Navigation

2 Descriptive Analytics

2.1 Data Warehouse Systems

2.2 Online Analytical Processing (OLAP)

Application Systems Pyramid

Typical Descriptive Questions

Source: Alfred Schlaucher, Oracle, p. 15

Frankfurt School Logo

Characteristics of Operational and Analytical Databases

Characteristics	Operational Data	Analytical Data
Target group	Operational employees	Management, Analysts, Divisional manager
Access frequency	high	slow to medium
Data volume per access	low to medium volume	high volume
Required response time	very short	short to medium
Level of data	detailed	aggregated, processed
Queries	predictable, periodic	unpredictable, ad hoc
Given period	current	past to future
Time horizon	1–3 months	several years up to decades

Contradiction between Operational Data Architecture and Analytical Information Needs

Source: Jung/Winter (2000): Data Warehousing Strategie: Erfahrungen, Methoden, Visionen, Berlin, p. 7.

Different Intensities of Use

The hardware load differs in operational application systems and data warehouse systems over time

=> A data warehouse relieves the operational systems

Source: Inmon: Building the Data Warehouse, 2002, p. 23.

Frankfurt School Logo

Definition: Data Warehouse

The term Data Warehouse was coined by Bill Inmon in 1990, which he defined in the following way:

"A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process."

He defined the terms in the sentence as follows:

Subject Oriented: Data that gives information about

The Data Warehouse Concept

Frankfurt School Logo

Frankfurt School Logo

Layers of a Data Warehouse System

Components of a Data Warehouse

Frankfurt School Logo

Frankfurt School Logo

Structure of the Database

The database represents the core of the data warehouse. It contains current and historical data from all relevant areas of the company. In contrast to the operational databases, the data here is not structured according to the business processes and functions of the company, but is aligned with the facts that determine the company. They are organized in such a way that they can be viewed from different dimensions.

Frequently used dimensions are:

Frankfurt School Logo

Examples of Aggregation Levels in a Data Warehouse

Frankfurt School Logo

ETL-Tools

The data in the data warehouse can normally only be accessed in a read-only manner. The only exception to this are ETL tools, which are responsible for transferring data from internal and external information sources. They automate the process of data acquisition from these sources.

The timeliness with which the transfer to the database is performed can be defined flexibly, e.g., within defined time intervals, immediately after each change,

Staging Area

The staging area is a temporary cache that performs the ETL process. It includes the extraction, transformation and loading of the data into the data warehouse.

Neither the operational systems nor the data warehouse are affected by the process. The data can be transferred incrementally from the operational systems without having to

Frankfurt School Logo

The Metadata Repository

Metadata supports data management and serves as a descriptor for an object that holds some data or information. In data warehouses, it is collectively organized in a catalog called the metadata repository.

Metadata is classified into three types:

- **Technical metadata:** Provides information about the structure of data, where it resides, and other

Data Warehouse and Data Mart

A single-subject data warehouse is typically referred to as a Data Mart, while Data Warehouses are generally enterprise in scope.

Comparison: Data Warehouse and Data Mart

	Data Mart	Data Warehouse
Application	Division/Department	Company
Data design	Optimized	Generalistic
Data volume	From gigabyte range	From terabyte range
Origin	Task oriented	Data model oriented
Requirements	Specific	Versatile

Architectural Variants of Data Warehouses in Practice

Frankfurt School Logo

Frankfurt School Logo

From Transactions to Reports

Example of a Dashboard

Source: https://www.bluemargin.com/power-bi-full-blog/topic/private-equity-data-intelligence

Star Schema (I)

The Star schema has a central fact table and exactly one dimension table for each dimension.

Source: Müller Lenz (2013), p. 60f

The cardinality between the fact table and a dimension table is n:1, i.e. a fact (e.g. sales: 67,000 euros) is described by one dimension expression for each dimension (e.g. time: 1st quarter=, region: Hesse, product group: books). A dimension characteristic (e.g. product group: books) is associated with 0, 1 or n facts.

Star Schema (II)

Frankfurt School Logo

Star Schema (III)

Star Schema Table Design (I)

Frankfurt School Logo

Star Schema Table Design (II)

Frankfurt School Logo

Source: https://www.guru99.com/star-snowflake-data-warehousing.html

Star Schema Example

Frankfurt School Logo

Exercise 1

An online wine retailer plans to design a data warehouse to collect key figures regarding its wine sales. The relevant part of the operational database consists of the following tables:

- **CUSTOMER** (ID, name, address, telephone, birthday, gender)
- **WINE** (ID, name, type, year, bottle_price, class)
- **CLASS** (ID, name, region)

Frankfurt School Logo

Queries on the Star Schema

Example:

In which years were the most cars purchased by female customers in Hesse in the 1st quarter?

select z.Year as Year, sum(v.Amount) as Total_Amount
from Branch f, Product p, Time z, Customer k, Sales v
where z.Quarter = 1 and k.gender = ‘f’ and p.Product_type = ‘Car’ and f.state = ‘Hesse’ and v.Date = z.Date and v.ProductID = p.ProductID and v.Branch = f.Branchname and v.CustomerID = k.CustomerID

Frankfurt School Logo

Snowflake Schema

In the Snowflake schema, the redundancies in the Star schema are resolved in the dimension tables and the hierarchy levels are each modeled with their own tables:

Source: https://www.guru99.com/star-snowflake-data-warehousing.html

Frankfurt School Logo

Star Schema vs. Snowflake Schema

– Speed of query processing is in favor of Star scheme

– Data volume tends to be lower with Snowflake schema

– Snowflake schema is easier to change (maintenance)

Frankfurt School Logo

Galaxy Schema

The Galaxy schema contains more than one fact table. The fact tables share some but not all dimensions. The Galaxy schema thus represents more than one data cube (multi-cubes)

Table Structure of the Galaxy Schema

Frankfurt School Logo

DW Planning – User Questions (I)

- What is the ratio of private to corporate customers?
- Does this quantity ratio correspond to the turnover ratio?
- Do customer characteristics such as occupational group, living situation, education, etc. play a role in turnover?
- What are the customers with the highest sales?
- What is the development of sales over time? Are there any shifts in relation to customer characteristics?
- Can additional customer segments be formed, e.g. if customer characteristics are combined?

DW Planning – User Questions (II)

Frankfurt School Logo

Source: Alfred Schlaucher, Oracle, p. 28

Frankfurt School Logo

Result: Star Schema

Source: Alfred Schlaucher, Oracle, p. 40

Exercise 2

Frankfurt School Logo

A restaurant chain wants to build a management information system. The application system architecture is to be aligned with the data warehouse concept. An OLAP system is chosen for report generation.

The company maintains various restaurants, which can be differentiated by region and by renovation status. The Leonardo-Campus restaurant belongs to the North region and is New from the renovation status. The restaurant Nordblick is also located in the North region, but its renovation status is Old. The Spätzleburg restaurant, on the other hand, is in the

Exercise 3

As an employee of the Controlling department, you will be given the task of setting up a data warehouse for sales and headcount analyses.

Flowers GmbH has two branches in Hesse, one in Rhineland-Palatinate and three in France. Product categories include cut flowers, garden flowers and bridal jewelry. Within garden flowers, the rose family includes the genera roses and wild roses. Products in the rose genus include thornless roses and roses with thorns. The company distinguishes between corporate and private

Navigation

Subject of OLAP

Online Analytical Processing (OLAP) is software technology focused on the analysis of dynamic, multidimensional data.

The goal is to provide executives with the information they need quickly, flexibly and interactively.

OLAP functionality can be characterized as dynamic, multidimensional analysis of consolidated enterprise data, which includes the following components:

- **Calculations of key figures over different dimensions and aggregation levels**

Facts and Dimensions

Frankfurt School Logo

Organizing Data as Hypercube

OLAP Operations: Slicing and Dicing

Frankfurt School Logo

OLAP Operations: Rotate

Frankfurt School Logo

OLAP Operations: Drill Down and Roll Up

Frankfurt School Logo

OLAP Operations: Nest

The result of a nest operation physically represents a two-dimensional matrix. It is extended by displaying different hierarchy levels of one or more dimensions on one axis (column or row) in nested form. In the following example, three dimensions are displayed:

OLAP as Excel plug-in

Architecture of an OLAP System

Frankfurt School Logo