

# 清华大学通过Informatica方案 搭建学校基础数据与数据共享平台



Informatica ETL工具，帮助清华大学建立起规范的数据处理流程，实现了对各种结构化数据的自动化抽取、转换，节约了运维管理成本，为系统集成和基础数据库提供服务。

——清华大学数据中心主管刘乃嘉

## 学校及项目简介

“清华”是一种深厚的文化底蕴，是一种浓郁青春气息，是一种强烈的进取精神！成立于1911年的清华大学（Tsinghua University）是中国著名高等学府，坐落于北京西北郊风景秀丽的清华园，是中国高层次人才培养和科学技术研究的重要基地。目前，清华大学设有19个学院，55个系，已成为一所具有理学、工学、文学、艺术学、历史学、哲学、经济学、管理学、法学、教育学和医学等学科的综合型、研究型、开放式大学。

在校园信息化建设方面，清华大学也一直走在行业的最前端。10年前，清华大学信息化就进入到全面的实质性应用阶段——数字化校园建设阶段，他们通过数据整合和集成，在全校信息模型的基础上，建立跨部门的基础数据和数据共享平台，使学校各个应用系统具有合理的数据分布并满足各个应用系统之间的数据交换需求，更好地为全校师生提供全方位信息服务。在建设基础数据和数据共享平台时，清华大学采用Informatica ETL数据集成工具实现了不同系统中数据的统一采集、抽取和清洗，并将数据以Web服务方式发布，方便的实现了系统之间的数据对接。

## 面临挑战：数据整合 需要提升数据抽取、 转换和加载性能

数字化校园是以网络为基础，利用先进信息化手段和工具，实现从校园环境、资源到活动（教、学、管理、服务、办公等）的全部数字化，是在传统校园基础上构建一个网络空间，以拓展现实校园的时间和空间维度，达到提高教育管理水平和效率的目的。建设数字化校园，首要解决的是各应用系统中数据的交换和共享问题。

经过多年建设，清华大学校园信息化应用已基本覆盖学校运行的各个方面。例如，学生的管理信息化已经覆盖学生的各个环节，包括招生、迎新、教务、办公、网络教学、奖助贷勤补、收费、毕业、学生离校、就业等，其中迎新、离校更是跨多个部门。学校信息系统的建立是一个渐进的过程，往往缺乏顶层和统一的信息标准，致使信息流通不畅、数据不能共享，形成“信息孤岛”和“应用孤岛”，而高校的运行管理特征使这种现象更加严重。

由于时间长，系统多，数据共享日积月累，数据交换已呈网状结构，随着数据共享需求的进一步增加，清华大学实施了数据整合工作，建设基础数据库，实



### 概述：

清华大学采用Informatica ETL数据集成工具建立跨部门的基础数据和数据共享平台，实现了不同应用中数据的统一采集、抽取和清洗，并将这些高质量的数据发布给Web服务，满足为全校师生提供全方位信息服务的需求。

### 业务需求：

- 解决信息孤岛，满足各部门应用系统间数据共享的要求
- 加快数字化校园建设，提升服务品质

### 挑战：

- 原有工具不能满足发展所需，技术支持不足，需要提升数据Web服务性能
- 数据环境复杂，保证ETL工具与各业务系统顺利对接

### 产品与方案：

- Informatica PowerCenter

### 收益：

- 构建起一个有效的数据交换枢纽，构建校级基础数据库
- 自动化数据处理流程提升数据抽取效率，降低运维成本

现各应用系统之间数据交换和共享，简化校园管理流程，为教工和学生提供更人性化的服务。原来负责数据抽取、转换、清洗功能的ETL工具已不能满足发展所需，清华大学需要更加清晰地梳理数据处理流程，采用更稳定可靠的ETL工具，来提升数据抽取、转换和加载性能，为全校师生提供更满意周到的服务。

## 客户收益：实现数据共享，提升学校服务品质和效率

目前，清华大学已成功部署了基础数据与数据共享平台，解决信息孤岛，满足了各部门应用系统间数据共享的要求，实现了学校/院系门户信息、户口办理、就业信息发布、宿舍管理、档案管理和校园一卡通等人事、后勤、财务部门应用系统的数据集成，打破原有各系统间界限和传统服务方式，为全校师生提供更加便捷、安全、个性的工作学习环境。

PowerCenter数据集成方案，帮助清华大学实现了数据的统一集中管理，建立了规范的数据处理流程，从复杂的数据源结构中，实现了对各种结构化数据的自动化清理、抽取、转换，节约了运维管理成本，保障了数据质量，为系统集成和各个应用系统之间的数据共享提供服务。

- 构建起一个有效的数据交换枢纽和完整数据统一集中管理环境。在数据流动的过程中，大大减轻了各供数系统的压力，确保业务系统的安全。
- 统一、适时的数据访问能够在正确时间提供可信的信息，满足全校师生、管理者、校友的各种需求。

- PowerCenter可将数据路径追溯到最初来源或其最终目标，随时查阅相关的详细的数据和文档。

## 解决方案：Informatica ETL工具易用、提效、高度可扩展

在选择数据集成方案时，清华大学综合比较、测试了多家ETL方面的通用性平台。Informatica一直专注于数据集成业务，专业化程度强，其数据集成产品PowerCenter是ETL领域最为成熟、且市场占有率极高的产品，允许用户快速、方便地访问、集成和传输数据，并且在易用性、数据处理效率、扩展性等方面有突出表现，是业界最突出的数据集成平台。

在具体实施过程中，清华大学对产品兼容性问题非常关心。在构建系统时需要满足不同时期建设的不同业务应用系统间的数据整合、平滑互联和信息共享，数据环境极为复杂，刘乃嘉非常担心ETL工具与正在上线运行的各业务系统的对接不顺畅。PowerCenter和业务系统的接口的耦合程度低，独立性强，允许现有系统无需改动或者能够将改动影响降到最低程度，即使面对最为复杂的数据环境也能及时、准确、经济有效地完成数据共享任务，完全保证了业务连续性。为了增加客户的信心，Informatica在项目建设过程中理论培训与现场实施并重，保证为客户解决出现的一切疑问与技术故障。

清华大学通过PowerCenter对各种源数据的整合，把各应用系统数据采集到一个临时缓冲区中，然后通过产品的数据清洗和标准化功能，把采集到得数据进行标准化处理，然后将高质量统一的数据发布为Web服务，供各业务系统调

用。作为ETL工具，PowerCenter为清华大学提供了统一的手段与方式进行内部不同类型数据的直接采集，降低了数据抽取的复杂性，具体功能包括：

- PowerCenter统一的开发调度监控界面提供了作业的可读可管理性，借助PowerCenter强大的工作流设计，很好的实现了所有作业的执行计划个性化编排；
- 通过PowerCenter的图形化界面和高复用的开发特性，使得开发与管理工作变得异常简单清晰，开发人员能够更快速、顺畅的操作，执行数据集成任务，分析数据关系，提高开发效率；
- Informatica PowerCenter的高性能的吞吐量和对复杂业务的强大计算优势，能够把标准化的数据进行复杂的计算后准确的展现给Web服务，然后应用于各个系统中。

清华大学已完成了基础数据与数据共享平台一期建设任务，各种数据信息已在全校园中“活”起来，便利师生，提高效率。未来，随着学校网络应用和用户规模的进一步扩大，清华大学还计划集成更多的应用，Informatica方案高度可扩展性能够根据数据源的调整，修改数据集成策略和处理任务，并且不需修改已经设计完的作业，节省硬件成本，并降低与数据停机有关的成本和风险。Informatica方案可谓是一个可持续发展的方案，协助百年清华数字化校园建设不断的深入。

  
**informatica**  
Put potential to work.™

中国客户支持热线：400 810 0900  
中国网站：[www.informatica.com.cn](http://www.informatica.com.cn)

### 全球总部

地址：美国加利福尼亚州红木城 Seaport 大街 2100 号  
邮编：94063 电话：650.385.5000 传真：650.385.5500

### 中国总部

地址：北京市朝阳区建国门外大街乙十二号双子座大厦西塔 EF 层 06 室  
邮编：100022 电话：+86-10-5675-2000 传真：+86-10-5675-2030