3.10 小结

传统的数据仓库架构直接移植到电子商务行业会出现“水土不服”的现象，原 因在于电子商务行业的数据源多而杂，需要处理和分析的数据量巨大，对数据分析 的实时性要求较高，并且需要支持的用户量巨大。要解决这些问题，需要对数据仓

库的整体架构做出调整和优化。传统 BI 的数据仓库主要使用关系型数据库来支 持离线分析和复杂查询，难以满足大数据处理和实时查询等需求，因此可以结合使 用关系型数据库、内存数据库和分布式数据库以满足电子商务行业对于商业智能 在数据处理、数据存储等方面的要求。

数据仓库按照数据流向可以分为三层结构，包括数据层、信息层和分析层。

数据层主要是数据源和标准数据接口。信息层包括 ODS、数据仓库、数据集 市、ETL、元数据。分析层主要是数据的应用，如数据挖掘、报表分析、数据支持等。

在数据仓库架构上，笔者建议在传统数据仓库的基础上，同时采用分布式数据 仓库来支持BI 大数据的存储与计算，可以显著地提升数据处理和数据搜索的性 能。在 Hadoop 分布式系统架构中，数据仓库工具 Hive 和基于内存的数据库 HBase 可以分别解决大数据批量任务和实时性查询需求。

正确和完备的数据模型，是BI 进行一切数据分析的基础，是决定数据仓库项 目成功与否的重要因素。目前常用的数据仓库数据模型为多维数据模型，这种模 型主要以星形模式和雪花形模式在关系数据库系统中存在，两种模式各有优缺点， 但在数据仓库中，通常不推荐“雪花化”,查询性能相对OLAP 系统来说更加被重 视，而雪花模式会降低数据仓库系统的性能。

ETL 负责将分布的、异构数据源中的数据如关系数据、平面数据文件等抽取

第三章 商业智能在电子商务的解决方 案

到临时中间层后进行清洗、转换、集成，最后加载到数据仓库或数据集市中，成为联 机分析处理、数据挖掘的基础。

存储在各种数据源中的数据通过 ETL 存储到数据仓库中，对数据仓库中的数 据进行多维建模，再利用多维分析技术满足决策支持或者满足在多维环境下特定 的查询和报表需求。多维分析技术的核心是“维”的概念，维指的是分析问题和观 察事物的角度。

在数据仓库或多维分析的基础上，对数据进行更深层次的分析和知识提取，需 要应用到数据挖掘技术。数据挖掘的价值体现于发现深度隐藏的规律或模式，数 据挖掘在电子商务行业有广阔的运用空间，如精细化营销、市场评估与预测、客户 分析(客户分群、流失预警、忠诚度分析等)、个性化推荐、网站异常预警与监控等。

BI 在电子商务的应用包括对内的运营决策支持和对外的网站用户支持。电 子商务对内决策分析系统为分析层、决策层提供决策分析平台，为操作层以及各业 务系统提供执行优化的智能帮助，如企业级 KPI 决策系统、部门级绩效考核系统、 业务监控和分析以及分析型工作流程。BI 对于网站用户的支持主要体现在个性 化服务、异常用户挖掘、网站效果分析与优化和提供数据分析增值与服务。

选择合适的软硬件产品是支持企业实施 BI的关键问题。BI 产品种类繁多，每 家厂商几乎都会宣称自己的产品是如何的好，在构建企业自身的 BI 系统时应该根 据企业自身的需求考虑软件硬件产品的选择。数据库方面 DB2、Oracle 都是很好 的产品，Teradata 和 EMC 公司的 Greenplum 在大数据处理方面有着很强的能力。 根据 Gantner 在2012年2月发布的数据仓库魔力象限显示，Teradata 在数据仓库 方面处于领导者地位。主流的 OLAP 工具有 BO 、Cognos 以及 Qilikview 等， Gartner 前端报表工具最具魅力象限连续几年纳入了Qlikview 这个产品，这个产品 代表了基于内存的报表分析技术，美观简单、用户体验极佳。数据挖掘产品领域， 有 SAS 、SPSS等两大厂家，而像 IBM 、Teradata 也都有自己的挖掘工具。做好数 据挖掘仅仅有工具是不够的，更需要数据挖掘人员对数据挖掘的理解和认识。

BI的实施策略，有代表性的是数据仓库支持 OLAP 理念的两位创始人 Bill Inmon 和 Ralph Kimball,Bill Inmon倡导自上而下实施策略，而 Ralph Kimball 则 提出不同看法，认为数据仓库应该实施自下而上的实施策略。笔者认为可以采用 第三种策略，那就是整体规划、分步实施，这种方式可以融合两种策略的优点，又可 以避免其中的短处。

BI 实施风险包含技术、业务、实施、推广等方面，但BI 最大风险还是来自于很 多人认为 BI 可有可无，有则应该更好，无也没关系的这样一个定位。笔者总结了 BI 实施风险主要有组织风险、技术风险、业务风险、实施风险。