

# SS3 Data Processing Lesson Note Second Term

## SIGMA TERM ENOTE

**CLASS : SS 3**

### **Parallel database**

A parallel database system seeks to improve performance through parallelization of various operations, such as loading data, building indexes and evaluating queries. Although data may be stored in a distributed fashion, the distribution is governed solely by performance considerations. Parallel databases improve processing and input/output speeds by using multiple CPUs and disks in parallel. Centralized and client-server database systems are not powerful enough to handle such applications. In parallel processing, many operations are performed simultaneously, as opposed to serial processing, in which the computational steps are performed sequentially.

### **Architectures of parallel Database**

#### **Shared memory architecture**

Where multiple processors share the main memory (RAM)space but each processor has its own disk (HDD). If many processes run simultaneously, the speed is reduced, the same as a computer when many parallel tasks run and the computer slows down.

#### **Shared disk architecture**

Where each node has its own main memory, but all nodes share mass storage, usually a storage area network. In practice, each node usually also has multiple processors.

#### **Shared nothing architecture**

Where each node has its own mass storage as well as main memory.

### **The Benefits of Parallel Database**

Parallel database technology can benefit certain kinds of applications by enabling:

- \*. Higher Performance
- \*. Higher Availability
- \*. Greater Flexibility
- \*. More Users

**Higher Performance** With more CPUs available to an application, higher speedup and scaleup can be attained.

**Higher Availability** Nodes are isolated from each other, so a failure at one node does not bring the whole system down.

**Greater Flexibility** An OPS environment is extremely flexible. Instances can be allocated or deallocated as necessary.

**More Users** Parallel database technology can make it possible to overcome memory limits, enabling a single system to serve thousands of users.

### **Distributed Database**

In a distributed database, data is stored in different systems across a network. For Example, in mainframes, personal computers, laptops, cell phones, etc.

#### **Advantages of distributed database:**

- 1) In a distributed database, data can be stored in different systems like personal computers, servers, mainframes, etc.
- 2) A user doesn't know where the data is located physically. Database presents the data to the user as if it were located locally.
- 3) Database can be accessed over different networks.
- 4) Data can be joined and updated from different tables which are located on different machines.
- 5) Even if a system fails the integrity of the distributed database is maintained.
- 6) A distributed database is secure.

#### **Disadvantages of distributed database:**

- 1) Since the data is accessed from a remote system, performance is reduced.
- 2) Static SQL cannot be used.
- 3) Network traffic is increased in a distributed database.
- 4) Database optimization is difficult in a distributed database.
- 5) Different data formats are used in different systems.

- 6) Different DBMS products are used in different systems which increases in complexity of the system.
- 7) Managing system catalog is a difficult task.
- 8) While recovering a failed system, the DBMS has to make sure that the recovered system is consistent with other systems.
- 9) Managing distributed deadlock is a difficult task.

### **Types of Distributed Databases**

Distributed databases can be broadly classified into homogeneous and heterogeneous distributed database

#### **Homogeneous Distributed Databases**

In a homogeneous distributed database, all the sites use identical DBMS and operating systems.

#### **Heterogeneous Distributed Databases**

In a heterogeneous distributed database, different sites have different operating systems, DBMS products and data models.

### **Architectural Models**

Some of the common architectural models are –

Client - Server Architecture for DDBMS

Peer - to - Peer Architecture for DDBMS

Middleware (distributed applications)

provides services for the various components of a distributed system

### **Networking**

Networking is the practice of linking two or more computing devices together for the purpose of sharing data. Networks are built with a mix of computer hardware and computer software.

### **Network Design**

Computer networks also differ in their design. The two basic forms of network design are called client/server and peer-to-peer. Client-server networks feature centralized server computers that store email, Web pages, files and or applications.

### **A Network Topology**

A network topology represents its layout or structure from the point of view of data flow.

### **Network Protocols**

Communication languages used by computer devices are called network protocol. Yet another way to classify computer networks is by the set of protocols they support. Networks often implement multiple protocols with each supporting specific applications. Popular protocols include TCP/IP, the most common protocol found on the Internet and in home networks.

### **The Internet**

The Internet is the global system of interconnected computer networks that use the Internet protocol suite (TCP/IP) to link devices worldwide. It is a network of networks that consists of private, public, academic, business, and government networks of local to global scope, linked by a broad array of electronic, wireless, and optical networking technologies. The Internet carries an extensive range of information resources and services, such as the inter-linked hypertext documents and applications of the World Wide Web (WWW), electronic mail, telephony, and peer-to-peer networks for file sharing.

### **An intranet**

An intranet is a private network that is contained within an enterprise. It may consist of many interlinked local area networks and also use leased lines in the wide area network. Typically, an intranet includes connections through one or more gateway

computers to the outside Internet. The main purpose of an intranet is to share company information and computing resources among employees. An intranet can also be used to facilitate working in groups and for teleconferences.

### **An extranet**

An extranet is a private network that uses Internet technology and the public telecommunication system to securely share part of a business's information or operations with suppliers, vendors, partners, customers, or other businesses. An extranet can be viewed as part of a company's intranet that is extended to users outside the company. It has also been described as a "state of mind" in which the Internet is perceived as a way to do business with other companies as well as to sell products to customers.

### **A personal area network-PAN**

Definition: A personal area network-PAN- is a computer network organized around an individual person. Personal area networks typically involve a mobile computer, a cell phone and/or a handheld computing device such as a PDA. You can use these networks to transfer files including email and calendar appointments, digital photos and music.

### **Local area network**

A local area network (LAN) is a network that connects computers and devices in a limited geographical area such as home, school, computer laboratory, office building, or closely positioned group of buildings. Each computer or device on the network is a node.

### **Metropolitan area network**

A Metropolitan area network (MAN) is a large computer network that usually spans a city or a large campus.

### **Wide area network**

A wide area network (WAN) is a computer network that covers a large geographic area such as a city, country, or spans even intercontinental distances, using a communications channel that combines many types of media such as telephone lines, cables, and air waves. A WAN often uses transmission facilities provided by common carriers, such as telephone companies.

### **Client-Server Network**

A client-server network is designed for end-users, called clients, to access resources such as files, songs, video collections, or some other service from a central computer called a server.

### **A Peer-To-Peer (P2p) Network**

In its simplest form, a peer-to-peer (P2P) network is created when two or more PCs are connected and share resources without going through a separate server computer. A P2P network can be an ad hoc connection—a couple of computers connected via a Universal Serial Bus to transfer files.

### **Network Topology**

A network topology is the pattern in which nodes (i.e., computers, printers, routers or other devices) are connected to a local area network (LAN) or other network via links (e.g., twisted pair copper wire cable or optical fiber cable).

There are four principal topologies used in LANs: bus, ring, star and mesh. The most widely used of these is bus, because it is employed by Ethernet, which is the dominant LAN architecture. In a bus topology all devices are connected to a central cable, called the bus or backbone. This topology is relatively inexpensive and easy to install for small networks.

In a ring topology each device is connected directly to two other devices, one on either side of it, to form a closed loop. This topology is relatively expensive and difficult to install, but it offers high bandwidth and can span large distances. A variation is the token ring, in which signals travel in only one direction around the loop, carried by a so-called token from node to node.

In a star topology all devices are connected directly to a central computer or server. Such networks are relatively easy to install and manage, but bottlenecks can occur because all data must pass through the central device.

The mesh topology can be either a full mesh or a partial mesh. In the former, each computer is connected directly to each of the others. In the latter, some computers are connected to most of the others, and some are connected only to those other nodes with which they exchange the most data.

The several basic network topologies can be combined in various ways to form hybrid topologies, such as a ring-star network or a tree network. The latter consists of two or more star networks connected to a linear bus.

The word topology comes from the Greek words *topos* meaning place and *logos* meaning study. It is a description of any locality in terms of its layout. Topology is a branch of mathematics concerned with properties of geometric figures that are distorted without tearing or bonding together.

## **Internet security**

Internet security is a branch of computer security, Its objective is to establish rules and measures to use against attacks over the Internet. [1]The Internet represents an insecure channel for exchanging information leading to a high risk of intrusion or fraud, such as phishing.

## **Malware**

Malware, short for malicious software, is any software used to disrupt computer operation, gather sensitive information, or gain access to private computer systems. Malware is defined by its malicious intent, acting against the requirements of the computer user, and does not include software that causes unintentional harm due to some deficiency.

## **A botnet**

A botnet is a network of zombie computers that have been taken over by a robot or bot that performs large-scale malicious acts for the creator of the botnet.

## **Computer Viruses**

Computer Viruses are programs that can replicate their structures or effects by infecting other files or structures on a computer. The common use of a virus is to take over a computer to steal data.

**\*. Computer worms**

\*. Computer worms are programs that can replicate themselves throughout a computer network, performing malicious tasks throughout.

**\*. Ransomware**

\*. Ransomware is a type of malware which restricts access to the computer system that it infects, and demands a ransom paid to the creator(s) of the malware in order for the restriction to be removed.

**\*. Spyware**

\*. Spyware refers to programs that surreptitiously monitor activity on a computer system and report that information to others without the user's consent.

**\*.A Trojan horse**

\*.A Trojan horse, commonly known as aTrojan, is a general term for malicious software that pretends to be harmless, so that a user willingly allows it to be downloaded onto the computer.

**Denial-of-service attacks**

A denial-of-service attack (DoS attack) or distributed denial-of-service attack (DDoS attack) is an attempt to make a computer resource unavailable to its intended users.

**Phishing**

Phishing occurs when the attacker pretends to be a trustworthy entity, either via email or web page. Victims are directed to fake web pages, which are dressed to look legitimate, via spoof emails, instant messenger/social media or other avenues. Often tactics such as email spoofingare used to make emails appear to be from legitimate senders

**Application Vulnerability**



Applications used to access Internet resources may contain security vulnerabilities such as memory safety bugs or flawed authentication checks. The most severe of these bugs can give network attackers full control over the computer.