

企业级信息系统数据迁移^①

吴淑玮, 闫训超, 曹 齐

(国网电力科学研究院, 南京 211106)

摘 要: 随着当今社会信息化程度的提高, 企业的关键业务对信息系统的依赖性越来越大, 而科技的进步以及企业 IT 环境随业务需求发生的改变, 企业常常面临着对现有信息系统相关数据的迁移. 而数据迁移会带来用户业务的中断, 时间的浪费意味着机会的丧失, 所以要求系统迁移必须要迅速、准确. 电力企业信息科技发展迅速, 信息系统更新换代频繁, 因此如何保证系统间数据迁移方便、快捷、安全、可靠执行, 是电力企业信息部门的重要研究课题. 以企业级数据管理的角度, 详细分析了数据迁移过程的各个环节, 并综合电力系统数据迁移经验, 提出电力生产系统数据迁移方案, 对企业级信息系统数据迁移的执行有一定的参考价值.

关键词: 数据迁移; 电力信息系统; 同构; 解决方案; 在线迁移

Enterprise Information System Data Migration Solution

WU Shu-Wei, YAN Xun-Chao, CAO Qi

(State Grid Electric Power Research Institute, Nanjing 211106, China)

Abstract: With the advancement of informationization in our society, the key business of enterprise rely on information system more and more. The progress of technology and change of IT environment of enterprise which is required by business, often force enterprises facing with the problem of existing information system data migration. Data migration brings interruption to the service of users, which is a waste of time that means losing opportunities, so it requires system data migration must be quick and accurate. Electric power enterprise information technology is developing rapidly, and the information systems update frequently, therefore how to ensure the system data migration convenient, fast, safe, reliable performance, is an important research topic of the information department in an electric power enterprise. In this paper, we made a detailed analysis of the data migration process in all aspects from the management of an enterprise data standpoint, integrated the experience of power system data migration, and put forward a solution of data migration of electric power production system, which has some reference value for the implementation of enterprise information system data migration.

Key words: data migration; electric information system; isomorphism; solution; Online migration

1 概述

根据国家电网公司“一强三优”现代公司建设战略和管理方式转变的推进, 以及电力信息化建设的发展, 原有信息系统逐步更新换代. 为保证在系统切换过程中的业务连续和基础支撑, 需要将作为原有信息化建设成果的业务数据迁移至新系统中.

数据迁移是指在系统软件开发中, 将具有实际业

务价值的业务数据, 依据功能需求或系统开发的要求, 在不同存储媒介、存储形式或计算机系统之间转移的过程. 数据迁移是系统开发经常涉及到的一项工作. 在企业级应用系统中, 新系统的开发, 新旧系统的升级换代, 以及正常的系统维护, 不可避免地涉及到大量的迁移工作. 而在一个以数据为核心的业务系统中, 数据的迁移更是无处不在.

① 收稿时间:2012-09-10;收到修改稿时间:2012-10-30

2 数据迁移类型和要素分析

衡量数据迁移执行成功与否,主要取决于四个方面:数据完整性、数据一致性、业务连续性、迁移安全性。要满足以上四个方面,则需要对迁移信息系统的环境和需求进行充分分析。

2.1 数据迁移类型

按照环境的不同,数据迁移可分为同构系统数据迁移和异构系统数据迁移。

数据迁移的同构系统主要指源系统和目标系统间操作系统、存储、数据库类型和版本、中间件类型和版本一致。同构系统数据迁移可能发生在系统硬件升级扩容、服务器更新维护、开发环境向运行环境转换或者数据库拆分合并等情况下。

产品平台更换、多种数据源集成、不同数据库之间进行转换等,可称为异构系统数据迁移,特点是底层支撑软硬件不一致,不支持存储复制、镜像技术,需要进行数据库数据转换清洗。

2.2 数据迁移要素分析

为确保迁移前后数据安全、完整、一致,满足迁移后业务流程连贯性,有必要在制定方案时分析数据迁移各阶段要素。

① 环境比对

环境比对主要对源系统和目标系统的操作系统类别、存储技术、数据库类型版本、中间件类型版本、数据字典结构等进行分析,环境的差异将决定数据迁移所应采取的技术手段和实施难度。

② 数据整理

在数据整理中,需要确认数据规模大小、收集业务数据、对垃圾数据进行清理、并考虑各种数据类型(如结构化、非结构化数据)的处理办法。

③ 迁移实施

执行迁移过程中,首先要对迁移数据进行测试,根据实际需求确定执行的方式(如停机迁移/在线迁移),同时对迁移过程数据文件进行加密以确保数据安全。

④ 结果检查

迁移执行后,对迁移过程中的迁移日志进行审查、对源数据库和目标数据库数据内容和数量进行对比,并在人工参与验证和校准后,对迁移中间文件进行处理封存。

3 数据迁移解决方案

3.1 数据迁移基础技术

基于 Oracle 数据库,目前有以下基础技术提供选择:

① Emp/Imp

导入/导出是 Oracle 幸存的最古老的两个命令行工具。Exp 将数据库内的各对象以二进制方式下载成 damp 文件,方便数据迁移。Imp 将 exp 下载的 damp 文件上载到数据库内。Oracle 数据库的导入/导出分为三种模式:表备份、用户备份和完全备份。在 Oracle 8i 后,新增了一种名为“表空间传输”的快速在数据库间移动数据的办法,是把一个数据库上的格式数据文件附加到另外一个数据库中,而不是把数据导出成 Dmp 文件,这在有些时候是非常管用的,因为传输表空间移动数据就象复制文件一样快。

② DBLink

物理上存放于网络的多个 Oracle 数据库,逻辑上可以看成是一个单一的大型数据库,用户可以通过网络对异地数据库中的数据进行存取,而服务器之间的协同处理对于工作站用户及应用程序而言是完全透明的,开发人员无需关心网络的链接细节、数据在网络节点中的具体分布情况和服务器间的协调工作过程。通过 DBLink 即可实现这种效果,即从一个数据库服务器访问另外一个数据库服务器的表或视图,实现基本等同于本机访问的效果。

③ SQL Plus Copy

在 SQL Plus 中的 copy 命令,可以完成远程数据库、本地数据库或 Oracle 数据库与非 Oracle 数据库之间的数据复制。其性能与导入/导出相同。通常, copy 命令用于 Oracle 数据库与非 Oracle 数据库之间复制数据。如果在 Oracle 数据库之间复制数据,应该使用 create table as 和 insert 的 SQL 命令。

④ SQL Loader

SQL Loader 是 Oracle 提供的一个实用程序,用以将外部文件(Txt、Excel、CSV)装入 Oracle 数据库。SQL Loader 可以筛选进行装入的数据,同时允许装入用于修改数据的 Oracle SQL 函数。除了变量、定长数据和流式数据以外,它还可以装入面向对象的数据、LOB 数据和集合。

⑤ 第三方工具

利用一些第三方的工具,或开发配套工具实现数据迁移,例如 DSG,特点代价高,实施也相对比较复杂。

3.2 数据迁移技术方案

对于业务数据的迁移,目前可归纳为如下四种方法:

- 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.

3.3 方案比对

方案	停机时间	性能	资源	难度
磁盘阵列复制	复制切换时间	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
逻辑卷镜像	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
数据备份恢复	导出导入时间	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
开发配套工具	不需要	中等,可切实满足需求	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 电力生产系统数据迁移方案

由于国家电网公司管理方式的转变,区域公司作为国家电网公司区域分部进行功能定位调整,区域公司原有生产系统数据需按属地化划转到区域内各省(区)公司.该迁移属于同构系统数据迁移,过程中采用配套迁移工具,以避免人为参与带来的失误.

4.1 数据迁移方式

生产系统迁移的数据分为静态基础数据和动态业务数据.对于数据移交工作的不同要求,可大致划分为以下两种数据移交模式:

① 分步迁移,先静后动

按照分步实施的原则,先将静态基础数据进行迁移,在供电公司系统中完成静态基础数据的导入后,相关业务应用切换到供电公司系统中开展,新产生的各类记录在新系统中完成录入操作;同时,在不影响生产业务正常开展的前提下,逐步开展动态历史数据的迁移工作,待全部历史数据迁移到供电公司系统中,工作

宣告结束.

② 整体迁移,同步开展

按照一次性整体迁移的原则,首先确定数据迁移工作开始及截止时间节点,在数据迁移操作正式启动前,各类新产生的业务数据仍在原区域公司系统中维护,待所有数据一次性全部迁移至供电公司系统后,正式切换至供电公司系统开展业务应用.

本次数据移交工作按照“静态基础数据优先移交、动态历史数据逐步迁移”的方式,统一安排各省公司数据移交时间和步骤,保证了整体迁移的进度和质量.

4.2 迁移环节分析

在数据迁移过程经历了以下阶段:制定方案、迁移准备、数据迁出、数据迁入、迁移完成.

① 制定方案

制定方案阶段,国网总部、区域公司和各省(区)公司进行沟通协调,对数据迁移整体要求提出指导意见,项目组根据相关要求编写方案,为数据迁移提供了理论依据.

② 迁移准备

迁移准备阶段的主要内容有:迁移工具开发、数据模型比对、区域公司源数据库清理、移交数据梳理划分、特殊数据预处理、迁移工具使用环境准备等.其中,迁移工具开发,在提供迁移工具成果的同时,将进行一系列的工具体测试工作;数据模型比对将需要对迁出的区域公司系统和迁入的各省公司系统的模型版本进行对比,确保迁移后数据一致;垃圾数据清理可避免冗余数据导入各省(区)公司数据库;移交数据梳理划分为后期迁移制定了数据迁入的范围;特殊数据预处理是指未结束的流程统一走完流程等需要在迁出前处理的操作;迁移工具使用环境准备工作中,需要有同时连接源数据库和目标数据库的网络环境.

迁移准备阶段为数据迁移的准确无误地进行提供了基础支持和保障.

③ 数据迁出

生产系统的数据迁移采用数据库整体迁出,分别迁入的方法进行迁移.这样,迁入到各省(区)公司只需要迁移工具的一次操作,避免多次迁移,缩短迁移时间和业务中断时间.

在数据迁出阶段,区域公司对经过准备阶段处理的数据库进行备份,加密后移交到各省(区)公司,在各省(区)公司搭建源数据库进行迁移,以创造源和目标

数据库相互可连接的条件.

④ 数据迁入

当各省(区)公司接收到加密文件后进行解密,部署到本地源数据库,同时根据迁移准备中的迁移划分范围对相关数据执行迁移.

在整个迁移过程中,为实现业务连贯性的目标,采用了在线迁移的技术,以减少中断各省(区)公司生产系统的时间.并在迁移完成后,刷新服务器缓存、重启系统.

⑤ 特殊处理

数据迁移的特殊处理主要指外键关联关系的处理,考虑迁移后关联的完整性,对相关数据整体迁移后进行调整.

⑥ 迁移完成

迁移操作完成后,将进行一系列的分析、对比、验证、校正及后续处理等工作.

通过分析迁移日志,有针对性的去排查迁移过程中出现的问题和错误;并将源数据库和目标数据库进行数据对比,验证数据迁移的完整性、一致性;再辅以人工验证业务流程的连续性等;最后对迁移分析验证过程中的问题进行相应的数据校正或还原.

在完整无误的数据迁移后,删除区域公司提供的数据库备份文件和相关数据库.原有区域公司生产系统,在一定时间内继续提供服务,方便各省(区)公司开展数据查询统计、数据核对、差异分析.之后,将只保留数据库和应用的备份文件,并进行锁定封存.

4.3 迁移工具处理流程

数据迁移工具主要处理环节说明如下:

① 明确数据范围

数据迁移离不开数据,且只迁移符合业务划分范围的数据.

② 制定迁移规则

迁移规则包括定义所需源数据库、目标数据库、迁移的数据表、过滤条件、数据间的映射关系、迁移方式选择、容错记录数、提交行数等信息.

③ 分析规则、预览数据

通过分析所制定的规则,提供迁移前数据和迁移后数据的预览,从而使在执行迁移前,对迁移数据范围和迁移效果有一个初步判断,减少迁移出错的发生.

④ 执行迁移

数据迁移分为两种情况:数据库之间直连迁移和

中间文件迁移.

当源数据库与目标数据库可同时访问时,可通过数据库直连迁移方式进行迁移.该方式的优点是操作安全、方便、快捷,同时可在线迁移,从而使系统停机时间大大缩短,保证了系统的持续运行.当源数据库与目标数据库网络不联通时,需经过中间文件进行迁移.考虑到数据的安全,需要对中间文件进行加密解密操作.

⑤ 迁移日志

迁移日志提供记录数据迁移过程的操作信息,还可监视迁移过程中的错误事件,为分析迁移正确性提供依据.迁移日志记录内容大致有:源数据库迁出记录数、数据迁移开始时间、数据迁移结束时间、目标数据库迁入记录数、迁移错误信息等.

⑥ 正确性验证

正确性验证通过三个方面来进行:迁移日志分析;数据库对比;业务数据人工验证.其中,日志分析和数据库对比由计算机自动分析计算.

数据迁移工具流程如图 1 所示:

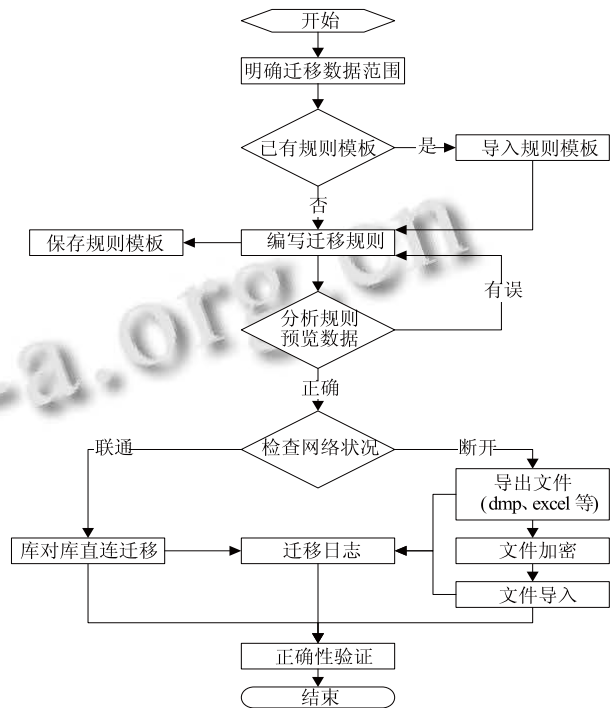


图 1 数据迁移工具流程

4.4 数据迁移风险预控

① 停机风险

数据迁移可能会带来比较长时间的系统不在线,这就意味着用户的业务会有比较长的中断时间.在本

次迁移过程中,采用在线迁移的技术,只在迁移结束后,对目标系统刷新缓存或重新启动应用.从而大大降低源系统和目标系统的中断时间.

② 数据风险

迁移的数据风险主要包括数据遗漏、数据不完整、数据不一致、数据泄密等等.

针对数据遗漏,根据迁移日志和数据库比对,验证数据迁移范围;针对数据不完整和不一致,主要通过严格分析数据迁移规则,对迁移前后数据进行预览,测试等;针对数据泄密,通过中间文件加密操作,从而保障传输的过程安全可靠.

③ 可靠性风险

数据迁移可靠性风险主要包括业务连续、出错恢复等.

数据迁移的业务连续风险主要指业务流程信息的连续性.一方面,通过技术手段保障数据的一致完整;另一方面,从业务上规定当前未处理完成的流程数据的迁移规则,如迁移后将根据省公司的流程定义重新启动新的业务流程.数据迁移的出错恢复,主要通过分析数据迁移日志,对迁移错误数据进行恢复重导或直接纠正等.

(上接第 35 页)

门公钥解密得到的数字摘要进行比较验证,如果一致,则认为审核的汇总表确实是安全的,没有遭受到攻击以及相互的欺骗.

由于该退税审核子系统是税务管理信息系统的一个子系统,整个子系统都建立在前期工程的 PKI 架构之上.因此公钥证书的管理相对容易.

7 总结语

随着计算机网络的发展,过去依赖于手工签名的各种业务在网上都可以使用电子数字签名代替,它是实现电子政务、电子商务、电子出版等系统安全的重要保证.在网络安全应用中发挥着越来越重要的作用.

参考文献

- 1 胡铮.网络与信息安全.北京:清华大学出版社,2006.
- 2 Tanenbaum A.S.计算机网络.潘爱民.第 4 版.北京:清华大学出版社,2004.

5 结语

数据迁移是在信息化发展过程中保障信息化建设成果的必要手段.在迁移过程中,为确保数据完整无误、业务流程连贯,应充分分析源系统和目标系统迁移的需求和差异,坚持工具化,使迁移过程有据可查,同时迁移后的用户验证也是不可或缺的关键步骤.

本文在论述了数据迁移的相关要素和技术方法,结合电力生产系统数据迁移方案实施经验,提出基于工具化自动迁移方案,对风险进行了详尽分析,以保障迁移环节顺利进行.相关数据迁移方法论在国家电网公司“三集五大”建设实践中得到了充分论证.

参考文献

- 1 李喆,万小健.企业级信息系统数据迁移方法.计算机系统应用,2011,20(1):182-184.
- 2 楼宏良,胡建,殷云飞.海量数据库迁移与升级,2012,28(6):152-154.
- 3 赵钦,周丹.政府办公自动化信息系统数据迁移解决方案,2008,(4).
- 3 贺雪晨.信息对抗与网络安全.北京:清华大学出版社,2006.
- 4 杨波.网络安全理论与应用.北京:电子工业出版社,2002.
- 5 曹建国,王丹,王威.基于 RSA 公钥密码安全性的研究.计算机科技与发展,2007,17(1):172-173.
- 6 叶生勤.公钥密码理论与技术的研究现状及发展趋势.计算机工程,2006,32(17):4-6.
- 7 高鹏飞.PKI 安全体系在电子公文系统中的应用研究.信息与电脑,2011,(9):84-85.
- 8 石志坚,谭全权,段海龙.RSA 算法实现数字签名的研究与应用.微型电脑应用,2008,24(6):50-51.
- 9 易红军,余名高.MD5 算法与数字签名.计算机与数字工程,(34):44.
- 10 陆宗跃.基于公钥基础设施技术 CA 认证中的密钥算法分析与认证.湖北第二师范学院学报,2011,28(8):51-53.