Database in Computing Systems

W. Brett McKenzie
Roger Williams University, USA

INTRODUCTION

"Big ideas" drive the disciplines. In biology, the insights of Darwin generated evolutionary theory. In chemistry, Mendeleev's vision of the organization of elements predicted subsequent discoveries. In computing, the database and associated database management systems (DBMS) are one of the "big ideas". The database was conceptually possible prior to the development of the computer, but it was the digital computer that made the database the common tool it is today. The core idea of the database is distinguishing between the data description and the data itself. Among other things, this idea makes the Web possible and has made manageable new fields for discovery, such as modeling the human genome.

BACKGROUND

The modern database is an invention of the 1960's that took the computer hardware developments of the 1970's and the infrastructure of technology companies in the 1980s to become the multi-billion dollar industry of the 21st century. The foundations for the database, however, reach back to the 1890s.

The Article 1, Section 2 of the U.S. Constitution requires a population "Enumeration…" which has become the decennial census. Approaching the census of 1890, when the U.S. population had risen to more than 65 million, presented a problem because the prior census, when the population was almost half as much, had taken seven years to tabulate. The government addressed the concern by contracting with Herman Hollerith to mechanically tabulate the results using punch card machines he had designed. Hollerith successfully completed tabulating the census in less than three years, some even say in three months (Austrian, 1982). He subsequently formed the Hollerith Tabulation company to promote his invention and method of managing data. Through mergers and growth, the company, renamed Computing Tabulating Recording (CTR) Company became International Business Machines (IBM) in 1924.

Two core concepts stemmed from Hollerith's mechanical tabulators. First, data could be represented numerically by location on a punch card. Second, a machine with a wiring panel could be used to tabulate different data. The first concept allows for more than mechanical computation and permits representation of personal data, such as a name, numerically. The second introduces the notion of a general purpose machine that may be programmed—in the mechanical case, by switching the wiring pattern, in later digital computers, through a programming language.

Until the development of digital computers, database-like technologies remained in the punch card arena. Developments in national governments, such as the requirements of tracking wages for a Social Security system, required more sophisticated, specialized machines to record and tabulate information. It was the military, however, in its early Cold War defense projects such as SAGE that provided impetus to the issues of data management.

Database applications become the focus of interest in the computing industry in the late 1950s with the formation of the Conference on Data Systems Languages (CODASYL). While the first achievement of this group was the COBOL programming language for business, the group reformed in the mid-sixties to create extensions to COBOL to assist with processing records (Olle, 1978). At about the same time, data storage evolved from magnetic tape, which required serial access, to magnetic disks (forerunner of the computer hard drive) which permitted random access. The results of the conference were CODASYL compliant databases using a network data structure. Its network data structure allowed the representation of one-to-many relationships. The dominant database in this area was the integrated data store (IDS) based on ideas of Bachman (1982) and developed at General Electric. Almost simultaneous to these developments, IBM developed the information management system (IMS)—a hierarchical database independent of the CODASYL group (Blackman, 1998).

Copyright © 2008, IGI Global, distributing in print or electronic forms without written permission of IGI Global is prohibited.

This database was also able to represent one-to-many relationships and was largely a result of support for the Apollo project at the National Aeronautics and Space Administration (NASA).

As the network and hierarchical database models became commercial products and consolidated their positions in government and business, E. F. Codd, a mathematician and IBM researcher, discovered and promoted the relational model for databases. He was forced to defend his seminal paper, *A relational model of data for large shared data banks* (Codd, 1970), within IBM because the company had developed IMS as its primary database product.

The relational model challenged both the hierarchical and network models as both data models depend upon navigating paths to the data. Codd instead argued for a high-level, non-procedural language to access data, which itself was represented free from any machine constraints. The flexibility of the model resulted from its theoretical foundations in relational algebra, a mathematical sub-field of predicate calculus. In response to external interest in Codd's model, IBM developed System R in the mid-seventies as a demonstration project. The project created SEQUEL, an acronym for "structured English query language" to access the data. Later shortened to SQL (structured query language), it became an ANSI standard in 1986.

The relational database management system (RDBMS) using SQL has become the standard for databases in government and industry. The leading vendors include: Oracle, with its namesake products, first released in 1979; IBM with DB2, released in 1983, and includes NCR's Teradata for very large data stores; and the Microsoft SQL line of products and Access, desktop database for the Microsoft Office Suite. More recently, with the rise of Linux, open source databases, such as Postgres (build on Ingres, initially a government sponsored SQL project in the mid-seventies) or MySQL, have gained greater followings.

DATABASE DEFINITION

While the ability to store material is relatively trivial, witness the offices of some co-workers with their piles of papers, books, and magazines, to store data meaningfully is a challenge for two reasons. First, meaning derives from context; therefore the data must map to the real world. Second, raw data has little value. As a response to a question, however, its value increases because result has become information.

The value of this information increases with the complexity of the question. For example, one can store the following data—country, gross domestic product, and the population. The country, GDP value, and population are all pieces of data. The answer to the question, "Who has the highest GDP?" gives me some information. The answer to the question, "What is the GDP per person?" gives me information evaluated on two parameters: size and relative size per person. Furthermore, the second question responds with derived data, achieved by dividing the GDP by the population.

An entry in a database is known as a record. A record is made up of fields. A field is a characteristic of the item being specified by a record. For example, a person may be described by a first and last name, a height, weight, gender, and age. Each descriptor is a field. A record would describe a specific person, for example "Jane Doe, 150 cm, 50 kg, female, 27 years old." Records are collected in tables.

Depending upon the database model and the theoretical level of interest, the terms "table", "record", or "field" may be referred to as "entities" or "attributes". For the purposes of this article, tables will contain records, and records will contain fields. In the table, records will be the row entry while fields will be the column entry. The data is the value at the intersection of a specific row and column.

Inherent in modern databases is the notion of relationships. Often these relationships are referred to as parent-to-child relationships to indicate dependencies of data. In the first example, specific GDP and populations are "children" of the country or "parent" because they are dependent upon them. Additionally, a person could have a home address and a work address. The person's identity can be considered the "parent" and the address the "child". In this example, one parent record has two child records.

DATABASE MODELS

Flat File

The most common data model is the flat file. For example, a phone book that orders entries alphabetically with a corresponding phone number is a flat file database. An Excel spreadsheet, which is an ordered array of

4 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/database-computing-systems/16701

Related Content

Using a Social Network Game as a Teaching Tool for Visual Merchandising
Erica O'Tooleand Seung-Eun Lee (2015). *International Journal of Online Pedagogy and Course Design (pp. 1-16).*
www.irma-international.org/article/using-a-social-network-game-as-a-teaching-tool-for-visual-merchandising/127034

A Quest about eQuest and Blended Learning in Teacher Education: An Indian Study
A. Sivasankarand K. Mohanasundaram (2015). *Curriculum Design and Classroom Management: Concepts, Methodologies, Tools, and Applications (pp. 264-279).*
www.irma-international.org/chapter/a-quest-about-equest-and-blended-learning-in-teacher-education/126701

Effects of Gamification on Learning Outcomes, Satisfaction, Engagement, and Motivation in Virtual Learning Environments Between 2020 and 2022
Kexin Zhangand Zhonggen Yu (2022). *International Journal of Online Pedagogy and Course Design (pp. 1-18).*
www.irma-international.org/article/effects-of-gamification-on-learning-outcomes-satisfaction-engagement-and-motivation-in-virtual-learning-environments-between-2020-and-2022/306684

UDL and Technology Integration Models: Addressing Technological Determinism
Haidee A. Jacksonand Kara Rosenblatt (2024). *Unlocking Learning Potential With Universal Design in Online Learning Environments (pp. 98-125).*
www.irma-international.org/chapter/udl-and-technology-integration-models/342191

Podcasts as Learner-Created Content in Higher Education
Raphael Struck, Heikki Kynäslahti, Lasse Lipponen, Olli Vesterinenand Sanna Vahtivuori-Hänninen (2011). *International Journal of Online Pedagogy and Course Design (pp. 20-30).*
www.irma-international.org/article/podcasts-learner-created-content-higher/53547