

(PART 10)

# 备份和归档

# 备份和归档

学完本模块后，您将能够：

- 描述备份粒度
- 说明备份和恢复操作
- 描述各种备份目标
- 说明重复数据消除
- 描述虚拟化环境中的备份
- 说明数据归档

# 模块 10：备份和归档

## 第 1 课：备份概述

本课程将讲述下列主题：

- 备份粒度
- 备份方法
- 备份体系结构
- 备份和恢复操作

# 回顾

一个机构部署了100块磁盘组成的存储阵列。每一个磁盘都有50, 000小时的平均故障间隔时间(Mean Time Between Failure)？

- A. 50 小时
- B. 500小时
- C. 5, 000 小时
- D. 50, 000 小时

# 回顾

在CAS系统内，哪一项用来唯一地识别一个对象？

- A. Meta ID
- B. LUN ID
- C. 内容地址
- D. Object ID

# 回顾

ESXi 中的哪一个组件对“无法被虚拟化的操作系统特权指令”进行二进制转换 (Binary Translation)?

- A. VMFS
- B. 虚拟机内核
- C. 资源管理器
- D. 虚拟机监控器

# 什么是备份？

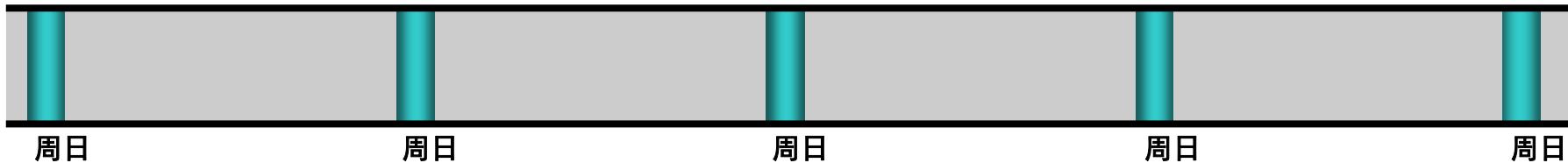
备份

它是仅仅为了恢复已丢失或遭受破坏的数据而创建并保留的生产数据的额外拷贝。

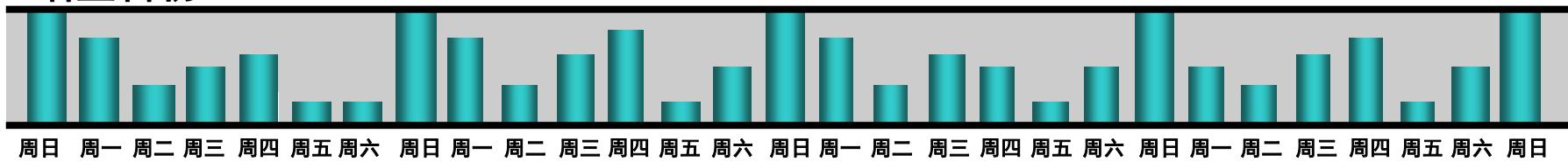
- 组织还会进行备份以遵从法规要求
- 执行备份有三个目的：
  - ▶ 灾难恢复
  - ▶ 操作恢复
  - ▶ 归档

# 备份粒度

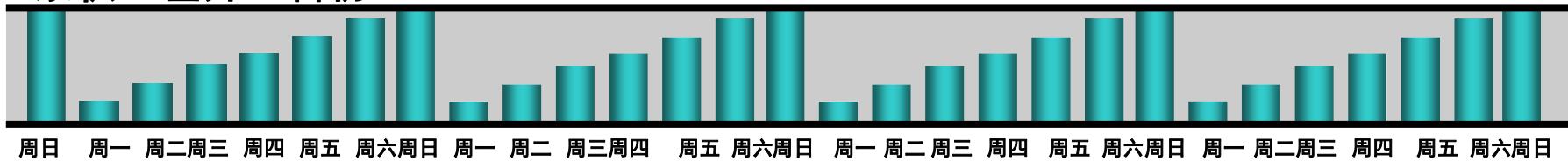
## 完整备份



## 增量备份

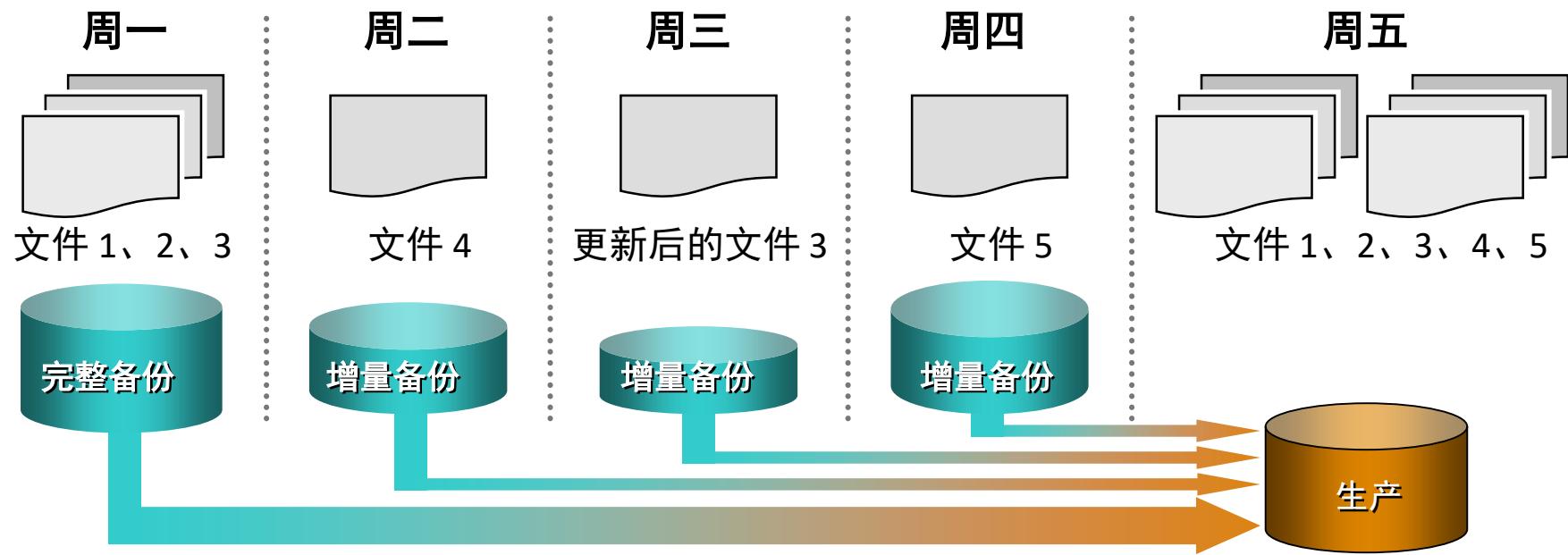


## 累积（差异）备份



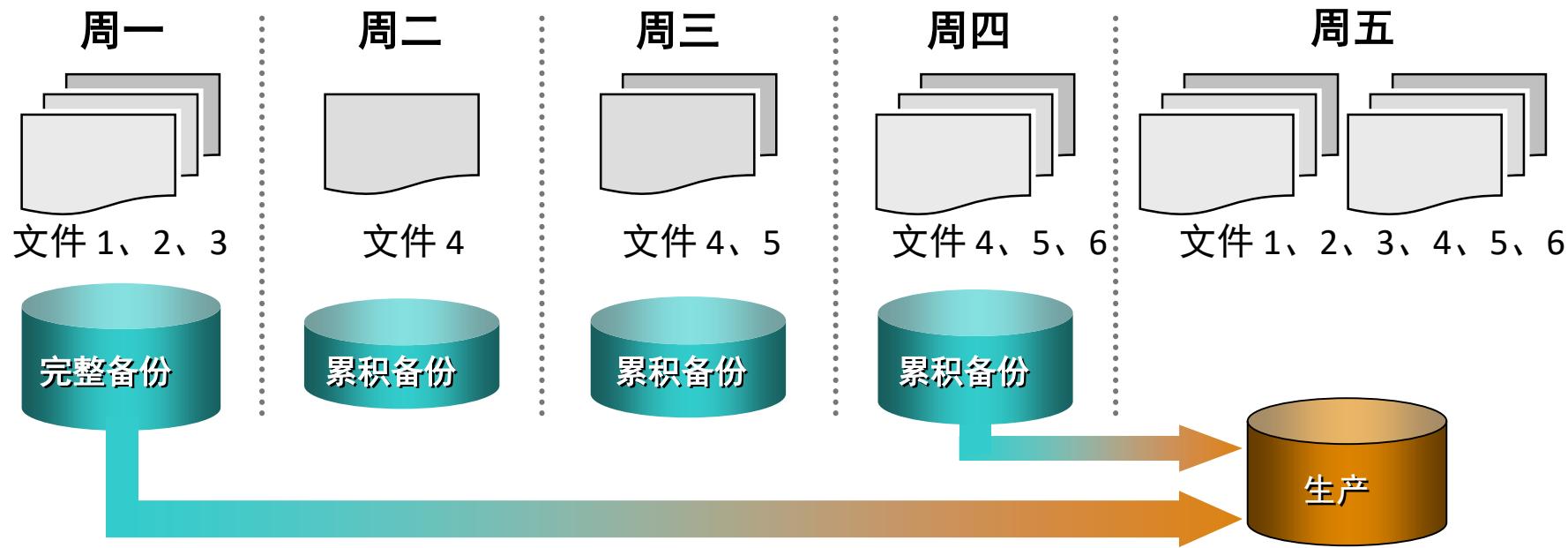
数据备份量

# 从增量备份恢复



- 要备份的文件数更少，因此，备份所用的时间更少且需要的存储空间更少
- 由于必须应用上次的完整备份以及后续所有的增量备份，因此恢复时间更长

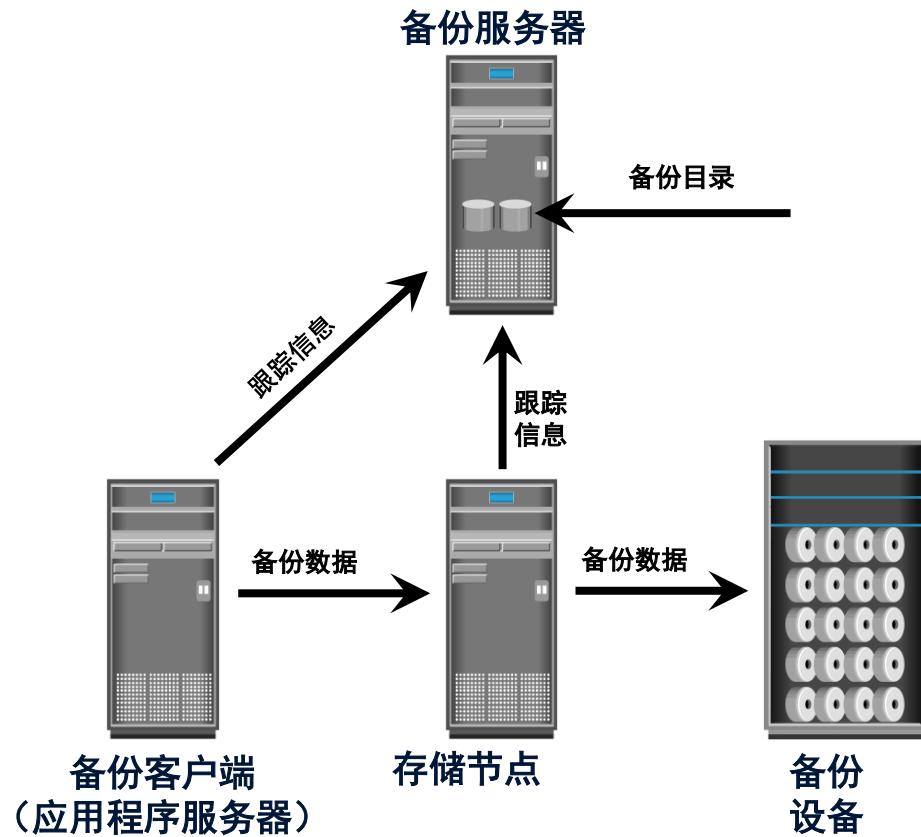
# 从累积备份恢复



- 要备份的文件更多，因此，备份时间更长，所需的存储空间更多
- 由于只须应用上次完整备份和上次累积备份，因此恢复速度更快

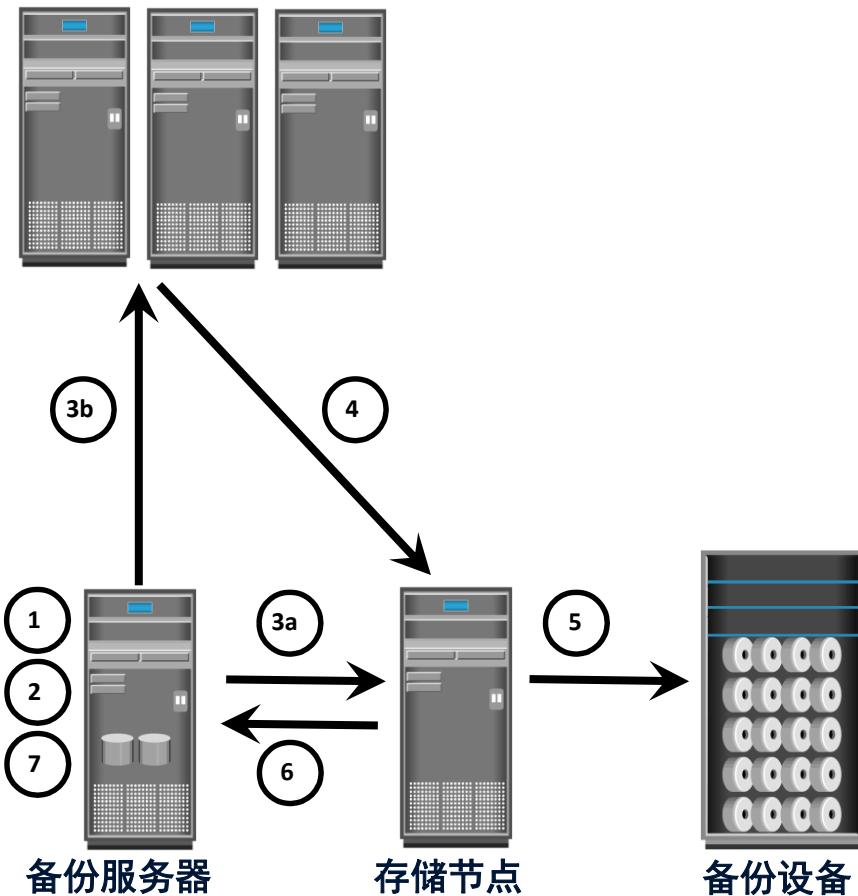
# 备份体系结构

- 备份客户端
  - ▶ 收集要备份的数据并将其发送到存储节点
- 备份服务器
  - ▶ 管理备份操作和维护备份目录
- 存储节点
  - ▶ 负责将数据写入备份设备
  - ▶ 管理备份设备



# 备份操作

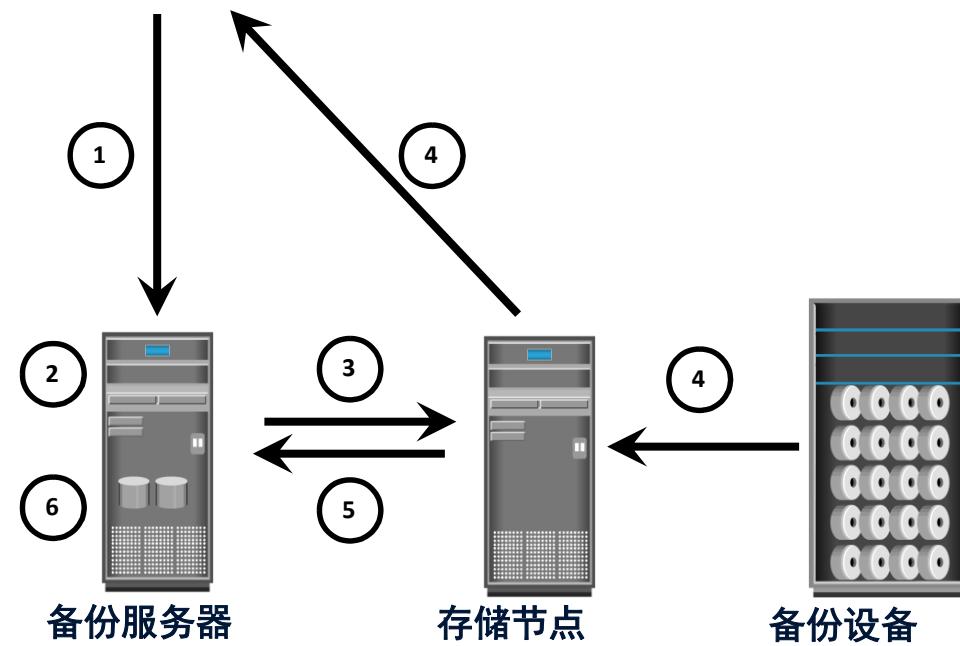
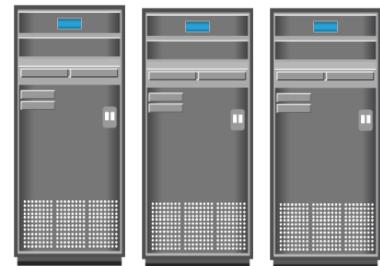
应用程序服务器  
(备份客户端)



- 1 备份服务器启动预定的备份流程。
- 2 备份服务器从备份目录检索备份相关的信息。
- 3a 备份服务器指示存储节点在备份设备中加载备份介质。
- 3b 备份服务器指示备份客户端将要备份的数据发送到存储节点。
- 4 备份客户端将数据发送到存储节点并在备份服务器上更新备份目录。
- 5 存储节点将数据发送到备份设备。
- 6 存储节点将元数据和介质信息发送到备份服务器。
- 7 备份服务器更新备份目录。

# 恢复操作

应用程序服务器  
(备份客户端)



- ① 备份客户端请求备份服务器进行数据恢复。
- ② 备份服务器扫描备份目录，以识别要恢复的数据和要接收数据的客户端。
- ③ 备份服务器指示存储节点在备份设备中加载备份介质。
- ④ 随后会读取数据，并将其发送到备份客户端。
- ⑤ 存储节点将恢复元数据发送到备份服务器。
- ⑥ 备份服务器更新备份目录。

# 备份方法

- 两种备份方法根据执行备份时应用程序的状态进行备份
  - ▶ 热备份或在线备份
    - ▶ 备份期间应用程序处于启动和运行状态，同时用户在访问其数据
    - ▶ 打开文件代理可用于备份打开文件
  - ▶ 冷备份或离线备份
    - ▶ 要求在备份过程中关闭应用程序
- 裸机恢复
  - ▶ OS、硬件和应用程序配置均已适当备份，可进行完整系统恢复
  - ▶ 服务器配置备份 (SCB) 也可将服务器恢复到不同硬件上

# 服务器配置备份

- 根据用户定义的时间表创建和备份服务器配置文件
  - ▶ 配置文件用于在生成服务器故障时配置恢复服务器
  - ▶ 配置文件包括 OS 配置、网络配置、安全配置、注册表设置、应用程序配置
- 使用两种类型的配置文件
  - ▶ 基本配置文件
    - ▶ 包含恢复服务器所需的 OS 的关键要素
  - ▶ 扩展配置文件
    - ▶ 通常比基本配置文件大且包含重建应用程序环境所需的所有信息

# 备份/恢复的主要注意事项

- 客户业务需要确定：
  - ▶ 恢复要求是什么 – RPO 还是 RTO？
  - ▶ 哪些数据需要备份？
  - ▶ 数据备份的频率如何？
  - ▶ 备份需要花多长时间？
  - ▶ 要生成多少拷贝？
  - ▶ 备份拷贝要保留多长时间？
  - ▶ 文件的位置、大小和数量？

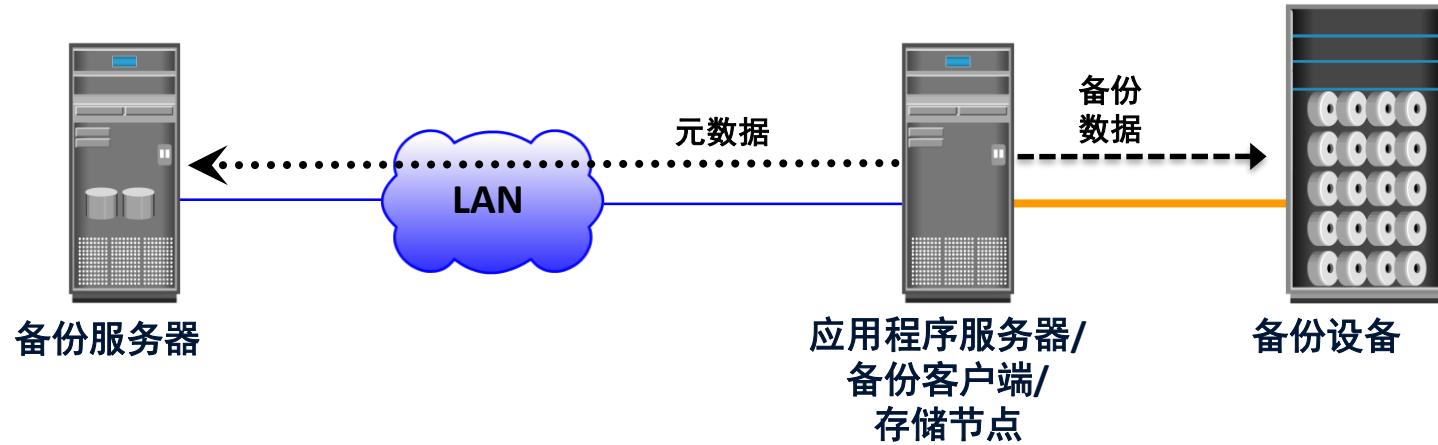
# 模块 10：备份和归档

## 第 2 课：备份拓扑以及在 NAS 环境中备份

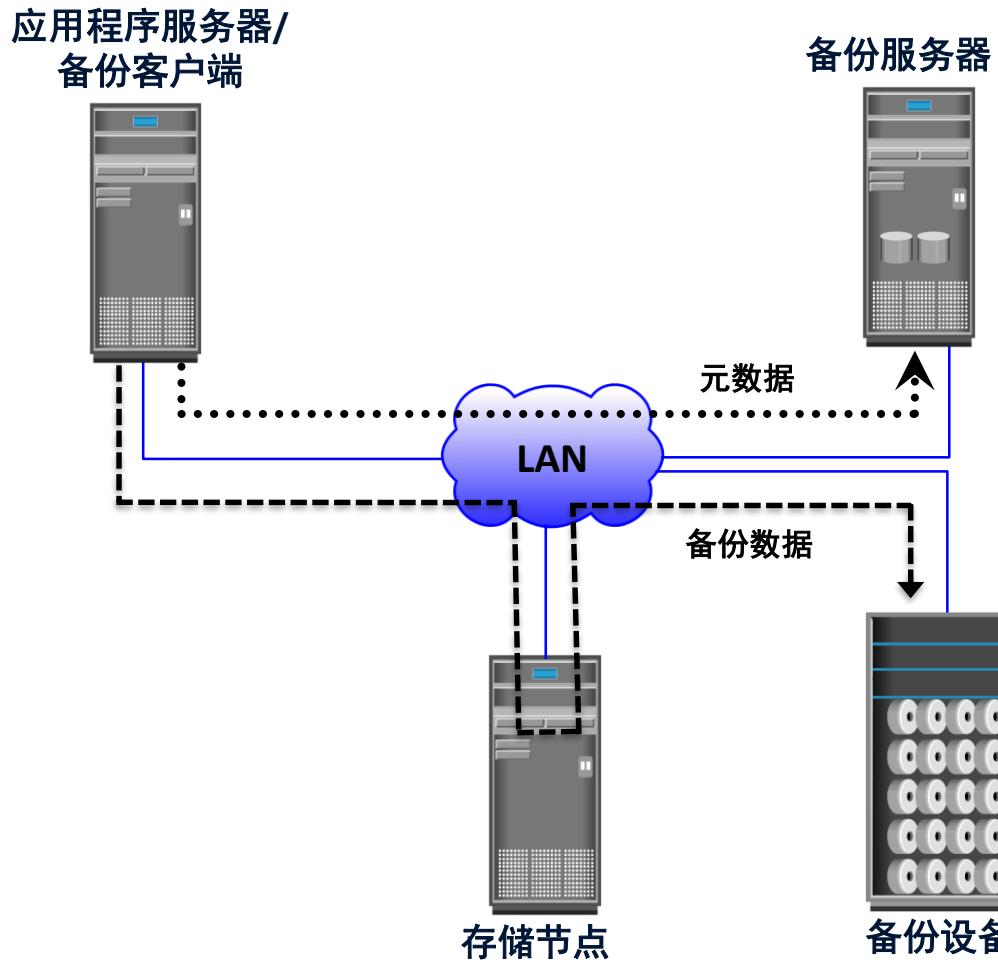
本课程将讲述下列主题：

- 通用备份拓扑
- 在 NAS 环境中备份

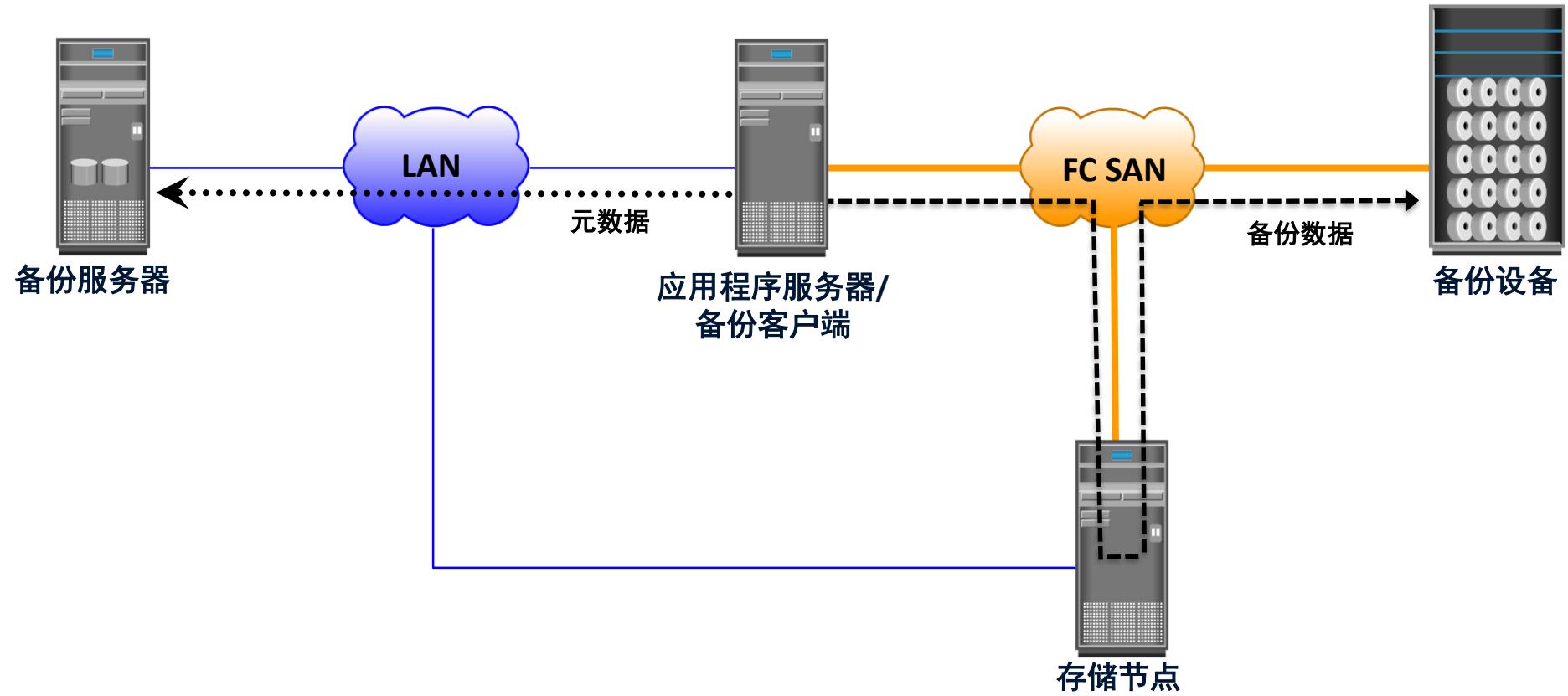
# 直连备份



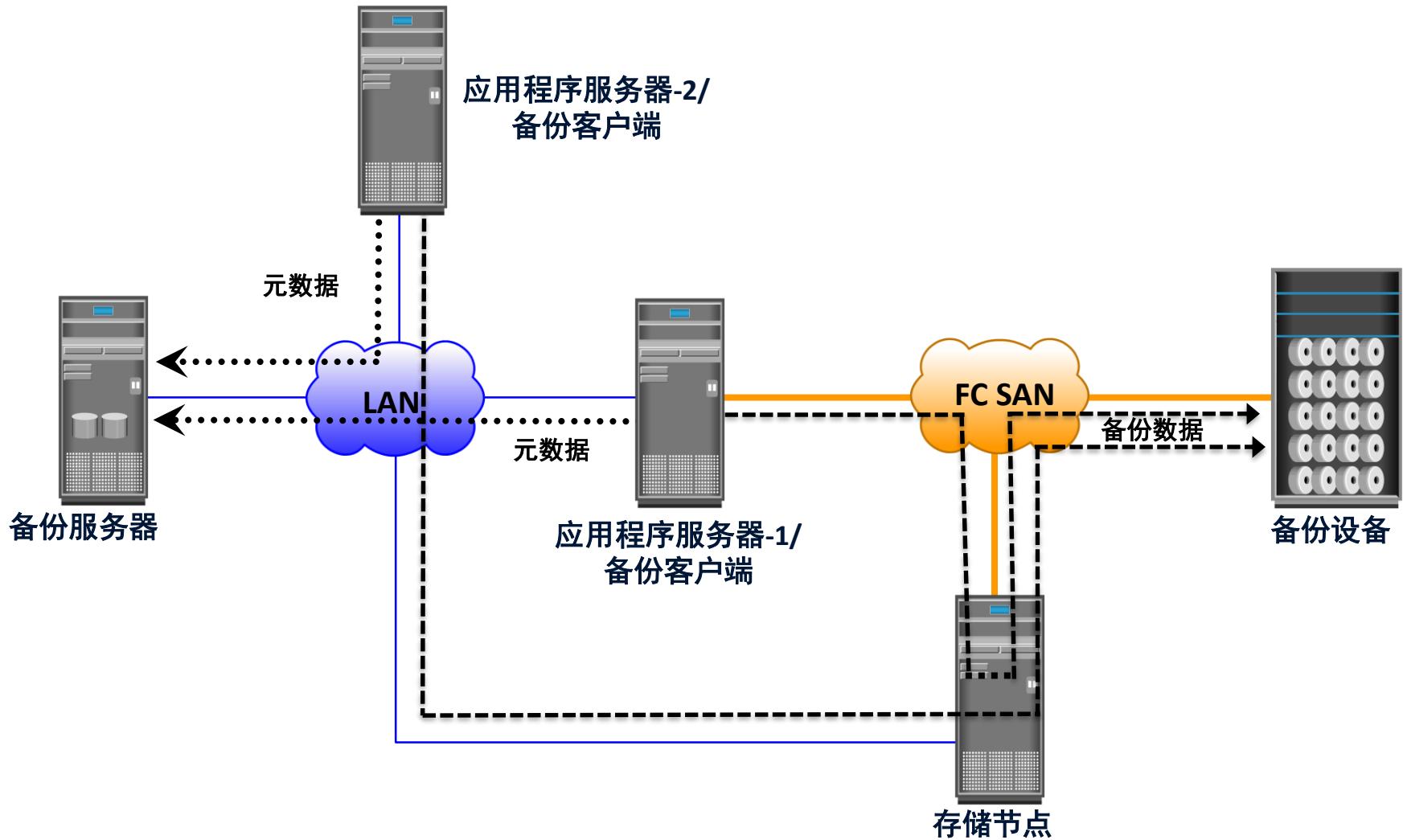
# 基于 LAN 的备份



# 基于 SAN 的备份



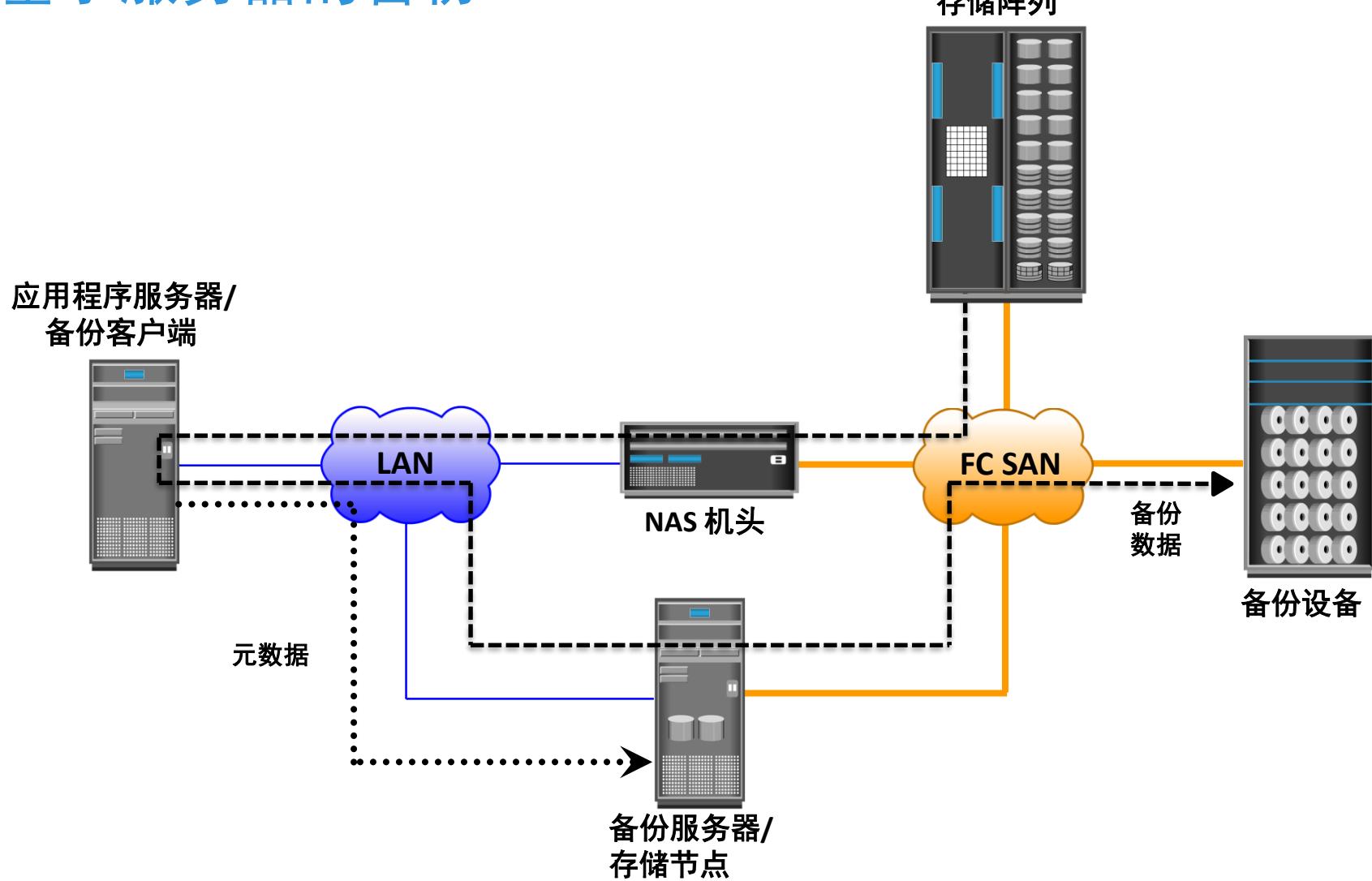
# 混合备份拓扑



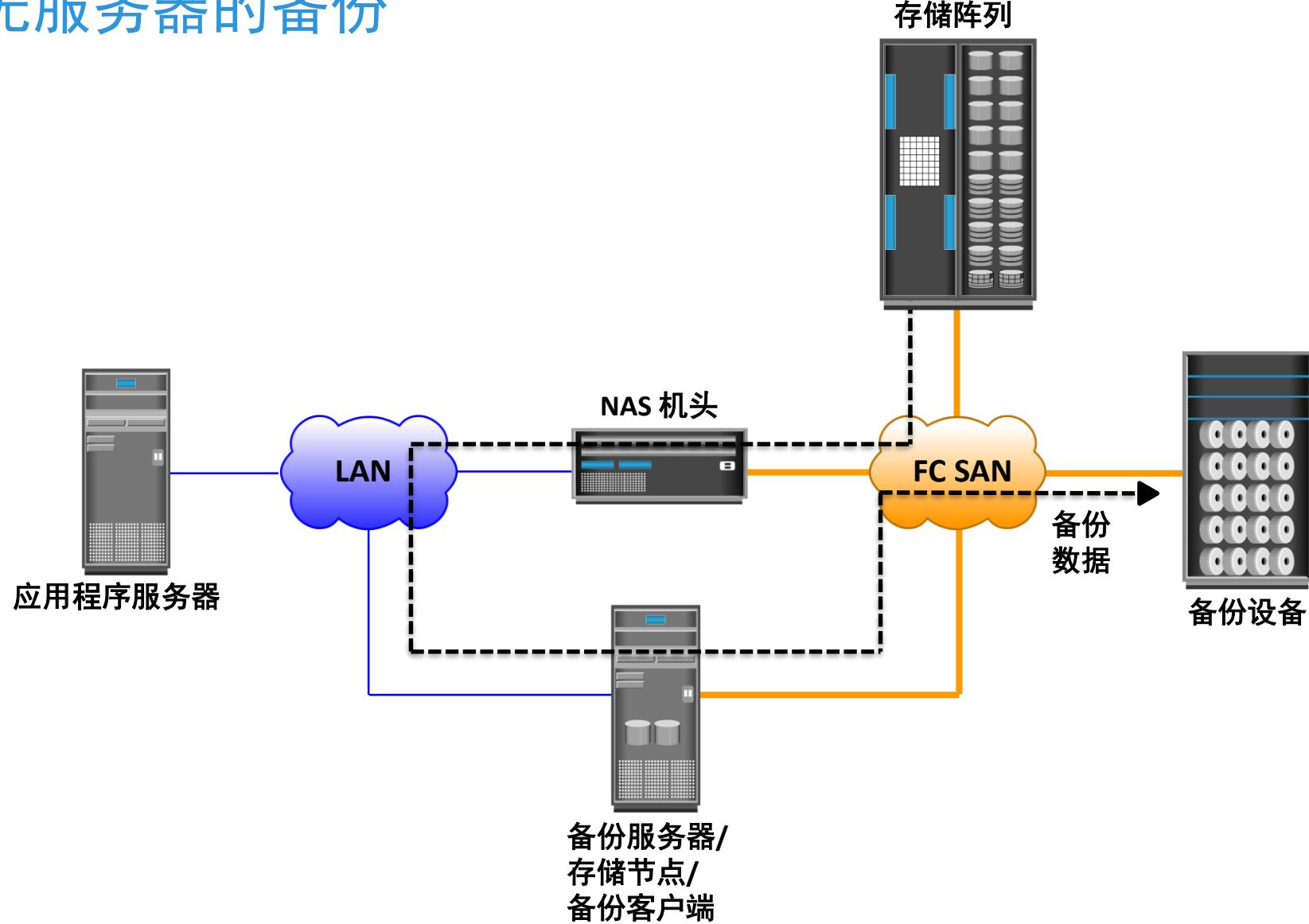
# 在 NAS 环境中备份

- NAS 环境中的常见备份实施有：
  - ▶ 基于服务器的备份
  - ▶ 无服务器的备份
  - ▶ NDMP 双向备份
  - ▶ NDMP 3 向备份

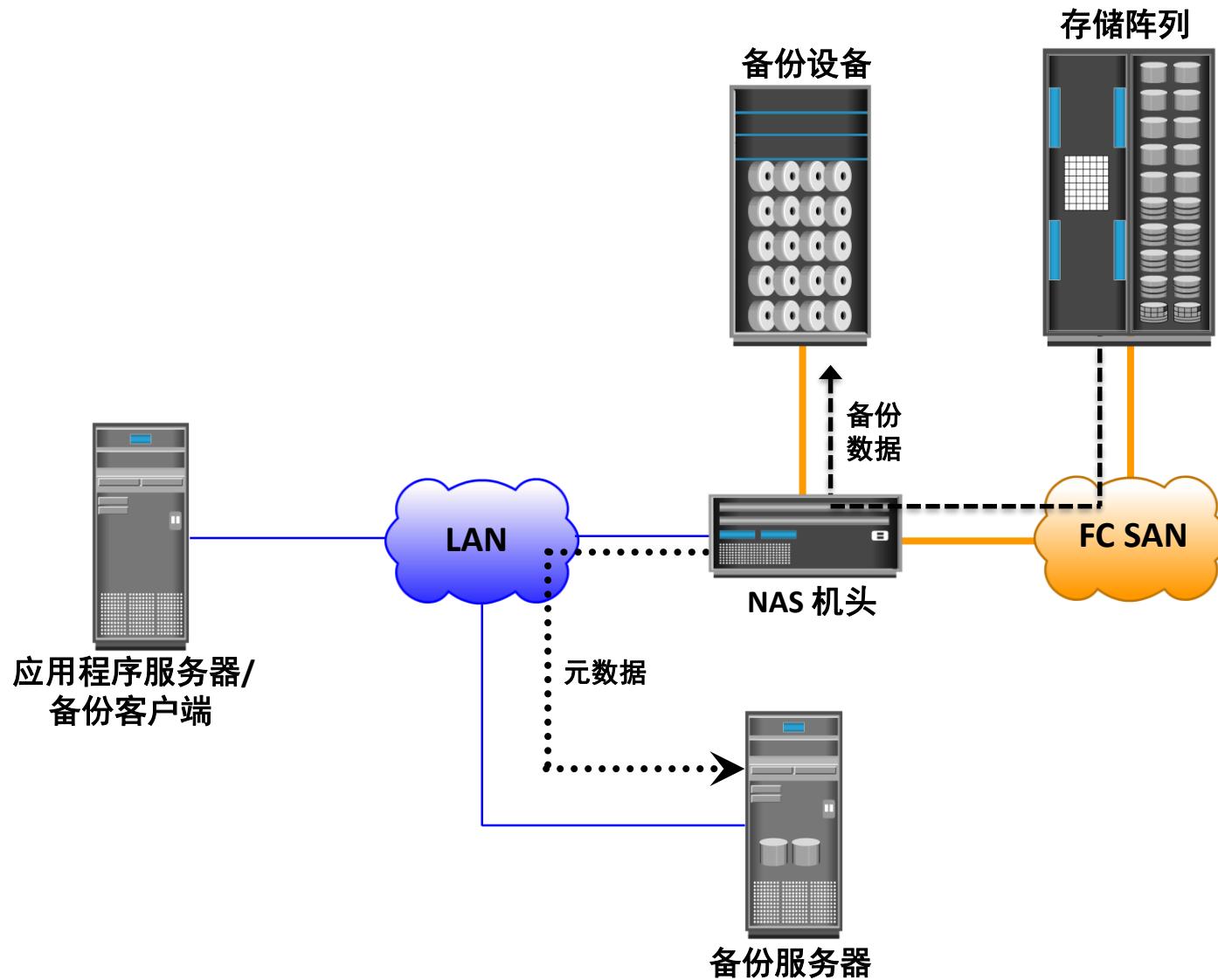
# 基于服务器的备份



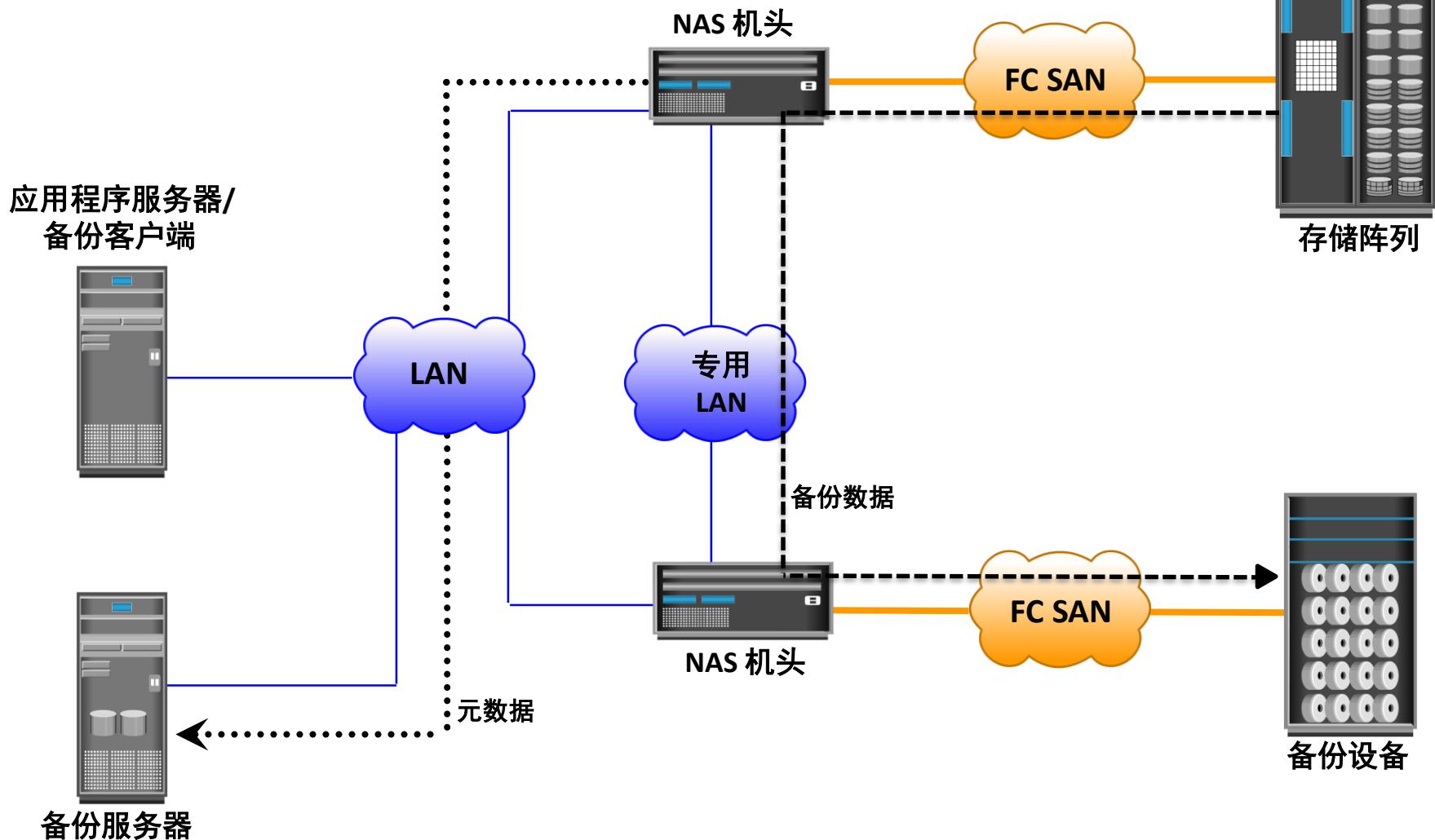
# 无服务器的备份



# NDMP 双向备份



# NDMP 3 向备份



# 模块 10：备份和归档

## 第 3 课：备份目标

本课程将讲述下列主题：

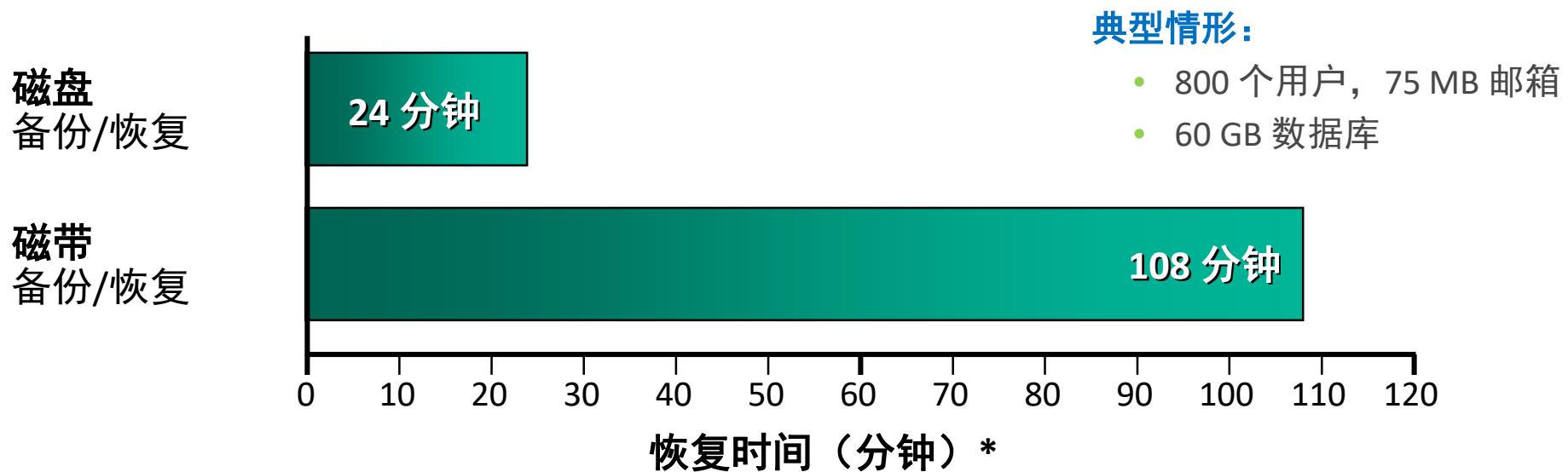
- 磁带备份
- 磁盘备份
- 虚拟磁带备份

# 磁带备份

- 传统低成本解决方案
- 磁带机用于从磁带读取数据/向磁带写入数据
- 顺序/线性访问
- 多数据流会提高介质性能
  - ▶ 从单一磁带的多数据流写入数据
- 磁带限制
  - ▶ 备份和恢复操作由于连读访问速度缓慢
  - ▶ 磁带的磨损
  - ▶ 运送/处理难题
  - ▶ 磁带存储需要受控环境
  - ▶ 导致“来回走带影响”或“后端差错”

# 磁盘备份

- 备份和恢复总体性能提高
  - ▶ 可随机访问
- 更可靠
- 可由多个主机同时访问

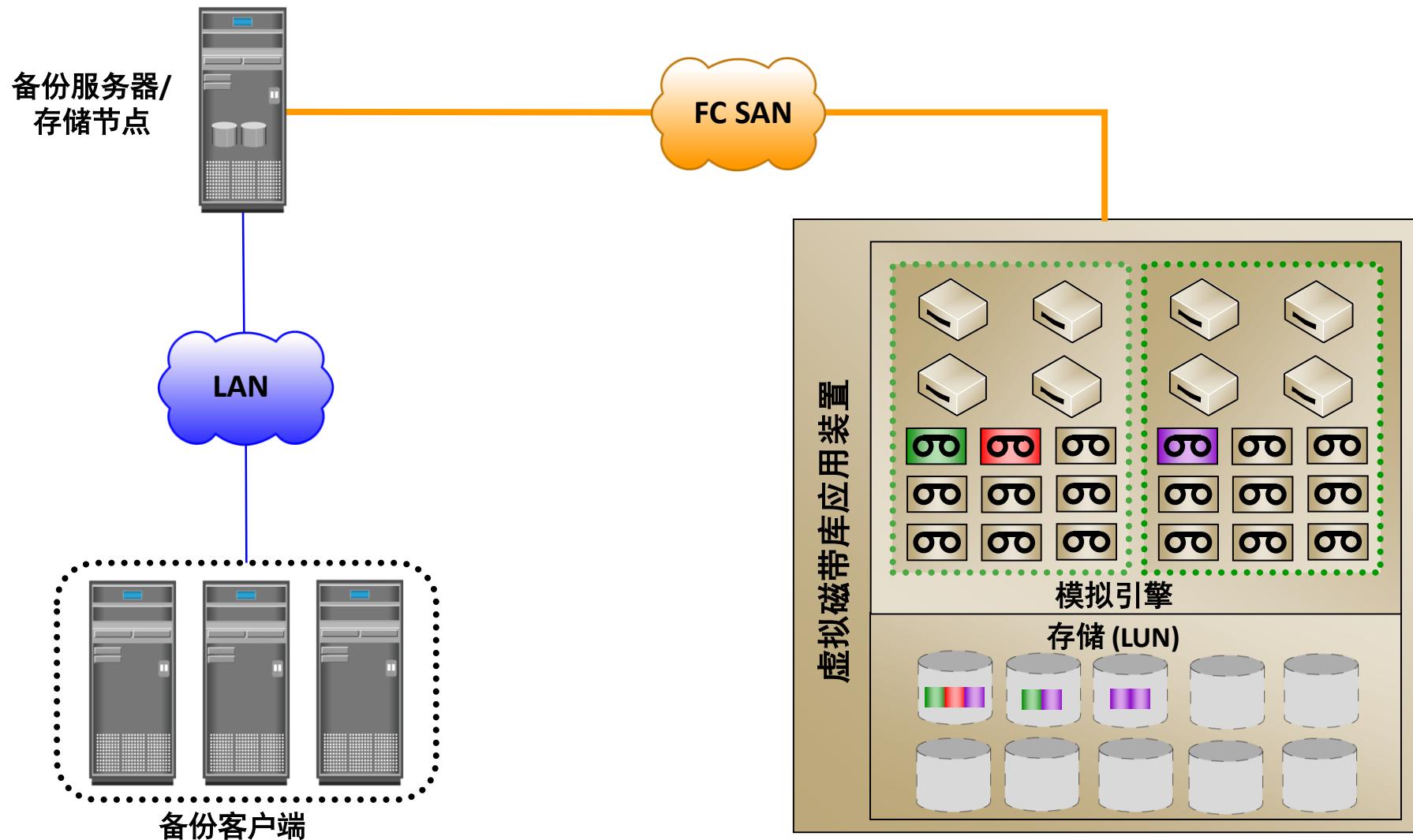


来源：EMC Engineering 和 EMC IT

# 虚拟磁带备份

- 模拟的并将作为磁带提供给备份软件的磁盘
- 不需要在早期备份软件中添加任何模块或进行任何更改
- 提供与物理磁带相比更佳的单数据流性能和可靠性
- 在线且随机的磁盘访问
  - ▶ 提供更快的备份和恢复

# 虚拟磁带库



# 备份目标比较

	磁带	磁盘	虚拟磁带
异地复制功能	否	是	是
可靠性	无固有保护方法	RAID, 备盘	RAID, 备盘
性能	低	高	高
使用	仅用于备份	多种用途（备份和生产）	仅用于备份

# 模块 10：备份和归档

## 第 4 课：重复数据消除

本课程将讲述下列主题：

- 重复数据消除概述
- 重复数据消除方法
- 重复数据消除实施
- 重复数据消除的主要优势

# 什么是重复数据消除？

## 重复数据消除

它是识别并消除冗余数据的过程。

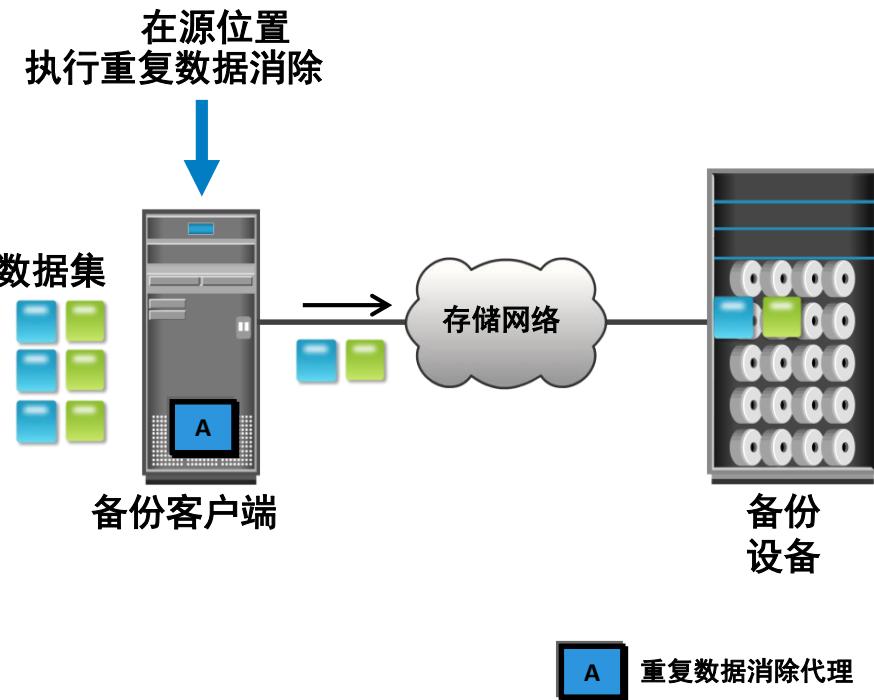
- 重复数据消除方法
  - ▶ 文件级
  - ▶ 子文件级
- 重复数据消除实施
  - ▶ 基于源的重复数据消除
  - ▶ 基于目标的重复数据消除

# 重复数据消除方法

- 文件级重复数据消除（单实例存储）
  - ▶ 检测并删除相同文件的冗余拷贝
  - ▶ 存储文件后，对同一文件的其他所有引用都将引用原始拷贝
- 子文件重复数据消除
  - ▶ 在文件内部和文件之间检测冗余数据
  - ▶ 两种方法
    - ▶ 固定长度数据块
    - ▶ 可变长度段

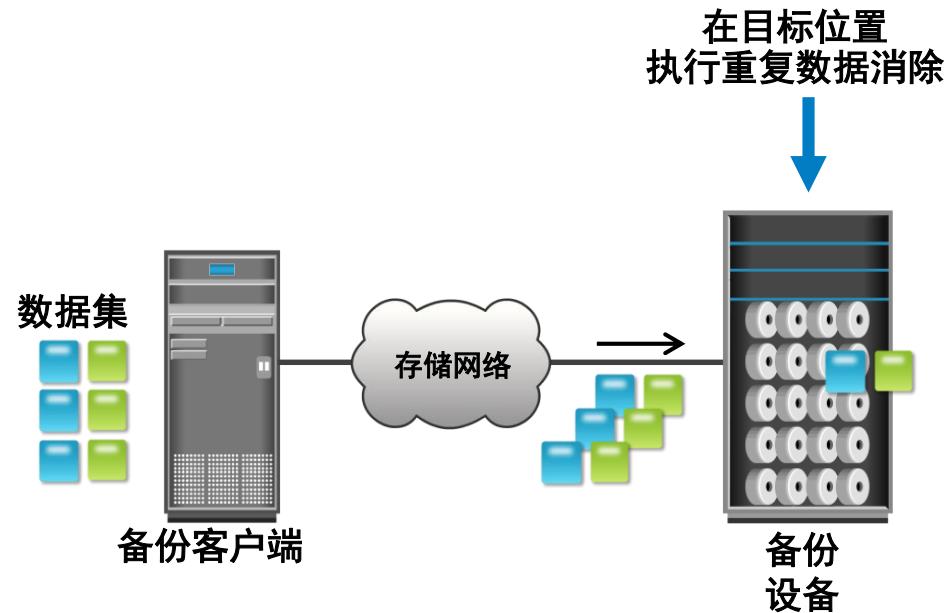
# 重复数据消除实施 – 基于源的

- 数据在源位置进行重复数据消除（备份客户端）
- 备份客户端仅通过网络发送新的唯一段
- 存储容量和网络带宽需求减少
- 备份客户端的开销增加



# 重复数据消除实施 – 基于目标的

- 数据在目标位置进行重复数据消除
  - ▶ 线内
  - ▶ 处理后
- 从重复数据消除过程中卸载备份客户端
- 所有备份数据都会遍历网络



# 重复数据消除 – 主要优势

- 降低了基础架构成本
  - ▶ 通过消除冗余数据，容纳备份映像所需的存储更少
- 实现更长时间的保留
  - ▶ 减少了每日备份的冗余内容量，因此，用户可以扩展其保留策略
- 缩短备份时段
  - ▶ 要备份的数据更少，这将缩短备份时段
- 减少了备份带宽需求
  - ▶ 基于源的重复数据消除可在通过网络发送冗余数据之前消除冗余数据

# 使用情形：远程办公室/分支办公室备份

- 跨多个位置保护企业分支和远程办公室的数据对企业而言至关重要
- 将远程办公室的数据备份到集中化的数据中心受到限制，原因为于
  - ▶ 通过网络发送大量数据所涉及的时间和成本
- 基于磁盘的备份解决方案与基于源的重复数据消除配合使用可消除集中备份远程办公室数据的难题
  - ▶ 减少网络带宽需求
  - ▶ 缩短备份时段

# 模块 10：备份和归档

## 第 5 课：虚拟化环境中的备份

本课程将讲述下列主题：

- 传统备份方法
- 基于映像的备份

# 虚拟化环境中的备份概述

- 备份选项
  - ▶ 传统备份方法
  - ▶ 基于映像的备份方法
- 备份优化
  - ▶ 重复数据消除

# 传统备份方法

- 虚拟机上的备份代理
  - ▶ 需要在运行在虚拟机管理程序上的每个虚拟机上安装备份代理
  - ▶ 只能备份虚拟磁盘数据
  - ▶ 不会捕获虚拟机文件，例如虚拟机交换文件、配置文件
  - ▶ 虚拟机恢复的挑战

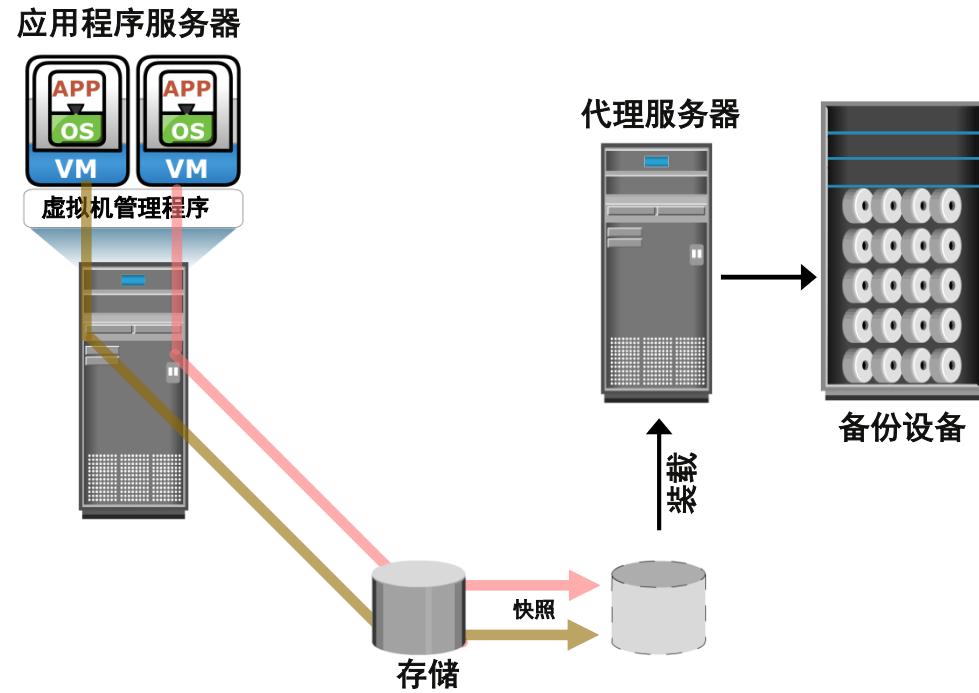
- 虚拟机管理程序上的备份代理
  - ▶ 只需在虚拟机管理程序上安装备份代理
  - ▶ 备份所有虚拟机文件



 = 备份代理

# 基于映像的备份

- 创建来宾操作系统、其数据、虚拟机状态和配置的拷贝
  - ▶ 备份另存为一个文件 – “映像”
  - ▶ 在代理服务器上装载映像
  - ▶ 可减轻虚拟机管理程序的备份处理负担
- 支持快速恢复虚拟机



# 模块 10：备份和归档

## 第 6 课：数据归档

本课程将讲述下列主题：

- 固定内容
- 数据归档
- 归档解决方案体系结构

# 固定内容

- 固定内容增长每年超过 90%
  - ▶ 大量新创建的信息归为此类别
  - ▶ 新法规要求保留和数据保护

## 固定内容示例

### 电子文档

- 合同和索赔
- 电子邮件附件
- 财务电子表格
- CAD/CAM 设计
- 演示文稿

### 数字记录

- 文档
  - 支票、证券交易
  - 历史资料保存
- 照片
  - 人物照片/专业摄影
- 调查
  - 地震、天文学、地理学

### 富媒体

- 医疗
  - X 光片、MRI、CT 扫描
- 视频
  - 新闻/媒体、电影
  - 安全性督察
- 音频
  - 语音邮件
  - 无线电

# 数据归档

- 存储固定内容的存储库
- 使组织能够将其数据保留很长一段时间，以便
  - ▶ 满足法规遵从性要求
  - ▶ 制定新的收入策略
- 可以多种形式实施归档
  - ▶ 在线
  - ▶ 近线
  - ▶ 离线

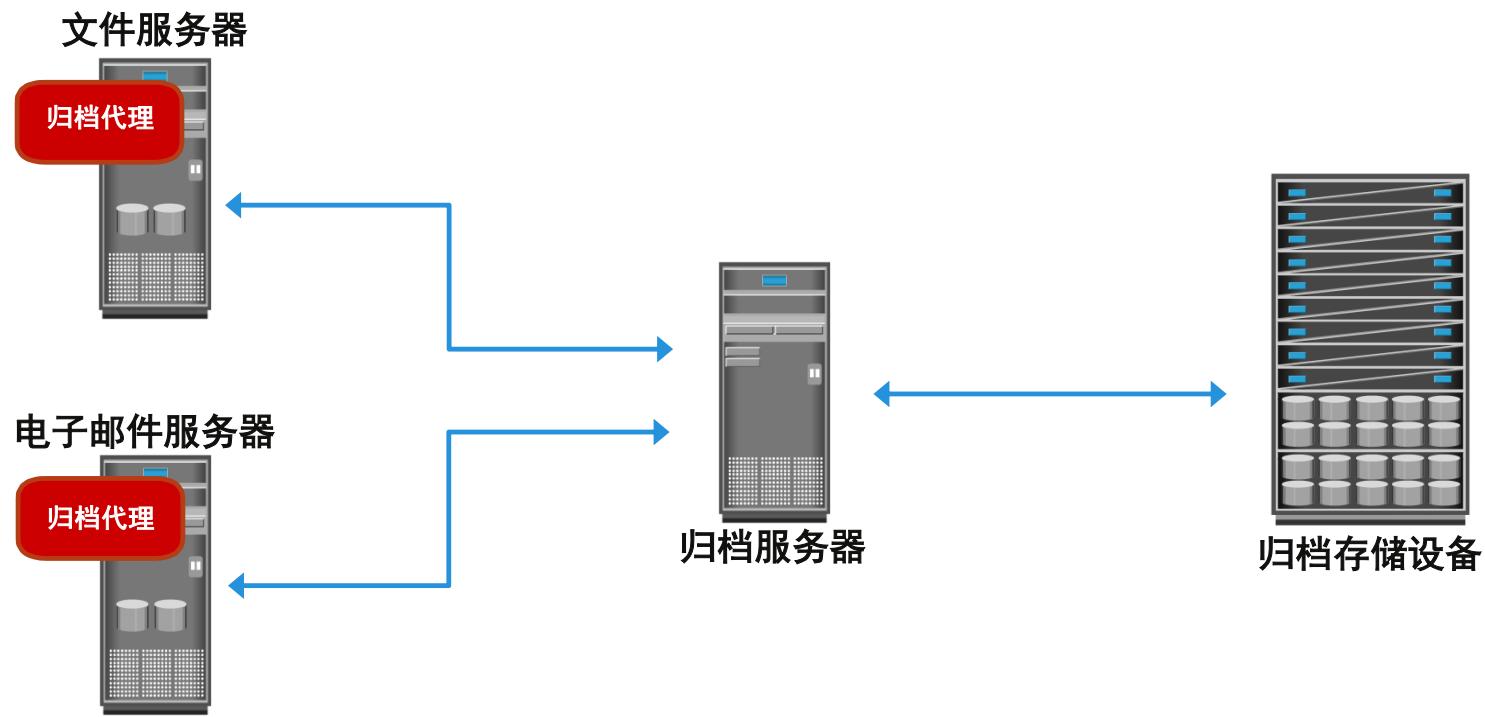
# 传统归档解决方案的挑战

- 磁带和光学介质均易磨损
  - ▶ 涉及运营、管理和维护开销
- 无法智能识别重复数据
  - ▶ 可能会多次归档相同内容
- 不足以实现长期保留（数年到数十年）
- 无法提供对固定内容的在线且快速的访问

# 内容寻址存储 – 一个归档解决方案

- 成为传统归档解决方案的替代方案的基于磁盘的存储
- 提供对归档数据的在线访问
- 使组织能够满足所需的 SLA 要求
- 提供存储归档数据所需的功能
  - ▶ 内容真实性和内容完整性
  - ▶ 位置独立性
  - ▶ 单实例存储
  - ▶ 强制执行保留
  - ▶ 数据保护

# 归档解决方案体系结构



## 使用情形：电子邮件归档

- 根据策略将电子邮件从主存储移动到归档存储
- 在主存储上节约空间
- 支持在归档文件中将电子邮件保留更长时间以满足法规要求
- 为最终用户提供几乎无限的邮箱空间
- 文件归档是可从归档解决方案中受益的另一种使用情形

# 模块 10：备份和归档

## 付诸实践的概念

- EMC NetWorker
- EMC Avamar
- EMC Data Domain

# EMC NetWorker

- 集中化、自动化和加速整个企业的数据备份和恢复操作
- 主要功能
  - ▶ 支持 Windows、UNIX、Linux 等异构平台，还支持虚拟环境
  - ▶ 支持各种备份目标 – 磁带、磁盘和虚拟磁带
  - ▶ 支持数据的多路复用（或多数据流）
  - ▶ 通过分别与 EMC Avamar 和 EMC Data Domain 集成提供基于源的和基于目标的重复数据消除功能
  - ▶ 云备份选项支持将数据备份到云

# EMC Avamar

- 可提供基于源的重复数据消除的基于磁盘的备份和恢复解决方案
- 三个主要组件包括 Avamar 服务器、Avamar 备份客户端和 Avamar Administrator
- Avamar 服务器包括
  - ▶ 仅软件、Avamar Data Store、Avamar Virtual Edition

# EMC Data Domain

- 基于目标的重复数据消除解决方案
- 提供技术优势
  - ▶ 数据无损体系结构
  - ▶ Data Domain Stream-Informed Segment Layout (SISL) 扩展体系结构
  - ▶ 支持本机复制技术
  - ▶ 全局压缩
- EMC Data Domain Archiver
  - ▶ 用于长期保留备份和归档数据的解决方案
  - ▶ 采用内部分层方法设计
  - ▶ 支持重复数据消除技术

# 模块 10：总结

本模块涵盖以下要点：

- 备份粒度
- 备份和恢复操作
- 备份拓扑
- 备份目标
- 重复数据消除
- 虚拟化环境中的备份
- 数据归档

# 知识测验 – 1

- 关于增量备份，以下哪项是正确的？
  - A. 恢复只需上次完整备份和上次增量备份
  - B. 恢复只需上次增量备份
  - C. 复制自从上次完整备份或增量备份以来有更改的数据 
  - D. 复制自从上次完整备份以来有更改的数据
- 虚拟化环境中基于映像的备份与传统备份方法相比有何优势？
  - A. 减轻虚拟机管理程序的备份处理负担 
  - B. 更快，因为它仅复制虚拟机磁盘数据
  - C. 更快，因为它仅复制虚拟机配置数据
  - D. 所需的空间是整个备份数据的一小部分

## 知识测验 – 2

- 以下哪一项准确描述了备份服务器的角色？
  - A. 收集要备份的数据并将其发送到存储节点
  - B. 负责将客户端发送的数据写入备份设备
  - C. 管理备份操作和维护备份目录 ❤
  - D. 控制磁带库中的机械臂
- 在磁带备份环境中，“来回走带”是什么意思？
  - A. 从多数据流将数据写入单一磁带
  - B. 模拟磁盘驱动器并作为磁带提供给备份软件的过程
  - C. 当备份数据流中断时磁带机进行的重复的往复移动 ❤
  - D. 删除备份数据中冗余内容的过程

## 知识测验 – 3

- 基于源的重复数据消除有何优势？
  - A. 提高了备份客户端的性能
  - B. 提高了备份服务器的性能
  - C. 将备份时段减至零
  - D. 减少了备份的网络带宽需求 ❤

# 练习：备份/恢复

- 当前情况

- 在每个周日执行完整备份，在其余日期执行增量备份
- 备份期间必须关闭数据库
- 备份数据的多个冗余拷贝
- 网络带宽限制

- 业务需求

- 不必在备份期间关闭数据库
- 需要更快备份和恢复
- 消除备份数据的冗余拷贝

- 任务

- 提出解决方案并论证