

Introducción y conceptos

Alvaro Ospina Sanjuan
alvaro.ospina@upb.edu.co

Antecedentes – Inicio de la computación

- La idea de distribuir la computación no es nueva
 - Sistemas mainframe con terminales brutas
 - Bancos de cintas para almacenamiento de la información
- Conforme la tecnología avanza el ideal de la distribución cambia

Antecedentes – Avances de los 80s

- Desarrollo de microprocesadores
 - De una máquina de 10 millones de dólares que ejecuta una instrucción por minuto, se pasa a máquinas de 1000 dólares que ejecutan 10 millones de instrucciones por segundo.
- Invención de redes
 - Posibilidad de conectar uno o más computadoras entre sí

Antecedentes - Consecuencias

- Es relativamente fácil y económico implantar una serie de equipos con múltiples procesadores en red
- Necesidad de sincronizar los trabajos de cada componente
 - Dos entidades trabajan más rápido que una sola
- Dos desarrollos importantes que han cambiando de manera dramática e irrevocable el panorama de la computación: la expansión dramática en el uso de Internet y el diseño de computadores multinúcleo

Aplicación en red

- Es un sistema distribuido implementado en una red
 - Sistemas de reservas (Aéreas, Cines)
 - Sistema de consulta de saldos
 - WWW
 - Sistema distribuido Multimedia (Radio)
 - DNS
 - Sensores para monitorizar cambios estructurales de infraestructuras (como puentes, montañas)

Sistema Distribuido

Sistema cuyos **componentes están ubicados en diferentes redes**, que se comunican y coordinan sus acciones mediante el **paso de mensajes** que dan al usuario la impresión de constituir un **único sistema coherente**.

Beneficios

- Los principales beneficios de los sistemas distribuidos se dan principalmente en las siguientes 3 áreas:
 - Rendimiento
 - escalabilidad
 - disponibilidad
- Permiten lograr un mejor rendimiento a un menor costo.

Implicaciones/Desafíos

- Un sistema distribuido consta de diversos componentes (computadoras/nodos) autónomos
- Los usuarios (personas o programas) creen que interactúan con un computador único
- Colección de nodos ¿Cómo gestionar la pertenencia a un grupo? ¿Cómo saber si realmente se está comunicando con un (no) miembro autorizado?

Otras definiciones

- Un sistema distribuido es un sistema en el que al fallar una computadora que ni se sabía que existía puede hacer que el sistema NO quede completamente inservible

• Leslie Lamport

- Un sistema distribuido es aquel que NO deja de trabajar cuando un equipo completamente desconocido falla

• Sape Mullender

Objetivos para distribuir

- Accesibilidad
 - Recursos fácilmente accesibles
- Transparencia
 - Ocultar que los recursos están distribuidos
- Apertura
 - Estandares
- Escalabilidad
 - A su tamaño, ubicación, administración

Obj1 Accesibilidad

- Facilidad de acceso a los sistemas remotos sin importar que sean estos
 - Economía
 - Eficiencia
 - Facilidad
- Problemas de seguridad

Obj 2 Transparencia

- Un sistema es transparente si es capaz de mostrarse a los usuarios como una sola computadora
- No es siempre requerida

Tipos de transparencia

- **De acceso**
 - Oculta diferencias de representación de los datos
- **Ubicación**
 - Oculta la localización
 - Nomenclaturas
- **Migración**
 - Oculta que un recurso pueda moverse

Tipos de transparencia

- **Reubicación (movilidad)**
 - Migración mientras está en uso
- **Replicación**
 - Oculta el número de copias de un recurso
- **Concurrencia**
 - Oculta que un recurso puede ser compartido
- **Falla**
 - Oculta la falla o recuperación de un recurso

Notas sobre la transparencia

- No es siempre deseada sobretodo ante el reporte de errores
- Puede presentar problemas en sistemas globales con varios usos horarios
- Las diferentes replicas de un recurso tienen problemas con la transparencia

Obj 3 Grado de apertura

- Oferta de servicios de acuerdo a reglas estándar
 - P.e. redes de computadoras con protocolos estándar
- Especificación por medio de interfaces
 - **Lenguaje de definición de interface o IDL**
 - Deben ser completas y neutrales
 - Deben garantizar la portabilidad y la interoperatibilidad

Obj 4 Escalabilidad

- Se mide en tres dimensiones
 - Con respecto a su tamaño
 - Geográficamente
 - Administrativamente
- Por lo general al ser escalable en un parámetro pierde rendimiento

Algoritmos Descentralizados

- Ninguna máquina tiene información completa con respecto al estado del sistema
- Las máquinas toman decisiones con respecto a la información local
- La falla de una sola máquina no daña el algoritmo
- No existen suposiciones con respecto a un reloj global

Problemas de las escalabilidad

- Información confidencial
- Bloqueos en la comunicación síncrona que puede tener alta latencia
- Comunicación no fiable

Técnicas de escalamiento

- Ocultar las latencias
 - Evitar las esperas de respuestas a peticiones remotas
 - Verificación completa de formularios
- Ocultar la distribución
 - División de componentes en partes más pequeñas
 - Zonas de DNS
- Ocultar la replicación
 - Uso de cache

Suposiciones erróneas

- Declarados por Peter Deutch
 1. La red es confiable
 2. La red es segura
 3. La red es homogénea
 4. La topología no cambia
 5. La latencia es igual a cero
 6. El ancho de banda es infinito
 7. El costo de transporte es igual a cero
 8. Existe un administrador
 9. Utiliza el cluster para agilizar

Hasta ahora..temenos que los Sistemas Distribuidos

- Conjunto de entidades que se comunican entre ellos a través de mensajes, los cuales son enviados sobre vías de comunicación.
- Entidades:
 - procesos, computadoras, redes computadoras
 - dispositivos, procesadores etc..

Características de los SD

- Ausencia de memoria común
- Sincronización del trabajo
- Ausencia de un estado global perceptible por un observador
- Comunicación a través de mensajes

Comunicación

- Componentes separados lógicamente y físicamente, por lo que necesitan comunicarse entre ellos para interactuar
- Comunicación involucra las operaciones siguientes:
 - La transferencia de datos
 - La sincronización de la recepción con la emisión
- Dos enfoques en sistemas distribuidos:
 - Paso de mensajes
 - Llamado de procedimiento remoto
- Existen dos modelos:
 - Modelo de comunicación par a par
 - Modelo de comunicación grupal

Otros Aspectos importantes

- Tolerancia a fallas
- Confiabilidad
- Disponibilidad
- Concurrencia
- Reconfiguración

Tolerancia a fallos

- Sistema distribuido que puede seguir funcionando, (tal vez con un menor desempeño), a pesar de que uno de sus componentes no este funcionando, (sistemas robustos)
- Se tienen dos enfoques:
 - Redundancia hardware
 - uso de componentes redundantes
 - Recuperación software
 - diseño programas para recuperarse de fallas

Confiabilidad

- Datos transmitidos a través vías comunicación
- Posibilidad de pérdida y modificación de los datos

Disponibilidad

- Disponibilidad de un sistema es medida de acuerdo a la proporción del tiempo en el cual esta disponible para su uso
- Falla en una sola computadora multiusuario da como resultado la no-disponibilidad del sistema para todos sus usuarios
- Cuando uno de los componentes falla en un sistema distribuido solo el trabajo que estaba usando el componente es afectado
- Usuario puede moverse a otra estación si la que usa falla, o un servidor puede reiniciarse en otra computadora

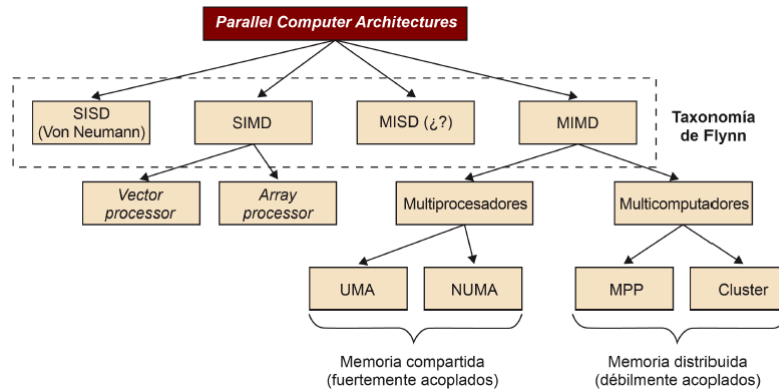
Concurrencia

- Varios procesos se encuentran sobre una sola computadora
- Ejecución intercalada en el caso de un solo procesador y simultanea si existen n procesadores
- Ejecución paralela posible debido a:
 - Varios usuarios invocan comandos o interactúan con programas de aplicación
 - Varios procesos servidores se corren concurrentemente

Reconfiguración

- La expansión del sistema no debe de afectar al sistema
- Habilidad para acomodar cambios en cualquier escala se conoce como reconfigurabilidad

Clasificación de Flynn



Clasificación de las arquitecturas según Andrew S. Tanenbaum

Acoplamiento

Fuertemente

- Corto retraso envío de mensajes
- Transmisión de datos alta
- Equipos conectados mediante cables insertados en tarjetas
- Suelen utilizarse más como sistemas paralelos.
- Intercambio de datos a la velocidad de la memoria.

Débilmente

- Grandes retrasos envío de mensajes
- Transmisión de datos baja
- Dos computadoras conectadas mediante la red
- Suelen utilizarse como sistemas distribuidos (<> problemas no relacionados entre si)
- Algunas fibras pueden funcionar a la velocidad de memoria.

En resumen para el curso: Tipos de Sistemas Distribuidos

- SD de cómputo
 - Clúster
 - Grid (Malla)
- SD de información
 - Procesamiento de transacciones
 - Integración de aplicaciones empresariales
- SD embebidos