Distributed System – Theory

1. What is a distributed system?

Definitions

- *"A distributed system is a collection of independent computers that appear to the users of the system as a single computer" (Tanenbaum, 1994)*

- Aspects:
  - hardware: the machines are autonomous.
  - software: the users think of the system as a single computer

Example 1

- the system of a large bank with hundreds of branch offices all over the world
  - Each office has a master computer to store local accounts and handle local transactions
  - each computer has the ability to talk to all other branch computers and with a central computer at headquarters
  - transactions can be done without regard to where a customer or account is
  - the users do not notice any difference between this system and the old centralized mainframe that it replaced

Example 2

- a factory full of robots
  - each robot: a powerful computer for handling vision, planning, communication, & other tasks.
  - all robots act like peripheral devices attached to the same central computer
  - a robot on the assembly line notices that a part it is supposed to install is defective, it asks another robot in the parts department to bring it a replacement

Example 3

- a network of workstations in a university or company department
  - a pool of processors in the machine room that are not assigned to specific users but are allocated dynamically as needed
  - a single file system with all files accessible from all machines in the same way and using the same path name
  - when a user typed a command, the system could look for the best place to execute that command, possibly on
    - the user's own workstation,
    - an idle workstation belonging to someone else,
    - one of the unassigned processors in the machine room
  - the system as a whole looked and acted like a classical single-processor timesharing system

Modern definition (no general agreement)

- *A distributed system is an information-processing system that contains a number of independent computers that cooperate with one another over a communications network in order to achieve a specific objective.*

- Aspects:
    - Computers are linked to one another over a communications network that enables an exchange of messages between computers.
    - Objective of this message exchange is to achieve a cooperation between computers for the purpose of attaining a common goal.

Views

- Physical view: computers as nodes of the communications network along with details about the communications network itself

- Logical view:
  - applications aspects
  - interpreted as a set of cooperating processes
  - logical distribution is independent of the physical one: eg: processes do not necessarily have to be linked over a network but instead can all be found on one computer.

Distributed vs. Parallel

The Computing Continuum

Loosely Coupled ← Special Purpose "SETI / Google" | "Grids" | Clusters | Highly Parallel → Tightly Coupled

Advantages of Distributed Systems over Centralized Systems (1)

- Decentralization is a more economic:
  - networked computing systems offer a better price/performance ratio than centralized systems
  - redundancy increases availability when parts of a system fail
  - applications that can easily be run simultaneously also offer benefits in terms of faster performance vis-à-vis centralized solutions
  - distributed systems can be extended through the addition of components, thereby providing better scalability compared to centralized systems.

Advantages of Distributed Systems over Centralized Systems (2)

Item	Description
Economics	Better price/performance rate for networked computers than central.
Speed	A distributed system may have more total computing power than one
Inherent distribution	Some applications involve spatially separated machines
Reliability	If one machine crashes, the system as a whole can still survive
Incremental growth	Computing power can be added in small increments

Advantages of Distributed Systems over Independent PCs

Item	Description
Data sharing	Allow many users access to a common data base
Device sharing	Allow many users to share expensive devices
Communication	Make human-to-human communication easier
Flexibility	Spread the workload over available machines in the most cost effective way

Advantages of Distributed Computing Environment over Standalone Application

1. ***Higher performance:*** Applications can execute in parallel and distribute the load across multiple servers.
2. ***Collaboration:***   Multiple applications can be connected through standard distributed computing mechanisms.
3. ***Higher reliability & availability:*** Applications or servers can be clustered in multiple machines.
4. ***Scalability:***   By deploying reusable distributed components on powerful servers.
5. ***Extensibility:*** Dynamic (re)configuration of applications distributed across network.
6. ***Higher productivity & lower development cycle time:*** Breaking up large problems into smaller ones, these individual components can be developed by smaller development teams in isolation.
7. ***Reuse.*** Services that can potentially be used by multiple client applications.
8. ***Reduced cost.*** Due to the reuse of once developed components that are accessible over the network

Disadvantages of Distributed Systems

Item	Description
Software	Complexity of programming distributed systems
Networking	The network can saturate or cause other problems
Security	Easy access also applies to secret data