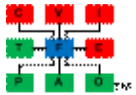

DATABASE TYPES

Introduction:

At the beginning of the database engineering phase of the data engineering process lifecycle the enterprise logical data model (ELDM), or a subset of it, is to be used to create a physical database for a business processing application.

The question that must be asked before the physical database implementation is begun is:

- What is the purpose of the database?

Is the purpose to capture and maintain data, or is the purpose to report or search the data and discover relationships between the values of the data?

Fortunately, the Relational Data Model (Codd) gives five different normal forms (Kent) for creating the physical database. The first normal form is best for searching and reporting the data. The second normal form is best for capturing data in a restricted environment such as banking ATM transactions. The third normal form is best of OLTP applications and the fifth normal form is best for enterprise data warehousing.

Each of the normal forms is a tradeoff between the amount of storage space used and the processing time required to update or report the database. Of the five normal forms, the 1st and 5th normal forms represent the extremes of the tradeoffs. The details of the tradeoffs are shown in Table 1.

database type	advantage	storage required	update processing time	report processing time
5th normal form	no update anomalies when changing data	minimum	minimum	maximum
1st normal form	no join processing for reporting data	maximum	maximum	minimum

Table 1: Database Type Tradeoffs

To make data capture as efficient as possible, the database should be in 5th normal form. Implementing the database as a completely normalized structure minimizes the storage space required, minimizes the processing time to update the database (at a single location) and maximizes the application reporting time

To make data reporting as efficient as possible, the database should be in 1st normal form. Implementing the database as a 1st normal form table minimizes the application reporting time, maximizes the storage space required and maximizes the time required to update the database (at multiple locations).

Assumption:

All the different types of database implementations use the same logical data model. The database implementation has been chosen to enhance performance and ease of use for a single purpose.

Database Types:

When looking at the Data Lifecycle, each stage has a different use and purpose and each purpose should be using a different database type.

In the Data Creation / Sourcing stage, the purpose is to capture the data as it is being created.

In the Consolidation stage, the purpose is to reconcile the data from the many data sourcing applications.

In the Consumption stage, the purpose is to report the data from the results of the data consolidation.

In summary, each data lifecycle stage has a different purpose with different interface requirements because of the tools used and data governance compliance requirements.

data life cycle stage (data architecture)	purpose	database interface	data governance compliance	comment
data creation / sourcing	real time data capture	database stored procedures	internal to database (in stored procedure)	subset of ELDM database for an OLTP application
consolidation	data reconciliation	command line from ETL tool	external to database (in ETL process)	implemented ELDM
consumption	data search and reporting	command line from OLAP tool	external to database (in data mart loading process)	Subset of ELDM database for an OLAP application Essentially a materialized view on the ELDM

Table 2: Data Lifecycle Database Type Summary

REFERENCES:

Codd, Edgar, “A relational model for large shared databanks”, Communications of the ACM, Vol. 13, No. 6, Jun 1970

Kent, William, “A Simple Guide to Five Normal Forms in Relational Database Theory”, http://www.bkent.net/Doc/simple5.htm, Sep 1982