Database Technology

Introduction

This course is about database systems in general and not any particular database system. It is always possible to make a case that a particular database system is the market leader at the moment and so the course should concentrate on that one. But we are not going to do that. System specific knowledge goes out of date too quickly even in the field of database technology which changes far more slowly than most fields within computer science. Instead we will cover the ways all database systems work and why they have to work that way.

We will start with a definition of what a database is and where the idea of database systems came from in the first place and go on to look at how they work now and eventually how they are likely to work in the future.

So, to start from first principles; a database is a computer system which contains lots of organized information. The thing that organizes the information and keeps it organized is a **database system**. Data is stored using a database system so that it can be retrieved later. Since the database system organizes the information it understands where it is and how to find it. People who want to extract information from the database must be able to talk to the database system in order to specify the information they need. Thus the users of a database must know what data is stored in it and must have a language for communicating with the database system about this data. In practice this means that the users of a database system must have a mental picture of what data is stored and a vocabulary to describe it. That language is influenced by the history of database systems.

The technology of large scale data storage, and eventually database systems, grew out of punched cards (1880s) then magnetic tape (1950s). The first widely used language for manipulating large quantities of data was COBOL (1960). COBOL worked with data whose structure (and terminology) came from punched cards.

In the late 1960s IBM produced IMS (Information Management System). This was a true database system and was based on the idea that the data should be organized in some way which related more to its meaning than to a storage medium that was no longer used. It was a good idea but in practice the system was very difficult to use. In 1970ish the same committee that produced COBOL produced CODASYL. This introduced some important concepts that are still essential today but it was in some senses a standardization and unification of several ad hoc systems that various large scale data users had developed to deal with their particular problems. As such is was firmly based on existing technology and the current understanding of the sort of data business organizations wanted to store.

At almost the same time Codd produced his relational theory. This had a purely theoretical and mathematical basis and nothing to do with practicalities at all. Or to put it another way it was not limited by the technology of the time. Codd's **relational theory** is the basis of the vast majority of database systems used today and also the basis of much of this course.