

Normalisation

- The process of designing a relational database.
- Aims to produce the best and most effective design.

Normalisation Considerations

- Remove redundant or duplicated components.
- Ensure data in linked tables is consistent.
- Allow complex queries to be carried out.
- Ensure records can be added or removed without problems.

First Normal Form

- Attributes may contain a single value only.

Second Normal Form

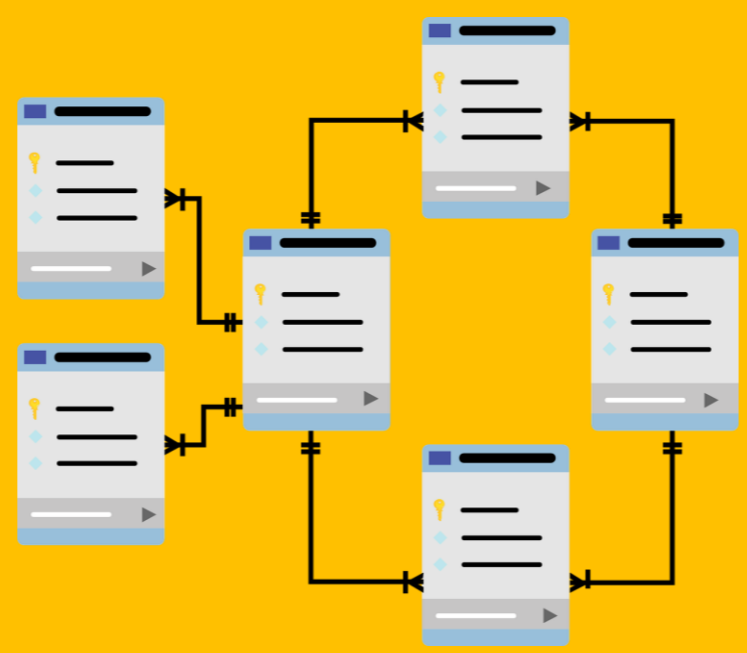
- In First Normal Form.
- Partial dependencies are not allowed.

Third Normal Form.

- In Second Normal Form.
- Non key dependencies are not allowed.

Run Length Encoding

- A lossless compression method.
- Repeated values are replaced with a single instance of the value and the number of times the value occurs.
- It relies on all consecutive pieces of data being the same.
- It offers poor reduction in file size if there is little repetition.



Indexing

- Stores the position of each record when records are ordered by a certain attribute.
- The primary key is automatically indexed.
- Allows data to be found and accessed quickly

Capturing Data

- There are many ways to capture the data needed for a database.
- The most appropriate way will depend on the type and quantity of data needed and available resources.
- Data may be manually entered by a human or scanned in using optical character recognition, sensors or barcodes.

Selecting, Managing and Exchanging Data

- Data may be selected based around set criteria
- Only data matching the criteria is input to the data
- SQL can be used to sort, structure and filter the data
- Data may need to be transferred between systems or organisations
- This is know as data exchange
- This can be accomplished using EDI (Electronic Data Exchange)

Entity Relationship Modelling

- One to One – Each entity can be associated with one other entity only.
- One to Many – A single table many entities in another table.
- Many to Many – Many entities in one table are linked to many in another table.

Unit 1.3 Exchanging Data Page 1

Referential Integrity

- Ensures consistency.
- Ensures that information is not removed if it is needed elsewhere in the database.

Transaction Processing

- A single operation executed on data.
- Must be processed in line with ACID

ACID

- Atomicity, Consistency, Isolation, Durability.
- Atomicity - the whole transaction must be processed.
- Consistency - transactions must maintain the referential integrity rules between linked tables.
- Isolation - executing transactions at the same time must produce the same result as if they were executed one after the other.
- Durability - when a transaction has been executed it will not be undone.

Record Locking

- Prevents records being accessed by more than one transaction at the same time.
- Prevents inconsistencies and data loss.
- Can result in deadlock

Redundancy

- Multiple copies of the data are kept in different physical locations.
- If data in one copy is lost or damaged it can be retrieved from another copy.

SQL Commands

- SELECT - returns fields from a table.
- FROM - specifies the table or tables.
- WHERE - specifies the search criteria.
- LIKE – used to specify wildcard criteria in conjunction with the % character.
- AND, OR – match more than one criteria.
- JOIN - allows rows from multiple tables to be returned and defines how the tables are linked
- INSERT INTO - inserts a new record in an existing table.
- DELETE - delete a record from a table.
- DROP – delete an entire table.

SQL Examples

```
SELECT CustomerName,Address FROM Customers
WHERE CustomerName LIKE '%Smith%'

SELECT CustomerName,Address FROM Customers
WHERE CustomerName LIKE '%Smith%' AND
CustomerAddress LIKE '%Road%'

SELECT Orders.OrderID,
Customers.CustomerName, Orders.OrderDate
FROM Orders
JOIN Customers ON
Orders.CustomerID=Customers.CustomerID

INSERT INTO Customers (CustomerName,
ContactName, Address, City, PostalCode,
Country)
VALUES ('Cardinal', 'Tom B. Erichsen',
'Skagen 21', 'Stavanger', '4006', 'Norway')

DELETE FROM Customers WHERE
CustomerName='Alfreds Futterkiste'

DROP TABLE Shippers
```

Databases

- An entity is item about which information is stored such as books, or customers.
- Attributes are the categories in which data is collected such as height or name.

Flat File Database

- Consists of a single file.
- Usually based around a single entity.
- Only one table.

Relational Database

- Uses many tables to store data about different entities.
- These tables are linked together.

Primary Key

- A unique identifier, different for each object in the database.
- Usually and ID number or other unique ID.

Foreign Key

- Used to link two tables together.
- The primary key from a different table.

Secondary Key

- Used to enable searching or sorting.
- Usually a common field like name.

Dictionary Encoding

- A lossy compression method.
- Commonly used data is replaced with an index.
- The compressed data is stored with a dictionary.
- The dictionary can restore original data.
- The dictionary links the commonly used data to the index.

Hashing

- Turns an input into a value of a fixed size.
- The input is known as a key.
- The output is known as a hash.
- The hash cannot be turned into the key.
- A hash table stores keys and their matching values.
- They can be used to lookup data in an array.
- They are used in situations where lots of data needs to be looked up in a constant time.
- Algorithms which perform this task are called hash functions.
- The output of a hash function should be smaller than the input.
- If two inputs produce the same hash this is known as a hash collision.
- Using a second hash function and storing items together with the hash helps to overcome collisions.
- Good hash functions are quick to run and have a low rate of collision.

Encryption

- Used to keep data secure.
- Used when transmitting or storing data in ways where others may have access to it.
- Scrambles the data to prevent it being easily read.
- Encryption keys are used to encrypt and decrypt data.

Symmetric Encryption

- The same private key is used by the sender and receiver.
- The same key is used to encrypt and decrypt data.
- A key exchange process is used to share the key.
- Data can be read should the key be intercepted.

Asymmetric Encryption

- Uses two different keys.
- The public key is used to decrypt data and can be shared anywhere.
- The private key is used to encrypt data and must be kept securely.
- The two keys are known as a key pair and are related to each other.
- Encrypting a message using the public key verifies that it was sent and encrypted by the owner of the key.

Search Engines

- Search a database of web addresses to find resources based on criteria set by the user.
- Rely on an index of pages through which they search.
- Web Crawlers build the index by traversing The Internet exploring all links on the page.
- Crawlers collect keywords, phrases and metadata from pages.

Network Topologies

Bus Network

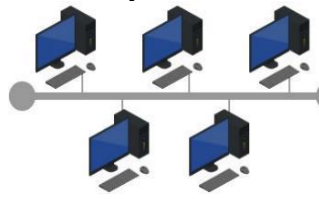
All devices are connected to a single cable (called the bus)
A terminator is at each end of the cable.

Advantages:

- Easy to install extra devices.
- Cheap to install as it doesn't require much cable.

Disadvantages

- If the cable fails or is damaged the whole network will fail.
- Performance becomes slower as additional devices are connected due to data collisions.
- Each device receives all data, a security risk



Star Network

All nodes are connected to one or more central switches.
Often used with wireless networks.

Advantages:

- Every device has its own connection so failure of one node will not affect others.
- New devices can be added by simply connecting them to the switch.
- Usually have higher performance as a message is passed only to its intended recipient.

Disadvantages:

- If the switch fails it takes out the whole network.
- Requires a lot of cable so can be expensive.



Mesh Networks

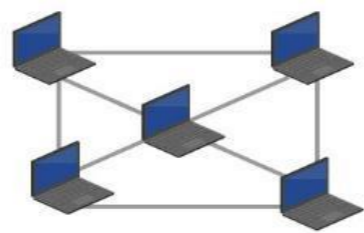
No central connection point, with each device connecting directly to others.
Full mesh networks have every device connected to every other device.
Partial mesh networks have each device connected to several others but not necessarily every other device.

Advantages

- Messages can be received more quickly.
- Messages have many possible routes they can take.
- Multiple connections mean that no device should be isolated
- Each device can talk to more than one node at the same time.
- Devices can be added without interruption.

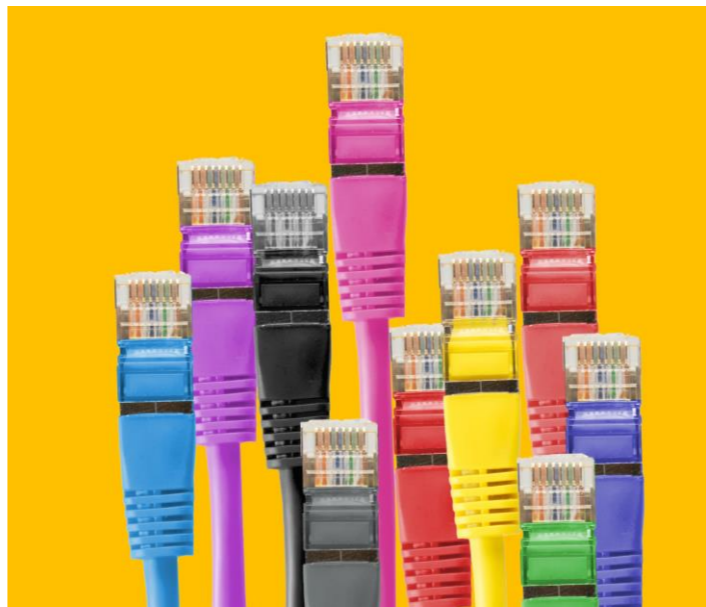
Disadvantages

- Can be impractical and expensive to setup.
- Require a lot of maintenance



Computer Networks

- A network is two or more computers connected together for the purposes of transmitting data.
- The physical topology defines the physical layout of the network
- The logical topology defines the way data flows through the network
- A protocol is a set of rules for communication between devices.
- They allow devices from different vendors to communicate
- A LAN (local area network) covers a small physical area.
- A WAN (wide area network) covers a large physical area.



Internet Protocols

TCP/IP Stack

- Transfer Control Protocol / Internet Protocol.
- A group (stack) of protocols which work together.
- Controls the flow of data packets through the network.

DNS

- Domain Name System
- Allows websites and other network devices to be identified by a human readable name.
- DNS Server converts the name to an IP Address.
- A hierarchy.
- Each domain name is separated by a dot.
- The names to the right are highest in the hierarchy.

Application Layer

- Top of the stack.
- Specifies the required protocol needed by the application the user is using.

Transport Layer

- Uses TCP to establish a connection through the network between the source and recipient devices.
- Splits data into packets labelled with a packet number.
- Requests retransmission of any packets lost during transit.

Network Layer

- Adds a source and destination IP Address to packets.
- Routers use this address to forward packets through the network to their destination.

Link Layer

- The physical connection between devices.
- Uses a MAC Address to communicate.

LANs and WANs

- LAN – Local Area Network – covers a small area.
- WAN – Wide Area Network – covers a large area.

Unit 1.3 Exchanging Data Page 2

Client Server Network

- Clients connect to a central server.
- The server is a powerful computer central to the network.
- It holds all the data.
- More secure setup.
- Clients do not need to be backed up.
- Data and resources can be shared easily.
- Expensive to setup.
- More secure.

Peer to Peer Network

- Computers are connected directly to each other.
- Computers share data with one another.
- Quick, cheap and easy to setup.
- Less secure.
- Easier to maintain.

Compression

- Reduces the space needed to store or transmit a file.
- Important when sharing files over a network or The Internet and when dealing with limited storage space.
- Increased the number of files which can be sent or received.
- Lossy compression removes some information whilst compressing the file. Original cannot be retrieved.
- Lossless compression reduces the size of the file without losing any information. Original can be retrieved.

Network Hardware

NIC

- Network Interface Card
- May be wired or wireless.
- Allows a device to connect to a network.
- Has a unique MAC (Media Access Control) address assigned to it.

Switches

- Controls the flow of data through the network.
- Used in star topologies.

Wireless Access Points (WAPs)

- Allows devices to connect wirelessly to a network.
- Used in mesh networks.
- Often used with a router to allow devices Internet access.

Routers

- Used to connect two or more networks together.
- Often used between a home/office network and an ISP to allow Internet access.

Gateway

- Used to connect networks using different protocols.
- Translates protocols to allow devices to communicate.
- Changes the packet headers.

Packet and Circuit Switching

- Packet Switching
- Data is split into packets.
- Packets are sent across the network.
- Packets may take different routes through the network.
- Circuit Switching
- A direct link is created between devices.
- The link is maintained for the entire conversation.
- Both devices must transfer data at the same rate.

Server Side Processing

- Client sends all data to the server for processing. Examples include SQL and PHP.
- It requires no plugins on the client.
- Servers can usually perform large or complex calculations more quickly.
- It is not browser dependent.
- It is more secure

Proxy Server

- Sits between a user and the resource they are accessing.
- Protects users' privacy.
- Caches frequently accessed websites to increase performance.
- Reduces web traffic.
- Uses rules to block access to sensitive information.

PageRank Algorithm

- Ranks each web page
 - Higher ranked pages appear first when results are shown.
 - Rank based on the number of incoming links on the page and the rank of these pages.
 - This is stored in a directed graph.
 - The sites are nodes and the links between the pages are the arcs.
- $$\text{PageRank}(x) = (1-d) + d[(\text{PageRank}(T1) \div \text{Count}(T1)) + \dots + (\text{PageRank}(Tn) \div \text{Count}(Tn))]$$

Client Side Processing

- Client processes the data locally.
- Examples include JavaScript.
- Web pages can immediately respond to actions.
- Code executes more quickly.
- It gives more control over the behaviour and look of websites

Firewalls

- Prevent unauthorised access to the network.
- Has two NICs.
- Data enters one NIC and is compared to a set of rules.
- Traffic which matches the rules is passed out the other NIC.