

# 海量非结构化数据 “双活/多活” 存储 的技术难点、方案与应用

李明宇 @ OStorage



## 主要内容

- 非结构化数据“双活”存储需求
- 海量非结构化数据“双活”存储带来的挑战
- 基于对象存储和全分布式架构的解决方案
- 金融和医疗领域的应用案例
- 从“双活”到“多活”——不仅仅是增加一个备份

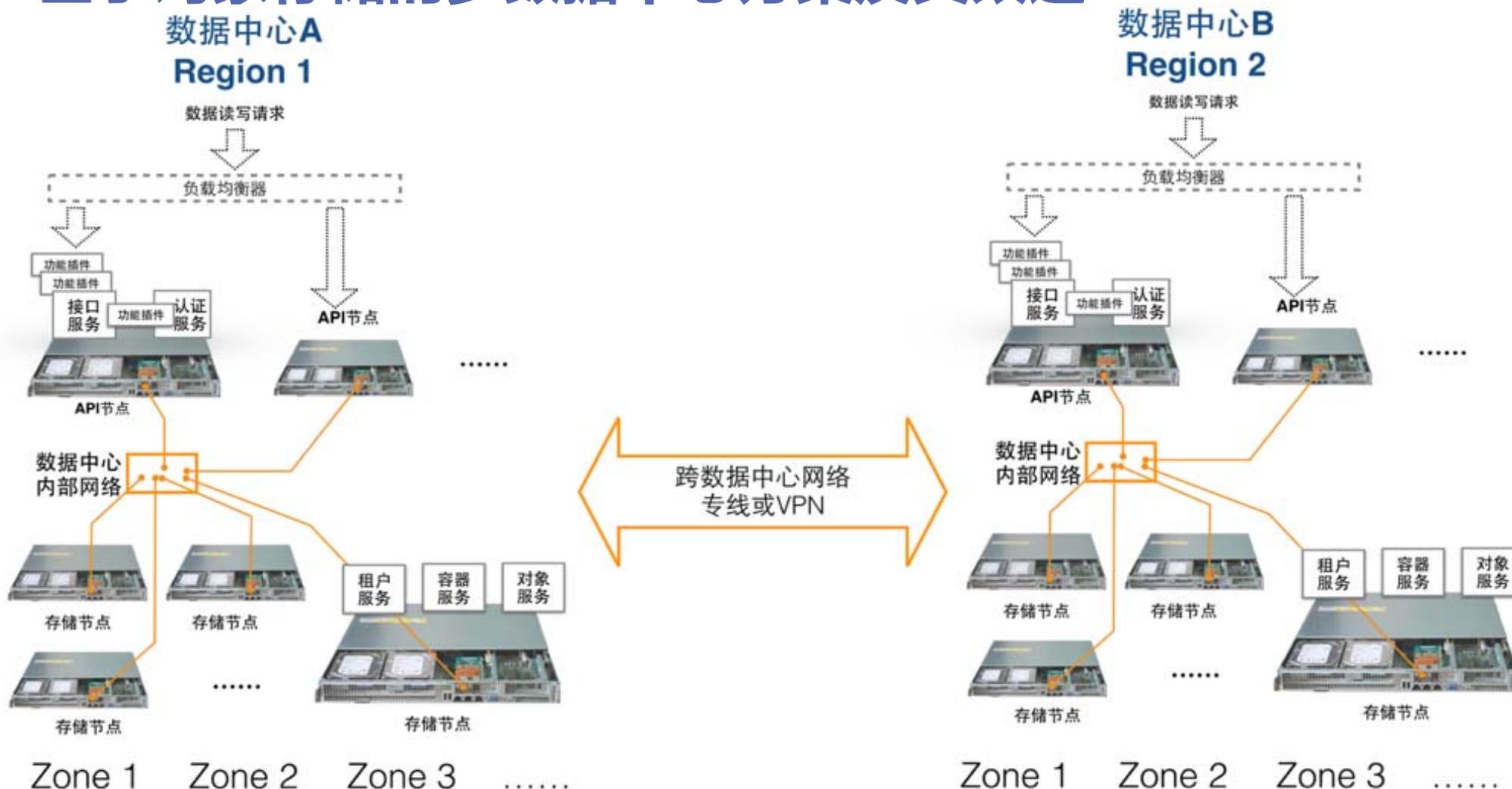
## 非结构化数据“双活”存储要达到的目标

- 金融票据、医疗影像.....
- 同城为主，双数据中心，光纤互联为主
  - 网络延迟在毫秒级
- 没有主、从之分，主要考虑网络状态
  - 网络正常时，两个数据中心同步写入
  - 网络异常时，两边均可以独立运行
    - 网络恢复正常后，数据在可控时间内自动完成同步
  - 临界状态 —— 网络拥塞
- RPO、RTO接近于0

# 海量非结构化数据存储带来的挑战

- “海量” 在各行业已经变得很常见
  - 总量亿级以上
  - 每年增量千万级以上
- 文件系统方案
  - 随机写带来的开销
  - 文件系统metadata维护的问题
  - 海量数据遍历周期问题
- 对象存储方案
  - 弱一致性问题

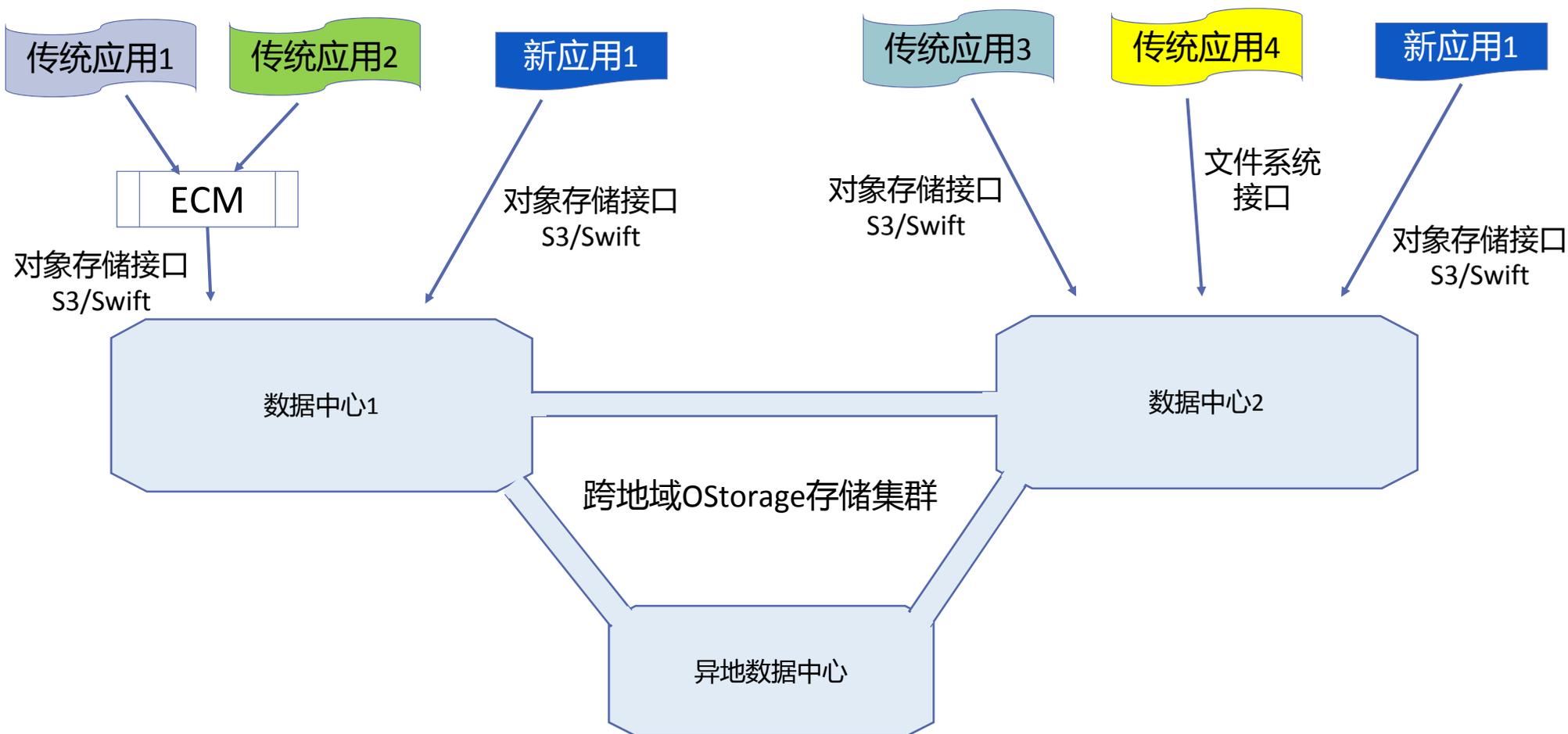
# 基于对象存储的多数据中心方案及其改进



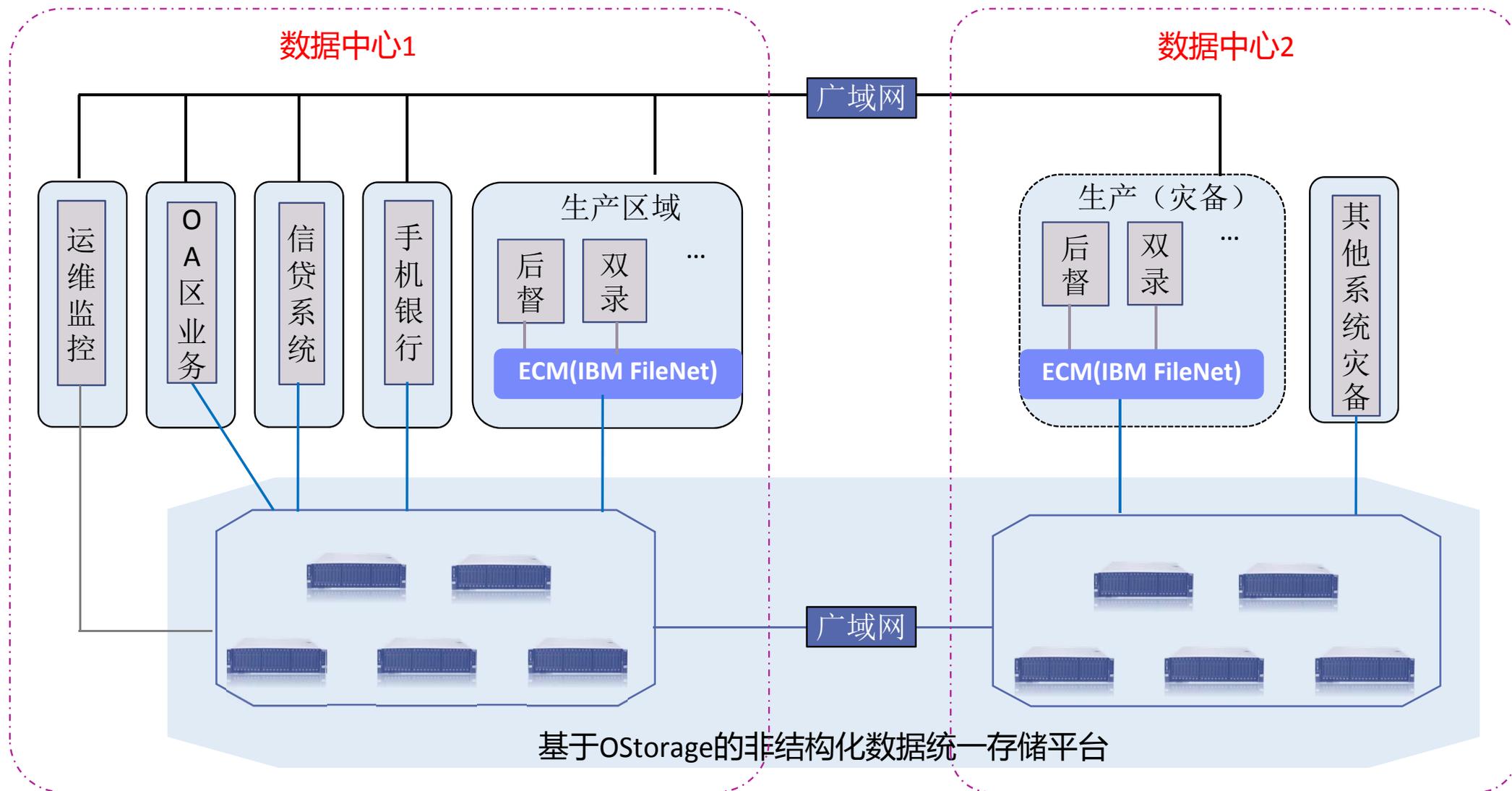
多数据中心方案不等同于  
“多活/双活”方案RPO未知



# 金融行业应用：银行非结构化数据集中存储平台示意图



# 金融行业应用：银行非结构化数据集中存储平台典型部署模式

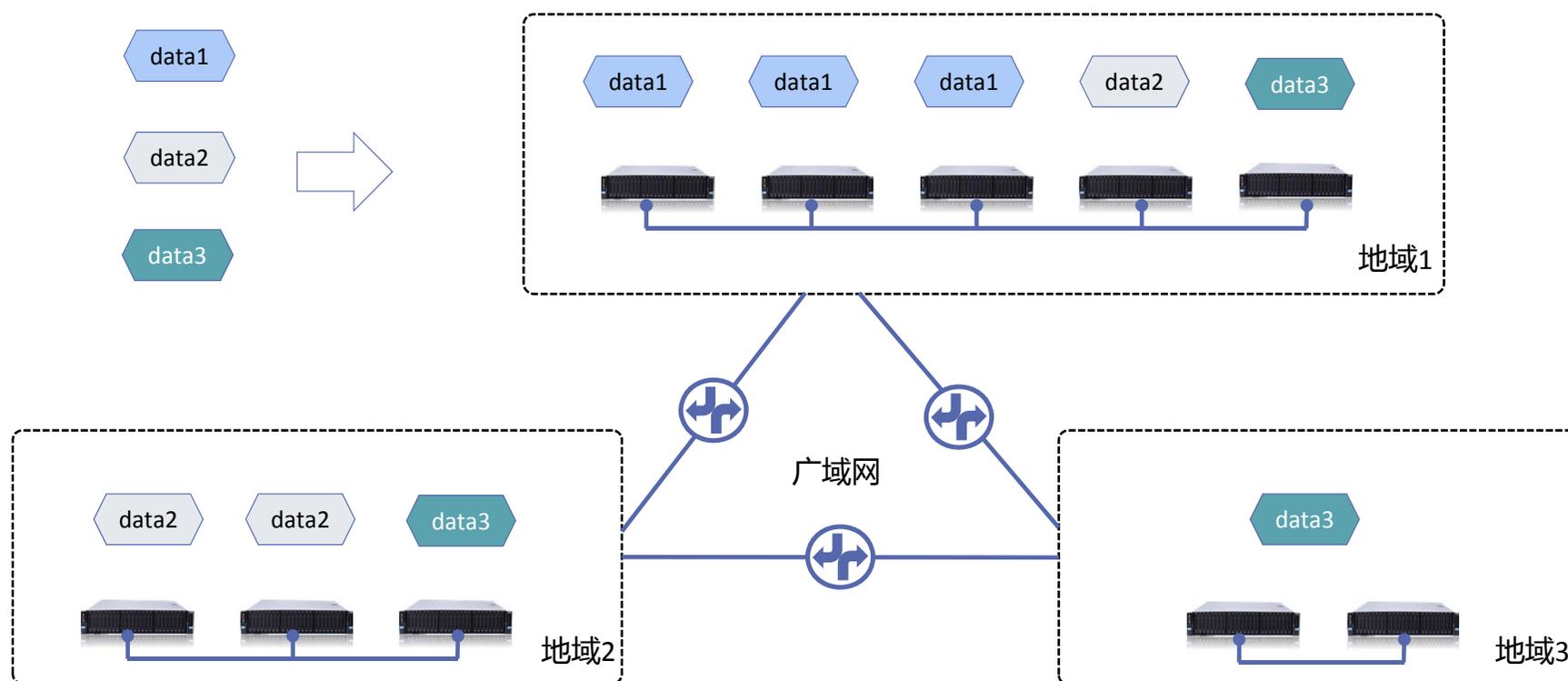


# 医疗行业应用：医疗影像存储



# 从“双活”到“多活”——不仅仅是增加一个备份

- 跨广域的多地组成单一存储集群，数据可按策略分布
- 集群实现近乎永久的持久性，永远在线的可用性
- 实现数据中心多活，自动容灾与多地域的本地访问





- 诞生于后互联网时代的对象存储，具有独特优势，是技术发展的必然选择
- 稳定的结构，简单直接的操作模式，分布式和横向扩展，应对未来海量数据的挑战
- 基于经过证明的稳定的对象存储技术核心，降低安全隐患，实现自主可控
- 采用通用服务器硬件的软件定义存储，可以大幅度降低企业成本
- 对象存储作为软件定义存储的代表，成为企业信息系统赢得未来的重要基础