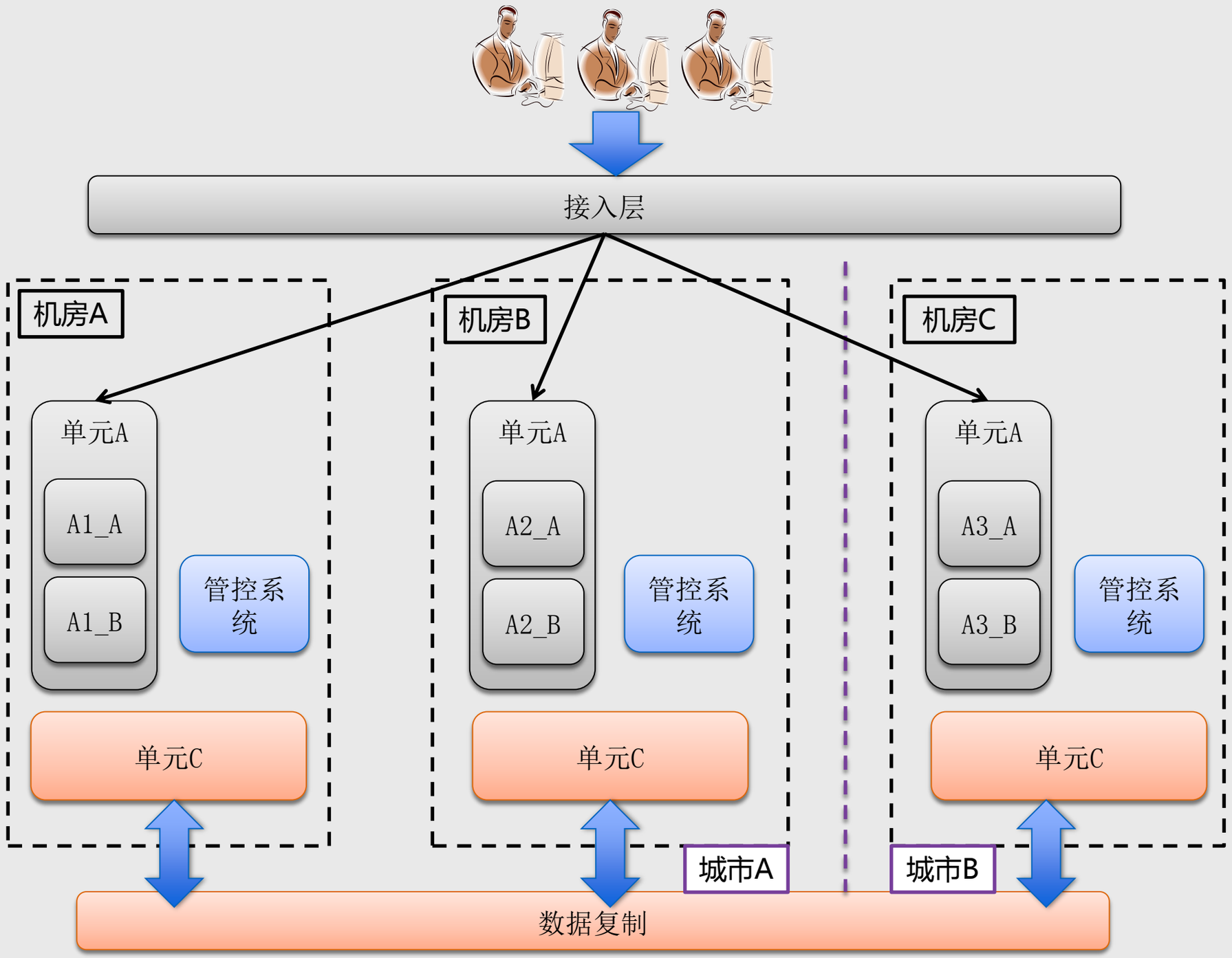
- 城市IDC资源限制
- 异地IDC容灾

异地单元部
署

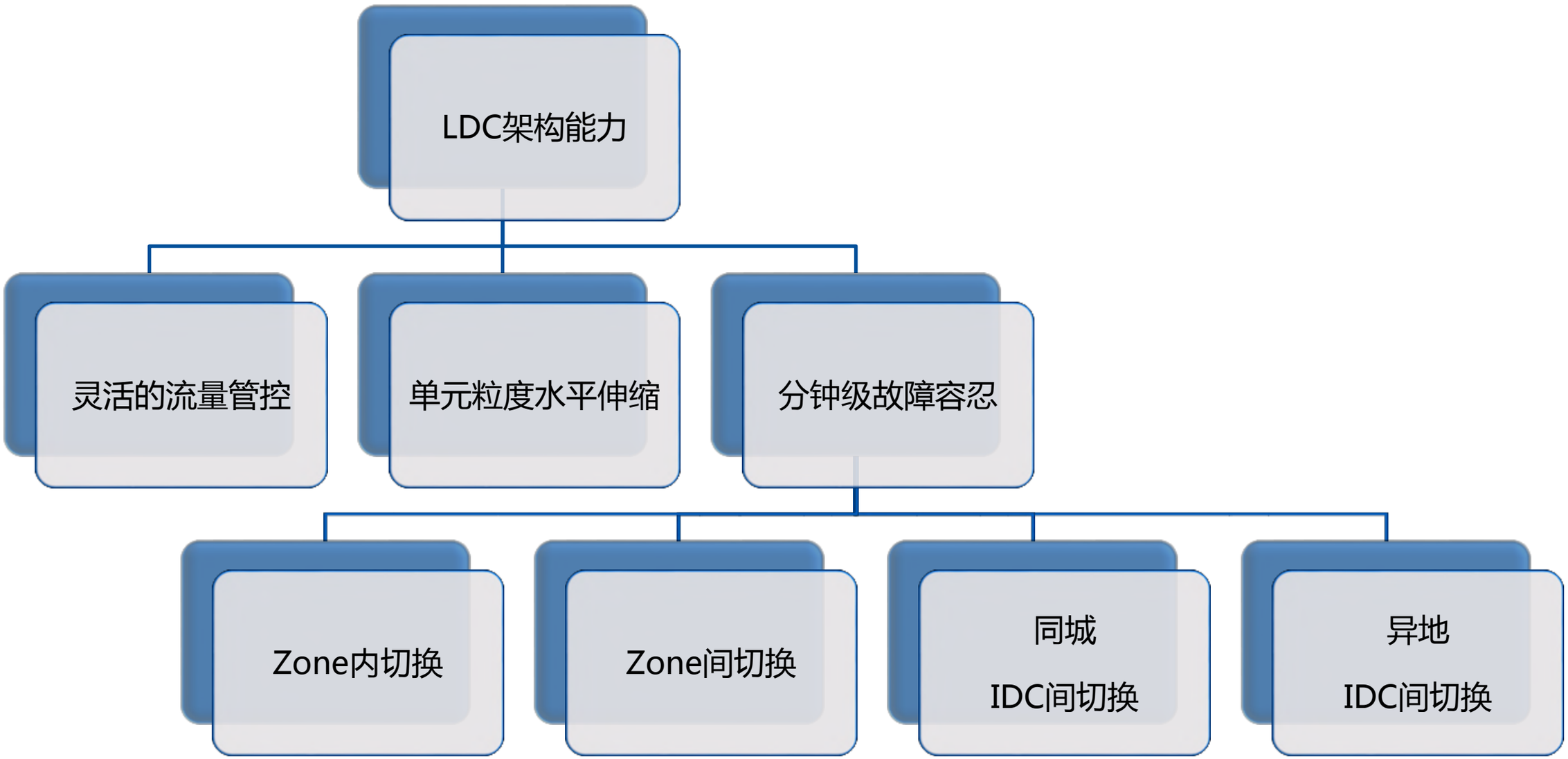
蓝绿发布


100%
流量调配
50%
50%
A.
Blue
老
A.
Green
老
100%
流量调配
0%-100%
100%-0%
A.
Blue
新
A.
Green
老
100%
流量调配
50%
50%
A.
Blue
新
A.
Green
新

支付宝单元化全局架构


接入层
机房A
机房B
机房C
单元A
A1_A
A1_B
管控系统
单元C
单元A
A2_A
A2_B
管控系统
单元C
单元A
A3_A
A3_B
管控系统
单元C
城市A
城市B
数据复制

单元化下的高可用及容灾能力



Alibaba Group 阿里巴巴集团 | 阿里技术保障 Alibaba Infrastructure Service

Thank you

Q&A

@阿里技术保障

@阿里技术保障