****

****

**数据库新技术课程报告**

|  |  |
| --- | --- |
| **论文题目** | 数据仓库发展现状 |
| **作者姓名** | 刘子凡 |
| **学 号** | DZ1533016 |
| **学科、专业名称** | 计算机科学与技术 |

**2018年5月18日**

目 录

[第一章 绪论 3](#_Toc515555783)

[1.1 背景介绍 3](#_Toc515555784)

# 绪论

## 背景介绍

人类进入信息时代以来，特别是近20年来，数据规模日益扩大，数据量呈爆炸式增长。例如电商管理者需要从日积月累的交易信息中发现商品畅销的规律以及预测未来的趋势；医生需要从大量病人的电子病历中发现某些疾病的起因、症状；天文学家需要从天文望远镜获取的观测数据（其规模）中发现新的遥远天体及其运动规律等。这些数据的共同特点有：一是数据量巨大，一般都是TB级别以上；二是均以结构化的形式存储在数据库中，包含了大量潜在的、有价值的知识，有的已被发现，而有的还未被发现。这样的知识获取，仅仅依靠数据库管理系统的查询检索机制和统计分析方法，已经远远不能满足实际需求。为了有效地管理和利用这些海量数据、发现其中潜在的知识，数据仓库应运而生。

### 诞生

数据仓库是建立在传统事务型数据库基础之上，为企业决策支持系统以及数据挖掘系统提供数据源的技术。迄今为止，国外的数据仓库已经发展了二十几年，虽然国内起步较晚，但是发展十分迅速，目前已有众多的大型公司或企业正在建设或已经建设不同规模的数据仓库。

传统数据库和数据仓库的最根本区别在于其侧重点的不同。数据处理分为两大类：事务型处理，又称联机事务处理（Online Transaction Processing,OLTP）；分析型处理，又称联机分析处理（Online Analytical Processing,OLAP）。事务型处理以传统数据库为中心进行企业日常的业务处理；分析型处理以数据仓库为中心分析数据背后的关联和规律，为企业决策提供可靠、有效的科学依据。事务型处理和分析型处理的分离，划清了数据处理的分析型环境与事务型环境之间的界限，从而由原来以单一数据库为中心的数据环境，演变为以数据库为中心的事务处理系统和以数据仓库为基础的分析处理系统。企业的生产环境也从以数据库为中心演化为以数据库和数据仓库为中心。因此，在事务处理环境中直接构建分析处理应用是不合适的，要提高分析和决策的效率和有效性，分析性处理及其数据必须与事务型处理及其数据相分离，必须把分析型数据从事务型处理环境中提取出来，按照决策支持的需要重新组织，建立相对独立的分析处理环境，数据仓库正是为了构建这种新的分析处理环境而出现的一种数据存储、组织和处理技术。

### 定义

数据仓库的概念最早出现在20世纪80年代。1993年