

CSCI-UA.60-1

Database Design

and Implementation

Database Management

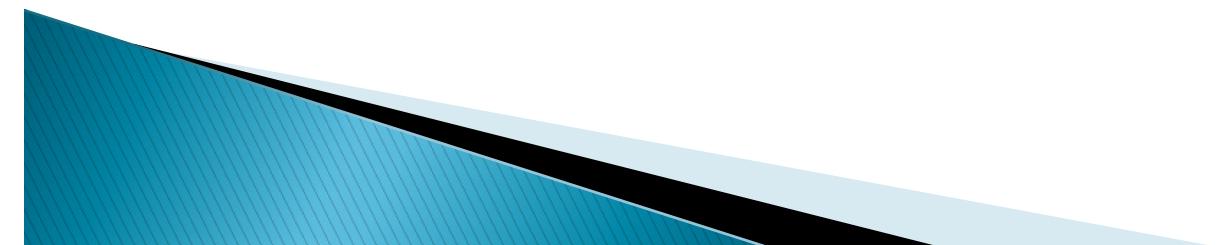
CSCI-UA.60-1
Deena Engel
deena.engel@nyu.edu

Database Administration: Tasks

- ▶ Monitoring and maintaining the database system
- ▶ Managing database security
- ▶ Providing/implementing database backup and recovery
- ▶ Optimizing performance
- ▶ Providing/maintaining metadata (data about the system; documentation)
- ▶ Managing the data warehouse

Managing the Database Structures:

- ▶ Configuration Control
 - Includes changes to the structure of the tables, indexes, relationships
 - ... and must be done without jeopardizing the integrity of the data!!
- ▶ Documentation
 - CASE [computer-aided software engineering] tools
 - Required to maintain and to provide for using historical data



Processing Rights and Access Privileges

- ▶ Different rights are generally assigned to specific users; to user groups; and to administrators
- ▶ The rights can be categorized as follows:
 - Insert queries – allowing a user to add data
 - Select queries allowing a user to run reports and do analysis on the data
 - Update queries – allowing a user to modify and/or delete data
 - Alter queries – allowing a user to modify the database/table structures

Processing Rights (permissions) can be implemented on many levels:

- ▶ Network / OS / Server level
- ▶ DBMS level (access to the DBMS)
- ▶ Application level
 - Users, user groups, and administrators can be further restricted to specific modules in a specific application and/or specific tables and files as well as restricted to specific queries within a given set of tables or files

Processing Rights: Art Gallery

Data Staff	Customers	Sales	Paintings	Artists
Sales Staff	Insert query Update/Delete Select query	Insert query Update/Delete Select query	Select query	Select query
Curator	Select query	Select query	Insert query Update/Delete Select query	Insert query Update/Delete Select query
Gallery Owner	Insert query Update/Delete Select query	Insert query Update/Delete Select query	Insert query Update/Delete Select query	Insert query Update/Delete Select query
DBMS Admin.	Modify structures; Grant rights	Modify structures; Grant rights	Modify structures; Grant rights	Modify structures; Grant rights

What are the ACID properties?

These are important transaction properties in a relational database management system (RDBMS):

Atomicity – *Each transaction is treated as a ‘unit’. For example, in the event of failure, all of the steps can be rolled back.*

Consistency – *Each transaction brings the database from one valid state to another.*

Isolation – *This ensures that actions that happen concurrently result in the same state as if they were handled sequentially.*

Durability – *Once an action has been committed, it will remain even if the system fails thereafter.*

Database Recovery

- ▶ When the Computer system fails ...
 - Checkpoints: Keep logs of all transactions that occur after a given checkpoint in time so that transactions can be repeated if necessary
 - Keep logs of all backups and contents
 - Rollback: remove all elements of incomplete LUWs



Database Security

- ▶ Run the DBMS behind a firewall!!
- ▶ Apply *all* patches to the OS and DBMS
- ▶ Use the least functionality possible
 - Disable all unnecessary access and network protocols; restrict user access; disable guest users, etc
- ▶ Protect the computer itself (lock it up!!!)
- ▶ Manage accounts and user groups carefully
 - E.g. forbid null passwords
- ▶ Plan for security emergencies



Data Repositories

- ▶ Metadata
 - data about the databases, applications, users, and other application components
 - data dictionaries provide comprehensive information about the tables (including the fields names, data types, default values or other important information and definitions of the field in case the name is ambiguous in any way); indexes (including the specific fields used in compound index keys) and relationships
 - Database policies and standards such as naming conventions should also be documented
- ▶ Version control software
- ▶ Code libraries
- ▶ Active repositories: use tools that automatically update documentation as the systems evolve

Data Warehousing

- ▶ Large DBMS systems typically contain two categories of data:
 - operational information (transactions)
 - these are typically shorter queries using smaller amounts of information but are run very frequently
 - the data audience would include customers and employees in many roles
 - analytical information (for decision support)
 - these are typically long queries run against large amounts of information but they are not typically as time-sensitive as operational queries
 - the data audience would be a more narrow range of employees responsible for data analysis and decision support



Data Warehouse

- ▶ Since it is generally difficult if not impossible to simultaneously optimize a database for both operational (transaction) and analytical queries, a data warehouse is one standard solution to this problem.
 - A Data warehouse is an electronic repository of data for a given organization.
 - Another goal is to separate out old data which is no longer needed on a daily basis, but to keep it available for research purposes.
 - The cost for maintaining “old” data can be extremely high. Every organization must decide how and when to make use of this kind of tool.

Data Warehouse: Data

- ▶ Data in a data warehouse are not necessarily normalized. In some situations, it is more cost-effective to de-normalize copies of tables so that non-programmers can search them more easily. Furthermore, many companies store tables with summary statistics and other “abbreviated” (read: “aggregated”) versions of the data in a digital warehouse environment for easier searching and reporting.
- ▶ Terms and descriptions you might hear used to describe a **data warehouse**:
 - structured repository
 - subject-oriented repository
 - historical data repository
 - enterprise-wide data for the purpose of analytical study.