CSCI 466/566

Introduction to Database Concepts

Jon Lehuta

Northern Illinois University

January 15, 2018

Northern Illinois University

Introduction to Database Concepts

- What is a Database?
- Database Management System (DBMS)
- Other Capabilities of DBMS Systems
- Leveled Architecture of a DBMS
- Basic Database Terminology

What is a Database?

- A collection of stored operational data used by the application systems of some particular enterprise.
- From the book - "a collection of related data"

Enterprise?

Enterprise

- a generic term for any reasonably large-scale commercial, scientific, technical, or other application
  - manufacturing
  - financial
  - medical
  - university
  - government

Operational Data

Operational data is

- Data maintained about the operation of an enterprise
  - products
  - accounts
  - patients
  - students
  - plans
- Notice that this **DOES NOT** include input/output data

Database Management System (DBMS)

A Database Management System (DBMS) is

- a collection of programs that enables users to create and maintain a database
- a general-purpose software system that facilitates
  - definition of databases
  - construction of databases
  - manipulation of data within a database
  - sharing of data between users/applications

Defining a Database

For the data being stored in the database, defining the database specifies

- the data types
- the structure(s)
- the constraints

Constructing a Database

- The process of storing the data itself on some storage device
- The storage device is controlled by the DBMS

Manipulating a Database

Includes functions that

- retrieve specific information in a query
- update the database to include changes
- generate reports from the data

Sharing a Database

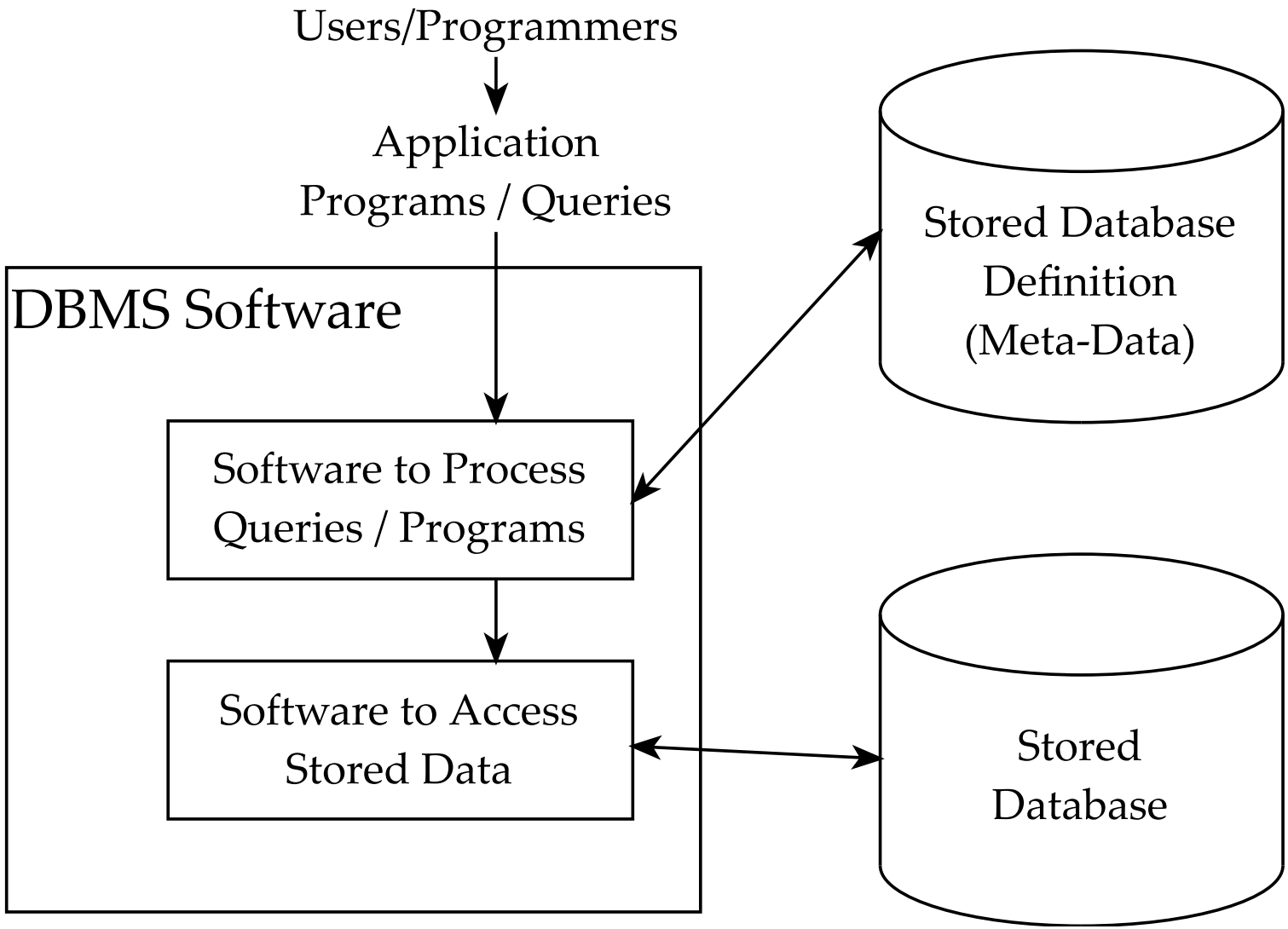
- Allows multiple users and programs to access the database at the same time
- Any conflicts between applications are handled by the DBMS

Other Important Functions of a Database

Other important functions provided by a DBMS include

- Protection
  - System protection
  - Security protection
- Maintenance
  - Allows updates to be performed easily

Simplified Database System Environment



What is a Database System?

Main characteristics of a database system are:

- Self-describing nature of a database system
- Insulation between programs and data, and data abstraction
- Support for multiple views of the data
- Sharing of data and multi-user transaction processing

Other Capabilities of DBMS Systems

Support for at least one data model through which the user can view the data

- There is at least one abstract model of data that allows the user to see the "information" in the database
- Relational, hierarchical, network, inverted list, or object-oriented

Other Capabilities of DBMS Systems

Support for at least one data model through which the user can view the data

- efficient file access which allows us to "find the boss of Susie Jones"
- allows us to "navigate" within the data
- allows us to combine values in 2 or more databases to obtain "information"

Other Capabilities of DBMS Systems

Support for high-level languages that allow the user to define the structure of the data, access that data, and manipulate it

- Data Definition Language (DDL)
- Data Manipulation Language (DML)
- Data Control Language (DCL)
- query language access data
- operations such as add, delete, and replace

Transaction Management

Transaction management is a feature that provides correct, concurrent access to the database, possibly by many users at the same time

- ability to simultaneously manage large numbers of *transactions*
  - procedures operating on the database
  - often transactions come from around the world
  - "lock-out" mechanisms

Access Control

Access control is the ability to limit access to data by unauthorized users along with the capability to check the validity of the data

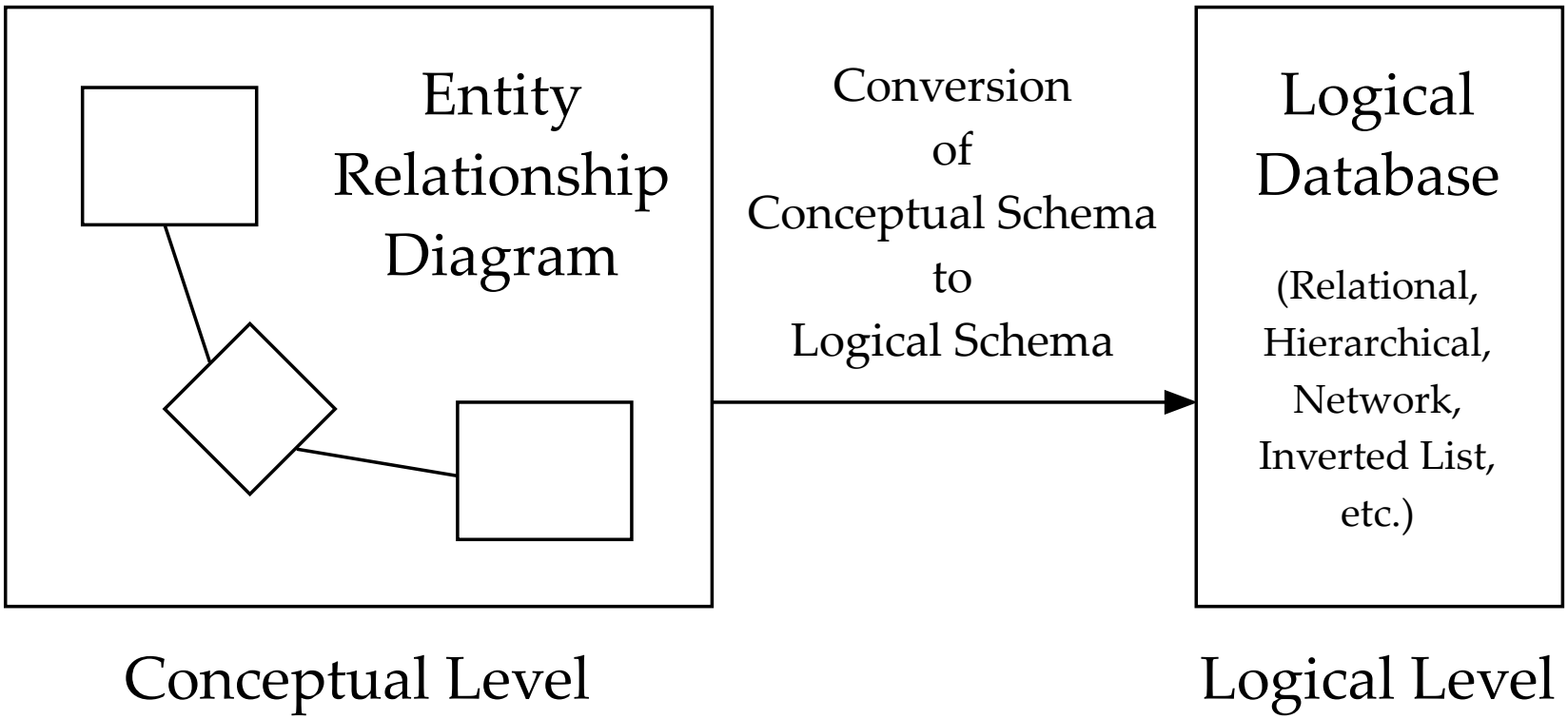
- protect against loss when database crashes
- prevent unauthorized access to portions of the data

Resiliency

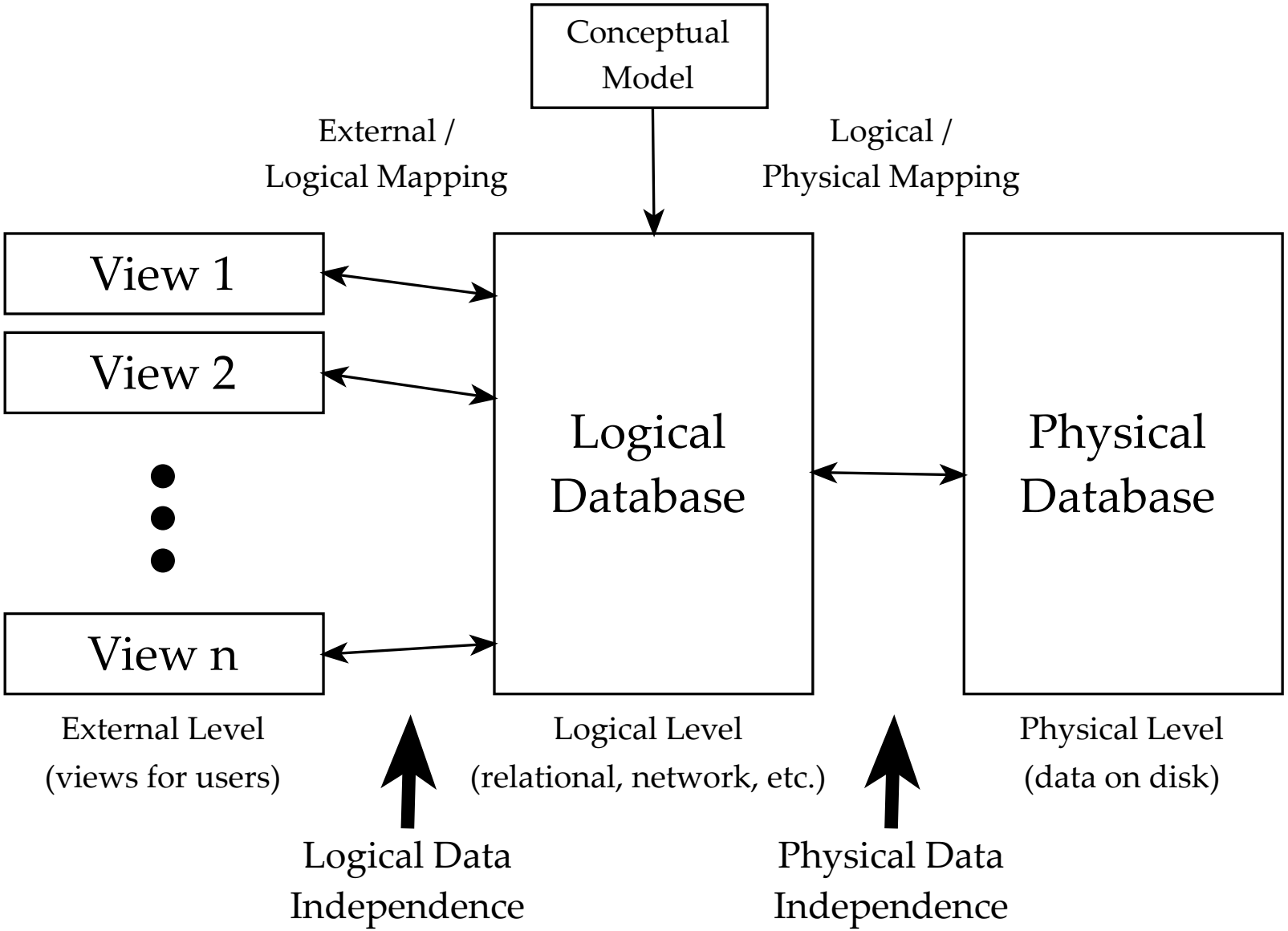
Resiliency is the ability to recover from system failures without losing data

- Ideally, should be able to recover from **any** type of failure.
  - sabotage
  - acts of God
  - hardware failure
  - software failure
  - etc.
- Obviously, some of these would require more than just software - offsite backups, etc.

Use of Conceptual Modeling



Leveled Architecture of a DBMS



External Level

a view or sub-schema

- portion of the logical database

may be in a higher level language

Logical Level

- abstraction of the real world as it pertains to the users of the database
- DBMS provides a data definition language (DDL) to describe the logical schema in terms of a specific data model such as relational, hierarchical, network, inverted list, etc.

Physical Level

- the collection of files and indices
- stored on secondary storage device (HDD, SSD, etc.)
- this is the actual data

Instance

An instance of the database is the actual contents of the data

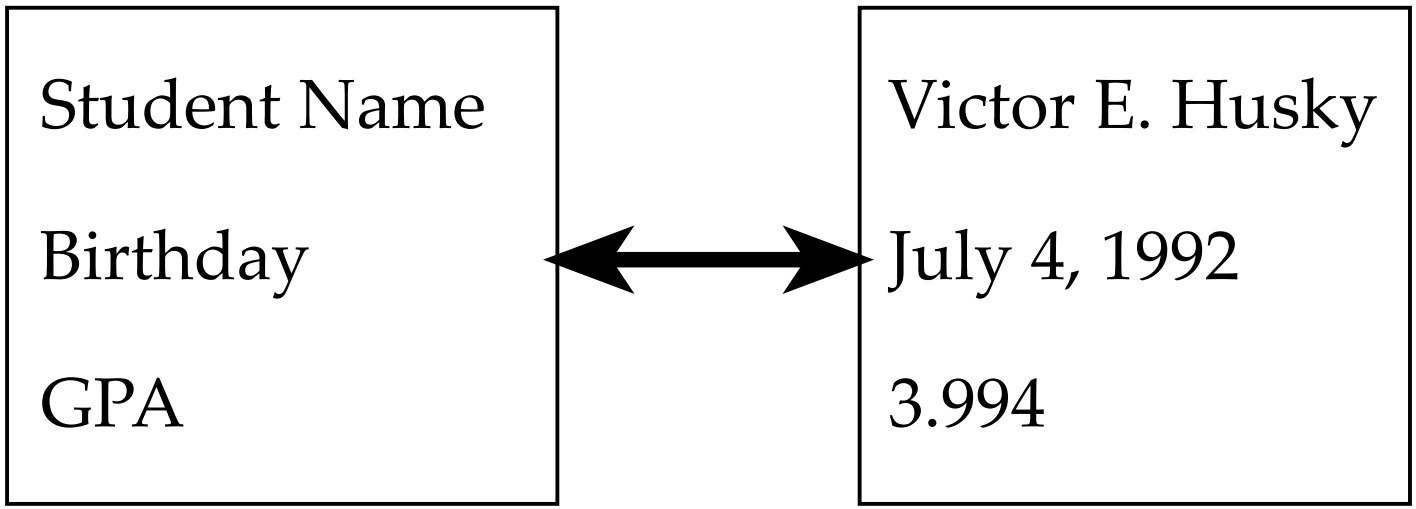
- the *extension* of the database
- current state of the database
- a snapshot of the data at a given point in time

Schema

The *schema* of a database is the data about what the data represents.

- plan of the database
- logical plan
- physical plan
- the *intention* of the database

Schema vs. Instance



Schema

description of what data can be stored

Instance

the actual data that is stored

Data Independence

Data Independence is a property of an appropriately designed database system

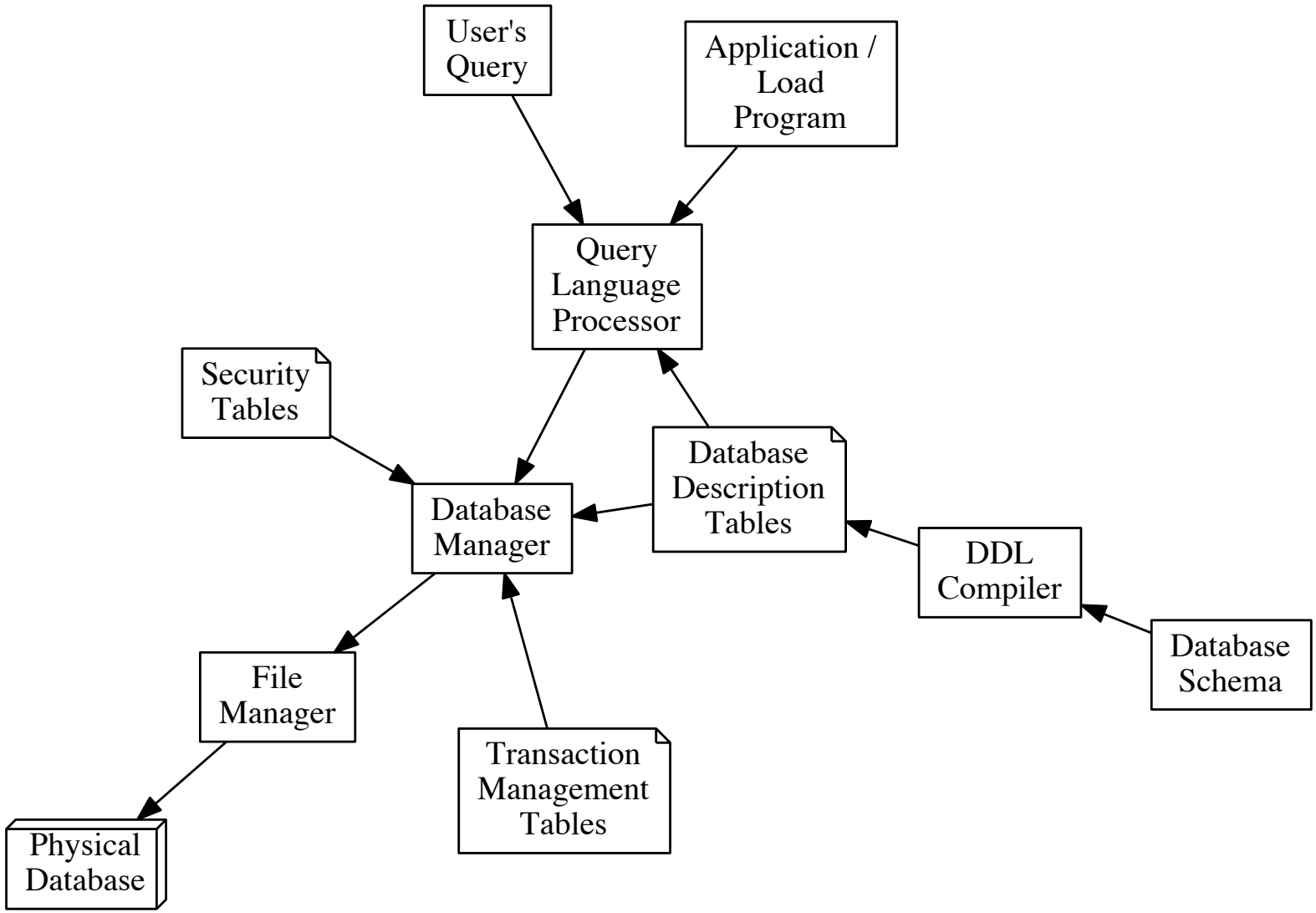
- has to do with the mapping of logical level to physical level, and logical to external
- physical data independence
  - physical schema can be changed without modifying logical schema
- logical data independence
  - logical schema can be changed without having to modify any of the external views

DCL, DDL, DML

- may be completely separate (in IMS)
- may be intermixed (DB2)
- Host language
  - application program in which DML commands are embedded such as COBOL or PL/I

Northern Illinois University

DBMS Components



Overall DBMS Usage Scenario

- Database Administrator (DBA) define the conceptual, logical, and physical levels using DDL
- DBMS software stores instances of these in schemas
- User defines views (External Schema) in DDL
- User accesses database using DML

Advantages of a Database

- Controlled redundancy
- Reduced inconsistency in the data
- Shared access to data
- Standards enforced
- Security restrictions maintained
- Integrity maintained more easily
- Provides capability for backup and recovery
- Permitting inferences and actions using rules

Disadvantages of a Database

- Increased complexity needed to implement concurrency control
- Increased complexity needed for centralized access control
- Security needed to allow the sharing of data
- Necessary redundancies can cause complexity when updating

Data vs. Information

Discuss data and information: what is the difference?