www.juergen-konicek.de

OnLine Analytical Processing (OLAP) and Reports

The BI deployment provides the main elements DWH and data marts for collecting and storing data. But the OLAP is the key technology of BI and it is used for business analysis and decision support.

Abbrevations:

APIs: Application Programming Interfaces

BI: Business Intelligence

CASE: Computer Aided Software-Engineering

CDIF: Case Data Interchange Format

CLI: Call Level Interface

CWH: Common Warehouse Model

CORBA: Common Object Request Broker Archtiecture

DBMS: Database Management System

DB2 UDB: DB2 Universal Database

DC: Direct Costs

DDL: Data Definition Language

DOM: Document Object Model

DTD: Document Type Definition

DWH: Data Warehouse

ECMA: European Computer Manufacture's Association

EIA: Electronic Industries Associations

ERM: Entity-Relationship Model

ETL: Extraction Transformation Loading

ERP: Enterprise Resource Planning

GIF: Graphics Interchange Format

IDL: Interface Definition Language

IRDS: Information Resource Dictionary System

ISO: International Organization for Standardization

According to [OMG02] the term OLAP is defined as follows:

"*OnLine Analytical Processing (OLAP) is an analysis technique in which business data originating from multiple, diverse operational sources is consolidated and exposed in a multidimensional format that allows business analysts to explore it a retrieval-friendly environment*" [OMG02, page 122].

With OLAP it is possible to aggregate multiple table queries faster than with normalised entities in an operational database [CBS02, page 943]. Based on these facts, a dimensional database is a better model for querying, but worse for operational use [WEnc04, OLAP].OLAP structured data enable both growth and trend analysis. The multidimensional format of OLAP data are very often represented by data cubes as displayed in figure 1. Such a data cube is divided into dimensions and measures appropriate to the DWH or data mart schemas in 2.2. In general, dimensions describes an analysis view of an application area that can be defined [BuG01, page 517]. In contrast, measures are the content (cell) of a data cube [BuG-01, page 517] in dependency on a chosen analysis view (dimension). The dimensions are represented by every edge of the cube such as item, person and date in figure 1. A set of dimension values identify only one unique measure of a cube cell. For instance the dimension values **person1001**, **item1277** and **05/12/2000** describe the measure actual revenue **342,82** in the corresponding cell as shown in figure 1.

MOF: Meta Object Facility

mUML: Multidimensional UML

MS: Microsoft

OC: Overhead Costs

OCL: Object Constraint Language

ODBC: Open Database Connectivity

ODBMS: Object-Orientated Database Management System

OLAP: Online Analytical Processing

OMS: Object Management System

OOD: object-oriented design

OMG: Object Management Group

PCTE: Portable Common Tool Enviroment

RPCS: Remote Prcedure Calls

SQL: Structured Query Language

SVG: Scalable Vector Graphics

UML: Unified Modeling Language

UDFs: User Defined Functions

WWW: World Wide Web

W3C: World Wide Web Consortium

XMI: XML Meta Data Intechange

XML: Extensible Markup Language

References:

[BuG01]
A. Bauer and H. Günzel: Data-Warehouse-Systeme - Architektur, Entwicklung, Anwendung. 1. Aufl. - Heidelberg; dpunkt-Verlag (2001).

[CBS02]
T. Connolly, C. Begg and A. Strachan: Datenbanksysteme; Eine praktische Anleitung zu Design, Implementierung und Management. Addison-Wesley (2002).

[OMG02]
J. Poole, D. Chang, D. Tolbert and D. Mellor: Common Warehouse Metamodel- An Introduction to the Standard for Data Warehouse Integration. published by John Wiley & Son, Inc., New York, (2002).

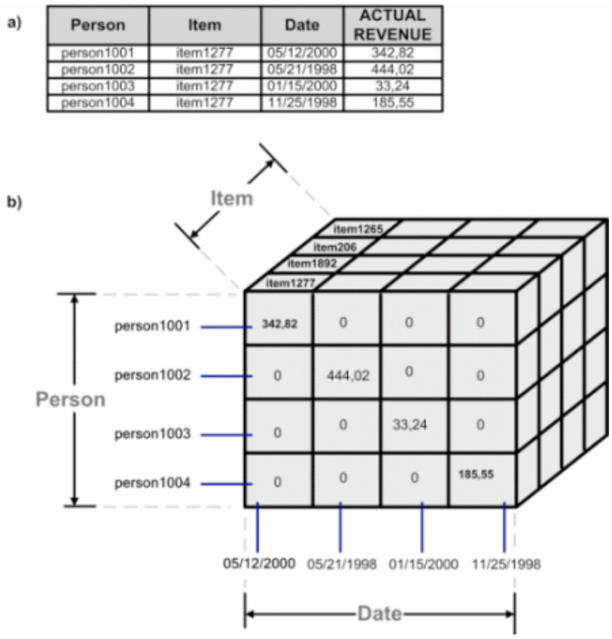

Figure 1: Multidimensional data representation in a table a) with four arrays in comparison to a OLAP cube b) (adapted from [CBS02, page 945]).

Referring to the displayed OLAP cube in figure 1, it is possible to apply essential navigation operations for ad hoc analysis:

• **Rotation/piovoting** to new dimensional comparisons in the viewing area.

• **Slice and dice** to view different dimensions such as actual revenue per person by item.

• **Rollup** generates a result set showing aggregates for a hierarchy of values.

• **Drill down/drill across** enable to drill through from one level of detail to another.

[Kon04]
Jürgen Konicek: Metadata within Business Intelligence Solutions, Thesis - University of Applied Science of Ulm (Commercial Information Systems) in cooperation with the Napier University, Edinburgh (UK) (School of Computing) TERM 2003/04 (2004).

[WEnc04]
Wikipedia, the free encyclopedi.URL: http://en.wikipedia.org/wiki/Main Page [5. April 2004]

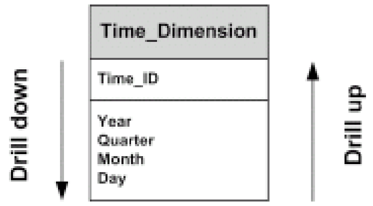

Figure 2: Example for a dimension hierarchy and navigation along a dimension table.

For a general description of a data cube in particular with regard to the data cube in figure 1, a formal notation is introduced in the following. According to the notation in [BuG01], the data cube in figure 4 can be stated here in the form given in [BuG01, page 222-223]:

• In general a dimension *D* with m dimension values can be defined formally as a ***m*** tuple (a sorted list of ***m*** values):

$$D = \{x^{D}_{1}, x^{D}_{2}, \ldots, x^{D}_{m}\}$$

• Consequently, $|D| = m$ defines the number of dimension values of the dimension D. For instance the *Time_Dimension* (cf. figure 2) with five dimension values (*Time ID, Year, Quarter, Month* and *Day*) can be generally characterised as $|D_{Time\ Dimension}| = 5$.

• Furthermore, a (data) cube *C* can be specified in the following notation:

$$C = ((D_1, D_2, \ldots, D_n), (M_1 : Typ_1, \ldots, M_m : Typ_m))$$

where:

• D_i represents the different dimensions,

• M_i : Typ_i describes the characteristic of the cube along with its data type.

Based on this notation, the data cube in figure 1 can be generally defined as follows:

$C_{(Revenue)}$= ((Person, Item, Date), (Revenue : real))

By Jürgen Konicek