## 上海证券交易所案例

### 项目背景

作为上海证券交易所(以下简称“上证所”)波士顿咨询规划的新一代项目之一—“新一代信息系统”，也是上证所信息管理的核心项目数据仓库项目（以下简称“该项目”）在交易所领导的关心和支持下，以及全体项目组成员的努力下，经过所领导精心部署和有计划有步骤的安排，建成了目前国内金融行业数据容量最大、数据内容最完整的数据仓库，同时也是亚太区最大的证券业数据仓库项目。

上证所的企业级数据仓库建成了一个涵盖从数据清洗、数据存储、数据整合、基本查询、统计应用、信息服务到高端的分析、挖掘和预测的企业智能信息平台，具体成功完成目标情况如下：

* 体系架构：目前数据仓库建立了双中心灾备系统，双系统可以互为主/备切换对外提供服务，实现了同城异地双中心运营管理；数据仓库系统包含了数据加载及输出、数据仓库存储、数据仓库门户(Portal)、元数据管理(MDS)、数据仓库管理(DWM)、非结构化数据管理等子系统。
* 数据存储；整个数据仓库的数据涵盖了1990年上证所建所以来的全部历史数据及最新增量数据，范围包括交易核心数据和交易相关数据，类型则包含结构化数据和非结构化数据，密级分类刚包含机密数据、秘密数据、开公数据。
* 信息服务：数据仓库系统目前为上证所仅次于核心交易系统的运营管理系统，面向所内、监管部门、证券市场提供各种数据服务，如：上证所月报、证监会数据报送、TopView数据加工、公开信息披露等；同时数据仓库门户系统包含了查询、报表、OLAP等应用达600多个，部署了九个应用专区，用户数超过250人，遍及12个所内部门。
* 高端分析：数据仓库系统作为所内SAS分析的基础数据库提供数据挖掘分析的信息平台，同时支持市场监管、新产品研发的核心数据统计分析，为规范市场和产品创新提供了有力的决策支持。
* 系统管理：建成了数据仓库管理系统、元数据管理系统、非结构化数据管理系统等辅助数据仓库管理，同时制定了一系列配套的项目和系统开发维护管理制度。

### 项目成果

1. **建设实施历程和成果**
   * **概述**

上证所数据仓库的建设到目前为止在大阶段上经过了数据仓库一期、数据仓库二期、信息增值服务建设、数据仓库灾备系统建设四个过程，整个建设过程遵循合理的规划、科学的部署原则，完成了数据集中存储、应用平台搭建、信息集中服务、安全运营保障的建设历程。

* + **数据仓库一期建设**
* 建设目标

数据仓库一期的建设目标是完成上证所基本信息平台的搭建。通过集中、统一的数据中心对所有结构化信息和非结构化信息的整合机制，最大程度地保证了上证所自建所以来所积累的全部历史信息，以及以后新增的所有有价值数据的完整性和可用性。

* 建设过程

上证所于2002年启动数据仓库一期(交易历史数据存储项目)的建设，经过将近一年的持续努力，正式建成了国内证券行业数据容量最大、数据内容最完整的TB级数据仓库一期工程。数据仓库一期的建成标志着上证所构建起了基本信息平台，通过集中、统一的数据中心对所有结构化信息和非结构化信息的整合机制，最大程度地保证了上证所自建所以来所积累的全部历史信息，以及以后新增的所有有价值数据的完整性和可用性。

交易历史数据存储项目于2002年10月正式进入项目实施阶段，上证所抽调精兵强将，组成项目组，制定了项目管理、文档规范、安全保密、变更控制、汇报制度等；专门设立项目管理委员会，所有重大事项均通过管理委员会进行决策。根据项目内容，划分了介质、数据分析、加载、应用等多条主线。全部实施过程经过了信息调查分析、业务需求分析、逻辑数据模型设计、体系结构设计、物理数据库设计、ETL设计开发、历史数据加载、备份、业务应用设计开发、系统测试、验收移交、用户培训及知识转移等各个环节。

由于前期做了充分的准备，采用科学的工作方法。在2003年初，上证所就建立了自动数据加载机制，并开始对2003年新产生的每日增量交易数据进行抽取、转换、加载；一方面不断清洗历史存量数据，另一方面根据用户的反馈，不断完善各种功能。2003年11月，开发工作全部完成，其后的测试、验收、全面试运行均取得了圆满成功。

* 建设成果

在科学的逻辑数据模型基础上，上证所通过恢复历史备份的数千张光盘和磁带，完成了1990年以来所有交易历史数据的整合与存储；通过对多个交易相关业务子系统的处理，完成了交易相关上市公司、会员、基金、债券、指数、板块、统计数据等信息的整合；通过设计对非结构化数据的整合方案，完成了交易所8大业务部门长期积累的数百万页非结构化文档信息的存储；并且通过建立了数据自动加载机制，完成了对于以后日常业务增量信息的自动加载处理。

* + **数据仓库二期建设**
* 建设目标

在完成数据仓库一期建设的基础上，于2005年启动数据仓库二期建设，数据仓库二期建设目标是在数据仓库一期的基础上开发信息服务平台及应用服务体系，使信息产生价值为所内外提供数据加工、统计分析、市场监管、决策支持服务，为进一步发挥数据仓库的价值服务打下良好的基础。

* 建设过程

数据仓库二期项目于2005年1月从业务部门的集中需求访谈开始，向所内9大业务部门共进行了32次的需求访谈，访问了业务部门领导和业务人员共40余人次，共收集了511个原始需求。项目组对原始需求进行了归纳整理，建立了全所级的10大需求分类，并在部门原始需求的基础上归纳整理形成311个全所级作为应用开发的业务需求。

在2005年6月底完成了第一阶段应用的开发工作，与此同时面向业务部门进行了大规模的应用推广工作，向业务部门使用人员全面地介绍数据仓库系统的操作和分析的使用方法，经过推广工作体现了系统各项开发技术都已成熟，业务部门对数据仓库系统有了全面的认识。

项目于2005年10完成了第二阶段大规模的开发工作，于此同时元数据管理子项目、非结构化数据管理子项目、数据仓库管理子项目也已经完成，并在生产环境下完成了全面的部署。2005年10月至11月期间进行了大规模的集成测试工作，对系统非功能性和功能性的各项指标全面进行了测试，为系统的投产打下了良好的基础。

项目于2005年12月之后进入试运行，于2006年2月转入了正式运行。系统运行的情况完全达到了项目的各项指标的要求，同时也作为上证所建成15周年的所庆献礼。

* 建设成果

数据仓库二期项目的建设完成了前端应用开发共完成了600多个包含报表、查询、多维分析、KPI等多种形式的应用，为上证所各业务部门提供统计报表、数据查询、关联分析等信息服务；数据仓库二期的Portal集成了系统的各个子项目的功能，包括：前端应用、元数据管理、非结构化数据管理、数据仓库管理等。通过统一的门户入口，用户根据权限分配情况可使用数据仓库二期提供的各项服务，实现单点登入的功能；对用户使用数据仓库系统的用户进行统一的管理；元数据管理子项目包含了元数据管理和数据管理，提供了完整的元数据管理和应用功能，包括：信息浏览、流程管理、在线帮助、影响分析、安全管理等内容。

* + **信息增值服务建设**
* 建设目标

通过数据仓库二期的建设，塔建的信息服务的基础平台已初步显现出数据仓库关联分析、数据统计的威力；上证所于2006年基于数据仓库平台逐步加大信息增值服务的开发，完成了Top-View信息产品加工、统计报表移植、债券统计系统移植、异动股票统计、交易公开信息披露统计、计费系统移植、权证信息披露等。

* 建设过程

在完成基础数据存储平台、应用服务平台搭建的基础上，上证所不断扩大数据仓库系统的应用支持服务；面向所外提供各种信息发布和增值服务支持，如：异动股票信息发布、权证信息披露、交易公开信息披露、Top-View信息产品数据加工、上证所统计月报/季报/年鉴等，使上证所数据仓库系统与证券市场息息相连，特别是Top-View信息产品数据加工通过向市场投资者提供信息增值服务为上证所带来了可观的经济创收；面向市场监管单位提供证监会数据报送、证监会统计报告、证监局外派机构监管信息统计等；面向所内除了提供前端应用查询、分析外，向所内提供的数据加工服务有：成交概貌统计、指数概况统计、债券信息统计、会员计费统计、决策数据支持等，同时向SAS信息挖掘系统提供数据分析支持。

* 建设成果

通过信息增值服务的建设，上证所一方面基于数据仓库统一视图的数据管理提供最权威的数据统计信息，确保数据统计口径的一致性和真实性，另一方面减少了重复投资代价。在确保数据仓库系统安全稳定运行的基础上，所内相关的统计报表系统逐步迁移到数据仓库的应用平台上，通过单一的数据中心向辐射提供各方面的信息支持服务，通过单一的门户网站集成所内相关的应用系统向所内相关人员提供信息查询服务。

* + **数据仓库灾备系统建设**

由于上证所数据仓库系统越来越成为企业运营不可缺失的支持平台，一旦数据仓库系统停止服务将有可能影响整个证券市场的正常运转，因此建立双中心的运营保障数据仓库体系对确保证券市场稳定发展是非常必要的。

于2007年上证所开始着手数据仓库灾备系统的建设，通过建立两套数据仓库系统实现双中心可主/备切换，达到同城异地灾备的目的，从而确保数据仓库系统提供持续稳定的对外服务，目前该项目在建设过程中。

通过双加载实现主从系统数据的同步，为确保双系统数据的一致性在系统之间增加数据一致性检查，在数据加载的事前、事中、事后进行数据的一致性检查，确保双系统的数据相一致，建制两套对等的应用服务系统使其可任意切换数据库服务，以松偶合方式增强系统的容错能力，任何一套数据仓库系统都可独立运行，通过简单的设置实现主/从系统的切换，达到互为备份的目的。

1. **数据仓库未来展望**

上证所数据仓库系统在上证所领导的规划下以科学的发展观逐步实现新一代信息系统的建设，根据上证所信息系统长期发展的战略规划，结合Teradata数据仓库技术不断的发展，从广度和深度上不断拓展和加深数据仓库系统的建设。

在现有数据仓库服务对象和服务内容的基础上，为更大地发挥上证所数据仓库的作用，扩大数据仓库服务的对象和内容是未来规划需要开展的工作；在服务对象方面，在考虑信息安全的前提下，进行面向全市场提供数据仓库服务的探研，使市场参与者能够直接或间接受益于数据仓库的服务；在服务内容方面将不断丰富上市公司、会员公司、证券市场的服务内容，为更好地服务于市场提供全面的服务。

随着上证所一新代交易系统的推出，EAI消息总线的投入使用，结合Teradata数据仓库系统已具有的 AEI(Active Enterprise Intelligence)数据仓库技术，这些都为上证所ADW(Active Data Warehouse)的建设具备了良好的基础，为再度提升上证所数据仓库的服务能力，使其符合AEI型的数据仓库已成为可能，从而能够更加紧密地跟踪市场的变化、提高服务效率和效能、加强市场监管力度，为更好地服务市场和监管市场都将起到重大的作用。