A Guide to Understanding *DATABASES*, April 2000

There are no particular prerequisites for the Part IB course **Databases**, though it's a great help to feel confident about sets, functions and relations. The course is about the description, representation and manipulation of (large) collections of persistent data. Practical data management is currently dominated by the ***Relational Model*** of Data, a term introduced by E F Codd in a famous paper in 1970. The dominance has arisen for a number of reasons, but probably the principal one is that there is a widely accepted standard, SQL 1992, for a language to handle both Data Definition and Data Manipulation through Codd's relational model. But a fundamental problem is that data occurrences in the real world frequently don't conform naturally to a tabular view, and that creates stress between the twin goals at application level: to represent the data naturally, and to manipulate it using efficient and well-understood tools.

The aim of the course is to benefit the Database user and the Database administrator (DBA) rather than the implementer of a Database Management System (DBMS). DBMS are huge software constructs which account for a significant slice of the total software revenue: not just in terms of direct sales, but because many applications are built on top of DBMS. Efficient and usable DBMS require great sophistication in such areas as index management, storage allocation and concurrency control and recovery; the integrity of the data base of large enterprises is essential to their very survival. The present course does no more than scratch the surface of these topics; material on database concurrency control has already been presented in the Part IB course **Concurrent Systems**.

The course summary is given below. The core is the treatment of the relational data model; the intention is to emphasise the weaknesses as well as the strengths of the model and its standard language SQL2. The model is motivated by the need for *data independence*, and the consequent advantages emerge clearly; but many simple practical examples (repeating *order lines* in an *invoice*, for instance) cannot be handled within the model in a natural way.

The final section of the course looks at current proposals for extending the model to handle non-tabular data conveniently. A lively debate has been going on for the past ten years or so. The issue is whether salvation should be sought in the *Object-Relational Model*, which is founded on the relational model, or a completely new *Object-Orientated* approach is desirable (OODB). The SQL3 standard for the former was a long time in gestation, and shows all the signs of being designed by committee (it was). The course places more emphasis on the proposals of the *Object Data Management Group* (ODMG), which build on work by the *Object Management Group* (OMG) in the **Distributed Systems** context.

The best book for the entire course is the 1997 text by **Ullman & Widom**, which is the first to appear that adopts a broadly similar perspective to the lectures. The new 7th edition of the book by **Chris Date** is excellent, and the 3rd editions of popular texts by **Korth & Silberschatz** and **Elmasri & Navathe** are still good, and there is a lot to be gained from them. In addition the bibliography contains a number of books and papers that relate to specific aspects of the course, all of which may be thought to improve the mind. The two ***Manifestos*** sowed the seeds for the debate mentioned in the previous paragraph; there are copies in the book locker, they are both quite short, and you'll find that they're rather fun to read.

There are no notes for this course, though I shall distribute copies of the foils from time to time. Past examination papers offer much the best guide to what you might expect this year. Roughly speaking, there are three types of examination question: requests for rather discursive notes about such topics as the *ANSI-SPARC Architecture* or *data independence*; more specific ones that address aspects of some Data Model, a typical example being *Relational Normalisation*; and design problems, in which a "real-life" situation is presented in a rather muddled way, and you're asked to supply (part of) a database design. For what it's worth, successful answers to design questions have usually presented the evolution from an informal specification to a more formal relational schema in short note form, developing any normalisations along the way.

Database: Bibliography

Course reading:

** Jeffrey Ullman, Jennifer Widom
"A First Course in Database Systems", Prentice–Hall, 1997.

* C J Date **"An Introduction to Database Systems"** (7th edn.), Addison–Wesley, 2000.

* R Elmasri, S B Navathe
"Fundamentals of Database Systems" (3rd edn.), Addison–Wesley, 2000.

* A Silberschatz, H F Korth, S Sudarshan
"Database System Concepts" (3rd edn.), McGraw–Hill, 1996.

Background:

William T Kent **"Data and Reality"**, North–Holland Book Company, 1978.

Michael Stonebraker and Joseph Hellerstein (ed)
"Readings in Database Systems" (3rd edn.) Morgan Kaufmann, 1998.

Data Manipulation Languages:

* C J Date, H Darwen **"A Guide to the SQL Standard"** (4th edn.) Addison–Wesley, 1997.

Peter M D Gray **"Logic, Algebra and Databases"** Ellis-Horwood, 1984.

Object Data Management:

* R G G Cattell **"Object Data Management"** Addison–Wesley, 1991. (out-of-print)

* R G G Cattell et al. **"The Object Data Standard: ODMG 3.0"** Morgan–Kaufmann, 2000.

** M P Atkinson, F Bancilhon, D DeWitt, K Dittrich, D Maier & S B Zdonik
"The Object–Oriented Database System Manifesto" in the Proceedings of the First International Conference on Deductive and Object–Oriented Databases, Kyoto, 1989.

** M Stonebraker et al. ("The Committee for Advanced DBMS Function")
"Third Generation Database System Manifesto" ACM SIGMOD, Vol 19 #3, 1990.

Doug Barry and T Stanienda, **"Solving the Java Object Storage Problem"**,
IEEE Computer, Vol 31, No 11, pp. 33-40, November 1998.

Concurrency and Distribution:

J M Bacon **"Concurrent Systems" (Part 3)**. (2nd edn.) Addison–Wesley, 1997.

J N Gray, A Reutter
"Transaction Processing: Concepts & Techniques", Morgan-Kaufmann, 1993.