

摘 要

随着信息管理系统的应用的增多，企业的业务系统中产生了越来越多的数据孤岛。怎样把数据孤岛利用起来，于是就产生了 ETL 概念。ETL 指的是数据抽取、数据转换和数据加载，他们在构建数据仓库过程中占到了极其重要的地位。在如今社会中，信息在企业中的地位日趋重要，是企业进行科学的管理以及决策分析的基础。目前，大部门企业在构建业务系统和办公自动化系统中花费了大量的时间和资金，用来记录业务处理各类数据。在普遍情况中，企业的海量数据每 2 年左右就会翻倍增长，数据中隐藏着大量的商业价值，但是往往企业所能运用的仅仅只占总量的 4% 左右。企业并没有将这些数据资源加以最大化地利用，从而浪费了更多的资金以及时间，更在制定商业决策时候失去了先机。于是，企业如何运用技术手段，将这些业务数据更好加以利用，已成为提高企业竞争力的瓶颈。

本文以国内某企业为研究背景，根据公司决策层的需求，分析和设计了适合公司决策支持的数据仓库物理结构和逻辑模型，并对数据抽取、转换、装载(ETL)和数据清洗等技术进行了具体地应用实践，设计并实现了可满足决策分析所需的数据仓库及其相关的 ETL 工具。本论文的主要研究内容和成果概述如下：

首先，在调研该企业的组织机构、业务系统和企业管理流程的基础上，分析和设计了适合该公企业实际需要的数据仓库的五大模块，即财务模块、供应链模块、成本模块、生产制造模块和人力资源模块。采用星型和雪花模型结合的方式，设计了五大模块的逻辑和物理模型。根据企业具体情况，在操作数据存储(ODS)下创建了数据仓库。

其次，设计了一个 ETL 工具以解决半结构化数据的抽取。运用 DOM 对象对数据源进行分析，从企业实际出发设计的 ETL 工具，能够将企业财务软件导出的 XML 格式数据加载到数据仓库中，也解决了商用的 ETL 工具无法对 XML 文档进行直接抽取、加载的弊端，解决了企业数据分析的实际困难。

最后，设计了基于结构化数据的 ETL 工具。在该企业实际运行的业务系统中，基于 ORACLE 和 SQL SERVER 的结构化数据占了大多数，基于结构化数据的 ETL

工具，给用户开放了自行设计清洗函数接口，解决了商用 ETL 工具清洗函数的不可扩展问题。另外，并且在元数据中保存了数据提取的 SQL 脚本，降低了同种 SQL 多次执行时重编译的时间开销。

关键词： ETL，ERP，数据仓库，XML，ODS

Abstract

With the increasing application of the information management system, enterprises in the business system has had more and more data island. How to use these data island, this created the concept of ETL. ETL is a data extraction (Extract), data conversion (Transform) and data loading (Loading), is to build a data warehouse is extremely important part. In today's society, information is the important resource of enterprise, is the enterprise with scientific management, decision analysis based. At present, most companies spend a lot of time and money to build online transaction processing management system and office automation system, used to store the transaction related data. According to statistics, the amount of data every 2 to 3 years time will grow exponentially, these data contains a huge commercial value, and enterprise concern often only accounts for the total amount of data from 2% to 4%. Therefore, the enterprise still not maximize the use of existing data resources, so waste more time and money, but also lost the best opportunity to make critical business decisions. Then, how the enterprises through a variety of technical means, and to convert data into information, knowledge, has become the main bottleneck to improve its core competitiveness.

In this paper, a domestic enterprise as the research background, according to the company's decision-making needs, analysis and design for the company's decision support data warehouse physical structure and the logical model, data extraction, transformation and loading (ETL), and the data cleaning technique for specific application practice, the final design and implementation can meet the decision analysis of the desired data warehouse and related ETL tools. The present paper main research contents and results are summarized as follows:

First of all, in the investigation of the enterprise organization, business system and management on the basis of analysis and design for the company, decision-making support required by the data warehouse in the five major themes, namely the financial subject, theme, the theme of supply chain cost, manufacturing and human resources theme theme. Adopt star type model and snowflake model combining, established the five major themes of the logical model and physical model. According to the company's specific situation, in the operation data storage (ODS) environment to achieve data warehouse.

Secondly, the design of a XML based semi structured data ETL tools. Use DOM object analysis of XML data source, from the enterprise actual situation the design and implementation of semi structured data ETL tools, solve the company's financial software system interface XML semi-structured data is loaded into a data warehouse of the problems, but also solve the commercially available ETL tools can not be directly extracted and loaded into the data warehouse XML document malpractice, to meet the actual needs of enterprises.

Then, the design of a structured data ETL tools. In the company's existing business systems, most of the data are based on ORACLE 9 and SQL SERVER 2005 structured data through structured data, design and implementation of ETL tool, to the user to reserve the custom data cleaning function interface, make up the commercial ETL tool cleaning function can not be extended defects. In addition, in support of ETL tools operation metadata to retain the data extracted from the SQL text, reduces the similar SQL is executed again to recompile brings time overhead.

Keywords: ETL, ERP, DW, XML, ODS

目 录

摘 要	I
Abstract	II
目 录	I
第一章 绪论	1
1.1 相关背景	1
1.1.1 课题背景	1
1.1.2 项目背景	1
1.2 国内外研究现状	2
1.3 研究意义	3
1.4 论文研究内容	4
1.4.1 研究思路	4
1.4.2 主要工作	4
1.5 论文章节安排	5
第二章 相关技术研究	8
2.1 ETL 技术介绍	8
2.1.1 数据抽取	8
2.1.2 数据清洗和转换	8
2.1.3 数据加载	9
2.2 增量抽取机制	9
2.2.1 触发器方法	9
2.2.2 时间戳方法	9
2.2.3 全表删除插入方法	10
2.2.4 全表对比方法	10
2.2.5 日志表方法	10
2.2.6 系统日志分析方法	10
2.3 分布式增量 ETL 研究	10
2.3.1 数据流分割	11
2.3.2 并行数据转换处理	11
2.3.3 系统日志分析方法	12
2.4 数据交换平台关键技术介绍	12
2.4.1 XML 技术介绍	12
2.4.2 Web Service 技术介绍	15
2.4.3 数据交换平台系统结构介绍	18
2.5 本章小结	21
第三章 企业 ETL 需求分析	22
3.1 项目需求	22

3.2 项目范围	22
3.2 项目成功标准	23
3.3.1 功能标准	23
3.3.2 技术标准	23
3.3.3 项目风险及相应措施	23
第四章 决策系统项目总体方案	24
4.1 决策系统平台概述	24
4.2 决策系统架构	24
4.3 决策系统业务规划	25
4.3.1 财务分析主题	25
4.3.2 供应链分析主题	27
4.3.3 成本分析主题	28
4.3.4 生产制造分析主题	29
4.3.5 人力资源分析主题	30
第五章 决策系统主题实现设计	32
5.1 主题实现项目流程图	32
5.2 主题技术实现描述	32
5.3 ODS 层设计	32
5.3.1 数据源到 ODS 层表设计	32
5.3.2 ODS 层表结构设计概述	35
5.3.3 ODS 层数据来源描述	35
5.3.4 数据源到 ODS 层表设计	36
5.4 DW 层设计	36
5.4.1 DW 层设计概述	36
5.4.2 维度表设计	36
5.4.3 公共表设计	38
5.4.4 ETL 过程监控表设计	38
5.4.5 ETL 调度起始时间控制表设计	39
5.4.6 ODS 至事实数据表验证 SQL 清洗表设计	39
5.4.7 ODS 至维度表数据验证 SQL 清洗表设计	39
5.4.7 事实表数据质量分析 SQL 清单	39
5.4.8 ODS 至事实表数据验证结果表设计	40
5.4.9 ODS 至维度表数据验证结果表设计	40
5.4.10 ERP 至 ODS 数据验证结果表设计	41
5.5 半结构化数据 ETL 工具设计	41
5.5.1 半结构化数据及其特点	41
5.5.2 XML 数据接口 DOM	42
5.5.3 半结构化数据 ETL 元数据的逻辑模型设计	44
5.5.4 算法描述	44
5.5.5 算法实现	45
5.6 结构化数据 ETL 工具设计	48
5.6.1 结构化数据 ETL 元数据的逻辑模型设计	48
5.6.2 结构化数据 ETL 工具的系统结构	48

5.6.3 自定义数据清洗方法在 ETL 工具中的实现.....	49
第六章 总结与展望.....	50
6.1 总结	50
6.2 展望	51
致 谢.....	52
参考文献.....	53
附录.....	56

第一章 绪论

1.1 相关背景

1.1.1 课题背景

自从数据库相关技术产生以来，越来越多的企业使用计算机来替代纸质文档记录生产活动中产生的业务信息，以提高工作的效率。但是，系统开发人员往往只是针对独立的问题来设计应用程序以完成特定的目的，从而导致这些系统程序之间由于缺少必然的联系不能协同工作，使得在现有的独立系统基础上无法实现企业高层次的信息共享。企业如果要从这些积累下的信息数据中提取所需要的信息，就必须要将这些相互独立的信息数据集成起来，创建数据仓库，从而达到各系统信息共享的目的，这样企业才能迅速、更准确的对当前和过往数据进行分析，对今后事态发展进行预测，才能为做出理想的决策提供依据^[1,2]。

企业的日常业务数据是建立数据仓库的基础，而建立数据仓库的目的是为了能够从数据仓库中筛选出所需要的信息以更好的支持企业决策。而建立数据仓库的整个过程，它是面向主题、集成并且不可更新的，是不断变化的，需要从基本框架的建立入手，不断完善和丰富整个系统^[3]。数据集成是建立数据仓库的难点，而在数据集成的过程之中又将解决三个主要问题：数据的更新同步，数据源之间各类数据的不一致表现，如何融合不同的数据源^[4]。

ETL 即数据提取、转换和装载，是数据仓库的核心技术之一^[5]，是构建数据仓库的重要一环。数据仓库项目中 60%-80%的工作量是用来进行 ETL 过程的建立与执行的^[6,7,8,9]。将数据从不同的数据源中采集到各自的目标数据库的过程，就是所谓的 ELT 过程，ETL 能够管理数据的质量，并且将其贯穿于整个方案的全过程中。数据仓库建设过程的主要部分就是 ETL，其效率和数据质量直接影响数据仓库的建设和企业决策分析的有效性。

1.1.2 项目背景

-----公司主要生产、销售给类油封制品、密封罩、保护圈、减震橡胶等其他工业橡胶制品。产品广泛应用于汽车船舶、航空航天、铁路车辆、建筑机械、农用机械、石油化工、家用电器等领域。公司成立于 1995 年，由于受当时条件限制，公司对企业信息化建设仅仅是从应用层面上展开，没有以企业整体角度进行规划，导致企业如今同时存在财务、生产、销售、采购、仓存、人事、客户管理等多套业务系统，并且各个业务系统的品牌和平台各不相同。一方面经过这么多年的使用，存储了海量的业务数据，随着数据量的剧增导致企业业务系统的响应速度越来越慢；另一方面企业管理层人员所需要的报表五花八门，有业务数据也有有财务数据的，但是同样的数据经过各自系统的加工之后就会存在很大差异，给企业决策带来了困难。这些运行在各个业务部门，满足部门级应用的业务软件系统给企业带来的问题表现为：

- ① 整体系统数据响应速度越来越慢。
- ② 缺乏分析数据一致性。来源于不同口径的统计数据出现了不一致现象，带来了决策上的不便。
- ③ 多数据源并存，数据统计分析困难。
- ④ 历史数据管理越来越繁杂。
- ⑤ 目前系统对企业的宏观管理缺乏数据支撑。

为解决这些问题，加强企业的决策力，把握未来企业的发展方向，公司高层决定使用 ETL 技术建立企业级数据仓库满足企业的财务、供应链、生产、成本、人事分析的需要。

1.2 国内外研究现状

从 20 世纪 70 年代中期开始人们就进行了数据集成的研究，其发展可归纳两个过程，过程一针对多数据库系统^[10,11]的研究。这个过程可分为三类。第一，物理上分布而逻辑上集中的结构；第二，逻辑和物理均分布的结构，每个结点采用邦联的独立模式，而非采用全局模式；三是多语言数据集成。

20 世纪 80 年代至 90 年代，随着计算机及网络的日益普及，传统的数据集成方案对人们获取更多数据的要求已经无法满足，这时数据仓库理论开始被人们接受，相关技术随之产生和应用。数据仓库的集成工具：ETL 技术的研究也随着数据仓库技术的发展日益活跃，研究主要涉及以下方面^[12,13]：

- (1)为避免扫描整个庞大的数据集而进行的高效数据检测算法的研究；
- (2)在数据自动化清洗和检测中增加人工判断过程为了提高处理的精度；
- (3)并行处理海量数据集的研究；
- (4)消除合并数据冗余的研究；
- (5)数据清洗框架通用性的研究；
- (6)数据仓库数据更新的研究；
- (7)ETL 工作流的研究；
- (8)ETL 过程优化的研究。

由于 ETL 在数据仓库中的重要地位，其已成为各类软件厂商占领市场的战略要地。众多国外数据仓库项目被成功实施，软件厂商纷纷推出了自己的相关产品，如 IBM 的 Warehouse Manager，Microsoft 的 DTS，Oracle 的 Warehouse Builder，Informatica 的 DataStage 等^[14,15,16]。

目前国内对 ETL 的研究也有了一定的基础^[17,18]。在实际系统的开发应用中，一部分数据集成工具只能在该系统下使用，仅仅是针对该特定系统开发的；另一部分对数据的抽取、转换和加载的过程是通过脚本的编写来实现的。无论是上面的哪种，都存在需要克服的弊端，比如灵活度很低，其次设计过程的代价很大，无论是设计过程的周期，还是对人力、物力的消耗，是一个耗时耗力的过程。此外，处理异构数据库系统的转换工具功能相对简单，在处理复杂数据时达不到好的效果。^[19,20]。

1.3 研究意义

数据集成在数据仓库的建立以及系统集成应用中占重要地位，数据的一致性、完整性、有效性，对数据仓库的建立和数据挖掘的成败起着关键作用。ETL

作为数据集成的解决方案，有着重要的实际意义。

ETL 研究理论意义

(1)ETL 工具研究使得不同数据源间的异构性得以解决。数据仓库中的数据都来自多个不同的数据源，每个数据源又往往是异构的，如：不同数据库数据、XML 文件甚至是平板文件。因此，在数据集成过程之中，如果针对不同的数据源进行设计、编码就会极大延长开发得周期，增加开发成本。支持异构数据源作为 ETL 工具的一个特点，它能够从多种数据源中读取或加载数据，这样，开发人员就能够节省时间、提高效率从而投入更多的时间在数据挖掘的方面。

(2)ETL 工具便于数据清洗。由于数据源多样性的原因，导致了大量“脏数据”(Dirty Data)的存在^[21]。这些所谓的“脏数据”会在被数据仓库装载的过程中增加数据仓库的负载，在数据质量无法保证的情况下，那么就会对数据仓库的后续操作和数据挖掘产生严重的影响。ETL 作为数据源与数据仓库间的一座桥梁，以克服“脏数据”对数据仓库的影响，提高处理效果。

ETL 研究应用意义

(1)加快企业各个系统及数据的整合，以便数据仓库为企业决策提供支持。如前文所讲，大多数企业尽管具备了各类计算机业务处理系统，积累了大量的基础数据，但数据系统之间相互隔离，信息共享度低，网络、信息资源利用度不高，信息化水平不高。数据仓库、数据挖掘、ETL 等技术能够帮助企业从积累的大量数据中挖掘信息、提炼知识，因此，这些技术必会慢慢普及；而作为建立数据仓库所需的 ETL 技术，也会随之发展。

(2)将原先异构数据源中的数据进行同一输出。如在电子商务领域，基于 Internet 的 XML 方式正逐步取代传统的 EDI 方式。由于 XML 的结构非常严格，使得对 XML 的分析变得容易。在研究中，元数据管理和转换规则的描述非常适合用 XML 来实现，而且 XML 能够成为不同应用程序数据的中间形式^[22]，这为解决数据源异构问题提供了便利。数据集成工具必须要能够提供数据交换接口：根据用户定义的数据格式，经过 ETL 工具在多个的数据源中抽取数据，并存入定义好的 XML

文件中，为电子商务做好数据的准备。

1.4 论文研究内容

1.4.1 研究思路

本课题来源于客户商业智能系统的开发项目，主要目的是利用现有数据资源建立一个数据交换平台，为企业能够快速、有效地对业务数据进行分析提供保证。因此本文主要研究了以下问题：

1、数据增量抽取法。在上一次数据抽取基础上，只对数据表中新增或者修改的数据进行抽取，从而降低 ETL 后续数据处理的消耗。快速准确捕获变化的数据而又不对现有系统增加太大压力是增量抽取能够实施的关键，所以，增量抽取的算法相对全量抽取来说更为复杂。捕获方法的要求有两点：一是准确，能够将业务系统中的数据变化准确及时的捕获到；二是性能，要实现组件各部分的独立运行而不能增加对整个系统的负荷。

2、分布式增量数据 ETL 方法。串行和并行是数据转换过程中的两种基本方式，串行不同于并行，串行方式直接按照顺序按部就班执行就可以，但对执行效率而言没有太大优势，而并行就可以解决这个弊端，并行方式是将串行方式进行分割，转换成多个过程，这就提高了执行的效率，在这里我们通过分布式的处理方法，就能够真正提高执行的效率，同时加入正确的协同处理机制，保证处理以后的数据具有规范性和正确性。

3、基于增量数据 ETL 分布式数据转换平台。数据转换平台能够在不同的数据库之间、不同数据格式之间，进行数据转换。它主要面对的问题在于：安全性、高效性、完整性、一致性、异构性等，本文主要在完整性和高效性上进行了研究。

1.4.2 主要工作

本文以提高增量 ETL 执行效率为目标，针对以上问题进行研究，主要取得了以下成果：

1、研究了增量数据抽取方法。通过对增量数据抽取方法和全量数据抽取方法的比较，并根据企业的业务特点，提出新的提取算法，这种算法是根据数据抽取前后数据属性的变化来建立的，属性的挑选是根据变化明显的特征进行的。根据比较的结果来进行增量数据的抽取，由于增量数据抽取的属性相比较全量数据抽取要大大减少，从而达到减轻数据抽取复杂度，提高数据抽取效率的目的。

2、设计了一个 ETL 工具以解决半结构化数据的抽取。运用 DOM 对象对数据源进行分析，从企业实际出发设计的 ETL 工具，能够将企业财务软件导出的 XML 格式数据加载到数据仓库中，也解决了商用的 ETL 工具无法对 XML 文档进行直接抽取、加载的弊端，解决了企业数据分析的实际困难。

3、设计了基于结构化数据的 ETL 工具。在该企业实际运行的业务系统中，基于 ORACLE 和 SQL SERVER 的结构化数据占了大多数，基于结构化数据的 ETL 工具，给用户开放了自行设计清洗函数接口，解决了商用 ETL 工具清洗函数的不可扩展问题。另外，并且在元数据中保存了数据提取的 SQL 脚本，降低了同种 SQL 多次执行时重编译的时间开销。

1.5 论文章节安排

论文的结构安排如下：

第一章 绪论。介绍了课题背景，阐述了各种数据交换的方法特点，对 ETL 的研究现状进行了分析，结合作者在攻读硕士学位期间的理论学习和学术研究工作，确定论文研究思路。

第二章 相关技术的研究。对当前流行的增量 ETL 技术进行了介绍，介绍了数据交换平台的关键技术，研究了增量 ETL 技术发展趋势，并分析了其特点。

第三章 对企业 ETL 的需求进行了研究和分析，为企业数据分析系统设计提供了基础。

第四章 对企业 ETL 分析系统的系统架构、功能模块进行了介绍。

第五章 对设计的系统各个功能进行了详细的介绍和分析，并结合文章的创新要求提出了自己的观点并加以实现，从而完成了本文的重点工作。

第六章 总结与展望。阐述了全文的工作方法，总结了研究成果的优缺点，以及后续工作的努力方向。

第三章 企业 ETL 需求分析

3.1 项目需求

企业的管理者非常重视数据分析对于企业管理经营的支持作用。自 2001 年起就设计大量的 Excel 报表用于分析企业的各核心业务流程以及经营绩效，同时，在金蝶 K/3 系统应用之后，也自定义开发了多张管理报表，用于决策支持。但是，无论是自己设计的大量 Excel 报表还是金蝶 K/3 的报表，都无法提供系统的，关联的，快速的，高效的数据分析功能。具体表现如下：

- a) 大量 Excel 报表，录入填报工作量重，无法获得及时数据。金蝶 K/3 系统提供的报表有限，大量数据闲置，无法释放价值。
- b) 尽管企业的 Excel 报表设计非常科学，但是缺乏关联，无法形成分析体系。
- c) 不同部门对于数据有着不同的理解和标准，未形成一致的数据标准。
- d) 多个企业数据汇总不及时，没有形成全面监控体系。

- e) 虽有企业全面绩效考核愿望，但是由于数据的收集和计算困难，一直没有执行。
- f) 数据重用效率不高，没有形成跨业务领域数据支持。

综上，该公司需要一套能够实时监控业务，穿透和多维度分析经营，协助全面绩效管理的数据分析平台。

3.2 项目范围

项目实施是一个迭代开发的过程，最终用户看到数据分析结果时总会不断产生更多的新需求，所以，整体规划，分步实现是 BI 项目需要遵守的最优原则。

最终目标是实现全集团各个业务的数据共享和整合的分析平台，实现跨业务领域的数据库重用和端到端的数据支持。但是，部分分公司的信息化基础薄弱，所以，项目的实施应该从信息化程度最高的分公司开始。本期详细实施范围列示如下。

- a) 目标组织机构：———有限公司
- b) 目标业务流程：财务、供应链、成本、生产、人力资源及企业高管平台
- c) 数据源系统（抽取范围）：金蝶 K/3 系统，其他规范 Excel 文档模板
- d) 源数据获取时间范围：2005 年至今的企业业务数据。（其中金蝶 K/3 系统数据从 2006 年开始，以往数据从规范 Excel 模板中抽取）

3.2 项目成功标准

3.3.1 功能标准

- a) ETL 功能，能够自动完成从 HIS 系统和 EXCEL 文档的数据抽取和整合工作。
- b) 多维数据分析功能和穿透查询功能（OLAP），按照需求描述的维度支持多维分析和穿透查询功能。
- c) 简单易用的 DashBoard，定制 DashBoard 的展现（需求描述的报表列表）
- d) 系统以 B/S 架构实现，允许授权用户通过浏览器访问门户。
- e) 系统支持不同权限的用户访问，并可以实现数据按照用户权限动态加载和展示。

3.3.2 技术标准

- a) 处理效率，本系统的响应速度应满足全公司并发用户数量内业务和管理人员同时使用系统的要求，查询结果响应时间不超过 10 秒。
- b) 可用性，系统稳定运行时间超过总体使用时间的 90%，故障间隔时间应大于 60 天。
- c) 可维护性，提供系统维护说明文档，并移交系统开发说明书以及元数据，以便日常维护

3.3.3 项目风险及相应措施

- a) 源系统数据精细度不一致的风险，建议选择最低颗粒度的原子数据以支撑更加灵活的分析。
- b) 系统外的平面文件数据源（Excel）只有汇总信息没有明细信息的风险，建议将汇总信息分解至最低粒度。
- c) 数据质量和准确度的风险，校准原始数据质量不在本次实施范围之类。
- d) 用户应用的风险，系统如果最终无人使用将丧失价值，建议在实施和推广过程中不断向机构内部强化概念，鼓励应用。

第四章 决策系统项目总体方案

4.1 决策系统平台概述

金蝶 K/3ERP 系统是一个基于企业业务发展、评估和决策的管理分析系统，它既可贯穿于生产制造的整个监督管理过程，更重要的是能够在整个业务流程的各个环节中进行分析判断和归纳总结，给企业决策者提供可靠的决策信息。该系统建立和预置多个业务分析主题，包括财务分析主题、成本分析主题、供应链分析主题、生产制造分析主题以及人力资源分析主题。金蝶 K/3ERP 是面向企业高层管理者的决策信息平台，通过标准化、图形化的关键绩效指标监控、业务报表分析、覆盖企业经营全过程的即时智能查询、以及日常工作集中处理等功能，帮助管理者快速做出企业经营管理的正确决策，为企业提供全面和可靠的分析解决方案。这个系统具有数据采集灵活、集成度高、数据处理效率高、响应时间快等特点，并且用户界面十分人性化，操作简洁方便。

4.2 决策系统系统架构

根据企业的实际需求，金蝶商业智能解决方案选择金蝶 K/3 商业智能产品实现系统的搭建。系统技术架构如下图 4-1：



图 4-1 系统架构

金蝶 K/3 商业智能在一个综合解决方案里提供三个组成部分：

- 快速查询引擎：数据被加载到内存中允许立刻查询，或子集数据来展示与特定用户有关的数据。此外，给用户显示了一次选择中不包括的数据。

-

按需计算引擎：图表、曲线图以及金蝶 K/3 商业智能里的各种表格都是多维分析。也就是说，他们跨越一个或多个维度（例如：总销售额按地区划分），展现了一个或多个量度（例如：关键绩效指标，表达式等）。最大的不同是，这些计算是用户点击驱动的，而绝不提前进行。

● 可视化交互式用户界面（UI）：金蝶 K/3 商业智能提供数百种可能图表、表格的类型和样式；有操纵维度的列表框，统计框，和许多其他用户界面元素。每一个用户界面元素都可以点击查询。

4.3 决策系统系统业务规划

金蝶 K/3 商业智能能在一个平台上，提供传统上要求复杂而昂贵的一套软件产品才能提供的所有功能。提供灵活的分析能力，强大的分析应用，和简单的报告打印。这要求组织机构要部署到每一个人——专业分析师做特定详细报告，主管要有一个关键业务信息的仪表盘，最高上司分析输出成绩。此外，金蝶 K/3 商业智能可以让组织机构消除剩余的纸报告，并以业务需求驱动的报告取代之，其网络控制如图 4-2。

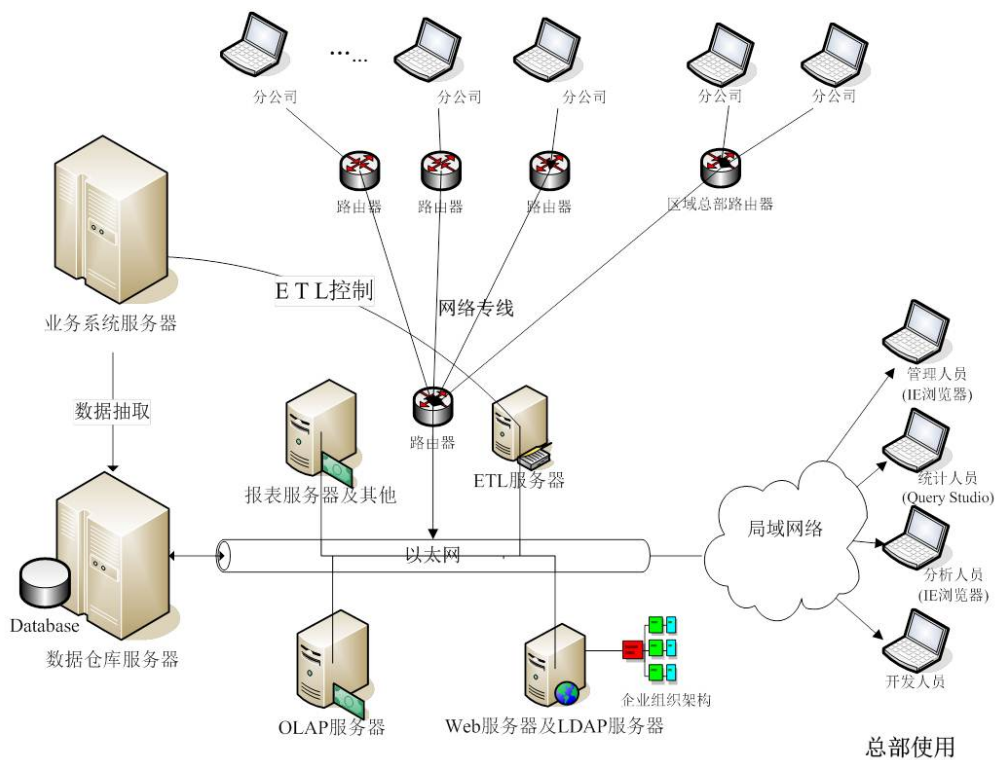


图 4-2 网络控制结构

4.3.1 财务分析主题

财务分析主题通过对各类财务报表进行数据分析和评估，提高企业的投入分配能力、盈利能力、资金运营能力。该主题通过制定诸如利润报表、资金报表、债务表等多项报表，比如盈利及偿债能力，该主题通过波动比率、速动比率、应用账款周转率、资产收益率、固定资产回报率五项指标来进行分析，并和上一周期的数据进行比对，得到有效

的分析数据为决策者提供决策支持。如图（4-3，4-4）。

金蝶K/3商业智能



图 4-3 KPI 分析主题



图 4-4 财务主题关键指标图

4.3.2 供应链分析主题

供应链分析主题对所有企业而言是个很便捷和实用的功能。该主题通过对客户、职员、物料以及部门的供应链进行分析，通过各项供应关系来分析供应链和库存以及材料周转周期的相互关系。同时也反映了销售、采购和库存的合理构成，帮助管理者通过多维分析得出有效的决策信息，使销售、采购和库存相关部门能够更加合理化地进行合理分配和良好协作。以便各部门之间及时反馈供应信息。如图（4-5,4-6）。

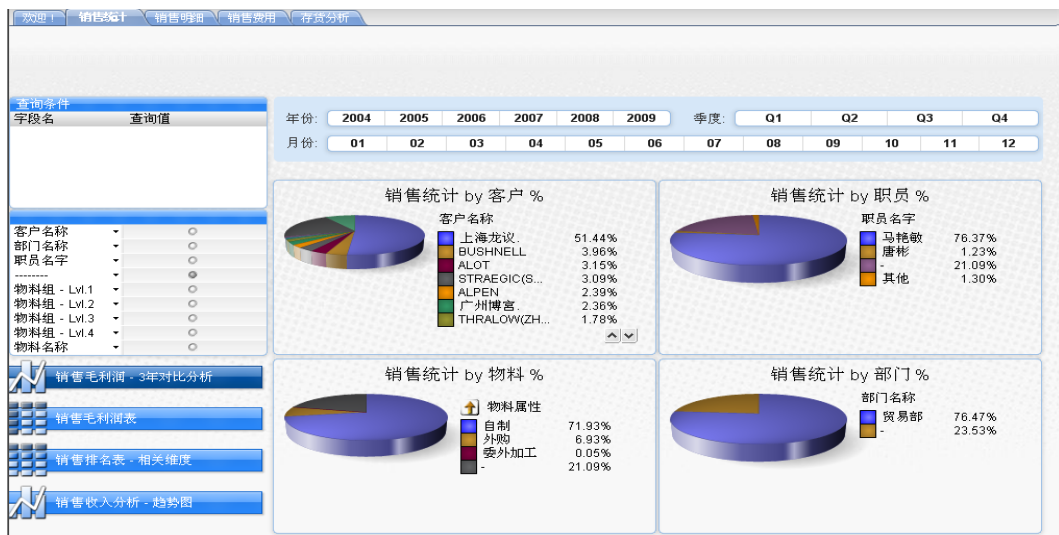


图4-5 销售数据分析



图4-6 存货周转数据分析

4.3.3 成本分析主题

成本分析主题主要是通过对成本结构、成本比较、成本趋势和成本排名的构成进行分析，在这个分析的过程中采用了直方图的直观表现形式，通过成本和成本增长率的比较来制定相应的成本控制措施。如图（4-7,4-8）。



图 4-7 成本结构分析图



图 4-8 成本明细分析图

4.3.4 生产制造分析主题

生产制造分析主题主要通过各项数据分析来检验产品的生产率和产品的质量，这个工作贯穿于产品生产的整个过程，并且对设备的使用情况、产品的质量情况、物料的使用情况、生产订单完成情况等等都有相关的统计和分析，这个分析在企业里面是一个基础工作也是保底工作，产品的质量才是市场竞争最重要的砝码。如图（4-9,4-10）。



图 4-9 产品生产率分析



图 4-9 产品质量分析

4.3.5 人力资源分析主题

人力资源分析主题主要就是通过对企业人力资源的统计分析，从另一方面而言就是对企业内部结构的一种分析。通过对人员结构和编制控制的分析，可以清晰地反映企业的内部结构情况，通过对薪酬的分析，可以得到企业人力资源成本的有效信息。一个企业的内部结构和人力成本，是根据企业的业务需求以及各部门在企业发展中扮演的角色所决定的，这些数据能够为企业的人事部门掌握人力资源情况提供了很好的决策。如图 4-11。

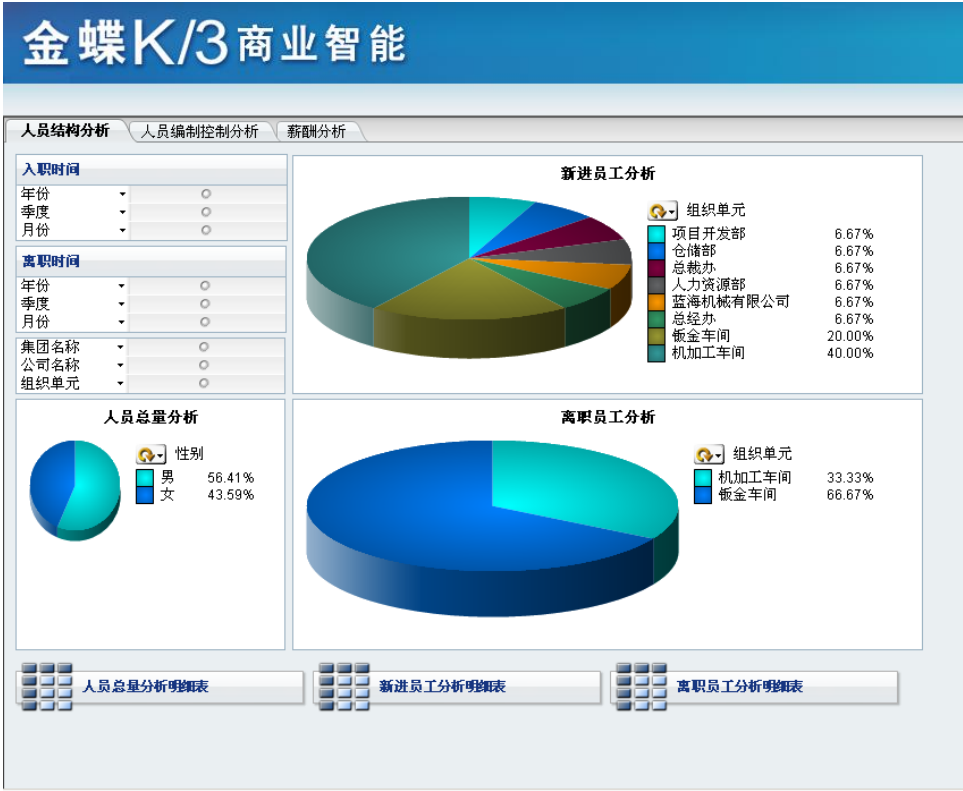


图 4-11 员工结构分析

第五章 决策系统主题实现设计

5.1 主题实现项目流程图

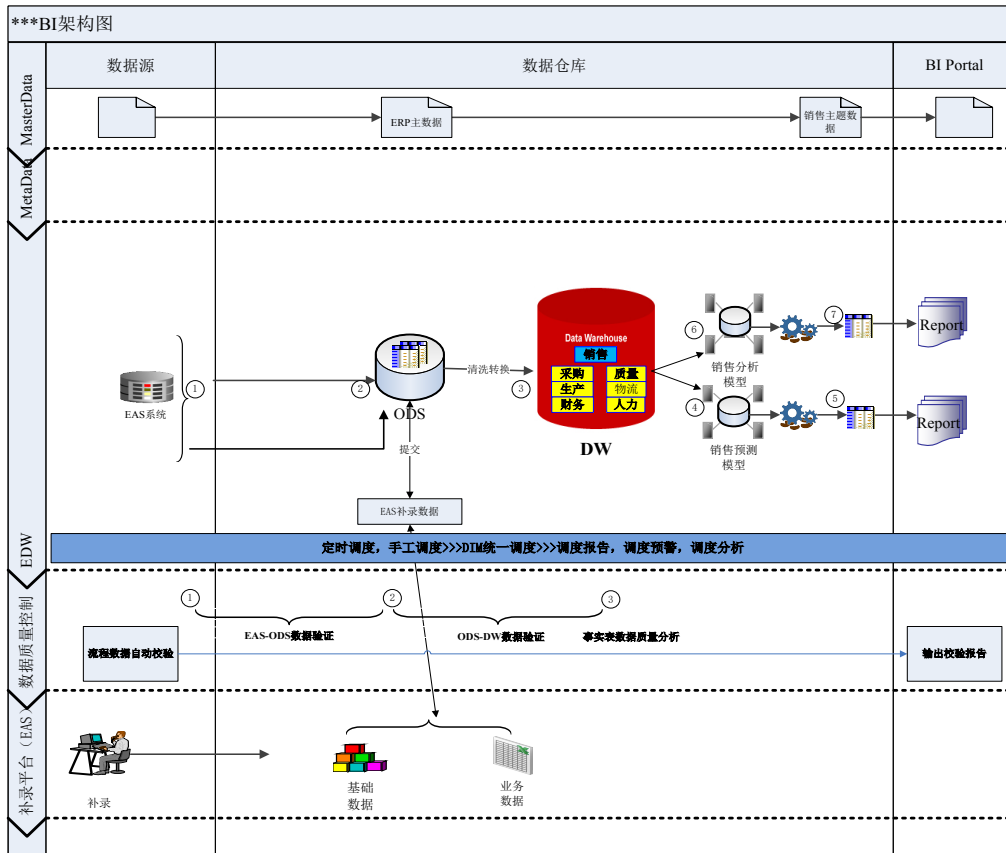


图 5-1 项目主题流程图

5.2 主题技术实现描述

从 ERP 系统获取各种基础及交易数据，通过 ETL 工具（Data Manager、Procedure）对其清洗、转换并计算出各种指标、维度加载到数据仓库。再利用 COGNOS 的 FRAMEWORK、TRANSFORMER 建立各种多维分析模型，最终通过 COGNOS 的 PORTAL 展现出各种多维分析的报表。

5.3 ODS 层设计

5.3.1 数据源到 ODS 层表设计

表 5-1 ODS 层表结构

源表		目标表	描述	序号	
EAS	基础表	t_bd_period	ODSEAS_t_bd_period	时间	1
		t_bd_customer	ODSEAS_t_bd_customer	客户	2
		t_bd_province	ODSEAS_t_bd_province	省份	3

	t_bd_country	ODSEAS_t_bd_country	国家	4
	t_org_sale	ODSEAS_t_org_sale	销售组织	5
	t_bd_saleperson	ODSEAS_t_bd_saleperson	销售人员	6
	t_org_storage	ODSEAS_t_org_storage	库存组织	7
	t_org_company	ODSEAS_t_org_company	财务组织	8
	t_bd_aircrew	ODSEAS_t_bd_aircrew	机组	9
	t_cmr_stellgroup	ODSEAS_t_cmr_stellgroup	钢组	10
	T_BD_Material	ODSEAS_T_BD_Material	钢号	11
	t_bd_gangzhongtype	ODSEAS_t_bd_gangzhongtype	三钢八大类	12
	T_ASS_NewGangZhongType	ODSEAS_T_ASS_NewGangZhongType	原冶金八大类	13
	T_BD_Industry	ODSEAS_T_BD_Industry	行业	14
	T_BD_CURRENCY	ODSEAS_T_BD_CURRENCY	币种	15
	T_DB_WAREHOUSE	ODSEAS_T_DB_WAREHOUSE	仓库（港口）	16
	t_sd_salecontract	ODSEAS_t_sd_salecontract	销售合同	17
	T_SCM_DeliveryType	ODSEAS_T_SCM_DeliveryType	运输方式	18
	T_BD_SettlementType	ODSEAS_T_BD_SettlementType	结算方式	19
	T_BD_PaymentType	ODSEAS_T_BD_PaymentType	付款方式	20
	T_BD_Supplier	ODSEAS_T_BD_Supplier	承运商	21
	t_bd_asstattrvalue	ODSEAS_t_bd_asstattrvalue	技术标准	22
	t_su_figure	ODSEAS_t_su_figure	形状	23
	t_bd_aircrew	ODSEAS_t_bd_aircrew	材质	24
	t_Bd_Measureunit	ODSEAS_t_Bd_Measureunit	计量单位	25
	t_bd_customercompanyinfo	ODSEAS_t_bd_customercompyinfo	客户财务资料	26
	T_BD_GeneralAsstActType	ODSEAS_T_BD_GeneralAsstActType	自定义核算项目	27
	T_BD_Person	ODSEAS_T_BD_Person	人员信息	28
	t_bd_accountview	ODSEAS_t_bd_accountview	科目信息	29
	T_BD_EXCHANGERATE	ODSEAS_T_BD_EXCHANGERATE	汇率	30
	T_CMR_DAYINTERESTRATE	ODSEAS_T_CMR_DAYINTERESTRATE	应收票据贴息日利率	31
	T_CMR_STELLGROUP	ODSEAS_T_CMR_STELLGROUP	钢号钢组规则	32
	T_EXU_PRODUCTIONLINE	ODSEAS_T_EXU_PRODUCTIONLINE	产品路线	33
	T_EXU_PRODUCTIONLINEENTRY	ODSEAS_T_EXU_PDLINEENTRY	产品路线分录	34
	t_cas_receivingbilltype	odseas_t_cas_receivingbilltype	收款类型	35
	t_org_admin	odseas_t_org_admin	部门	36
	T_BD_AssistantHG	ODSEAS_T_BD_ASSISTANTHG	辅助账横表	37
补录表	t_apd_contracttype	odsapd_contracttype	合同类别	38
	t_apd_deliverystatus	odsapd_deliverystatus	交货状态	39
	t_apd_groupclassify	odsapd_groupclassify	集团内部分类	40
	t_apd_importantproduct	odsapd_importantproduct	重点产品	41
	t_apd_materquality	odsapd_materquality	材质	42
	t_apd_materqualitytype	odsapd_materqualitytype	材质大类	43
	t_apd_pricetype	odsapd_pricetype	价格类型	44
	t_apd_processeffect	odsapd_processeffect	加工用途	45

	t_apd_productroute	odsapd_productroute	产品路线	46
	t_apd_representtype	odsapd_representtype	代表品种	47
	t_apd_saleavgtypeclassify	odsapd_saleavgtypeclassify	销售均价代表品种分类	48
	t_apd_salecalculate	odsapd_salecalculate	销售测算	49
	t_apd_saleexpensefore	odsapd_saleexpensefore	销售费用预算	50
	t_apd_saleperformgoals	odsapd_saleperformgoals	销售考核目标	51
	t_apd_sangang	odsapd_sangang	三钢大类	52
交易表	t_Sd_Salecontract	ODSEAS_t_Sd_Salecontract	销售合同号表头	53
	T_SD_SaleContractEntry	ODSEAS_T_SD_SaleContractEntry	销售合同号分录	54
	T_SD_SaleOrder	ODSEAS_T_SD_SaleOrder	销售订单表头	55
	T_SD_SaleOrderEntry	ODSEAS_T_SD_SaleOrderEntry	销售订单分录	56
	T_SD_PostRequisition	ODSEAS_T_SD_PostRequisition	发货通知单表头	57
	T_SD_PostRequisitionEntry	ODSEAS_t_sd_PostReqEntry	发货通知单表体	58
	t_st_lading	ODSEAS_t_st_lading	装运清单表头	59
	t_st_materialentry	ODSEAS_t_st_materialentry	装运清单汇总分录	60
	t_st_ladingentry	ODSEAS_t_st_ladingentry	装运清单明细分录	61
	T_ST_ProduceTaskBill	ODSEAS_T_ST_ProduceTaskBill	生产任务单表头	62
	T_ST_ProduceTaskBillEntry	ODSEAS_t_st_ProdTaskBillEntry	生产任务单表体	63
	T_IM_ManufactureRecBill	ODSEAS_T_IM_ManufactureRecBill	生产入库单表头	64
	T_IM_ManufactureRecBillEntry	ODSEAS_t_im_ManufatRecBillEntry	生产入库单表体	65
	T_ST_StorageForCheck	ODSEAS_T_ST_StorageForCheck	待检入库单表头	66
	T_ST_StorageForCheckEntry	ODSEAS_t_st_StorForCheckEntry	待检入库单汇总分录	67
	T_ST_StorageForCheckDetail	ODSEAS_t_st_StorForCheckDetail	待检入库单明细	68
	T_ST_DetailStorageAdjustBill	ODSEAS_t_st_DetailStorAdjBill	明细库调整单表头	69
	T_ST_DetailSABE	ODSEAS_T_ST_DetailSABE	明细库调整单表体	70
	t_st_producedetail	ODSEAS_t_st_producedetail	产品明细库	71
	T_IM_StockTransferBill	ODSEAS_T_IM_StockTransferBill	库存调拨单表头	72
	T_IM_StockTransferBillEntry	ODSEAS_t_im_StockTranBillEntry	库存调拨单汇总分录	73
	T_IV_StockTBDE	ODSEAS_T_IV_StockTBDE	库存调拨单明细	74
	T_IM_MoveIssueBill	ODSEAS_T_IM_MoveIssueBill	调拨出库表头	75
	T_IM_MoveIssueBillEntry	ODSEAS_T_IM_MoveIssueBillEntry	调拨出库汇总分录	76
	T_IV_MoveIssueBillDetailEntry	ODSEAS_t_IV_MoveIssBillDtEntry	调拨出库明细	77
	T_IM_MoveInWarehsBill	ODSEAS_T_IM_MoveInWarehsBill	调拨入库表头	78
	T_IM_MoveInWarehsBillentry	ODSEAS_t_IM_MoveInWhsBillentry	调拨入库汇总分录	79

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要
下载或阅读全文，请访问：

<https://d.book118.com/908137016076006051>