数据仓库发展现状

刘子凡 DZ1533016

一、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。