

Databases – lecture schedule, Easter Term 2001

(12 lectures)

The nature of a database.

Essential features of databases, examples. Formal data models: resilience to change, data independence. The role of metadata.

Mutation rate of data items. Benefits and cost of indexing.

The architecture of a DBMS.

ANSI/SPARC architecture for DBMS. The need for a Data Definition Language. Brief history of Data Models and DBMS. Corporate evolution and the need for Standards. Schema evolution.

Transaction processing, physical database design.

Recap of ACID properties of transactions. Rotating storage devices, space allocation, the need for synchronised atomic change.

Indexing and inverted files (B-trees assumed). Extensible hashing schemes.

Database design.

Entities and attributes. Mathematical foundations: sets and relations; relation product and natural join; theta-joins. Functional dependency between attribute sets. Single and multi-attribute keys.

Relationships between entities. Entity-Relationship (E/R) diagrams. Degree and multiplicity of relationships.

Entity types as classes. Object-based design methodology. UML.

Network model of data.

Historical review of DBMS. The CODASYL/DBTG proposals as an example of the ANSI/SPARC architecture.

Bachman diagrams. Storage diagrams and explicit traversal. Record identifiers (DB keys). Control of record placement. Lack of data independence.

Relational model: basic semantics.

Codd's 1970 paper. Data independence. Domains and data types. n -tuples over domains. First Normal Form.

Keys as user level tuple identifiers. Data independence. Describing entities using tuples. Describing inter-entity relationships by tuples. Research into efficient implementation. System-R. Duplicate elimination. **Relations** as *sets*, **Tables** as *bags*. Transaction support, Roll-Back. Work on query optimisation. SQL (SEQUEL).

Relational algebra.

PRTV and ISBL. Boolean operations on relations. Key-based semantics. Cartesian product. Select, Project and (equi-)Join. Theta-joins. Semi-joins. Examples. Techniques for implementing joins. Cardinality. Views as expressions in the relational algebra. INGRES, control of access via enforced views.

Relational schema design.

Functional dependence between attributes. Data redundancy for access; update and insertion anomalies. Second and Third Normal Form (3NF). Boyce-Codd Normal Form. Examples. Referential integrity. Multi-valued dependencies. Fourth Normal Form (4NF).

Relational calculus and SQL.

Relationally complete languages. The SQL 1992 standard (SQL2). Combining DDL and DML functionality. Multi-set semantics. Null values and Outer Joins. SQL2 Syntax and semantics. Aggregates and GROUP_BY. Simple examples.

Semantic considerations in the SQL data model.

Entity integrity. Modelling application datatypes, repeating groups. Collection types. Subdomains and inheritance.

Generalisation. Referential integrity and Foreign keys. The need for transitive closure. Fixed-point semantics. Deductive databases. DATALOG.

Distributed applications.

Legacy databases. Data mining. Corporate restructuring. Federation. Identifying entities, naming and scope. Data collection and data warehouses. Materialised views. Entity representations as objects. (O)ODBMS. The Object Data Management Group. The Object Database Standard ODMG 3.0. The ODMG object model.

The Object Database Standard ODMG 3.0.

The ODMG Object Model. ODMG/ODL as a Data Definition Language. Meta-objects and the ODL Schema Repository. Type system: atomic objects (user-defined), collection objects, literal types, structures.

NULL-extended types. Tables. Attributes. Object identifiers as values. Relationships and their inverses.

Operations and exceptions. Transactions. Lock types, Two-Phase Locking (2PL) as the default concurrency control policy. Databases.

OQL: an object query language extending the SQL query sublanguage. Update by method invocation. Navigation via path expressions. WHERE clauses. Definition as a view mechanism.

The final lecture of the course will take place on May 22nd 2001, when we are lucky enough to have Hugh Darwen as guest lecturer. Hugh served (for many years!) on the SQL3 Standard Committee.

The SQL1999 Standard (SQL3).

The SQL 1999 standard (SQL3). Extensions to the type system. Recursive Table expressions. Triggers. Persistent stored modules. User-defined routines. Support for Objects(?), REFs. Subtables. Conformance and Certification.

Ken Moody