

4.2_E

Data storage and databases

For humans, the challenge of storing and accessing data existed long before the development of digital databases. Some of the earliest examples of data storage come from the Middle East, where information about payments to workers was stored on clay tablets. Since then, humans found ways to store more data in smaller spaces, to make data storage more secure and to provide access to data over wider distances. Paper was a major step forward in these developments. It could be used to store large amounts of data, it was cheap, could be stored for a long time and could be transported over distances. However, paper could be damaged by fire, water or insects, it took up a lot of space and, once data was printed on paper, it could not be changed.

On early computers, data was limited and expensive, generally stored on printed cards, which took up as much space as paper storage and needed as much time to access. Fortunately, new technologies were developed such as magnetic tape, which cost less, could store more data and provided quicker access to it. In addition, these technologies meant that data could be changed very quickly and easily. As a result, by the early 1960s, database management systems (DBMS) were developed to make this easier. A core feature of these was the database engine. The engine manages the creation, reading, updating and deleting (CRUD) of data by users and makes sure that it is not corrupted. The DBMS also allows a database of administrators to control who can access different types of data and to produce reports from queries or logs of transactions.

The introduction of relational databases using Structured Query Language (SQL) in the late 1970s and early 1980s was another major step forward. A key feature of relational databases is the use of tables and queries. Tables are structured into fields and rows. Fields can only contain a specific type of data, such as text or numbers. Rows are made up of different fields of data. Relational databases are powerful because data in different tables can be linked together. Relational databases were also a major advance for database designers because they could reduce data duplication by using schemas (a plan for the structure of the whole database). SQL was a standardized language which could be used (with some variations) across many different types of computers.

The introduction of MySQL as an open-source database was another major development in the way databases were used. From the early 1980s onwards, digital databases were relatively easy to administer, secure, reliable and could manage large amounts of data. Unfortunately, they also cost a lot of money. In 1995, the developers of MySQL – a high-quality DBMS – made it available as a free and open-source product. Over time, the use of MySQL grew rapidly. It was combined with three other free open-source technologies (the Linux operating system, the Apache web server and the PHP programming language) as the LAMP stack and used to create dynamic websites. The LAMP stack was a major factor in the growth of the internet during the early decades of the millennium. As an open-source database, MySQL also had a significant role in the development of large e-commerce companies, such as Amazon, and social media companies, such as Facebook.

Ongoing developments such as cloud computing created additional challenges. One challenge was the need for a single database to run across large numbers of computers. In addition, there was a need for databases to handle unstructured data. These challenges led to the development of NoSQL databases. Some of these, such as Redis, provided faster access to data. Others, such as MongoDB, handled unstructured data more easily. Still others were graph databases, designed to make it easier to identify patterns in linked data. These technologies were designed for cloud computing so they were more efficient. Because they can manage less structured data, they make it easier to deal with very large amounts of different types of data.

Since the 1960s, database technologies have completely changed the way in which data is stored and accessed. Firstly, relational databases provided effective ways to structure data so it makes the best use of available hardware. SQL provided a common language for managing databases and quickly retrieving relevant data. In addition, by making high-quality database technology freely available, MySQL enabled the rapid growth of a more dynamic internet, allowing individuals across the world to easily access and to share data. Finally, NoSQL databases have responded to the development of cloud computing and the huge quantities and different types of data it produces. Managing these in the future is a major challenge for computing. Fortunately, as the past shows, there is significant potential to meet these challenges.

4.2_E

Answers

Data storage and databases

For humans, the challenge of storing and accessing data existed long before the development of digital databases. Some of the earliest examples of data storage come from the Middle East, where information about payments to workers was stored on clay tablets. Since then, humans found ways to store more data in smaller spaces, to make data storage more secure and to provide access to data over wider distances. Paper was a major step forward in these developments. It could be used to store large amounts of data, it was cheap, could be stored for a long time and could be transported over distances. However, paper could be damaged by fire, water or insects, it took up a lot of space and, once data was printed on paper, it could not be changed.

On early computers, data was limited and expensive, generally stored on printed cards, which took up as much space as paper storage and needed as much time to access. Fortunately, new technologies were developed such as magnetic tape, which cost less, could store more data and provided quicker access to it. In addition, these technologies meant that data could be changed very quickly and easily. As a result, by the early 1960s, database management systems (DBMS) were developed to make this easier. A core feature of these was the database engine. The engine manages the creation, reading, updating and deleting (CRUD) of data by users and makes sure that it is not corrupted. The DBMS also allows a database of administrators to control who can access different types of data and to produce reports from queries or logs of transactions.

The introduction of relational databases using Structured Query Language (SQL) in the late 1970s and early 1980s was another major step forward. A key feature of relational databases is the use of tables and queries. Tables are structured into fields and rows. Fields can only contain a specific type of data, such as text or numbers. Rows are made up of different fields of data. Relational databases are powerful because data in different tables can be linked together. Relational databases were also a major advance for database designers because they could reduce data duplication by using schemas (a plan for the structure of the whole database). SQL was a standardized language which could be used (with some variations) across many different types of computers.

The introduction of MySQL as an open-source database was another major development in the way databases were used. From the early 1980s onwards, digital databases were relatively easy to administer, secure, reliable and could manage large amounts of data. Unfortunately, they also cost a lot of money. In 1995, the developers of MySQL – a high-quality DBMS – made it available as a free and open-source product. Over time, the use of MySQL grew rapidly. It was combined with three other free open-source technologies (the Linux operating system, the Apache web server and the PHP programming language) as the LAMP stack and used to create dynamic websites. The LAMP stack was a major factor in the growth of the internet during the early decades of the millennium. As an open-source database, MySQL also had a significant role in the development of large e-commerce companies, such as Amazon, and social media companies, such as Facebook.

Ongoing developments such as cloud computing created additional challenges. One challenge was the need for a single database to run across large numbers of computers. In addition, there was a need for databases to handle unstructured data. These challenges led to the development of NoSQL databases. Some of these, such as Redis, provided faster access to data. Others, such as MongoDB, handled unstructured data more easily. Still others were graph databases, designed to make it easier to identify patterns in linked data. These technologies were designed for cloud computing so they were more efficient. Because they can manage less structured data, they make it easier to deal with very large amounts of different types of data.

Since the 1960s, database technologies have completely changed the way in which data is stored and accessed. Firstly, relational databases provided effective ways to structure data so it makes the best use of available hardware. SQL provided a common language for managing databases and quickly retrieving relevant data. In addition, by making high-quality database technology freely available, MySQL enabled the rapid growth of a more dynamic internet, allowing individuals across the world to easily access and to share data. Finally, NoSQL databases have responded to the development of cloud computing and the huge quantities and different types of data it produces. Managing these in the future is a major challenge for computing. Fortunately, as the past shows, there is significant potential to meet these challenges.