Advanced Topics in Databases

- Active Databases
- Temporal and Spatial Database
- Deductive Databases
- Distributed Databases
- XML
- Data Mining
- Data Warehousing
- OLAP
- Mobile Databases
- Multimedia Databases
- GIS
- Genome Data Management
Active Databases

• Active Behavior = ability to react to events
• dimensions:
  • event
  • condition
  • action
  • execution model
  • management
• Application Areas
  • composite objects
  • integrity constraints, business rules
  • derived data
Temporal Databases

- A **temporal database** is a database management system with built-in time aspects, e.g. a temporal data model and a temporal version of structured query language.
- More specifically the temporal aspects usually include valid-time and transaction-time. These attributes go together to form bitemporal data.
- Valid time denotes the time period during which a fact is true with respect to the real world.
- Transaction time is the time period during which a fact is stored in the database.
Spatial Databases

• offers *spatial data types*
• supports *spatial indexing*
  • find all cities in Bavaria
• supports *spatial joins*
  • for each river, find all cities within 50 KM
• managing space → large collections of simple geometric objects
• 2D: geography (GIS), VLSI design
• 3D: astronomy, brain maps, molecules
Deductive Databases

• logic programming + persistence
• prolog → datalog
• facts and rules
• inference engine
Distributed Databases

- data and processing (server) reside on multiple computers
  - transparent replication/distribution
  - increased reliability and availability
  - improved performance
  - easier expansion
- federated database system
  - shared global schema
- multidatabase system
  - interactively constructs shared schema
Distributed Transactions

- ensuring transaction properties is difficult, since multiple machines are involved
- 2 phase commit:
  - phase 1:
    - coordinator sends “prepare to commit” to all participants
    - participants force-write all logs
    - participants indicate “ready to commit” or “cannot commit”
  - phase 2:
    - if all participants are ready to commit, coordinator sends commit command
    - if any participant cannot commit, coordinator sends abort command
Structured, semi-structured, unstructured data

• structured data: fits a predefined format
  • relation schema
  • Java class

• semi-structured data: structure is flexible, but can be described at any particular time
  • structure is embedded in the data
  • aka self-describing data

• unstructured data: no discernable structure
  • raw text
XML

- XML has become popular for dealing with semi-structured data
- creates a hierarchical structure
- schema may be embedded with data or separate
- allows for dynamic databases where structure changes more quickly than can be handled with schema modifications
- persistence and queries can be specialized for XML structures
  - XPATH, XQUERY
<?xml version="1.0" standalone="yes"?>
<projects>
  <project>
    <Name>ProductX</Name>
    <Number>1</Number>
    <Location>Bellaire</Location>
    <DeptNo>5</DeptNo>
    <Worker>
      <SSN>123456789</SSN>
      <LastName>Smith</LastName>
      <hours>32.5</hours>
    </Worker>
    <Worker>
      <SSN>453453453</SSN>
      <FirstName>Joyce</FirstName>
      <hours>20.0</hours>
    </Worker>
  </project>
  <project>
    <Name>ProductY</Name>
    <Number>2</Number>
    <Location>Sugarland</Location>
    <DeptNo>5</DeptNo>
    <Worker>
      <SSN>123456789</SSN>
      <hours>7.5</hours>
    </Worker>
    <Worker>
      <SSN>453453453</SSN>
      <hours>20.0</hours>
    </Worker>
    <Worker>
      <SSN>333445555</SSN>
      <hours>10.0</hours>
    </Worker>
  </project>
  ...
</projects>
Data Mining

• KDD: Knowledge Discovery in Databases
  • 1. data selection
  • 2. data cleansing
  • 3. enrichment
  • 4. data transformation or encoding
  • 5. data mining
  • 6. reporting
Data Mining

• Mining discovers
  • association rules
  • sequential patterns
  • classification hierarchies
  • patterns within time series
  • clustering

• Goals
  • prediction
  • identification
  • classification
  • optimization
Data Warehousing

• Data Warehouses are
  • very large
  • multi-dimensional (i.e. temporal)
  • not transactional
  • the result of KDD step 4
OLAP

- Online Analytic Processing
  - analysis of complex data from a data warehouse
  - makes use of the multi-dimensional structure of the warehouse
  - provides the tools to knowledge workers
Mobile Databases

• Issues:
  • data distribution and replication
  • transaction models
  • query processing – how to handle incomplete or unavailable information?
  • recovery and fault tolerance
  • security
Multimedia Databases

- Applications:
  - repositories
  - presentation
  - collaboration

- issues
  - modeling: complex objects with “hidden” semantics
  - indexing
  - storage: generally very large objects, not suitable for storage as records in files
  - queries and retrievals: what do we match, what should be returned
GIS

- Specialization of spatial database systems
  - also has temporal aspects
  - highly dependent on complex range queries
Genome Data Management

- data characteristics
  - highly complex
  - highly variable
  - rapidly changing schema
  - multiple representations of same data
  - generally read-only
  - most users are not database savvy
  - context dependent
  - representation of complex queries is important