

## Sammanfattning

Redan på 1960-talet kom tankarna på att samla in data från transaktionssystem för att användas i beslutsstödjande syfte. För att möjliggöra beslutsstöd i en organisation måste ett datalager byggas och hänsyn tas till många olika tekniker för bland annat datainsamling, datalagring och rapporteringsmöjligheter. Syftet med uppsatsen är att beskriva vad ett datalager är ur ett tekniskt perspektiv, beskriva hur data lagras samt hur ett datalager kan effektiviseras för snabb dataåtkomst och möjliggöra skalbar lagring av data. Ett datalager är en ämnesindelad databas med historisk, detaljerad och integrerad data, som inte ändras efter inladdning, med syftet att utgöra ett stöd för ledningsbeslut. Vi beskriver en arkitektur för datalager där den detaljerade datan är normaliserad enligt relationsmodellen. Utifrån den datan lagras summerad eller på annat sätt aggregerad data för avdelningsspecifika behov i vad som kallas data marts. Data i en data mart är inte normaliserad utan lagras enligt dimensionsmodellen som frångår normaliseringsprinciperna för att optimera prestanda. Företag har ofta behov av att göra många olika jämförelser av data från ett relativt litet antal tabeller vilket dimensionsmodellen är anpassad för. Kombinationen av historisk och detaljerad data leder till enormt stora datavolymer vars lagringsbehov i praktiken ofta ökar exponentiellt. För att minska databasens svarstider krävs utökad indexering och partitionering men även aggregerad redundant data används i stor utsträckning. En av våra slutsatser är att det är viktigt att ha en pragmatisk inställning till designfrågorna, att i första hand lösa organisationernas problem och i andra hand följa teoretiska designprinciper för att möjliggöra flexibel användning i framtiden.

# **Retrieval, Storage and Processing of Data for Decision Support in Large Organisations**

- A Comprehensive Study of Data Warehousing

## **Abstract**

In the 1960's the thought of collecting transactional data for decision support arose. To enable decision support in an organisation a data warehouse must be built taking into consideration a variety of techniques for data retrieval, storage and reporting. Our purpose with this dissertation is to describe a data warehouse from a technical perspective, how data is stored, how a data warehouse can be streamlined for fast data retrieval and how to enable scalable data storage. A data warehouse is a subject-oriented, integrated, non volatile, time-variant data store supportive of managerial decisions. We describe an architecture for a data warehouse where the detailed data is normalised according to the relational model. The warehouse data is summarized or aggregated in a fashion designed to accommodate departmental needs and stored in what is called a data mart. Data in a data mart is not normalised but is stored according to the dimensional model which abandons the principles of normalisation in favour of optimized performance. Organisations often need to compare data from multiple tables for which the dimensional model is well suited. The combination of historical and detailed data leads to enormous volume requirements which, in practice, often increases the need for storage exponentially. This demands heavy use of indexing and partitioning of the database. To decrease response time aggregated redundant data is used to a great extent. One of our conclusions is the importance of a pragmatic approach to data warehouse design. The main priority must be to solve the organisation's problems. Following theoretical design principles is of secondary concern.