

电信数据仓库项目实施方法论探讨 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/66/2021_2022__E7_94_B5_E4_BF_A1_E6_95_B0_E6_c41_66269.htm

今天的商业环境中，电信行业所面临的经营环境是一个多变复杂的环境，通过有效管理，向管理要效益已经成为企业的选择。为了实现这一点，数据仓库已广泛应用于电信行业。许多数据仓库项目的实施不是很成功或者另客户不满意，原因很多，本文主要从数据仓库项目实施方法论方面进行探讨，希望对读者有所启发。项目规划电信数据仓库的数据来源于现有的各业务系统及外部数据源。由于电信业务发展迅速，因而存在业务系统不断地升级或改造的现象，加上外部数据来源不断拓宽，这都需要数据仓库作出相应的调整，以保证数据的完整性和一致性。建设数据仓库的根本目的是为管理层、决策层提供决策支持，并为操作层提供活动指引，随着时间的推移，越来越多的数据充实到数据仓库中，使得数据仓库蕴藏的信息越来越大，可分析的主题会越来越多。因此上数据仓库的建设不是一个产品，而是一个持续的过程。（见图1）长期规划需要分步实施任何IT系统如果缺乏缺乏长期的规划都将缺乏生命力。对于数据仓库也不例外，要近可能想的长远一些，为增量式建设提供好的技术架构、优化的数据模型。数据仓库建设根据生产系统的状况、数据质量状况和业务需求划分成不同阶段，每阶段都应努力达成阶段性目标。自下而上的建设思路从理论上讲，应该是先有数据仓库的概念，然后才有数据集市。按照这种模式建设数据仓库就是自上而下的建设思路。很多关于“建立企业数据模型，不考虑应用，先建

立数据平台”的说法，对于一个规模不大企业而言应该是一种好的选择。而对于象中国电信这样拥有很多业务支撑系统，业务系统模型差异很大的企业而言，采用这种模型建设花费的时间过长，加上业务发展迅速、市场变化大的情形，可以说不是一种理想的模式。自下而上的建设思路是先从数据集市入手，就某一个特定的主题先做数据集市，再建设数据仓库。这种模式能够比较快地让使用者见到效果，应该是一种比较理想的建设思路。短期见效的应用驱动面对激烈的市场竞争，在瞬息万变的商业环境中，企业的反应速度往往决定着企业的成败，IT系统需要为企业经营活动给予有力的支撑，并创造价值。如果数据仓库不能尽快让业务部门、管理者和公司领导近早看到结果，领导会失去耐心，对项目也会缺乏持续投入的热情，因此数据仓库项目需要尽快见效果。结构选择电信数据仓库设计通常有两种结构，即带有ODS的数据仓库应用系统体系结构和不带ODS的数据仓库应用系统体系结构。ODS（Operational Data Store）是数据仓库体系结构中的一个可选部分，ODS具备数据仓库的部分特征和OLTP系统的部分特征，它是面向主题的、集成的、当前或接近当前的、不断变化的数据。业务系统与数据仓库的隔离层一般的数据仓库应用系统都具有非常复杂的数据来源，这些数据存放在不同的地理位置、不同的数据库、不同的应用之中，从这些业务系统对数据进行抽取并不是一件容易的事。因此，ODS用于存放从业务系统直接抽取出来的数据，这些数据从数据结构、数据之间的逻辑关系上都与业务系统基本保持一致，因此在抽取过程中极大降低了数据转化的复杂性，而主要关注数据抽取的接口、数据量的大小、抽取方式等问题

。转移部分业务系统细节查询功能在数据仓库建立之前，大量的报表、分析是由业务系统直接支持的，在一些比较复杂的报表生成过程中，对业务系统的运行产生相当大的压力。ODS的数据从粒度、组织方式等各个方面都保持了与业务系统的一致，那么原来由业务系统产生的报表、细节数据的查询自然能够从ODS中进行，从而降低业务系统的查询压力。完成数据仓库中未完成的功能一般来说，在带有ODS的数据仓库体系结构中，DW层所存储的数据都是进行汇总过的数据，并不存储每笔交易产生的细节数据，但是在某些特殊的应用中，可能需要对交易细节数据进行查询，这时就需要把细节数据查询的功能转移到ODS来完成，而且ODS的数据模型按照面向主题的方式进行存储，可以方便地支持多维分析等查询功能。在一个没有ODS层的数据仓库应用系统体系结构中，数据仓库中存储的数据粒度是根据需要而确定的，但一般来说，最为细节的业务数据也是需要保留的，实际上也就相当于ODS，但与ODS所不同的是，这时的细节数据不是“当前、不断变化的数据”，而是“历史的，不再变化的”数据。目前电信内部有很多的子系统，如：营业前台、计费帐务、综合结算、1000号系统、联机采集等等。他们之间的信息的重叠和不一致显然存在，如客户资料、客户所使用的电信的业务信息、帐户信息等。如何综合应用这些数据，首要任务就是进行数据整合，ODS的特点恰恰可以担当起这付重任，而且ODS的诞生，也正是因为有了这种需求的存在。数据模型设计在数据仓库数据模型设计中有ER模型和维模型。实践表明，ER模型和维模型在数据仓库设计中并非绝对对立。对于部门数据集市，当数据量不大、报表较固定时可

以采用星型模式；对于企业级数据仓库（也就是存放细粒度数据层）考虑到系统的可扩展能力、投资成本和易于管理等多种因素，最好采用第三范式。通常在数据仓库中维表的数量比较多，为了减少I/O，有必要减少维表的数量。我个人的意见是对事实表和集市做规范化处理，建议不要对所有维表都做规范化处理，也就是更多地采用星型模型，而非雪花模型。分析主题的选择追求企业价值最大化是企业的目标。数据仓库建设的目标应该是增强内部管控，提升客户体验。也就是说帮助企业经营者了解自己的客户有哪些，自己的产品有哪些，收入是由哪些部分构成的，服务提供的成本有多少，企业在市场的份额如何，企业有哪些资源，企业的服务质量如何，有哪些改进机会等等。由此可以得出电信数据仓库的主要分析主题应包括：客户情况分析、业务发展分析、收益情况分析、市场竞争分析、服务质量分析、营销管理分析、大客户分析、新业务与数据业务分析和合作服务方分析等。分析主题确定后，就需要考虑从哪些角度、分析哪些指标了。例如：对于客户情况的分析，可以从地域、客户类型、客户在网时间等多角度分析客户的构成以及客户使用产品的情况。还可以通过建立分析模型对客户作更深入的分析。例如通过客户的社会属性、客户消费属性进行客户细分，也可以从客户使用产品的情况、客户的通话量等视角建立客户忠诚度分析模型来评估客户的忠诚度。采用客户细分的手段则是通过聚类算法，将客户分成不同的客户群，识别这些不同客户群的消费习惯和消费特征，有针对性地作新业务推介，实现针对性营销，达到销售成本最小化的目的。通过对客户价值的分析，将客户通过价值实现分级，实现大客户个性化

服务、商业客户专业化服务、公众客户标准化服务。通过建立客户分析模型可以实现客户赢回。例如：客户流失预测，通过离网调研和数据挖掘，捕捉客户离网前的特征，预测客户流失的概率。这对于优质客户的保护是十分重要和有效的。对于客户流失预测，从两个方面来看：一个是客户流失预警，一个是流失客户特征分析。客户流失预警，定义了统一的预警模型，例如客户使用异网IP业务、客户话务量突然减少、客户使用呼叫转移等可以做为模型的分析因子。根据预警模型，客户话单数据中自动匹配预警数据，预警模型可以按如下规则定义，并且可以灵活扩展。流失客户特征分析，通过决策数算法，分析流失客户特征，然后通过这些特征得到当前在网客户中匹配流失概率高的客户数据。通过客户流失分析获得流失客户数据和潜在流失客户数据，从而将这些数据分配给客户服务部门，整合销售服务资源，根据客户的需求，设计个性化的营销策略，快速反应，挽留流失概率高的客户。100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com