

Informatica 技术保障江苏电信规范化 ETL 开发

——江苏电信 EDA 企业数据架构建设



解决方案：

- Informatica PowerCenter

收益：

- 实现了数据的统一集中管理，提供数据共享服务
- 保障数据质量，支撑日常数据运营，实现精细化管理，提升全业务精确化营销能力
- 节约运维成本，提升数据管理效率

Informatica 产品帮助江苏电信构建起一个统一集中管理 ETL 平台，管理人员不需再维护 EDA 中多个系统的不同 ETL 平台，大大节约了运维管理成本，同时提升了数据管理的效率。高质量的数据信息很好地支撑了企业日常数据运营，实现精细化管理，提升全业务精确化营销能力。

——中国电信股份有限公司江苏分公司企业信息化部 数据仓库项目经理

中国电信股份有限公司江苏分公司

中国电信股份有限公司江苏分公司（简称江苏电信）成立于 2000 年 7 月，公司下辖 13 个地市级分公司、6 家省公司直属单位、3 个专业子公司以及 56 个县（市）电信局。江苏电信主要经营固定电话、移动通信、卫星通信、互联网接入及应用等综合信息服务。公司现有员工 2 万多人，收入规模居中国电信集团第二位。

业务挑战：企业 EDA 建设 ETL 发挥承前启后作用

为适应电信行业迅速变化市场竞争环境，电信企业往往建立了许多面向不同对象的数据业务系统来支撑决策，但这些系统相对独立，数据分散、缺乏一致性，反而制约了经营决策分析的应用支撑。为解决上述问题，江苏电信希望整合全业务数据，部署一种更高效的企业数据架构（EDA，Enterprise Data Architects），通过持续完善和实施 EDA 数据架构，融合企业前后端 90% 以上生产数据，支撑决策层领导、专业管理人员、营销策划和执行人员、数据管理人员以及各个生产系统等人员角色，围绕集团公司“发挥数据价值、提升企业智慧”的战略目标，提升数据应用水平。



江苏电信 EDA 架构包括数据仓库 ODS、客户行为分析、协同支撑系统、经营风险管控系统、客户维系系统以及本地数据集市几大部分。在 EDA 体系中，ETL（数据抽取、转换、装载）功能在各相关需要数据抽取的系统均存在，这使得运维人员维护管理困难，维护费用高昂。且从产品统一化，利于扩展的角度来看，将分散在各系统中的 ETL 功能进行整合，成为独立系统，可以服务于其它需要数据整合的系统，例如 ODS 和 EDW 可以使用同一个 ETL 整合系统。因此要求 ETL 整合平台在 ETL 工具的辅助下，需要达到系统可配置、可管控、可二次开发、自动化流程等要求。

挑战：

- 业务系统和应用系统独立，数据分散、缺乏一致性
- 各系统拥有自己的 ETL 平台，运维人员维护困难，维护费用高昂
- 缺少统一的数据整合平台，异构、低质量的数据无法支撑精确化管理、精确化营销

在仔细分析 EDA 项目建设需求后，江苏电信认为他们选择的 ETL 方案必须满足下面几点要求：

- ETL 处理过程必须包括有统一调度、监控和管理的功能；
- ETL 处理过程可提供图形化的操作界面，具有良好的易用性；
- 能够支持各种平台的异构数据源系统的数据；
- 数据抽取过程支持增量抽取、完全抽取等抽取策略，对于数据源系统支持异步抽取或同步抽取；
- 数据转换模块可进行二次开发，并通过对插件及外部程序的支持来处理复杂的异常情况，提供调试、跟踪功能。

解决之道：规范化的 ETL 开发管理是 EDA 成功的技术保障

江苏电信在选择 ETL 工具时，始终将目光锁定在 Informatica 的 PowerCenter 产品上，不仅因为方案完全满足上述要求，还因为 Informatica PowerCenter 在稳定性、架构设计、数据处理效率、扩展性等方面都经过了江苏电信长期的使用检验。据江苏电信企业信息化部郑勇先生介绍，早在 2005 年，江苏电信建立经营分析系统之时，就选择了这款产品作为 Oracle 数据仓库的 ETL 工具。随着业务扩展、系统增加，2009 年以 IBM DB2 搭建数据仓库时，依然沿续 Informatica PowerCenter 为 ETL 工具之路。

目前，ETL 面临的挑战是当接收数据时各源数据的异构性、低质量。江苏电信 EDA 的数据仓库共有 64 节点，源自 28 个源生产系统，如果 CRM 系统按照统计代码管理数据，计费系统按照账目数字管理，客服系统按照语音 ID 管理，当 ETL 需要对这三个系统进行集成以获得对客户的统一视图时，这一过程需要复杂的匹配规则、名称正常化与标准化。Informatica PowerCenter 可以提供对广泛的应用和数据源的支持，包括对 BSS 域中 CRM、计费、结算、渠道、客服等系统，OSS 域中资源管理、服务开通、综调、网元系统，以及财务、网管系统的支持，并且适用于电信的结构化和非结构化数据的处理和管理，把这些不同来源、格式、特点性质的数据有机集成起来，为各系统提供全面的数据共享。

在江苏电信 EDA 架构中，数据仓库为数据处理核心，数据分布和流转过程首先从各个联机事务处理环境、业务和应用源系统中提取出数据，由 Informatica PowerCenter 进行清理以提升数据质量，然后经过抽取、转换和装载，即 ETL 过程，导入到 ODS 仓库中，从而得到企业数据的一个全局视图，之后根据不同需求下发到各个不同的数据集中。Informatica 产品能够与存储过程相结合，与 ODS 实现数据同步，数据仓库负责运算，内部复杂的业务逻辑通过存储过程实现。这种方式大大优化了 ETL 执行过程，使存储过程的运行效率最高可达每秒 5900 万条记录，平均效率可达每秒 36 万条记录，数据汇总能够及时、准确完成，保证江苏电信能够每月 6 日完成月报数据提供，每日 8 点半之前完成日报数据提供。

行业独特性使其系统平台的二次开发不可避免，Informatica PowerCenter 提供了一个完全图形化的开发模式，江苏电信运维人员不需要进行复杂编码即可进行二次开发，而且对 Informatica 方案多年的使用经验，让他们对产品的开发机制非常了解，能够进行非常规范化的 ETL 开发管理。通过开发调试，ETL 中每步都具有详尽的日志信息，作为错误分析和代码优化的依据，并且，任何 ETL 过程都可以接收本地网参数、时间参数，可以进行小范围重新加载，为调度带来更大的灵活性。此外，Informatica PowerCenter 能够为所有数据提供端到端的血缘，代码中嵌入技术元数据，可以实现自动血缘分析，深化对复杂数据关系的认识。

客户收益：支撑日常数据运营，提升数据管理效率

目前，江苏电信已成功部署了高效的企业数据架构 EDA，运行状态良好，满足了企业数据共享的要求，支撑了精确化管理、精确化营销以及本地网开展数据应用，实现了企业的“发挥数据价值，提升企业智慧”核心目标。而由 Informatica PowerCenter 搭建的 ETL 平台，帮助江苏电信建立了自动化的数据处理流程，使数据整合能力得以加强，节约了运维管理成本，同时提升了数据管理的效率。

在实际应用中，江苏电信实现了数据的统一集中管理。Informatica 产品在应用与各子系统之间构建了高效稳定的数据整合层，实现了跨系统的数据采集，降低了数据抽取的复杂性，实现了完整的全公司数据统一集中管理环境，达到“数据纵向贯通、横向关联”的效果。目前，已完成 28 个源系统的数据整合，总数据量达到 120T。同时，实现了全面的数据共享，数据可下发至 13 个本地数据应用集市，为 CRM 系统、10000 号系统、网厅系统、综合调度系统、外呼系统、积分管理系统等多个生产系统提供数据共享服务。

一个稳定、统一、灵活的 ETL 调度平台让日常数据运营工作更加高效，不仅大大提升了工作效率，降低了管理成本。而且通过 Informatica PowerCenter 无编码、图形化的开发管理界面，还大大方便了运维人员的二次开发和产品维护，通过更具行业特点的插件开发进行跟踪、捕获，及时发现异常情况。

关于 Informatica

全球领先的独立企业数据集成软件提供商 Informatica（纳斯达克代码：INFA）让世界各地的企业获得最大化的数据投资回报，从而赢得竞争优势。全球众多知名企业依赖 Informatica 使用及管理其在本地的、云中的和社交网络上的信息资产。

中国客户支持热线：400 810 0900

中国网站：www.informatica.com.cn



INFORMATICA®

北京办事处

地址：北京市朝阳区建国门外大街乙 12 号 LG 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

上海办事处

地址：上海市浦东世纪大道 201 号渣打银行大厦 5 楼
邮编：200120 电话：86-21-6182 6825 传真：86-21-6182 6755