

Distributed Systems

Section # 1 Introduction to Distributed Systems and Characterization

Distributed Systems

- We define a distributed system as one in which hardware or software components located at networked computers that communicate and coordinate their actions only by passing messages. Computers that are connected by a network may be spatially separated by any distance.
- A distributed system is a collection of independent computers that appear to the users of the system as a single computer.

Consequences of Distributed Systems

1. **Concurrency:** The coordination of concurrently executing programs that share resources is a major challenge in the world of distributed systems. Usually this problem is addressed by message passing or synchronization.
2. **No Global Clock:** When programs need to cooperate, they coordinate their actions by exchanging messages. Close coordination often depends on a shared idea of the time at which the programs' actions occur. But it turns out that there are limits to the accuracy with which the computers in a network can synchronize their clocks – there is no single global notion of the correct time.
3. **Independent Failures:** Each component of the system can fail independently, leaving the others still running. And it is the responsibility of system designers to plan for the consequences of possible failures.
4. **Heterogeneity:** All the computers in a distributed system aren't similar, they may have different architectures, different OS etc.

Cluster

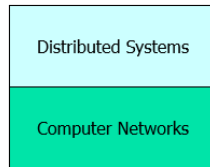
A type of parallel or distributed processing system, which consists of a collection of interconnected stand-alone computers cooperatively working together as a single, integrated computing resource.

Cloud

A type of parallel and distributed system consisting of a collection of interconnected and virtualised computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers.

Network vs. Distributed Systems

- **Networks:** A media for interconnecting local and wide area computers and exchange messages based on protocols. Network entities are visible, and they are explicitly addressed (IP address). Networks focuses on packets, routing, etc., whereas distributed systems focus on applications.
- **Distributed System:** existence of multiple autonomous computers is transparent. Every distributed system relies on services provided by a computer network.



Reasons for Distributed Systems

- **Functional Separation:** Existence of Computers with different capabilities and purposes. These can be shared between computers using distributed systems e.g. printers etc.
- **Inherent Distribution:**
 - **Information:** Different information is stored and maintained by different computer and needs to be accessed by others
 - **People:** Computer Supported collaborative work
 - **Retail:** Inventory chains for companies
- **Power Imbalance and Load Variation:** Computational load is shared between computers
- **Reliability:** Data is backed up at different locations
- **Economies:** Shared resources reduce the overall cost of systems

Characteristics of Distributed Systems

- **Parallel Activities:** Autonomous components executing numerous tasks at the same time
- **Communication via Message Passing:** The components don't have a shared memory, they communicate via message passing
- **Resource Sharing:** Hardware and software resources can be shared between computers
- **No Global State:** No single process can have knowledge of the current global state of the system
- **No Global Clock:** Distributed systems have no global clock, only limited precession for processes to synchronize their clocks

Section # 9 Name Services

In a Distributed System, a Naming Service is a specific service whose aim is to provide a consistent and uniform naming of resources, thus allowing other programs or services to localize them and obtain the required metadata for interacting with them.

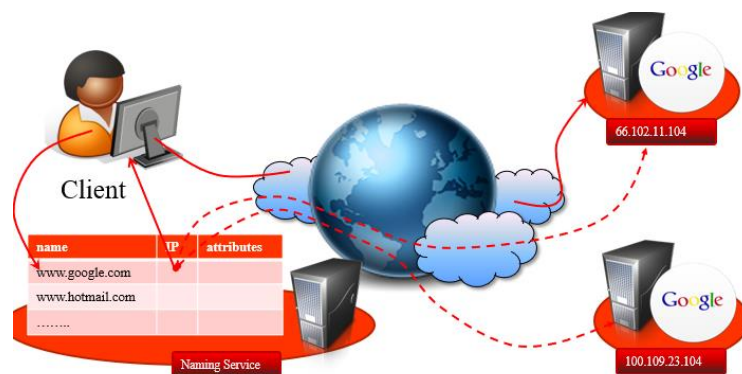
- In a distributed system, names are used to refer to a wide variety of resources such as:
 - Computers, services, remote objects, and files, as well as users.
- Naming is fundamental issue in DS design as it facilitates communication and resource sharing.
 - A name in the form of URL is needed to access a specific web page.
 - Processes cannot share particular resources managed by a computer system unless they can name them consistently
 - Users cannot communicate within one another via a DS unless they can name one another, with email address.
- Naming Services are not only useful to locate resources but also to gather additional information about them such as attributes

Advantages

- Resource localization
- Uniform naming
- Device independent address (e.g., you can move domain name/web site from one server to another server seamlessly).

Different Naming Services

- An URL facilitates the localization of a resource exposed on the Web.
 - e.g., *abc.net.au* means it is likely to be an Australian entity?
- A consistent and uniform naming helps processes in a distributed system to interoperate and manage resources.
 - e.g., *commercials use .com; non-profit organizations use .org*
- Users refers to each other by means of their names (i.e. email) rather than their system ids



Uniform Resource Identifier

Uniform Resource Identifiers (URI) offer a general solution for any type of resource. There two main classes:

- URL – Uniform Resource Locator (URL)
 - typed by the protocol field (http, ftp, nfs, etc.)

- part of the name is service-specific
 - resources cannot be moved between domains
- *URN – Uniform Resource Name (URN)*
 - requires a universal resource name lookup service - a DNS-like system for all resources
 - *format: urn:<nameSpace>:<name-within-namespcae>*
 - *examples:*
 - *urn:ISBN:021-61918-0*
 - *urn:cloudbus.unimelb.edu.au:TR2005-10*