CS 1555

Lecture 24

**Conceptual Database Design & ER Model (continued)**

EER Model: Enhanced ER model

- EER model introduced the concepts of superclass and subclass entity types in the ER model

Why EER model?

- To add more semantic clarity to the design

- Minimize NULL values

Inclusion constraints

- Disjoint constraint: the subclasses of a superclass are disjoint

- Non-disjoint constraints: specify that the subclasses are overlapping

Completeness constraints

- Total specialization: every entity in the superclass must be a member of some subclass

- Partial specialization: an entity may not be a member of a subclass

Union types or categories

- Collection of entities of distinct entity types

UML & OMT (Object Modeling Techniques)

- Describe software modules and their interactions including data requirements via diagrams

- Include class diagrams which are similar to EER diagrams

**Data Storage**

Storage hierarchy

- Primary storage: random access; volatile

- Secondary storage: random access; non-volatile

- Tertiary storage: non-volatile

Improving access time?

- I/O cost = secondary storage access + queuing delays

- Disk access = seek time + rotational delay + transfer time

- Queuing delays are reduced with scheduling

- Data access cost is reduced with data organization

- Store approaches: row stores, column stores

Row stores

- Store fields in one record contiguously on disk with word alignment

- Use small (e.g., 4K) disk blocks

Column stores

- Store attributes together

- Predominant way of storage for analytics (read-only)

- Problematic when you have to do updates

File types

- Unordered files

- Ordered files

- Clustered files

- Hash files

Unordered files (heap)

- Simplest file structure – records are stored in no particular order

- New records are inserted at the end of file

- Record insertion is efficient, search is not

- Periodic reorganization: records are packed by removing deleted records

Ordered/sequential files

- File records are kept sorted by the value of an ordering key which has unique value

- Insertion is expensive; search is efficient as long as it uses a non-ordering field

Clustered files

- Order files with ordering field that isn’t a key

- Ordering field does not have a unique value