

数据迁移方法概述

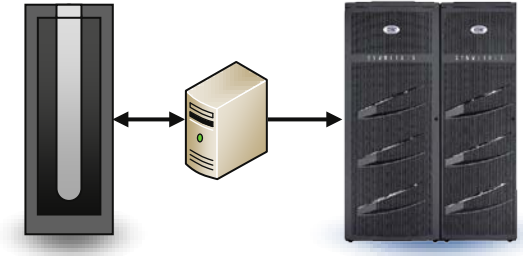
数据迁移方法的选择是建立在对以上各环节的具体分析基础之上,目前开放平台存储整合建设中可以采用的数据迁移方法主要下表所示的六种方法:

序号	数据迁移方法	方法简述	举例
1	逻辑卷数据镜像方法	对需要迁移的每个卷都做逻辑卷数据镜像	如 IBM LVM, Veritas VxVM
2	直接拷贝方法	利用操作系统命令直接拷贝要迁移的数据,然后复制到要迁移到的目的地	如 IBM AIX tar, dd, savevg, mksysb, cpio 等命令
3	备份恢复方法	利用备份管理软件对数据做备份,然后恢复到目的地	如 IBM TSM, EMC Legato NetWorker, Symantec Veritas NetBackup 等
4	数据库工具方法	使用数据库的自身工具对数据进行迁移	如 Oracle Export/Import, OracleDataGuard, GoldenGate, Oracle Logminer, Quest SharePlex
5	存储虚拟化的方法	通过存储虚拟化技术将数据从源端迁移到目的地	如 EMC 的 Invista、IBM 的 SVC 和 LSI StoreAge SVM 等
6	盘阵内复制方法	通过盘阵内的复制软件,将数据源卷复制到数据目标卷	如 EMC 的 TimeFinder, IBM FlashCopy, HDS ShadowImage 等
7	直接的阵列到阵列复制方法	通过盘阵复制软件对数据做迁移	如 EMC 的 SRDF, HDS 的 TrueCopy, IBM 的 Global Mirror 等
8	历史数据迁移方法	通过数据提取、转移、装载工具或定制程序进行装载	

1.1.1. 逻辑卷数据镜像方法

对于服务器操作系统已经采用逻辑卷管理器的系统,可以利用逻辑卷管理器的管理功能完成原有数据到新存储的迁移,支持联机迁移。

逻辑卷管理器可以理解为在物理存储设备和操作系统之间增加的一个逻辑存储管理层。



逻辑卷数据镜像方法

逻辑卷数据镜像方法的优点主要有:

支持任意存储系统之间的迁移

适合于主机存储的非经常性迁移

1.1.2. 直接拷贝方法

利用操作系统的一些命令可以实现磁盘数据的直接复制,适合于脱机迁移的环境。

1.1.3. 备份恢复方法

利用备份管理软件将原有数据备份到磁带然后恢复到新的存储设备中,对于联机要求高的环境,可以结合在线备份的方法,然后恢复到目的地。如 IBM TSM, EMC Legato NetWorker, Symantec Veritas NetBackup 等。

1.1.4. 数据库工具方法

对于数据库数据,还可使用数据库的自身工具,如 Oracle 自带复制工具 DataGuard,以及一些第三方的数据库复制工具,如 GoldenGate。

其中 Data Guard 是传输日志到在线数据迁移目标端,然后在在线数据迁移目标端解析成 SQL 语句执行,网络上传输的是归档日志,而以 Golden Gate、SharePlex 为代表的备份软件是在生产端解析日志,通过网络把解析出的 SQL 语句传输到在线数据迁移目标端执行。

相比之下,数据库工具方法更加节省带宽,操作更为灵活。

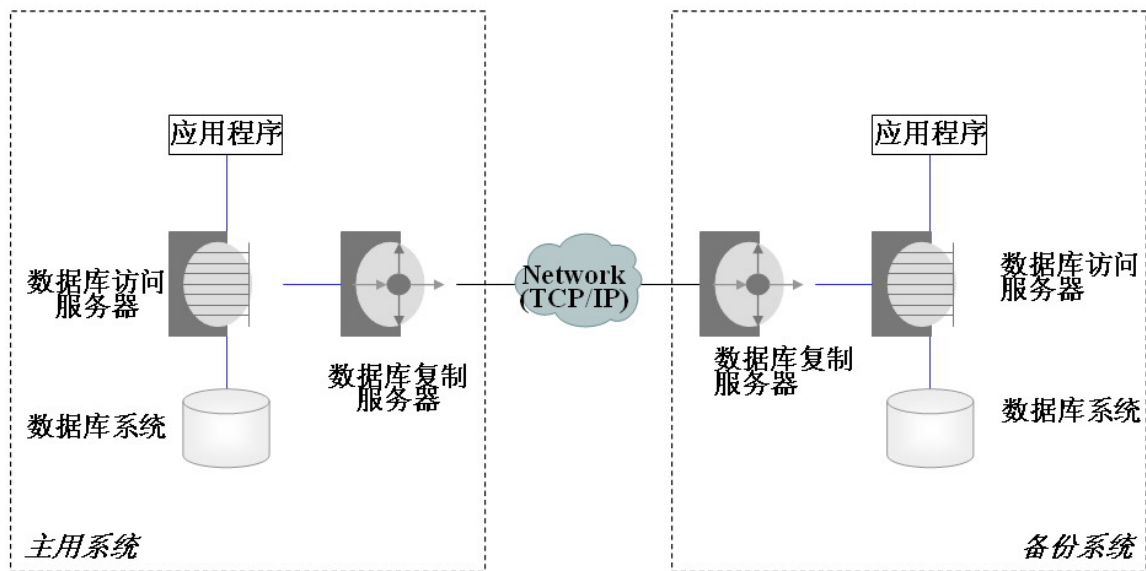
1) **实时复制:** 当源数据库内容被修改时,目标数据库内容实时地被修改,此种复制方

式对网络可靠性要求高。

2) **定时复制**：当源数据库内容被修改时，目标数据库内容会按照时间间隔，周期性地按照生产中心的更新情况进行刷新，时间间隔可长(几天或几个月)可短(几分钟或几秒钟)。

3) **存储转发复制**：当源数据库内容被修改时，源数据库服务器会先将修改操作 Log 存储于本地，待时机成熟再转发给目标数据库。远程数据库复制的实质是实现源、目标数据库的数据同步（实时或者准实时同步）。即是将主用系统数据库操作 Log 实时或周期性地复制到备用系统数据库中执行，实现二者数据的一致性。目标数据库复制对主机的性能有一定影响，可能增加对磁盘存储容量的需求（包括对 Log 的存储）。

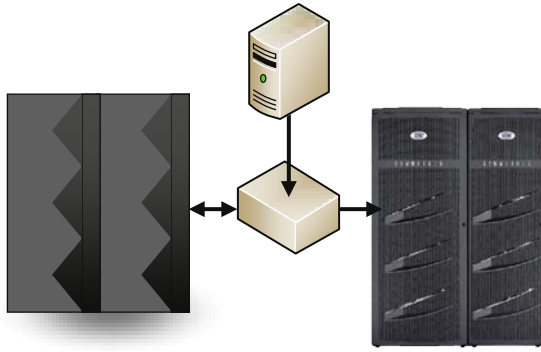
下图为远程数据库复制逻辑结构（举例）示意。



数据库工具方法的核心是数据库复制，需要主机同构，存储可异构。数据库、操作系统、中间件版本必须同构。

1.1.5. 存储虚拟化的方法

存储虚拟化的方法是在源和目的地增加必要的存储虚拟化设备，通过网络（包括 SAN 网络和 IP 网络、仅仅 SAN 网络或仅 IP 网络），将数据从源端迁移到目的地。采用这种技术的有 EMC 的 Invista、IBM 的 SVC 和 LSI StoreAge SVM 等。



通过网络的方法

通过网络的方法的主要优点有：

兼容现有主流存储设备，如：IBM，EMC，HDS，LSI Logic

支持不同厂商不同品牌存储设备间的数据迁移和容灾

适合于频繁移动数据的大型企业

1.1.6. 盘阵内复制方法

盘阵内复制方法是通过盘阵内的复制软件，将数据源卷复制到数据目标卷，主要步骤有：

定义源卷

定义目标卷（目标卷和源卷在同一盘阵内）

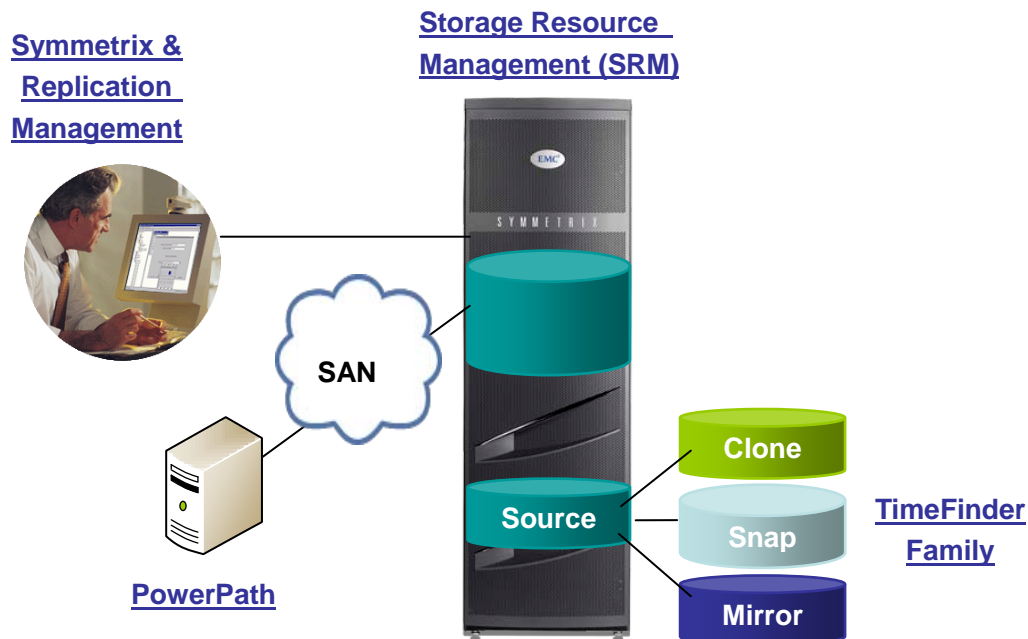
将源卷和目标卷配对

初始化同步目标卷，将源卷中的数据复制到目标卷，直到数据完全同步，两者进入同步状态，在此之前，有可能因种种原因需要临时中断同步过程，然后要进行再同步。

分离源卷和目标卷

将目标卷加载到服务器，和相关应用建立联系，驱动相关应用。

盘阵内复制方法的典型代表有 EMC 的 TimFinder，IBM FlashCopy，HDS ShadowImage 等。



从上图可以看出，源卷为 Source，目标卷有三种方式，一种是 Clone，一种是 Snap，另一种是 Mirror。

1.1.7. 直接的阵列到阵列复制方法



直接的阵列到阵列复制方法

这种方法是通过某种盘阵到盘阵的复制软件将数据从一个盘阵复制到另一个盘阵，即写数据到本地盘阵时，同步或异步地将数据写到远程盘阵。复制过程由磁盘阵列完成，不需要消耗服务器资源。

采用这种方法的有 EMC 的 SRDF，HDS 的 TrueCopy，IBM 的 Global Mirror 等。

主要步骤有：

定义源卷

定义目标卷（目标卷和源卷在不同盘阵内）

将源卷和目标卷配对

初始化同步目标卷，将源卷中的数据复制到目标卷，直到数据完全同步，两者进入同步状态，在此之前，有可能因种种原因需要临时中断同步过程，然后要进行再同步。

分离源卷和目标卷

将目标卷加载到服务器，和相关应用建立联系，驱动相关应用。

EMC 的 SANCopy 还支持异构盘阵之间的复制，不过 SANCopy 软件需要安装在 EMC CX 系列盘阵上，如果用户环境中没有 CX 系列盘阵，SANCopy 方案适合。

EMC 还有 Open Replicator for Symmetrix，支持“任意到 Symmetrix DMX”的迁移，适合于将数据移动到 Symmetrix 平台或从中移动数据。

各种数据迁移方法都有不同的特点，适用于不同的数据迁移需求，通常在有联机迁移要求且迁移数据量大的情况下，一般采用逻辑卷数据镜像方法或**直接的阵列到阵列复制方法**来实现数据迁移，相对简单、高效。

如果系统没有逻辑卷管理软件，可以考虑采用在线备份恢复的方式来实现，这种方式较前者步骤复杂，但使用可靠、成熟，在满足备份窗口要求的情况下，也是一种很好的选择。

对于迁移数据量不大的系统，可以考虑采用脱机迁移的方法，这种方式下，采用直接拷贝的方式就显得简单，快捷。

对那些需要在线不停机做数据迁移，同时又要求不占用业务系统服务器资源，我们建议采用**直接的阵列到阵列复制方法**。

在一个大型的存储体系规划建设过程中，会涉及到很多不同的应用，数据特点也各不相同，因此，在整体的数据迁移过程中，**往往会用到多种数据迁移方法**。

1.1.8. 历史数据数据迁移方法

在本项目新系统切换前，可以利用 ETL (Extract Transform Load) 工具把旧系统中的历史数据抽取、转换，并装载到新系统中去。其中 ETL 工具将购买成熟的产品，同时辅助以一些项目组自主开发的程序。

原系统数据库结构分析

进行数据移植首先要分析原系统数据库的数据存储物理结构。这部分工作由于是在原系统中进行，并且对旧系统中数据库结构的正确理解是对数据移植正确与否的决定因素之一，因此该部分工作应由熟悉原系统的原系统集成厂商或省局负责。

中间文件格式定义

中间文件格式是描述参与数据移植各方传递数据格式的规范说明文档。在中间文件格式

中对所有的数据都必须做出准确的无二义性的说明。

中间文件格式由项目组制定，原系统集成商和省局共同参与商定。

客户私有数据转换处理

若原系统存有客户私有数据，如密码等，由于新旧系统采用的安全机制不同，一些关键数据如客户 PIN 等私有数据在进行数据移植时需要进行转换，将旧系统的加密数据转换为新系统的加密数据。这一转换存在一个解密再加密的过程。为了保护客户私有数据的安全，转换应在同一程序中进行，任何数据移植过程生成的文件中不能带有明码。

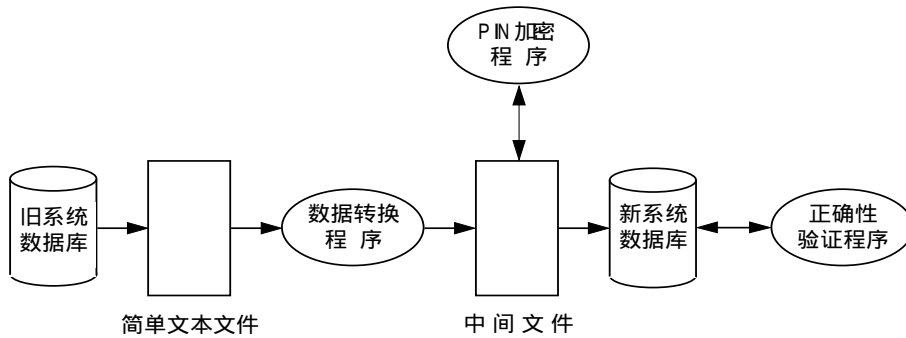
导出及预处理工具开发

导出工具的功能是在原系统中将数据从数据库中导出，并且生成与数据库结构类似的文本文件。预处理工具的功能是将导出的文本文件进行格式转换，并且生成中间文件格式的文本文件。

数据移植工具开发及测试

数据移植的功能是将中间格式的文本文件转换成可直接装载入新应用系统数据库中的格式。数据移植工具在进行数据格式转换的同时还将进行数据合法性和相关数据的一致性的检查。数据移植工具将不合法的数据生成错误文件，供数据修正参考。

数据移植的基本步骤如下图所示：



数据移植的基本步骤