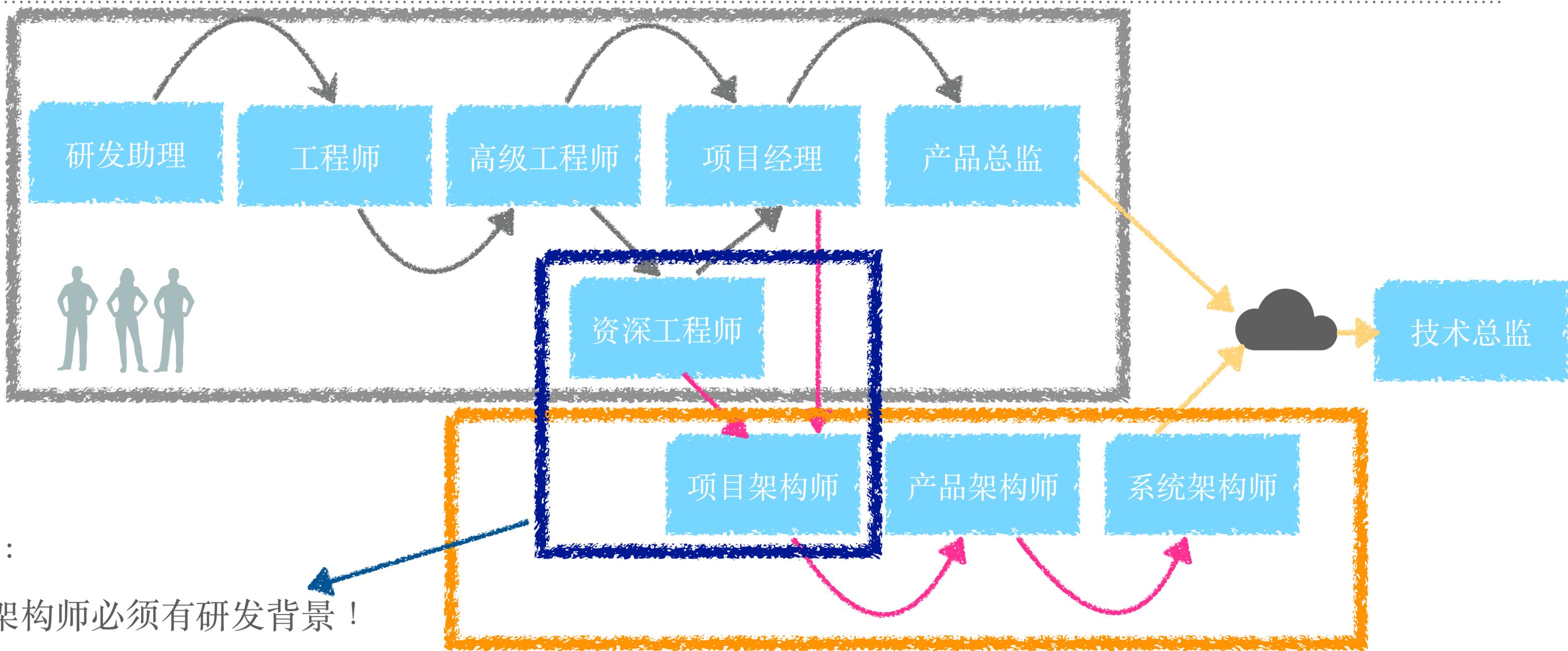




大数据平台架构设计

Speaker: 王胤然

IT职位体系中的架构师所处的位置



真相:

- 1、架构师必须有研发背景！
- 2、“培养”是GET架构师的途径！

OUTLINE

- 不可或缺的软件架构
- 如何完成一个大数据平台的架构设计工作
- 常见的大数据平台总体架构设计模式
- 大数据架构设计中的数据分布设计策略
- 大数据架构设计中的数据集成设计策略
- **Review**



不可或缺的软件架构



何为架构?

➤ 架构模式

- 没有任何一个架构是万能的，架构需要定义对应的架构场景
- 定义模块，梳理模块间的关系。解决代码耦合问题，避免需求变更带来的项目失败

➤ 架构实现

- 为了理解特定的架构模式，从而定义应用案例，帮助使用者理解架构模式
- 为了能够让架构实现到真正的案例中，往往会通过设计模式的实现手段来实现。在面向对象的设计思想下，往往会使用一种或多种“设计模式”来作为实现（单例、工厂、建造、原始模型等）

架构模式

名称

适用场景

模块定义

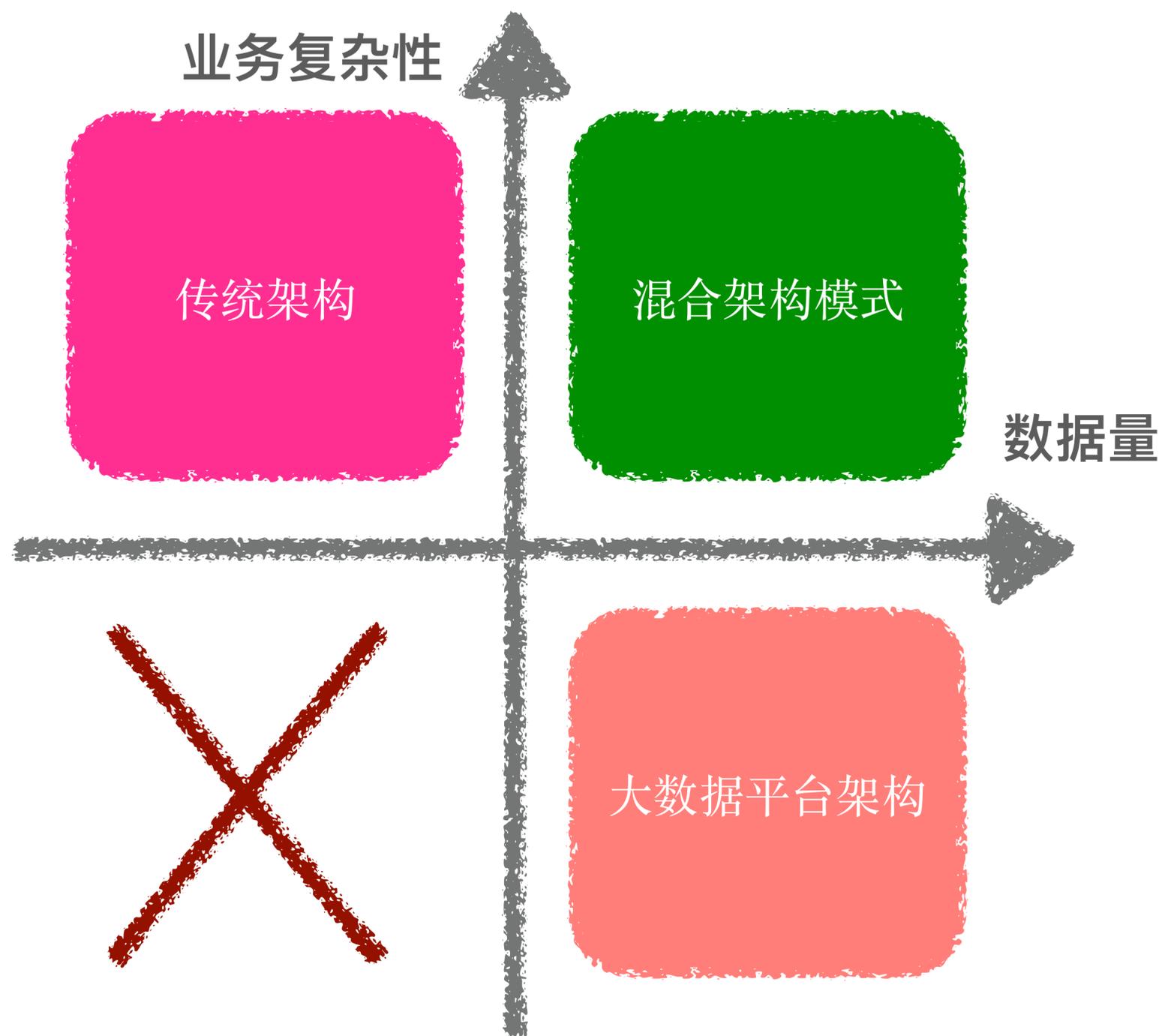
模块关系

架构实现

应用案例

设计模式

传统架构和大数据架构的差异性



种类	传统产品架构	大数据产品架构
数据量大小	一般	海量
数据处理方式	单机或伪分布式	完全分布式
适用业务场景	业务应用导向	数据价值导向

“

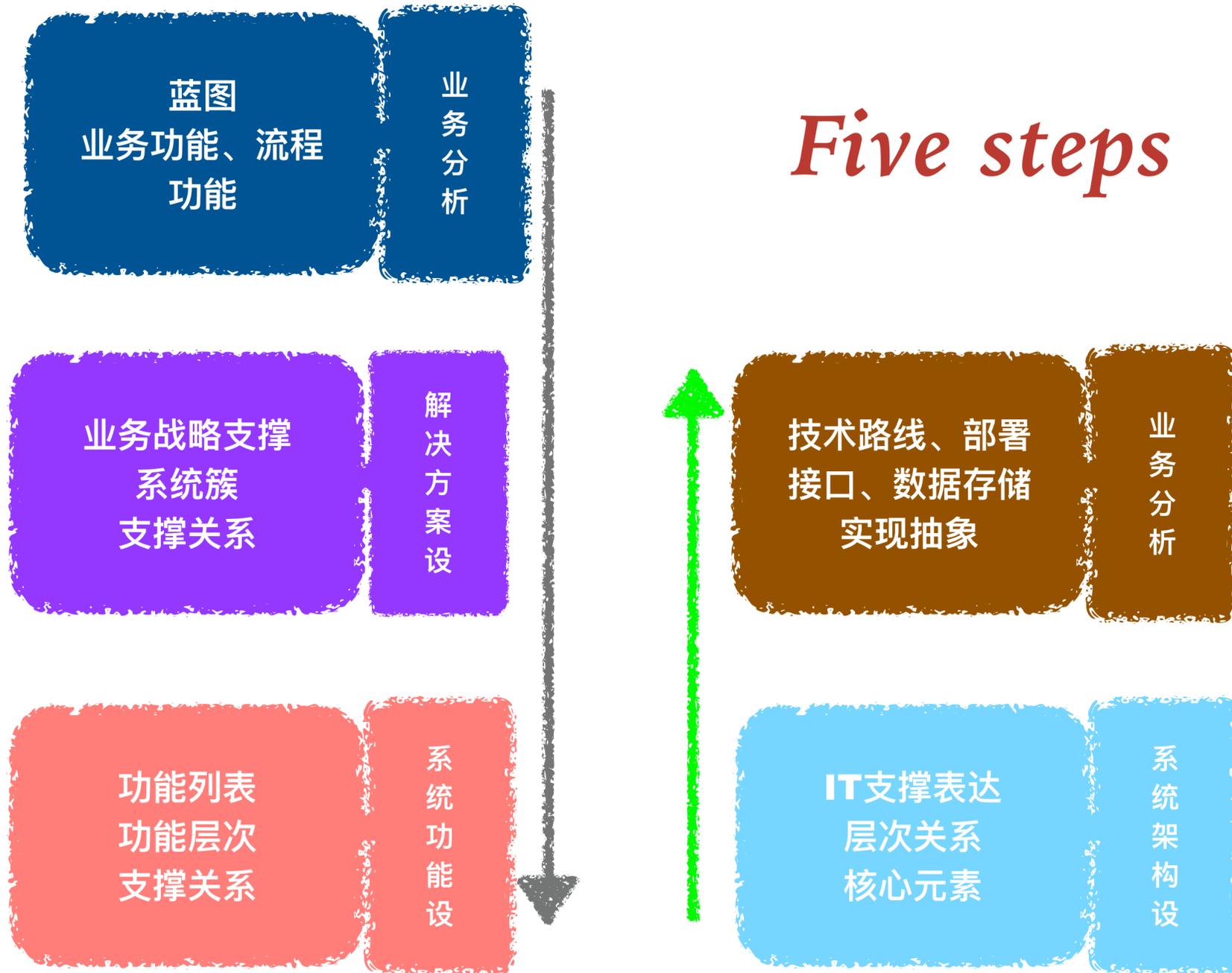
复杂的系统往往是既具有海量的数据，又需要实现繁杂的业务。混合架构，才是出路！

如何完成一个大数据平台的架构设计工作

Five steps! !



大数据平台架构设计过程



(1) 业务分析：针对目标行业的业务战略、蓝图、业务功能及流程进行分析，提出其中部分功能可以使用信息化进行处理，通过分析可以得出信息化要解决的问题。

(2) 解决方案设计：根据业务战略，形成行业信息化解决方案。它是一个系统簇，同时明确各系统间的支撑关系。

(3) 系统功能设计：明确信息化系统功能列表及功能层次（层次，例如经验决策层工，管理层功能，业务操作功能等），将功能散列在这些层次中，根据功能及应用特点形成一个或者多个子系统。

(4) 系统架构设计：针对某一系统明确系统IT支撑表达，层次化关系表达及功能、技术核心元素

(5) 技术体系设计：针对系统的接口、数据存储、技术路线、部署及实现抽象进行设计

常见的大数据平台总体 架构的设计模式



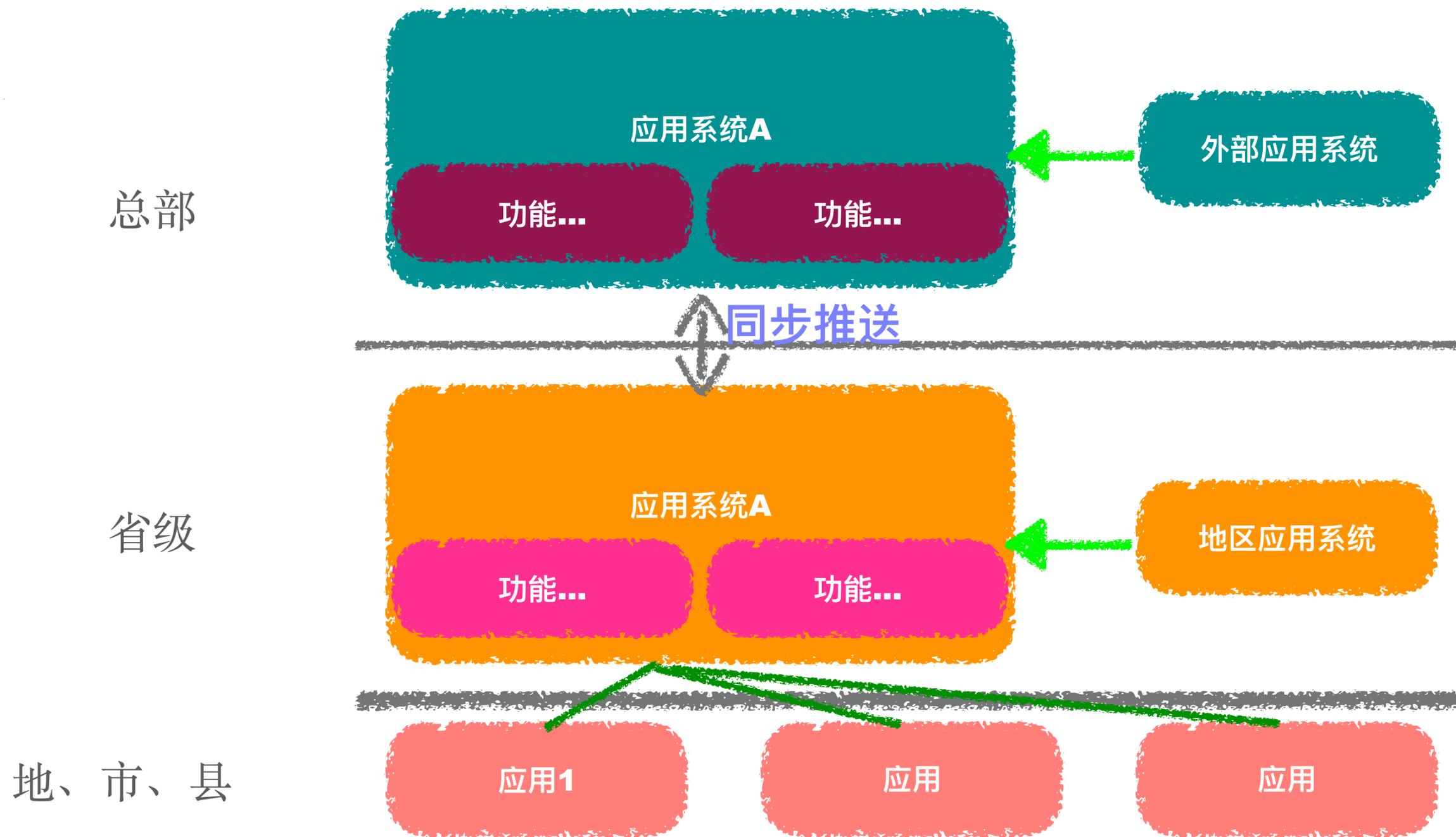
ASSF (ACCESS-SERVICE(BIZ)-STANDARD-FUNDATION)模式

- 访问-服务（业务功能）-标准-基础，对系统架构各个层次均有表达，但部署应用模式需要有单独说明，如下图方式组织系统总体架构：



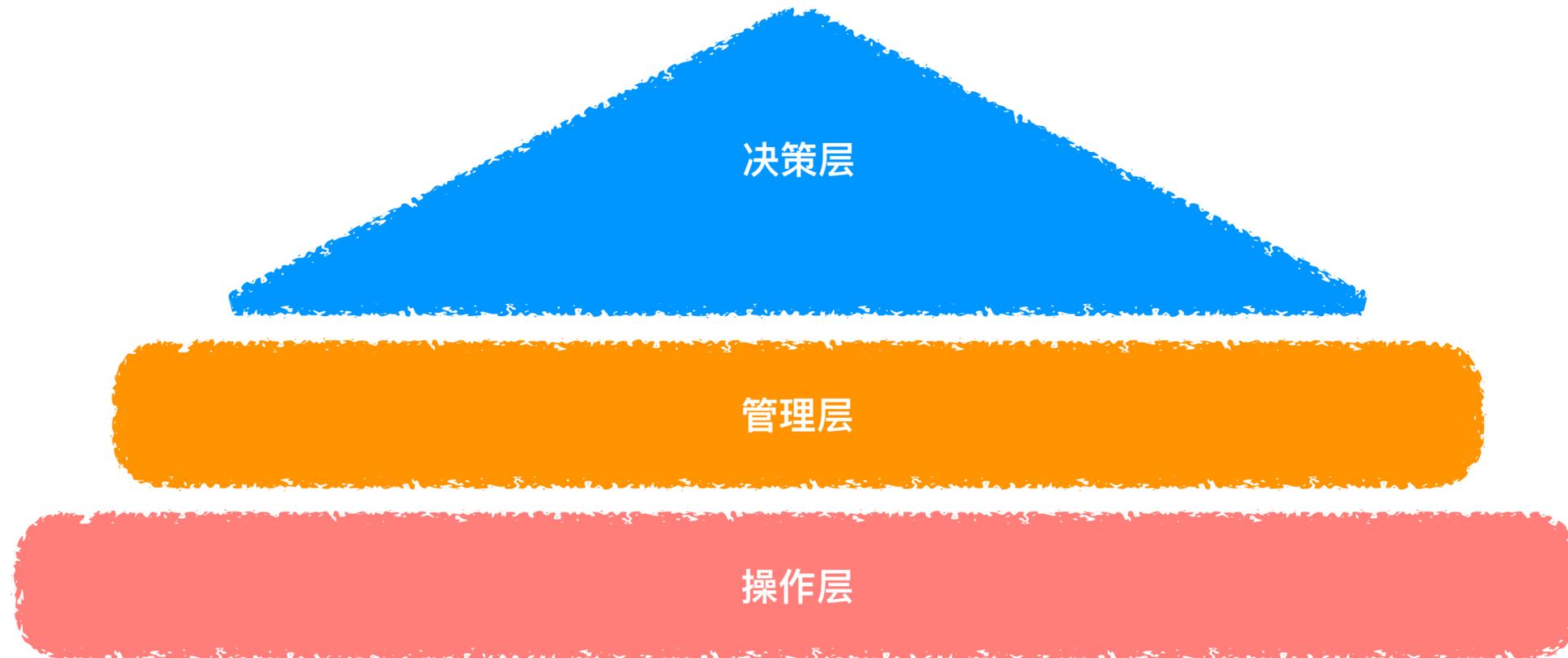
LOCATION模式

TIP: 区域视角，划分数据中心分级设计适用于“集团级应用”的开发，优势是：对于应用逻辑表达较为清晰



MANAGEMENT-LEVEL模式

- ▶ 决策层-管理层-操作层各个层次使用的功能。对于系统功能表达较为清晰，对于与客户达成一致性理解有较好效果



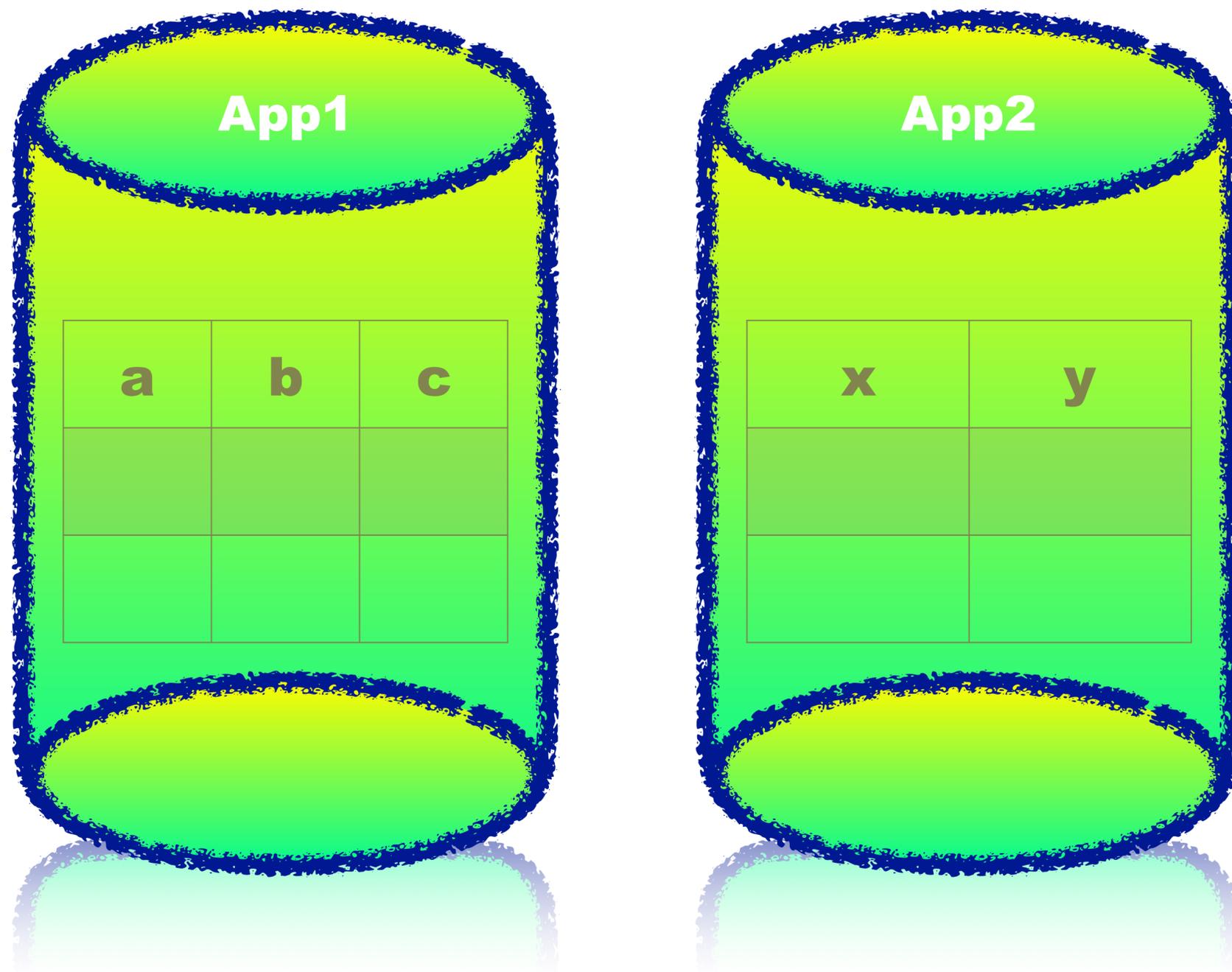
“

个人比较推荐ASSF模式做为主架构，同时制定Location模式与Three-management-level模式附加说明系统，从各个层次表达系统架构

大数据架构设计中的数 据分布设计策略



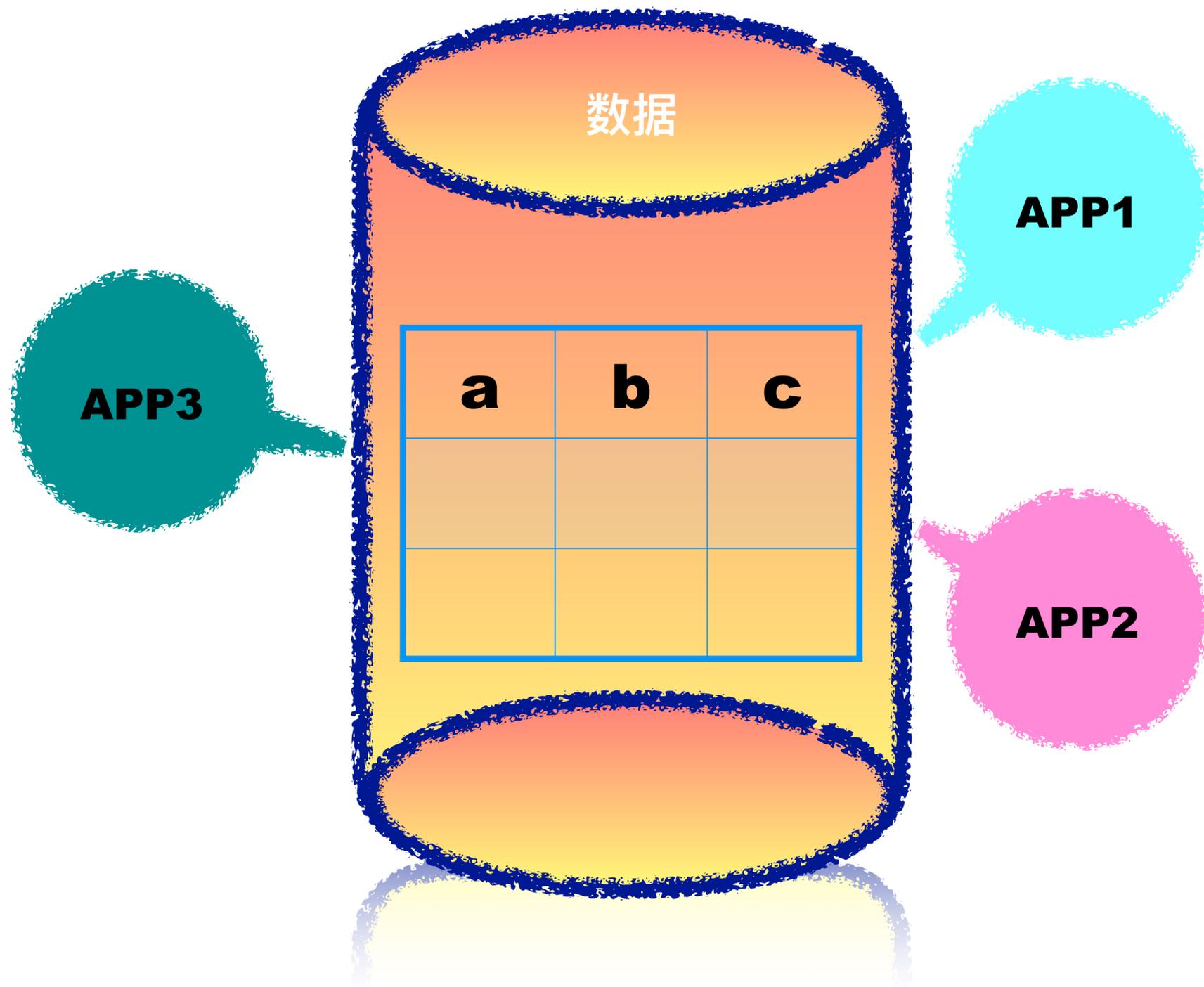
“独立”设计模式



适用场景：当一个大系统由相关的多个小系统组成，且不同小系统具有互不相同的数据库 **Schema** 定义。

优势：独立模式可管理性高，通信开销小

“集中”设计模式



适用场景：一个大系统必须支持来自不同地方的访问，集中存储分布访问

优势：数据进行集中化，统一格式存储。可管理性好、数据一致性高

“分区”设计模式

应用实例1

a	b	c
1	1	1
2	2	2

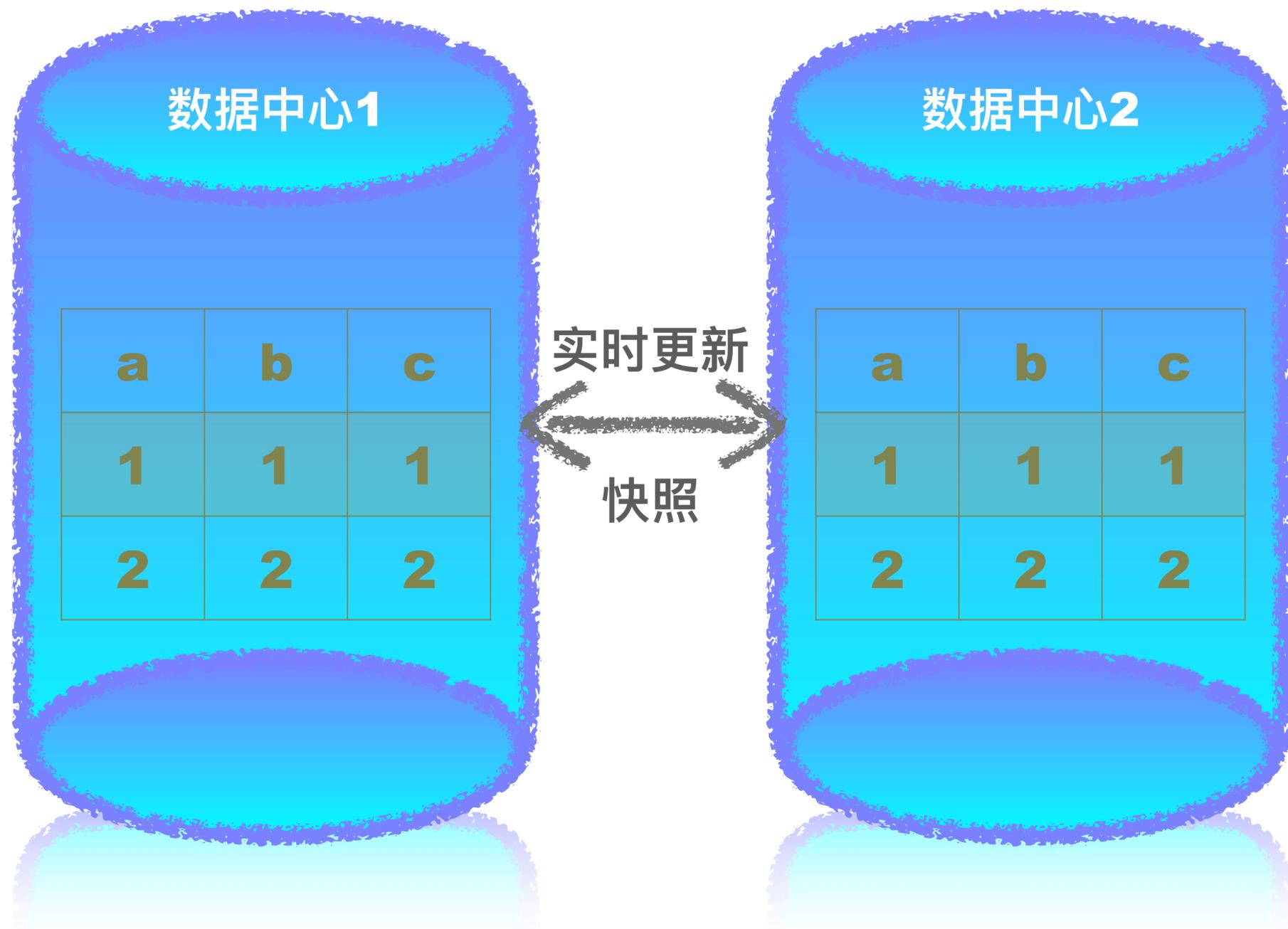
应用实例2

a	b	c
3	3	3
4	4	4

适用场景：为“地域分布广泛的用户”提供“相同服务”（区域数据中心设计）相同的数据模型，不同的数据值

优势：可伸缩性好。

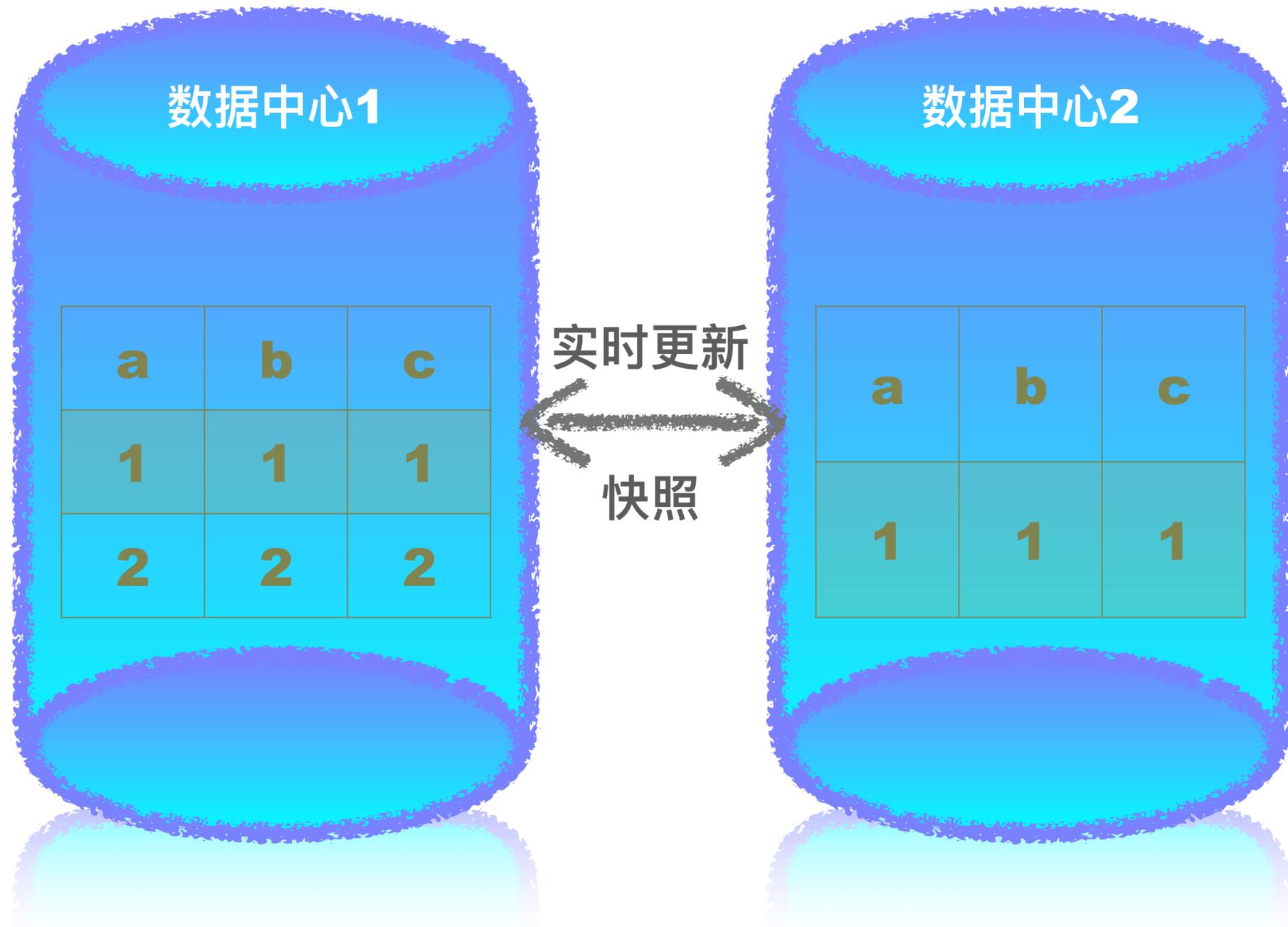
“复制”设计模式



适用场景：容灾，或异地灾备

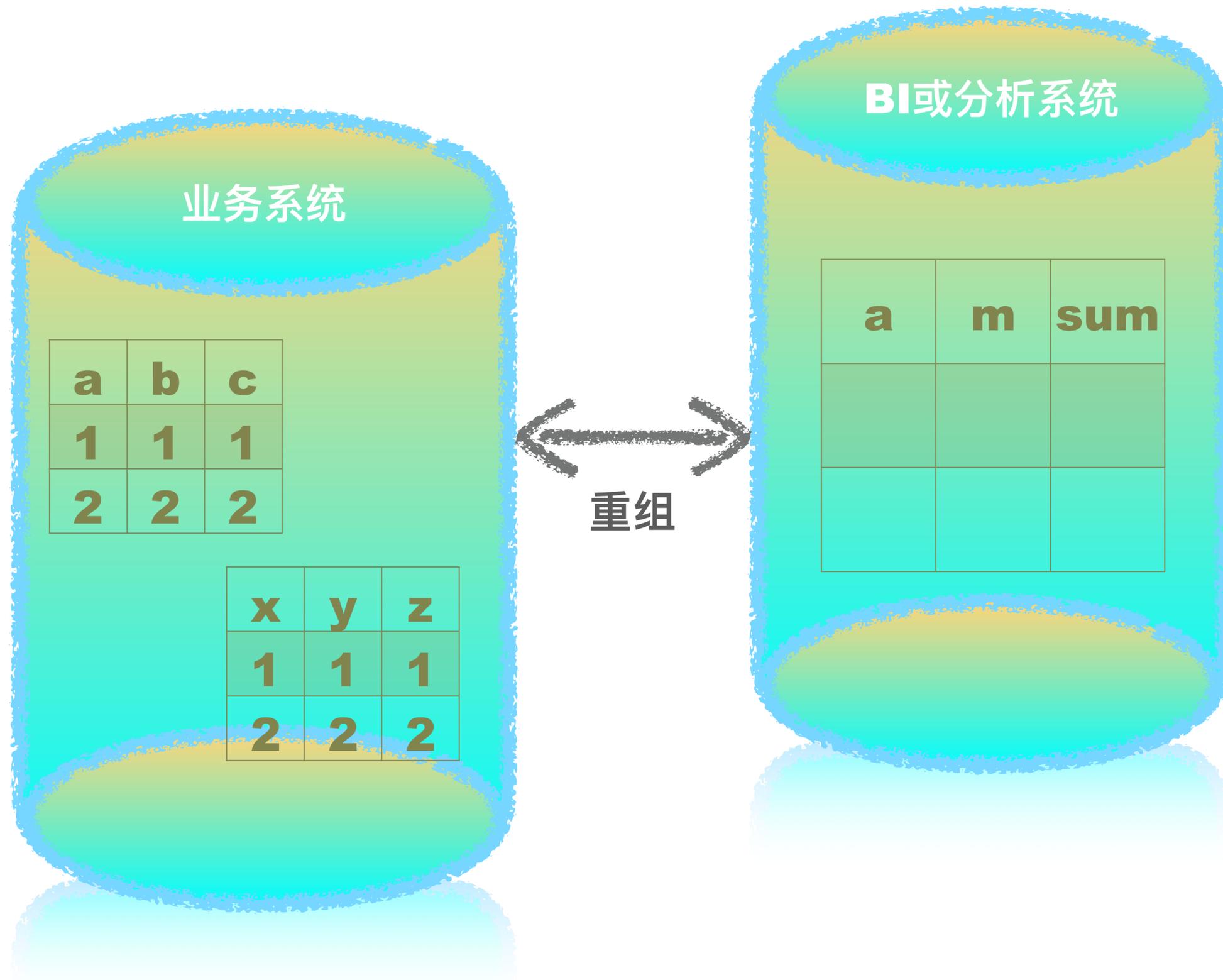
优势：提升数据的可靠性，主要是抗灾害能力，可以在一定程度上提升业务应用的访问吞吐

“子集”设计模式



“子集”设计模式是“复制”设计模式的特殊方式。就是某节点因功能或非功能考虑而保持全体数据的一个相对固定的子集

“重组”设计模式



不同数据节点因要支持的功能不同，而以不同的**Schema**保持数据，但本质上数据是同源的。重组以“重新组织”的格式进行传递和保持。

适用场景：**BI**或者高级分析系统

数据分布设计策略总结

	Schema相同	Schema不同
非复制方式	集中 (可管理性、数据一致性)	独立模式 (可管理性、通信开销)
	分区 (可伸缩性)	
复制方式	子集	重组
	复制 (可靠性)	

合适

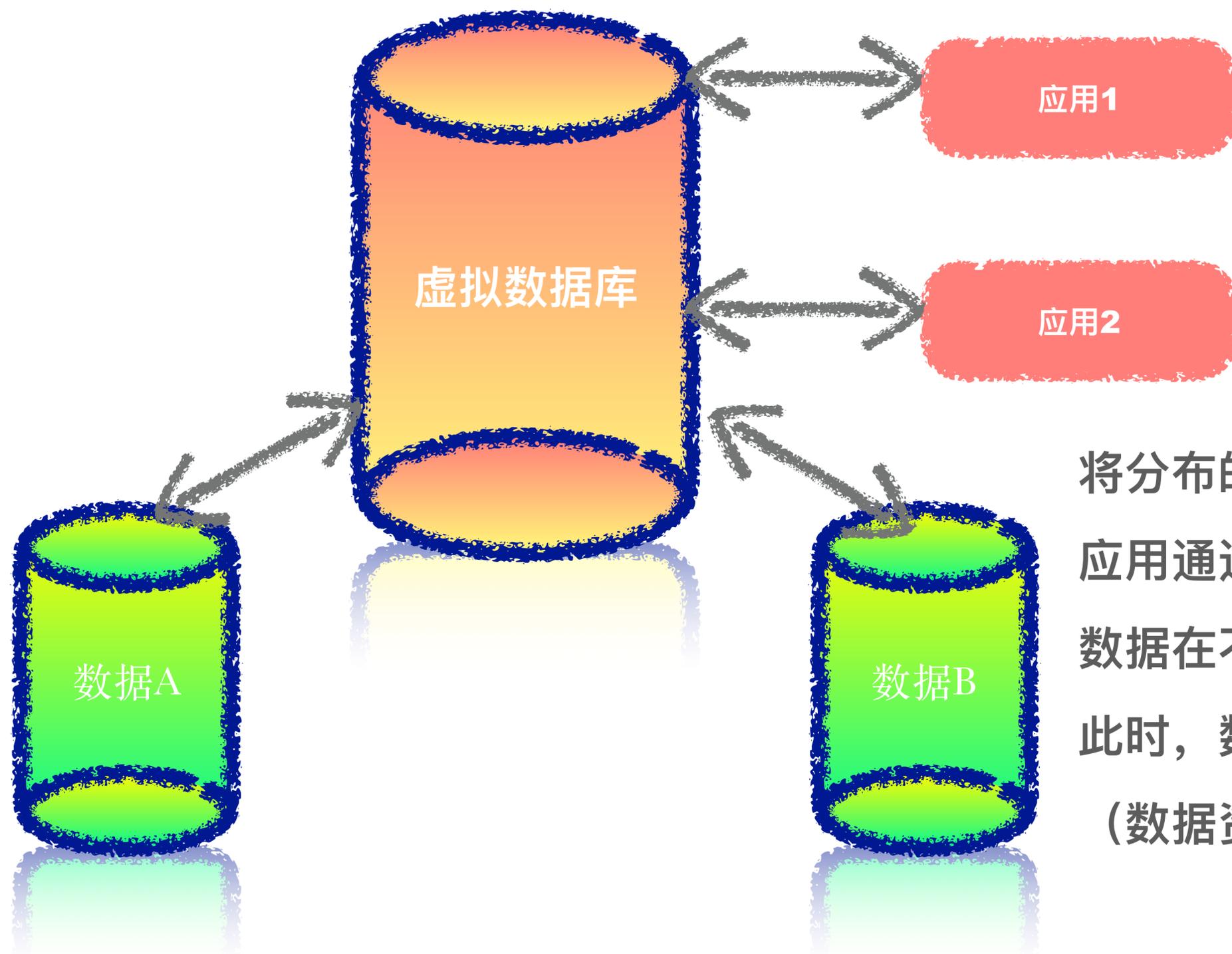
综合

优化

大数据架构设计中的数 据集成设计策略

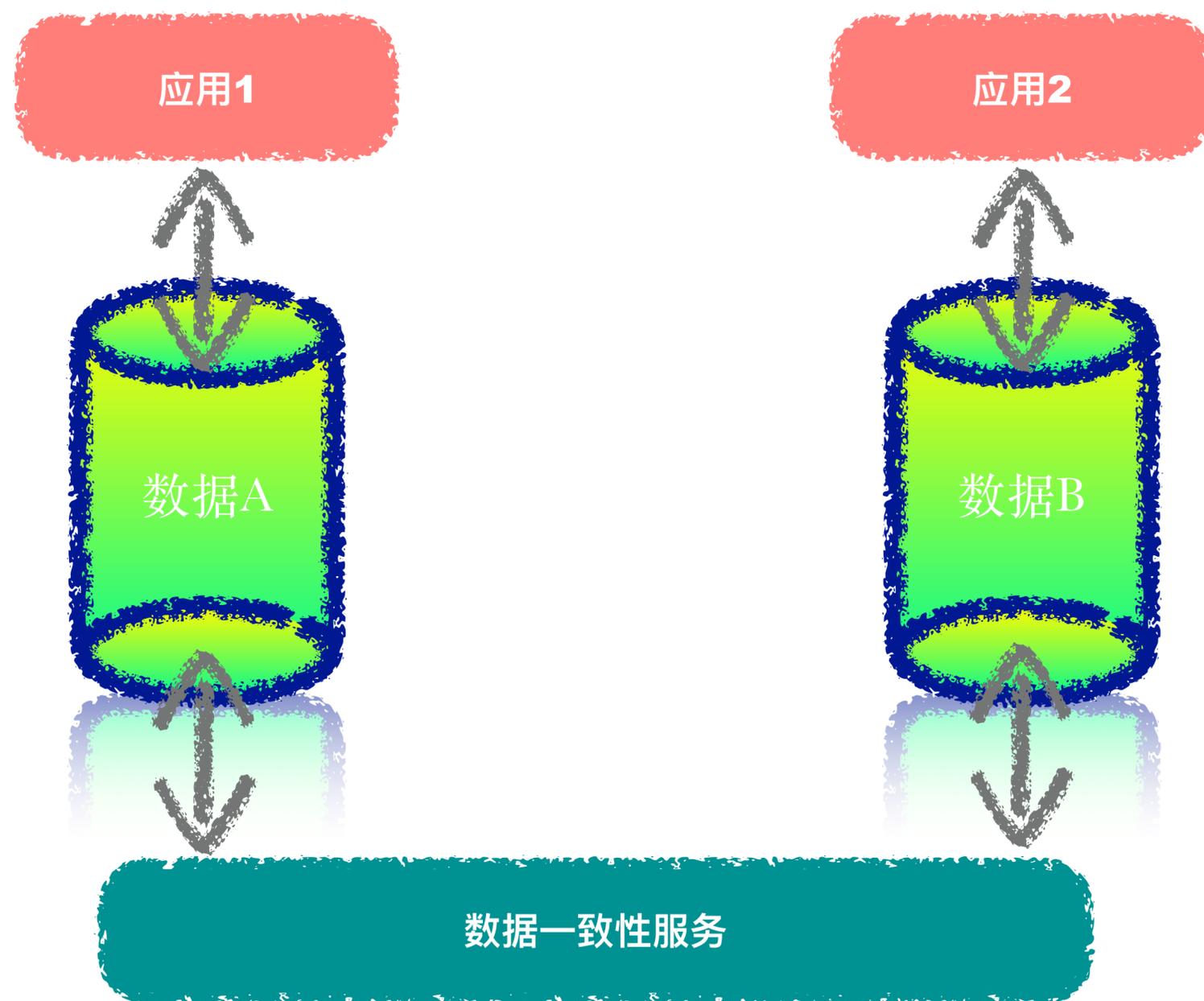


数据联邦模式(DATAFEDERATION)



将分布的数据逻辑集中，
应用通过访问整合平台的虚拟数据库进行数据访问，
数据在不同数据库实例中，
此时，数据整合平台做为数据访问通道。
(数据资源服务目录，烽火大数据平台)

数据复制模式(DATA REPLICATION)



采用数据复制模式，通过数据一致性服务实现多个数据源的数据一致性，各数据库均保留共享数据备份。

(阿里：**ADS-ODPS**)

REVIEW

- ▶ 大数据产品的架构设计，可以套用一般软件产品的架构设计模式和方法，不同点在于重点关注数据的特性选择设计模式。
- ▶ 复杂系统（企业级应用基本没有简单应用），采用混合架构设计模式才是解决之道。
- ▶ 万变不离其中，数据分布和数据集成的设计模式组合运用，一招鲜吃遍天
- ▶ 架构师不是画框图的！不是简单的把名词按照层级堆放！分层模式“！= ”架构设计
- ▶ 多实践，拍脑子想是解决不了任何问题的，因为你永远不知道复杂系统中会有什么“鬼”



架构迎接未来变化

IAS2017 • NANJING

THANKS BYE

2017.11

