

Chapter 4 – Databases and Data Warehouses

Organizations do three things with information:

1. Process information in the form of transactions – **Online transaction processing (OLTP)** involves gathering input information, processing that information, and updating existing information to reflect the gathered and processed information. Databases that support OLTP are most often referred to as operational databases
2. Use information to make decisions – **Online Analytical processing (OLAP)** is the manipulation of information to support decision-making. Built on the concept of data warehousing that contains information specifically for supporting decision-making tasks.
3. Manage information while using it – Organizations must manage information at the same time its knowledge workers are using it. It includes a variety of tasks such as who can view or use the information, how to back it up, how long to retain the information, and what storage technology to use. Choosing the appropriate technology to organize information so that knowledge workers can logically use it without having to know anything about its physical organization.

Physical – Technology terminal, storage media.

Logical – **Character** (smallest logical unit of information), **Field** (logical grouping of characters... person's name, product #, price), **Record** (logical grouping of fields), **File** (all logical associated records). When you bring together logical associated files, you create a database. **Data warehousing** is information from several databases and are central to today's logical view of information

What is a database?

A **database** is a collection of information that you organize and access according to the logical structure of that information.

Important characteristics of a database

1. A collection of information
2. A logical structure – A **data dictionary** contains the logical structure of information in a database. When you create a database, you first create a data dictionary. DD's contain important information or logical properties about your information.
3. Logical ties among the information – shows how files are related to each other. A **primary key** is a field in a database file that uniquely describes each other. A foreign key is a primary key of one file that also appears in another file.
4. Built-in Integrity constraints – rules that help ensure the quality of the information.

Database Management System – is software you use to specify the logical organization for a database and access it. A DBMS contains five important software components

1. DBMS engine – Providing the logical to physical bridge. Accepts logical requests from the various other dbms subsystems, converts them to their physical equivalent, and actually accesses the database and data dictionary, as they exist on a storage device. The **physical view** of information deals with how information is physically arranged, stored, and accessed on some type of secondary storage device. The **Logical View** focuses on how you as a knowledge worker need to arrange and access information to meet your particular business needs.

There are two advantages of separating the logical and the physical. First, the dbms handles all physical tasks, so you, as a database user, can concentrate on the logical information needs. Second, there is only one physical view of information, but there may be numerous knowledge workers who have different logical views.

2. Data Definition subsystem – helps you create and maintain the data dictionary and define the structure of the files in the database. (Often called the data definition language). Most important function is that it supports your ability to define the logical structure or properties of the information when you first create a database.
3. Data Manipulation subsystem – helps you add, change, delete information in a database and mine it for valuable information
 - View – allow you to see the content of a database file, make whatever changes you want, perform simple sorting, and query to find the location of specific information.
 - Report Generators – help you quickly define formats and what information you want to see in a report
 - Query by example tools – help you graphically design the answer to a question.
 - Structured query Language – SQL – Perform the query by creating a statement instead of clicking or dragging (sentenced-based rather than graphics-based). The basic form of a SQL statement is SELECT...FROM...WHERE.
4. Application Generation subsystem – Contains facilities to help you develop transaction-intensive applications. (Forms, entry tables)
5. Data Administration subsystem – Helps you manage the overall database environment by providing facilities for backup and recovery, security management, query optimization, reorganization, concurrency control and change management. Used by data administrators.
 - ⇒ Backup and recovery facilities
 - ⇒ Security management facilities – allows you to control who has access to what information
 - ⇒ Query optimization facilities – often takes queries from users and restructure them to minimize response time.

- ⇒ Reorganization facilities – maintains statistics concerning how the DBMS engine physically accesses information.
- ⇒ Concurrency control facilities – ensures validity of database updates when multiple users attempt to access and change the same information.
- ⇒ Change management facilities – allow you to access the impact of proposed structural changes.

Relational database Models – There are four database models in use today – The hierarchical, network, relational (most used), and object-oriented databases

Relational database model – uses a series of two dimensional tables or files to store information. Each table or relation stores information pertaining to a particular entity class. **Entity class** is a concept – typically people, place, or thing about which you wish to store information. An **instance** is an occurrence of an entity class that can be uniquely described

Composite primary key – primary key that uses more than one field to create a unique description

Atomic primary key – a primary key that uses only one field.

Object-Oriented Database Model

An **object** is software module containing information that describes an entity class along with a list of procedures that can act on the information describing the entity class.

OODM is a database model that brings together, stores, and allows you to work with both information and procedures that act on the information.

OODBMS is the DBMS software that allows you to develop and work with an object-oriented database.

Combining information and procedures more closely models how an organization works and the second key feature is that of reuse. Reuse simply means that, once you define a set of procedures for a given object, those procedures can also be used by other objects.

Knowledge worker, Computing and developing your own database

1. Define entity classes and primary keys
2. Defining relationships among entity classes – An entity relationship diagram is a graphic method of representing entity classes and their relationships. See p.148. **Normalization** is a process of ensuring that a rational database structure can be implemented as a series of two-dimensional tables (3 rules)
 - ⇒ Eliminate repeating groups or M:M relationships

- ⇒ Assure that each field in a relation depends only on the primary key of that relation
- ⇒ Remove all derived fields from the relations.

Intersection relation is a relation you create to eliminate a repeating group.

3. Defining information (fields) for each relation
4. Using a data definition language to create the database – take the structure you created in steps 1 to 3 and use a data definition language to actually create the relations. In this step you create a data dictionary.

Data Warehousing and Data Mining

Data warehouse – is a logical collection of information – gathered from many different operational databases – that supports business analysis activities and decision-making tasks.

Three key features:

1. Data warehouses combine information from different databases
2. Data warehouses are multidimensional. (Layers of columns and rows)
3. Data warehouses support decision-making, not transaction processing.

Data mining tools – are the software tools you use to query information in a data warehouse. They support the concept of olap.

- ⇒ Query-and-reporting tools – similar to QBE & SQL tools. Generate simple queries and reports
- ⇒ Intelligent agents – utilize various AI tools to form the basis for “information discovery”
- ⇒ Multidimensional Analysis Tools – slice and dice techniques that allow you to view multidimensional information from different perspectives.

Data warehousing is not necessarily the best technology for all businesses

1. Very expensive
2. Some organizations don't need data warehouses and data mining tools
3. Many IT departments suffer from supporting too many applications and application tools.

Managing the information Resource in an Organization

How will changes in technology affect organizing and managing information?

Changes in technology occur every day. The key is to decide whether those changes provide better ways of organizing and managing information. If they do, seriously consider them. If they don't, wait for one that will

What types of databases Models and Databases are most appropriate?

The two popular database models include the relational database model and the object-oriented database model. By type, databases include **centralized databases** that maintain information in one location, **distributed databases** that allocate information to different places, **external** or **online databases** that rest outside the organization, **knowledge worker databases** that support personal information needs, **operational databases** that support online transaction processing, and data warehouses that support online analytical processing. The choice is important

Who should oversee the organization's information?

Today, the responsibility of an organization's information rests with the **chief information officer** (CIO), who is responsible for **data administration** (planning, overseeing the development of, and monitoring the information resources) and **database administration** (the technical and operational aspects of managing information contained in organizational databases)

Is information ownership a consideration?

Information ownership deals with who is responsible for providing specific information and assuring the quality of that information. Ultimately, information ownership must rest on someone.

What are the ethics involved on Organizing and Managing Information?

Internally, organizations must still protect the privacy and information rights of their employees. We must all do our part to act ethically regarding the use of information.

How should database applications be developed and maintained?

Database and database applications can be developed and maintained in a variety of ways including knowledge worker development, development and maintenance from an organizational point of view, and outsourcing. The choice is a key one.