

企业级信息系统数据迁移方法^①

李 喆, 万小健

(中国电信陕西公司信息技术支撑中心, 西安 710065)

摘 要: 数据是业务运转的基础, 这些数据都是关乎着企业命脉的宝贵资产。随着业务不断发展, 数据量的增长是十分惊人的, 在庞大的数据量背后, 企业面临的数据管理问题主要涉及存储空间不足、数据安全等几个方面。本文从企业级数据管理的角度, 阐述企业级数据迁移中遇到的问题、数据迁移方法论、以及技术方案的特点比较等。

关键词: 数据迁移; 数据安全; 方法论; 信息系统

Enterprise Information System Data Migration Methods

LI Zhe, WAN Xiao-Jian

(Information Technology Support Center Shaanxi Branch of Chinatelecom Co.,Ltd, Xian, 710065, China)

Abstract: Data is the basis of business operations, these data are related to the lifeblood of the valuable assets of enterprises. With the continuous development of business, Increase the amount of data is very impressive. Behind the massive amount of data, data management issues enterprises are facing mainly data integration, data security and other aspects. From the perspective of enterprise data management, explained migration of enterprise-class data problems, data migration methodology and the choice of common technical solutions.

Keywords: data migration; data security; methodology; information system

1 概述

1.1 关于数据迁移

在信息化建设过程中, 随着技术的发展, 原有的信息系统不断被功能更强大的新系统所取代。从两层结构到三层结构, 从 Client/Server 到 Browser/Server; 在开发环境向运行环境转换、低版本数据库向高版本数据库转换、两个不同数据库之间进行转换以至系统硬件升级时, 数据都可能需要被转移并使之正常运行, 必然要面临一个数据迁移的问题。

1.2 数据迁移的难题

当今数据迁移的主要难题是进行一次成功的数据迁移的时间消耗和资金消耗越来越大。然而应用在数据量方面的需求不断增加, 同时企业的应用趋向于全年不停顿运行、对系统的可靠性、可用性要求不断提高, 维护时间窗口的不断减少等因素, 使得进行一次平滑的成功数据迁移似乎越来越不可能。

1.3 数据迁移相关因素

进行数据迁移项目计划时, 一些因素是必须考虑^[1]的:

- * 数据的保护
- * 维护时间窗口
- * 迁移技术
- * 计划和应用停顿的容忍程度
- * 测试需求
- * 数据迁移的时间跨度
- * 整个环境的复杂性

2 数据迁移方法论

数据迁移是一种准确完整地将物理上存在的多或单点数据转移集中至单一数据中心或单一存储系统的技术方案。方案涵盖了应用、主机及存储资源的重新配置, 并对系统资源进行合理的规划, 以保证在迁移

^① 收稿时间:2010-05-04;收到修改稿时间:2010-06-02

集中完成后,新系统环境可以达到业务系统的要求,从而避免由于迁移集中导致的业务中断,及系统性能下降,同时确保数据迁移集中后系统的安全性及可管理性。数据迁移一般分为两个阶段:迁移评估设计阶段和迁移实施阶段。

2.1 迁移评估设计阶段

在这个阶段,通过分析工具,问卷调查和现场访问等手段,收集掌握应用、主机、存储环境等信息。在了解将来的商业和技术目标对数据的要求^[2],以及理解当前的数据架构的基础上,设计未来满足要求的数据架构和数据迁移计划。

由以下子阶段组成:

(1) 项目初始化阶段

在最初的项目会议中确认项目目标和预期。在这一阶段中,项目基于业务的要求开始并定义。主要的负责人员被确认并联系,项目范围和定义经过业务负责人及项目人员的评论和确定。基于已经定义的项目范围,项目人员确认并排出相应的时间表。

(2) 现场技术访问阶段

在这一阶段中,项目组成员进一步了解和明确业务对数据迁移的商业和技术要求。同时对已有的应用、主机、存储等环境做更多的了解和熟悉。

(3) 收集数据阶段

这一阶段的目的是快速收集业务系统相关的计算环境信息和应用架构^[3]。将收集到的信息整理归档以备分析阶段使用。项目中涉及的每个系统都要进行数据收集。这个阶段是以后各个阶段的基础。

(4) 分析数据阶段

所有收集的信息必须经过分析和研究,然后形成数据迁移初步技术方案。回顾和分析所有收集的信息,与技术人员求证实施计划的可行性。

(5) 制定数据迁移计划阶段

完整的详尽的计划是实施数据迁移成功的关键。在普遍适用数据迁移技术的基础上,针对异构和同构的环境,进行多种技术的结合。

(6) 总结回顾阶段

阐述数据迁移计划,并与相关负责人共同分析数据迁移计划。在满意的基础上提供完整的计划文档。

2.2 迁移实施阶段

这个阶段是在迁移评估设计阶段之后进行的,依据数据迁移计划,使用适当的工具,将业务数据从原

有系统迁移到新系统。这个阶段主要由以下几个子阶段组成:

(1) 项目初始化阶段

项目成员和业务负责人共同分析数据迁移计划,明确使用的迁移方法和技术,准备相应所需的工具和设备。

(2) 迁移前测试和演习阶段

根据设计的数据迁移方案,建立一个模拟的数据迁移环境,然后在数据模拟迁移环境中测试数据迁移的效果。数据模拟迁移前也应按备份策略备份模拟数据,以便数据迁移后能按恢复策略进行恢复测试。在这个阶段可以使用小规模的数据量测试检验数据迁移计划的可行和完整。主要包括:数据一致性测试、应用软件执行功能测试、性能测试、数据备份和恢复测试等。

在成功完成数据模拟迁移测试的基础上,在正式进行数据迁移之前,进行一次迁移演习。在演习中,完全按照数据迁移计划,迁移生产系统。实际检测迁移计划的可行性和效果,为真正的数据迁移做好最完备的准备。

(3) 数据迁移阶段

根据数据迁移计划进行生产系统的数据迁移。这个阶段包括所有进行迁移所需工作,比如:进行系统全备份,安装临时软硬件,生成镜像等。

(4) 迁移后测试阶段

测试检验迁移后系统的运行情况,进行数据验证,为最后的验收做准备。根据设计的数据迁移测试方案测试数据迁移结果,也就是检查数据迁移后数据和应用软件是否正常,主要包括:数据一致性测试、应用软件执行功能测试、性能测试、数据备份和恢复测试等。

(5) 总结回顾阶段

根据系统运行正常的判据,检查验收数据迁移后的应用系统。交付完整项目文档,结束项目。

3 数据迁移整体设计

为保证在迁移集中完成后系统的安全性,可管理性,以及避免由于迁移集中导致的系统性能下降,目标系统环境必需涵盖应用、主机及存储设备的重新配置,并对系统资源进行合理的规划。

3.1 数据迁移的目的

进行的数据迁移是进行系统升级扩容、数据集中等工作的重要步骤。迁移一般要达到以下目标:

平滑的将原有系统内的数据完整地迁移到新系统中。最小限度的影响现有系统的运行，在最小化的维护时间窗口中完成数据迁移。

3.2 目标系统总体设计和分析

在进行初步分析的基础上，针对硬件设施改变与否，未来的选择可以分为两大类：同构环境和异构环境。

针对与现有系统同构，至少可以有以下一些技术手段可以选择：

- * 基于磁盘阵列远程数据复制技术的数据迁移。
- * 基于主机操作系统逻辑卷镜像技术的数据迁移。
- * 基于数据库备份和恢复技术的数据迁移。
- * 基于三方工具的数据迁移。

针对异构计算环境，可以使用上述后三种方法之一，具体方法的选择还需进一步详细了解现有系统运行环境：

4 数据迁移技术的选择

4.1 可选的数据迁移技术

对于业务数据的迁移，目前主要采用如下四种方法：

- ① 基于磁盘阵列远程数据复制技术的数据迁移。
- ② 基于主机操作系统逻辑卷镜像技术的数据迁移。
- ③ 基于数据库备份和恢复技术的数据迁移。
- ④ 基于三方工具的数据迁移。

4.1.1 基于磁盘阵列远程数据复制技术的数据迁移

主要采用两套磁盘阵列之间的数据复制技术来实现数据的迁移，如目前 HDS 的 TureCopy 技术，可以实现在两套磁盘阵列之间的数据迁移，并且此种方法不占用主机资源，对应用透明。但是源端和目标磁盘阵列必须是同厂家的同系列的产品。

4.1.2 基于主机操作系统逻辑卷镜像技术的数据迁移

主要利用主机操作系统内置的逻辑卷管理系统的逻辑卷镜像(LV Mirror)技术。此种方法存在如下优点：步骤简单，容易实现，速度快；不需要考虑到上层数据应用系统的内部的结构；可以在线进行，只需要较短的停机时间。

4.1.3 基于数据库备份和恢复技术的数据迁移

主要通过数据库自带的备份和恢复功能以及逻辑日志追加的技术，实现一个数据逐步迁移的方法，最

后达到把数据从原有的磁盘阵列完全迁移到新的磁盘阵列的目的。本方法数据迁移不成功时，不影响生产系统的正常运行，对技术要求较高。

4.1.4 基于三方工具的数据迁移

利用一些第三方的工具实现数据迁移，例如 DSG，特点需要另外购买，实施也相对比较复杂。

4.1.5 数据迁移技术方案比较

方案	停机窗口	性能	资源	难度
主机逻辑卷镜像技术	不需要	LUN 级灵活控制拷贝速度	需要消耗较少的主机端资源	采用文件系统命令难度很小且易控制
Oracle 备份和恢复	两次	中等	需要消耗一定的主机端资源	取决于对数据库的熟悉程度
磁盘阵列复制技术	一次	较快，但不能灵活调节	消耗阵列的控制器和缓存资源；主机 IO 增加一定的时延	需仔细配置 LUN 对应数据完整性
第三方软件 (DSG)	不需要	速度慢	5-10% 的主机系统资源,IO 消耗较大	需仔细配置表及相关数据完整性

4.2 数据迁移方法的选择

数据迁移方法的选择要依据应用的现状选择相应的迁移技术。例如：是否可以停机做数据迁移、可以停机时间、是否需要在线做数据迁移。通过对多种数据迁移方法的分析和比较，选择相应的数据迁移方案。

4.3 总结

数据迁移是信息化系统建设的重要工作。尽管技术是成熟的，但就企业系统而言，保证业务的连续可用、系统性能的影响等是数据迁移成败的最重要因素。为了保证实施过程的万无一失，应遵循如下原则^[3]：

- ① 需求分析和概要设计阶段是项目成功的基础、
- ② 详细设计阶段是项目成功的保证、
- ③ 实施和测试是项目成功的具体体现、
- ④ 系统移交是项目成功的总结与记录。

参考文献

- 1 Barker R,等著.舒继武,等译.存储区域网络精华-深入理解 SAN.北京:电子工业出版社,2004.
- 2 Whitten JL 等著.肖刚,等译.系统分析与设计方法.北京:机械工业出版社,2007.
- 3 张友生.系统分析与设计.北京:清华大学出版社,2004.