

基于 EAM 系统的核电厂数据导入方法^①

赵 堃, 孙为群, 张红兵

(核工业计算机应用研究所, 北京 100048)

摘 要: 数据作为企业信息化的重要组成部分, 历来在大型的企业级系统软件实施过程中占据关键的位置. 立足于核电厂 EAM 项目 AS 产品的实施背景, 通过数据源模型重构、模型间数据映射和视图导入程序开发等方法高效、准确、低成本的解决了企业级系统数据导入过程中存在的各种问题, 对于 ERP 等大型的企业软件数据导入工作也具备指导价值.

关键词: EAM; 基础数据; 主数据; 数据导入

Method for Nuclear Power Plant Data Loading Based on EAM

ZHAO Kun, SUN Wei-Qun, ZHANG Hong-Bing

(Computer Application Institute of Nuclear Industry, Beijing 100048, China)

Abstract: Data, as an important part of enterprise informatization, have always been in the key position in the process of a large enterprise system implementation. In the background of the nuclear power plant EAM project AS product of the implementation, through refactoring data source model, data mapping between models and views, import application development methods etc, this paper uses efficient, accurate and low-cost solution to solve all sorts of problems in the process of data import. This also has guidance value for large companies such as AS ERP software data import.

Key words: EAM; basic data; master data; data loading

1 本文背景及研究内容

1.1 中核核电 EAM 项目简介

EAM(Enterprise Asset Management)是面向资产密集型企业的信息化解决方案. 以提高资产可利用率、降低企业运行维护成本为目标, 通过信息化手段, 合理安排工作计划及相关资源与活动, 提高设备可利用率、优化安排企业各种资源, 从而提高企业的经济效益和企业的市场竞争力.

中核集团在 2012 年上线的 AS 系统是以企业资产及其维修管理为核心的商品化的 EAM 应用软件, 它主要包括: 工作管理、预防性维护管理、设计变更管理、行动跟踪管理、库存管理、采购管理、应付账管理、合同管理、运行管理、文档管理、辐射防护管理等基本功能模块, 以及决策分析、数据挖掘等可选模块.

1.2 中核核电 AS 系统数据导入面临的问题

AS 系统的主要作用是对核电维修信息、设备资产

信息和库存信息等进行有效整合, 而这些信息的载体和表达都要通过数据完成, 但数据准备和导入工作难度却最大.

首先, 基础数据涉及面广, 涵盖了核电厂中所有可见信息和不可见信息. 如物料基本信息, 设备基本数据, 会计科目, 工厂、仓库、车间等可见信息, 不可见信息如维修工单数据、排程计划数据等.

另外, 数据准备的工作量大, 以上各类信息数量从几个到几十万都有, 而每条记录包含的字段又可多达上百个, 通常造成项目延期的原因有 90%来自于数据整理.

再次, 数据的正确性是非常重要的, 系统的数据是许多程序正确运行的基础, 如果其中任何一个数据与实际不符, 将可能造成严重后果.

正是因为数据具有这些特征, 从而造成了收集准备工作量大、难组织, 数据导入要求准确率高, 一般需

^① 收稿时间:2013-06-13;收到修改稿时间:2013-07-19

要多个部门协调,投入的人力和时间都比较多,见效周期长等复杂性特点.

1.3 研究内容

目前的企业级软件数据导入方法有手工录入或使用软件自带导入工具导入.手工录入的方法,在动辄几十万上百万的数据面前,所花费的人力和时间成本将会很高,使用软件自带的导入工具有时候却不能满足和覆盖所有的业务数据导入范围.

本文将在核电 EAM 项目实施的基础上,探讨如何高效、准确、低成本的对核电的基础数据、主数据进行归类、收集、整理并通过分析和开发的手段有效的将数据导入到 AS 系统中的过程和方法,对目前的

大型的企业级软件实施如 ERP 等同样有很大的借鉴意义.

2 核电AS系统数据导入方法研究

数据导入问题实质上是要解决数据源和目标库之间的数据通道问题,目标系统的数据模型结构是固定的,因此确定和建立数据源的导入结构模型、建立导入模型与目标系统数据模型之间的数据映射完成数据通道成为整个工作的重点和关键,在此基础上再通过软件开发的技术手段驱动数据流动,就可以较好的完成系统数据的导入工作.基于此思路本文提出了以下系统数据的导入方案.

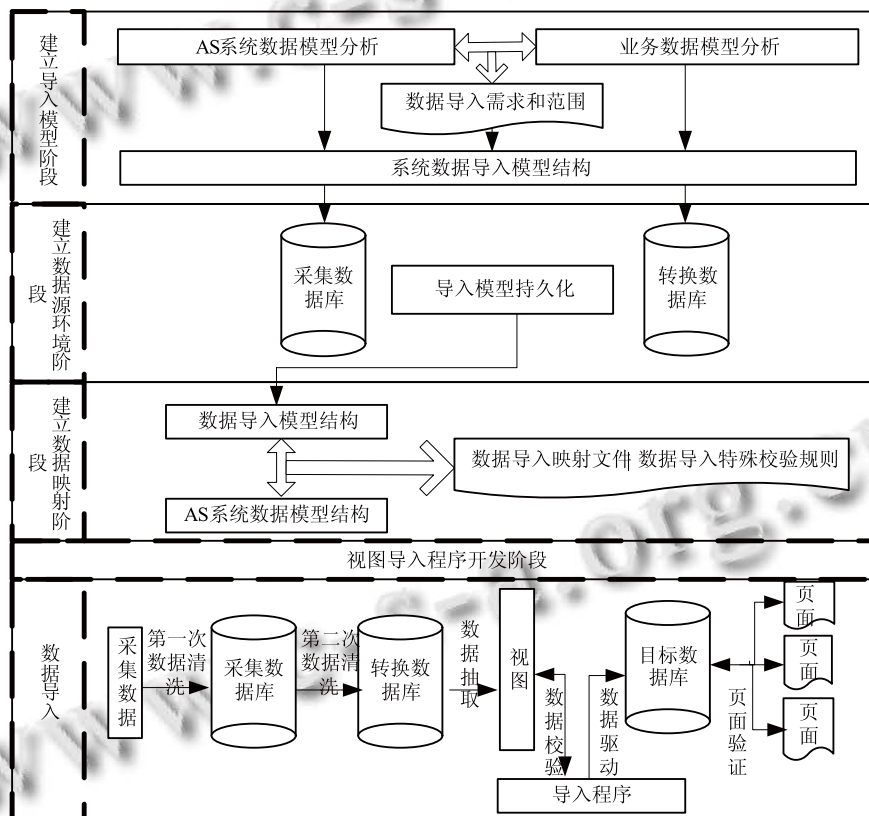


图 1 数据导入方案

2.1 建立系统数据源的导入模型

对于每一款企业级的系统软件来说,对业务对象信息的组织都有其内部的特殊的一套机制和模型,用户必须识别软件角度来讲支撑其数据模型所需要的必要信息,软件业务对象以及表结构的组织结构,这些信息一般的软件提供商会有详细的文件资料进行查阅,

但对于用户业务数据模型,从用户管理的角度出发,可能也存在支撑其业务流程的必要信息,因此通过系统的数据模型分析和用户业务的数据模型分析,用户才能最终确定数据导入的需求和范围,并根据两者的要求制定出合理的数据导入模型,供电厂的业务人员进行数据采集.

在制定数据导入模型过程中, 必须遵循以下几个原则和顺序:

① 按数据领域划分. 按照核电厂数据领域的划分, 用户可以将数据划分为文档数据、设备数据、物资/库存数据、辐射防护数据、工单数据等, 并为每一领域的数据进行编号, 如 EC 系列数据(设备数据)、DO 系列数据(文档数据).

② 按数据类型划分. 数据的类型一般可分为基础数据、主数据基本信息和主数据属性信息, 基础信息是指构成整个业务对象的基本元件信息, 如人员信息、库存位置、仓库信息等. 主数据是指系统间共享数据, 具备完备的生命周期管理的数据, 如设备数据、物资数据、文档数据等, 同时由于主数据信息庞大, 用户可以将某一领域内的主数据分为一个主数据基本信息和多个主数据参数信息来进行业务数据模型的组织. 如 EC10(设备基本信息)、EC20 设备参考文件信息、EC30 设备运行参数信息等.

③ 导入顺序划分原则. 对于通过以上原则划分好的数据模块来讲, 各个领域内的模块导入顺序和领域之间的导入顺序也是需要用户考虑的问题, 应该遵循基础信息在主数据信息之前、主数据基本信息在主数据参数信息之前, 静态主数据在动态主数据之前的原则.

2.2 建立数据源的数据库环境

通过数据导入模型的建立, 数据的导入需求、范围和结构都已经确定了, 这样用户可以通过在数据库中建立表结构的方法将数据模型固化下来, 作为数据源使用, 采集和转换数据库就是为了固化导入模型规范数据源而建立的, 两者在其表结构的组织上没有差异, 都属于源数据库的范畴, 但是在整个数据的导入过程中却承担着不同的功能和责任.

采集数据库由用户业务数据收集部门自行维护, 它的上游数据可能为第三方的系统业务数据、用户自行收集的电子版业务数据或纸质版的业务数据, 当所有的这些数据进入采集数据库前, 都应该经过用户手动的第一次数据清洗工作.

转换数据库由数据导入实施者或是专门的实施用户进行维护, 数据来源于采集数据库, 数据导入人员在此数据库上进行一些数据导入前的技术性第二次清理工作.

2.3 建立数据源导入模型与系统数据模型的映射

数据导入模型被固化到采集数据库和转换数据库后, 数据源的环境已经搭建好了, 此时用户需要梳理清楚数据导入模型与 AS 系统数据模型之间的映射关系, 同时进行记录和描述, 形成数据导入映射文件, 为下面的导入程序的开发做准备. 文件的编制应该包含所有的已确定的数据导入模型的内容, 并且也应该包括各模型数据之间的导入顺序. 映射文件内容如图 2 所示.

MF10 记录 ID / Record ID: 制造商 MANUFACTURER CODE
记录长度/Length of Record: 86 - 数据表 / Table(s): CVMFR

Asset Suite (MF10)					CNCC Source Data					
Key	Req	Target field	Type(Number of bytes)	Element (TIG.....)	Format	Source table/doc	Source field	Type(Number of bytes)	Value Provided	Comments
	Y	RECORD_ID	CHAR(004)	01185	CHAR				N	MF10 (alpha 'MF', numeric '10'), MF10 (字母'MF', 数字'10')
K	Y	MANUFACTURER_CODE	CHAR(24)	00126	CHAR	MFR_INFO	MANUFACTURER_CODE	CHAR(24)	Y	
	Y	MANUFACTURER_DESC	CHAR(50)	00405	CHAR	MFR_INFO	MANUFACTURER_DESC	CHAR(50)	Y	
	N	NPRDS_VENDOR_CODE	CHAR(8)	04456	CHAR				N	Default value = "N", 缺省值'N'

通过 MF10 导入数据后, 可在 AS 页面中查看数据, 下面是其中的一些页面 (仅供参考):

图 2 数据映射文件实例

上图左侧表示的 AS 系统数据模型数据库字段, 右侧表示数据导入模型数据库字段, 通过这样的文件组织形式, 可以清晰的将用户采集的业务数据和目标系统数据紧密的联系在一起, 实现了源和目标数据的可视、可查和可依据文件.

2.4 视图及导入程序的开发

经过数据导入模型建立及持久化、数据源与目标数据库映射活动后, 导入开发已具备条件, 数据导入程序的业务目标是提取用户完成收集、清理过程后的源数据, 按照数据映射规则进行数据转换, 转换后由导入程序导入到 AS 数据库中. 为此数据导入的开发阶段包含两个方面:

(1) 导入视图的建立. 用户所建立的源数据环境是综合了用户业务管理需求、数据采集便捷性和 AS 系统功能需求等多方面的因素形成, 在数据结构的组织上与目标数据结构会存在明显的差异, 为了消除这些差异, 用户可以目标数据库结构为出发点根据映射文件建立视图对源数据进行抽取, 形成满足条件的可导入数据.

(2) 导入程序的开发. 通过视图抽取形成满足目标数据库结构的源数据, 还不能满足导入的需求, 必须通过开发相应的导入程序来验证数据的有效性并驱动验证后的正确数据流入目标库. 其功能可简单描述如下:

源数据有效性验证功能: 包括字符类型校验和字符长度校验, 根据提取的源数据字段, 检查其类型和长度是否符合数据映射文件中的源数据信息定义.

数据映射功能: 包括灵活选择源数据库和目标数据、选择源数据表和目标数据表、选择源字段和目标字段、选择过滤条件、选择校验字段的校验类型等.

数据驱动功能: 引导和驱动导入程序验证后的正确数据流入目标数据库.

日志和错误处理功能: 对于数据导入过程中可能出现了数据连接错误、源数据校验错误、数据驱动错误进行详细记录.

2.5 AS 系统数据的导入

经过上述环节, 数据导入的条件已准备就绪, 数据导入的过程可分为以下几个步骤:

① 收集和清理源数据. 数据导入模型的建立就为用户的数据收集提供了规范化模板和数据标准, 用户应该完全按照导入模板的要求进行数据收集和清理工作, 确保数据进入源数据库环境前的数据质量.

② 冻结导入数据. 数据收集过程中经常存在业务数据不断调整的情况, 因此在采集数据进入源数据库环境之前, 必须冻结导入数据, 确保导入数据的准确和稳定性.

③ 抽取源数据为符合导入程序要求的数据格式. 数据导入程序对数据结构有要求, 源数据经过视图转

换才能满足这些要求. 转换过程中同时也应该按照映射文件对数据规则进行校验, 对于不符合的数据应生成错误报告, 尽快找出错误位置, 以便更正相关数据.

④ 导入数据到目标系统. 完成了数据的清洗和校验工作后, 就可以使用导入工具将数据导入到目标系统中.

⑤ 系统页面校验. 数据导入的结果是否成功和准确, 需要在数据导入之后进行页面流程验证才能确定, 用户可以根据导入的数据类型, 到系统的相关页面进行数据有效性检查, 如发现页面数据错误, 则需要重新导入.

3 结论

本文上述数据导入方案, 目前在中核集团核电 EAM 项目实施过程中正在使用, 得到了各方的认可, 经过项目验证此方案达到了高效、准确、低成本的数据导入目标, 同时方案中的数据源导入模型的建立和导入模型与系统模型间的映射方法, 对于大型的企业级系统数据导入活动也有极大的借鉴和指导意义.

参考文献

- 1 丁鑫, 张月祥, 王文清. 基于对象关系的通用数据导入算法. 计算机工程, 2008, (11).
- 2 张峰, 丁香乾, 胡瑞, 谢粤芳. 异构数据集成交换方案及其在 ASP 平台中的应用. 计算机工程与设计, 2006, (10).
- 3 吴优玲, 郑军红, 何利力. 元数据模型设计原理及具体应用. 工业控制计算机, 2012, (5).
- 4 容会, 于勇涛, 陈震霆, 王晓亮, 周绍景, 严敏. 元数据管理系统的研究与设计. 价值工程, 2012, 13.
- 5 景栋盛, 王芳. 电力企业数据模型建模方法研究. 计算机与现代化, 2011, (5).
- 6 高宁, 张宏军, 綦秀丽, 何健, 张睿. 面向实体的作战行动数据采集系统研究与设计. 指挥控制与仿真, 2013, (2).
- 7 黄刚, 王斌, 吴秀英. 元数据驱动的双向映射 ETL 模型研究. 计算机与数字工程, 2013, (3).