## A History of Database Systems

Database Management Systems – 2nd Ed.
Patricia Ward - CENGAGE Learning ( Fasttrack Series)
Page 2.



Systems accumulate data over time:

- Bank account transactions
- Sales transactions in a department store
- E-commerce sales
- Property rentals
- Facebook

This data must be stored for future reference.

Almost all systems have an underlying information system.



What is a data store?

- Somewhere data is stored
- ***Structured*** repository (easy storage and retrieval)
- Needed for the day to day running of a business

How have information systems evolved over the years?



## < 1979 : Paper-based systems



## 1980 : File-based Systems

Information stored in computer files on mainframe computer

Accessed via "Dummy Terminals"



## 1985-1999 : Relational Databases

- Central data store
- All users share the same data


## 2010: Cloud Computing




## From paper to computer files

Database systems developed because of the need to:

- store large volumes of data
- Retrieve data accurately
- Retrieve data quickly

Before databases, the data was stored in paper format in filing cabinets.



Consider a customer order processing system:

- **Order placed**
  - Paper
  - Phone
- **Existing customer**
  - Paper file retrieved to get customer details
- **New customer**
  - Customer details requested, recorded in paper file



- **The order form copied**
  - One copy in filing cabinet
  - Other copy sent to dispatch department
- **Stock Replenished**
  - File(s) detailing stock in warehouse retrieved
  - Purchase order (paper document) created to replenish stock
- **Invoice issued to customer**
  - Copy filed in filing cabinet

Lots of paper handled, generated and stored!



## Limitations of paper-based systems

- Generated large volumes of paper files which required storage.
- Recording of data was frequently duplicated
- Inconsistencies in data were common
- Data retrieval was tedious and often slow as files were often misplaced / lost
- Businesses needed a better way to manage their data



- Computerised file systems arrived in the early 1960's. Data could then be retrieved and processed much more quickly.
- These consisted of ***data files*** and ***application software***
  - **Data file**: data stored on magnetic tape; magnetic disk
  - ***Application software***: computer programs which processed the data files.

An application program was written for ***each*** task required. Consider our earlier ***Customer Order processing System*** example:

- Place an Order
- Check Stock Levels
- Issue Invoice
- Dispatch order
- Record Payment
- Replenish Stock
- List all dispatched orders
- List all orders for a specific customer
- ……..
- Lots of programs!

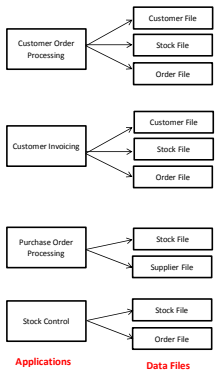


Although some applications might require similar data, data was stored in separate data files. See next slide.

Typically, each department within the organisation had their own set of data files and applications.



File-based approach to data






While file-based systems were a big improvement on paper-based systems, many limitations of file-based systems became apparent over time:



## Limitations of file-based systems

- Processing remained slow
- Data inconsistency was still a problem due to similar data being stored in different data files. Users often forgot to update data held in multiple data files (e.g. customer address)
- ***Shared files*** were introduced to resolve data inconsistency. However, shared files limited access to only one application at a time.



- Application programs required the file interface of any files accessed by the application to be defined in its source code.

  ***Interface Explosion***

  - This was time consuming, especially for applications which accessed multiple files.
  - Required much software maintenance if file definitions changed.

- File-based systems did **not** support data independence or structural independence
  - If an item of data in a file changed (size, data type) then all applications using that file had to be amended.
  - If the structure of the file changed (data item added/removed, order of data items changed) then all applications using that file had to be amended.



- Data queries required an application program to be written.

  Long wait time for even the simplest of queries while program written.

  No ad hoc querying facility.



## Early Database Systems

- By mid 1960's database systems were emerging
- Offered much improved file sharing
- Allowed simultaneous access to shared data by multiple users
- Querying facility
- Security/Access control
- Data integrity maintained (no inconsistencies)



- One such early system, the IBM product IMS, used **Hierarchical** data model (tree) to present data to users
- The **Network** data model emerged in the late 1960's
- Both models required skilled programmers to write the applications to create, access and change the data in the database
- Such systems tended to be used by large organisations



- The database approach used powerful software referred to as the database management system(DBMS) to control the data (database).
- The DBMS has several components which provide facilities for:
  - Querying data
  - Data security
  - Data integrity
  - Simultaneous data access (sharing)



- The underlying structure of the data is isolated from the actual data
- The description of the structure of the entire DB is called a **conceptual schema**
- Any changes to the logical structure of the data are made at the **conceptual** level. These changes are independent of the physical storage level and the end-user view.

## Relational Database Systems

- The study of databases was a major academic and research area by the 1970's
- The **Relational** model was proposed by E.F. Codd in 1970.
- Relational databases replaced the earlier hierarchical and network-based systems



- Data stored in relations (tables) as a set of tuples (rows) and attributes (columns)
- The relational data model is based on the mathematical principles of ***set theory*** and ***predicate logic***



- Initially, data manipulation was achieved using ***Relational Algebra*** and ***Relational Calculus***.

  x:\Cwoods\DBConcepts_CP\Lectures\Handouts\Handout_01_Rel_Algebra.pdf

- Structured Query Language (SQL) is now the de-facto language for querying all relational database systems



- Oracle's DBMS was the first commercial relational product released – 1980.
- The relational model used extensively for transaction processing systems
- Many vendors have expanded beyond the original DBMS software
- Most vendors now provide suite of software products and tools:
  - Form / Report generators
  - Code generators
  - Data warehousing tools