Udskrevet er dokumentet ikke dokumentstyret.

Datawarehouse architecture

Dokumentbrugere:	Dokumentansvarlig:	Redaktør:	Dokumentnummer:	Godkendt af:	Niveau:
KONC-POEM-MT	**D-PØMQansvar**	Version: **1.1**	**12. 6. 1.01**	**NAFR2711, PAMA2610 13.02.2009**	

1) Overview
2) Datatypes
3) Naming conventions

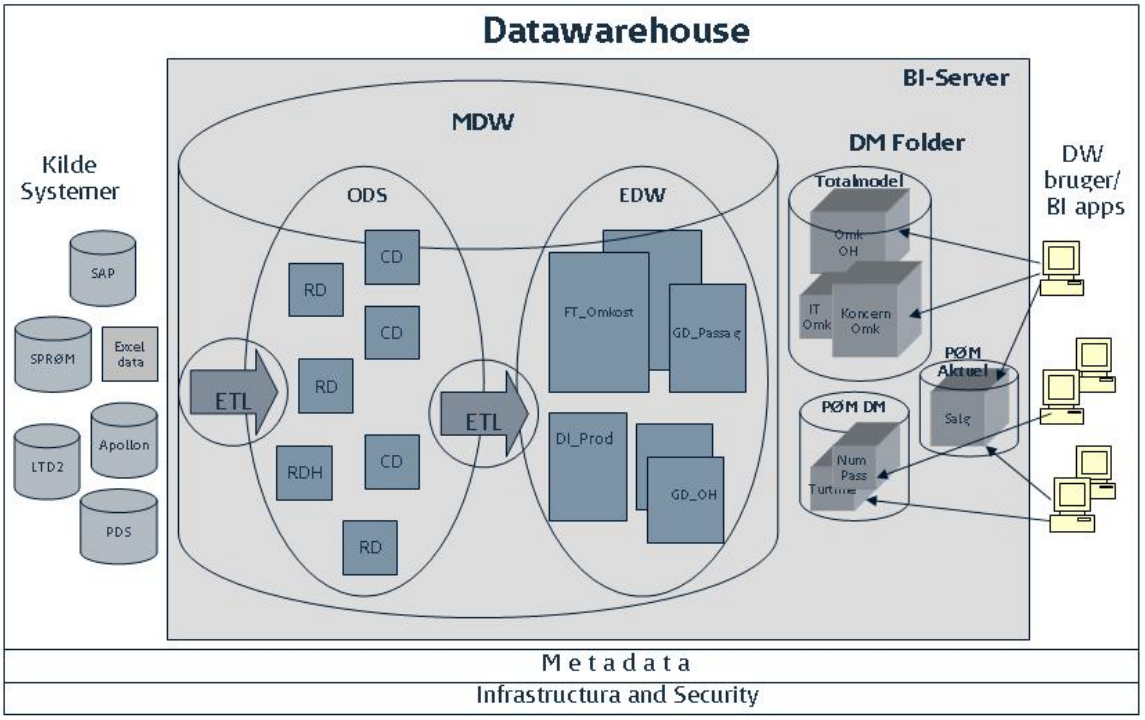
1) Overview

This part of the design refers to the datawarehouse database architecture. It shows how data is structured and organized in the databases of the DW.

There are 2 types of databases in the DW. One of the databases(MDW) contains "back-room" objects, such as the staging area, the extraction, transformation and loading objects, and the tables to keep enterprise datawarehouse data. In other words, objects utilized to prepare the front-room data of the DW.
The front-room data contains cubes accessed by the end-users. These are used for analysis, business intelligence and reporting and are stored in several different databases, each containing data on a well-defined businessarea. E.g one database stores cubes on Produktøkonomi-output and another database stores cubes on source-data.

Databases in the data warehouse:
1. MDW (back-room objects)
2. Severel DM-databases (front-room objects)

The following picture shows an overview of the DW architecture at a high level.



The circles in the picture represent the schemas in the database, and the rectangles or squares represent tables. The arrows represent processes, which move data from Data source systems to the ODS schema or from the ODS to the EDW schema. The ETL schema in the MDW database contains store procedures called by processes. The ETL schema also contains tables to log information about the ETL process.

The processes (packages) which move data between schemas are created in a separated "project file/files" generated by MS Integration Services, which is the tool utilized to build the ETL process.

2) Datatypes

In order to minimize problems with different datatypes in different tables we use a uniform set of datatypes throughout the datawarehouse. Notice however that GD-tables are using other datatypes than the rest of the datawarehouse. This is because the GD-tables are used in the SAS-environment where the ABC-model is being maintained and calculated. Therefore the datatypes in the GD-tables must comply with the standards defined in the SAS-environment.

For mapping datatypes from SQL-server, .Net and Integration Services you can refer to these pages on MSDN:
- SQL Server Data Types and Their .NET Framework Equivalents
- Mapping Data Types in the Data Flow

Datatypes used in the datawarehouse

SQL-server	IS	Comments
INT	INT32	Maxvalue: +/- 2.147.483.647
BIGINT	INT64	Used when INT is not big enough
DECIMAL(24,6)	Double	Can hold 18 characters before the decimalpoint and 6 after.
VARCHAR	String	
DATETIME	DATETIME	Format 'YYYY-MM-DD HH:MM:SS' where HH:MM:SS is optional
BIT	Boolean	Used for True/False
Float	Double	Used only for PØM-related tables

3) Naming conventions

Supplementing the naming conventions described per database and schema, there are other technical naming conventions that must be kept in mind when objects and structures are defined in the system.

Sourcesystems

Sourcesystems and sourcefiles are referenced with a short name. These are:

1. PDS
2. LTD2
3. FPC
4. SAP_ (Followed by a SAP sub source shortname. For example: SAP_BAL for SAP Balance)
5. ABM
6. SPR (Sprøm)
7. PRO (Protal)
8. RBS (ROSA, billetstatistik)
9. EX (Excel file. When a Excel file is a source, a short description of the file content must be part of the name, for example: Ex_BudIt to name an Excel file containing the budget of the IT area)

Tables

- The table prefixes are described in detail under the title: The MDW-database.
- The names of the tables written in Danish will not use the Danish vocals.
- Table names are plural, field names are singular.
- Table names should not contain spaces, and should be split_up_with_underscores. Limit the table name to 20 characters.
- If the table name contains serveral words, only the last one should be plural.
- Notice that the collation of a column always should be specified as <database default> - which is Danish_Norwegian_CI_AS.

Primary and Foreign Key Fields

- Colums that are primary keys should be prefixed with PK_. Colums that references a primary key should be prefixed with FK_ to denote it as a foreign key. It is recommended to name foreign keys similar to the primary key it references to. E.g. PK_Togsystemnummer are referenced from another table as FK_Togsystemnummer.
- Tables with compound PK's, use CK_ in place of PK_.

Stored procedures/scripts located in the Etl schema

- Named with a prefix that shows the scope of the stored procedure. E.g. ods_, edw_, grl_ (meaning General, this abreviation is used when the store procedure does not have a particular schema as scope). For more details on the etl-schema please refer to the relevant section in 'The MDW-database'-document.

Revisioner

Version	Godkendt	Revisions information
1	14.10.2008	
1.1	13.02.2009	