

Informe Tecnico 1: La arquitectura computacional en el contexto de automatización e integración de datos

Marcela Durán Leal¹, Maria Alejandra Fandiño², Jhuliana Bueno Duarte³

Facultad de Ingeniería y Ciencias Básicas
Universidad Central
Maestría Analítica de Datos
Automatización e Integración datos para IA
Bogotá, Colombia
`{1mduranl,2mfandinom,3jebuenod}@ucentral.edu.co`

August 26, 2023

1 Resumen o resumen ejecutivo

La sesión se centró en los principios básicos de la arquitectura computacional. Durante la clase, investigamos los conceptos clave que respaldan esta arquitectura, como los componentes fundamentales, la máquina de Turing y la arquitectura de Von Neumann. También analizamos las partes que constituyen los dispositivos computacionales, dividiéndolos en tres elementos esenciales: Hardware, Sistemas operativos y Aplicaciones / Servicios.

Profundizamos en el hardware, evaluando las funciones de la CPU, la RAM y el disco duro, y como cada uno de estos componentes impacta en el rendimiento global del computador. A su vez, revisamos cómo se soluciona a nivel global los límites de la capacidad de cómputo, a través de la conexión de dispositivos en la red, y los problemas que se pueden presentar durante la comunicación, como funcionan los medios físicos y su capacidad, a partir de esto revisamos el concepto del modelo TCP/IP y sus capas.

2 Índice o tabla de contenidos

- Introducción
- Contexto y antecedentes

- Metodología
- Desarrollo
- Resultados y análisis
- Conclusiones
- Referencias
- Anexos

3 Introducción

La arquitectura computacional es esencial en la automatización y la integración de datos. En esta sesión se exploró los fundamentos clave, desde los componentes básicos hasta la conectividad global. Abordamos la relación entre la capacidad de procesamiento y la transmisión de datos, así como la relevancia de la arquitectura en la eficiencia y escalabilidad de los sistemas.

4 Contexto y antecedentes

La importancia de la arquitectura computacional dentro del contexto de la automatización e integración de datos radica en que nos proporciona la base tecnológica necesaria para implementar con éxito sistemas de automatización e integración de datos, permitiendo que los procesos sean más eficientes, seguros y adaptables.

Principalmente podemos ver la relevancia en lo siguiente:

- **Eficiencia y Rendimiento:** Una arquitectura computacional adecuada permite diseñar sistemas que sean eficientes en términos de uso de recursos como el procesador, la memoria y el almacenamiento. Esto es crucial para obtener un rendimiento óptimo en el momento de procesar gran cantidad de información en un entorno de automatización.
- **Escalabilidad:** La automatización y la integración de datos a menudo implican el manejo de volúmenes crecientes de información, con una arquitectura bien diseñada se puede escalar de manera eficiente para manejar estas demandas sin sacrificar el rendimiento.
- **Integración:** Una arquitectura sólida facilita la integración de los sistemas y datos de manera fluida. Puede proporcionar interfaces estandarizadas, protocolos de comunicación y estructuras de datos comunes que permiten que diferentes sistemas intercambien información

5 Metodología

La sesión inició con una explicación acerca de que es la arquitectura computacional básica, en donde se abordaron los temas precedentes que dieron paso a la creación de los dispositivos.

A partir de allí, se explicaron los componentes de las computadoras modernas, su función y el impacto que tienen cada uno de ellos en el momento de evaluar el rendimiento de un dispositivo.

Para entender acerca del rendimiento hicimos un laboratorio en el que cerramos todo y revisamos el performance de la máquina. Presionando Ctrl + alt + Suprimir / Administrador de tareas entramos a ver el performance de la máquina.

Luego, empezamos a abrir pestañas de youtube en google monitoreando el rendimiento del equipo, observando que a mayor cantidad de pestañas abiertas era mayor el esfuerzo que hacía la máquina, ya que aumentaba el porcentaje de uso para la CPU y memoria. La recomendación para que el performance de los equipos se mantenga es que el límite de uso sea el 70

Retomamos la explicación acerca de cómo se soluciona a nivel global los límites de computación y la capacidad de cómputo, por lo que abordamos el tema de la conexión de varios dispositivos a la red, y el problema de comunicación de red iniciando por los medios físicos de transmisión, en donde vimos que se divide en par de cobre (LAN), fibra optica y redes inalámbricas.

Para entender el transporte de los datos a través de la red hicimos dos ejercicios. En uno por medio de speed test revisamos la velocidad de internet de los equipos de la universidad y la de los equipos celulares y comparar sus resultados.

Para entender lo anterior podríamos usar este ejemplo, tengo una app en mi celular para tomar una foto y la envío. El ancho de banda del celular 5MBPs sale a 10 GBps llega a la red submarina a 10GBps y en un servidor de Estados Unidos el servidor atiende a 1 GBps. Al final la velocidad es dada por la menor capacidad del sistema. Eso significa que en la transmisión influyen dos cosas, la capacidad de mi sistema y la capacidad de transmisión. La capa IP permite el movimiento de la información por la red de internet. Las ip se componen por cuartetos de números.

Para comprender lo anterior hicimos un ejercicio en el que indagamos cuánto se demora en conectar desde donde estábamos a los servidores de diferentes partes del mundo. Por medio de CMD usamos el comando tracert y luego incluimos la dirección a la que queremos acceder. El resultado es la ruta que utilizan los protocolos para acceder a la dirección IP que contiene la página,

así como los milisegundos que requiere para acceder. Con lo anterior entendemos que la información del nombre comercial se traduce a direcciones IP para comunicarse.

También usamos los comandos Ipconfig y nslookup, el primero sirve para revisar la dirección IP que tienes en la PC que estás usando, el segundo es para conocer la dirección real de internet.

Finalmente se terminó con una explicación acerca del modelo internacional para conectarse a internet y la funcionalidad de cada una de sus capas.

6 Desarrollo

Dentro del ámbito de comprensión de la arquitectura computacional, se exploraron diversos conceptos y se plantearon preguntas durante la realización de los laboratorios. A continuación, se presentan dichos cuestionamientos:

Arquitectura Computacional Básica - Von Neumann

La arquitectura computacional básica, también conocida como arquitectura de von Neumann, es el modelo fundamental que define cómo funcionan las computadoras modernas. Fue propuesto por el matemático y científico de la computación John von Neumann en la década de 1940. Esta arquitectura consta de varios componentes clave interconectados que trabajan en conjunto para ejecutar programas y procesar datos.

- **Unidad Central de Procesamiento (CPU):** La CPU es el cerebro de la computadora y se encarga de ejecutar instrucciones. Se compone de dos partes principales: la Unidad de Control y la Unidad Aritmético-Lógica (ALU). La Unidad de Control coordina el flujo de datos y controla la ejecución de instrucciones, mientras que la ALU realiza operaciones aritméticas y lógicas en los datos.
- **Memoria:** La memoria se utiliza para almacenar tanto las instrucciones del programa como los datos que se están procesando. En la arquitectura de von Neumann, tanto las instrucciones como los datos se almacenan en la misma memoria principal. Esta memoria es accesible por la CPU para leer y escribir datos.
- **Unidad de Entrada/Salida (E/S):** Esta unidad se encarga de la comunicación entre la computadora y el mundo exterior. Permite la entrada de datos desde dispositivos como teclados y mouse, así como la salida de datos hacia dispositivos como pantallas e impresoras.
- **Bus de Datos:** El bus de datos es un conjunto de líneas de comunicación que permiten que los datos se transfieran entre los componentes de la

computadora, como la CPU, la memoria y los dispositivos de E/S. Estas líneas transportan la información en forma de bits.

- **Bus de Direcciones:** El bus de direcciones se utiliza para especificar la ubicación de memoria o dispositivos de E/S a los que la CPU desea acceder. Indica la dirección en la que se leerán o escribirán los datos.
- **Unidad de Control:** La Unidad de Control es responsable de interpretar las instrucciones almacenadas en la memoria y coordinar las operaciones de la CPU en función de esas instrucciones. Controla el flujo de datos y señales en la computadora.

Máquina de Turing

Una máquina de Turing es un concepto en la teoría de la computabilidad y la ciencia de la computación propuesto por el matemático y lógico Alan Turing en 1936. Es un modelo teórico que se utiliza para estudiar la capacidad de las máquinas abstractas para realizar cálculos y resolver problemas. La máquina de Turing es esencialmente una representación matemática de un dispositivo de cálculo que puede realizar una secuencia de pasos definidos para manipular símbolos en una cinta de manera controlada.

La máquina de Turing consta de varios componentes clave:

- **Cinta infinita:** Es una cinta bidimensional dividida en celdas, cada una de las cuales puede contener un símbolo de un conjunto finito. La cinta es infinita en ambas direcciones y se utiliza para almacenar y manipular los datos.
- **Cabeza lectora/escritora:** Una cabeza móvil que puede leer o escribir símbolos en la cinta y moverse hacia la izquierda o hacia la derecha.
- **Estado:** La máquina puede estar en uno de varios estados, y el estado actual junto con el símbolo bajo la cabeza influye en las acciones que la máquina realizará a continuación.
- **Tabla de transiciones:** Esta tabla establece cómo la máquina de Turing debe comportarse en función del estado actual y el símbolo bajo la cabeza. Indica qué símbolo escribir en la cinta, cómo mover la cabeza y en qué estado pasar a continuación.

La idea detrás de la máquina de Turing es que puede simular cualquier algoritmo computacional, lo que implica que cualquier problema que pueda ser resuelto por un proceso algorítmico también puede ser resuelto por una máquina de Turing. Esto ayudó a Turing a definir la noción de computabilidad y demostrar que ciertos problemas son intrínsecamente insolubles por cualquier máquina de cálculo (como el famoso Problema de la Parada).

¿Cuál es la capacidad de procesamiento en una CPU?

La capacidad de procesamiento de una CPU (Unidad Central de Procesamiento) se refiere a su capacidad para ejecutar instrucciones y realizar operaciones en un período de tiempo determinado. Esta capacidad se mide en términos de la velocidad a la que la CPU puede ejecutar instrucciones y realizar cálculos. La velocidad de procesamiento de una CPU se expresa generalmente en Hertz (Hz) o en múltiplos de Hertz, como kilohercios (kHz), megahercios (MHz) o gigahercios (GHz).

¿Cuál es la velocidad de los procesamientos actuales?

Las CPUs para computadoras de consumo generalmente tenían velocidades que oscilaban entre varios gigahercios (GHz). Las CPUs de gama alta, como las utilizadas en estaciones de trabajo y servidores, podían superar los 4 GHz e incluso llegar a 5 GHz en algunos casos.

¿Cómo se mide los volúmenes de información?

Los volúmenes de información se miden comúnmente en unidades de medida específicas que representan diferentes niveles de tamaño y capacidad. Aquí hay algunas de las unidades de medida más utilizadas para medir volúmenes de información:

- **Bit (b):** Es la unidad más pequeña de información y puede tener dos valores posibles: 0 o 1. Se utiliza para representar la información a nivel binario.
- **Byte (B):** Un byte es una agrupación de 8 bits. Es la unidad básica de almacenamiento y representación de datos en muchos sistemas informáticos.
- **Kilobyte (KB):** Equivale a 1024 bytes. Aunque en el sistema decimal 1 kilobyte es igual a 1000 bytes, en el contexto informático se utiliza comúnmente la base binaria (1024) debido a la estructura interna de los sistemas de almacenamiento y memoria.
- **Megabyte (MB):** Equivale a 1024 kilobytes o 1,048,576 bytes. También se usa tanto en el sistema decimal como en el binario.
- **Gigabyte (GB):** Equivale a 1024 megabytes o 1,073,741,824 bytes. Esta unidad es ampliamente utilizada para medir la capacidad de almacenamiento de discos duros y otros dispositivos de almacenamiento.
- **Terabyte (TB):** Equivale a 1024 gigabytes o 1,099,511,627,776 bytes. Se utiliza para medir el almacenamiento en sistemas más grandes, como servidores y centros de datos.

- **Petabyte (PB):** Equivale a 1024 terabytes o 1,125,899,906,842,624 bytes. Se emplea para medir capacidades de almacenamiento a gran escala, como en centros de datos de empresas y organizaciones.
- **Exabyte (EB):** Equivale a 1024 petabytes o 1,152,921,504,606,846,976 bytes. Se utiliza para describir grandes volúmenes de datos a nivel global.
- **Zettabyte (ZB):** Equivale a 1024 exabytes o 1,180,591,620,717,411,303,424 bytes. A medida que los volúmenes de datos siguen creciendo, esta unidad se vuelve relevante para describir capacidades de almacenamiento muy grandes.
- **Yottabyte (YB):** Equivale a 1024 zettabytes o 1,208,925,819,614,629,174,706,176 bytes. Esta es una unidad extremadamente grande y suele utilizarse en discusiones teóricas sobre el crecimiento exponencial de la información.

¿Cómo se mide la capacidad de transmisión de información?

La capacidad de transmisión de información se mide a través del concepto de "tasa de transmisión" o "velocidad de bits". Se refiere a la cantidad de información que puede ser transmitida a través de un canal de comunicación en un período de tiempo dado. La tasa de transmisión se mide en bits por segundo (bps) o en múltiplos de bits por segundo, como kilobits por segundo (Kbps), megabits por segundo (Mbps) o gigabits por segundo (Gbps).

¿Qué hace cada etapa del modelo ICP/IP?

El modelo TCP/IP (Transmission Control Protocol/Internet Protocol) es un conjunto de protocolos de red que se utiliza para la comunicación en Internet y en muchas redes privadas. El modelo TCP/IP consta de cuatro capas, cada una de las cuales tiene funciones específicas para garantizar la comunicación eficiente y confiable. A continuación, se describen las funciones de cada una de las capas del modelo TCP/IP:

- **Capa de Acceso a la Red (Network Access Layer):** - También conocida como Capa de Enlace de Datos o Capa de Red. - Se encarga de la transferencia de datos a nivel de bits sobre el medio físico, como cables, Ethernet, Wi-Fi, etc. - Define los protocolos y estándares para la comunicación entre dispositivos en la misma red local (LAN).
- **Capa de Internet (Internet Layer):** - También conocida como Capa de Red. - El protocolo principal en esta capa es el Internet Protocol (IP), que permite el enrutamiento de paquetes entre redes. - Se encarga de direccionar y enrutar paquetes de datos a través de la red global (Internet) o redes interconectadas. - Realiza la fragmentación y reensamblaje de paquetes en caso de que el tamaño exceda la capacidad de la red.

- **Capa de Transporte (Transport Layer):** - Se encarga de la administración del flujo de datos y la entrega confiable de datos de extremo a extremo. - Los protocolos más comunes en esta capa son el Transmission Control Protocol (TCP) y el User Datagram Protocol (UDP). - TCP proporciona un flujo de datos confiable y garantiza que los datos lleguen en orden y sin errores, aunque puede introducir cierta latencia. - UDP es menos confiable pero más rápido, se utiliza en aplicaciones que no requieren una entrega perfecta, como transmisiones en tiempo real.
- **Capa de Aplicación (Application Layer):** - Es la capa superior del modelo y se relaciona con las aplicaciones y servicios utilizados por los usuarios finales. - Incluye una variedad de protocolos para servicios como correo electrónico (SMTP, POP3, IMAP), transferencia de archivos (FTP), navegación web (HTTP, HTTPS), resolución de nombres de dominio (DNS), entre otros. - Esta capa es la que interactúa directamente con los usuarios y las aplicaciones, proporcionando una interfaz para acceder a los servicios de red. Cada capa en el modelo TCP/IP tiene un conjunto de protocolos y estándares específicos que trabajan juntos para permitir la comunicación entre dispositivos en una red. La interacción entre estas capas garantiza una comunicación eficiente, confiable y escalable en el entorno de Internet y otras redes basadas en TCP/IP.

¿Cuál es el puerto utilizado por HTTPS?

HTTPS (Hypertext Transfer Protocol Secure) es un protocolo utilizado para transmitir datos de forma segura a través de Internet. Utiliza una capa adicional de seguridad proporcionada por SSL (Secure Sockets Layer) o su sucesor, TLS (Transport Layer Security). El puerto estándar utilizado por HTTPS es el puerto 443.

¿Cuál es el puerto utilizado por SMTP?

SMTP (Simple Mail Transfer Protocol) es un protocolo utilizado para transmitir datos de forma segura a través de Internet. Utiliza una capa adicional de seguridad proporcionada por SSL (Secure Sockets Layer) o su sucesor, TLS (Transport Layer Security). El puerto estándar utilizado por SMTP es el puerto 443.

¿Cual es el puerto IP utilizado por ORACLE?

Oracle Database utiliza varios puertos para diferentes funciones y servicios, dependiendo de la configuración y el tipo de acceso que estés utilizando. Algunos de los puertos más comunes utilizados por Oracle Database son:

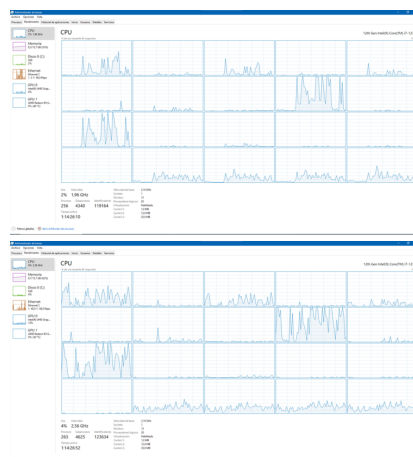
- **Puerto de Escucha (Listener):** El puerto de escucha es el puerto en el que el servidor de base de datos Oracle está esperando conexiones en-

trantes. El puerto de escucha por defecto es el 1521, pero esto puede variar según la configuración.

- **Puerto de Enterprise Manager Database Control:** Si estás utilizando Oracle Enterprise Manager Database Control para administrar tu base de datos, el puerto por defecto es el 1158.
- **Puerto de Oracle Enterprise Manager Cloud Control:** Si estás utilizando Oracle Enterprise Manager Cloud Control, el puerto por defecto es el 7800.
- **Puerto de Pluggable Database:** Si estás utilizando una arquitectura de base de datos con pluggable databases (PDBs), cada PDB puede tener su propio puerto de escucha.
- **Otros Puertos para Servicios Específicos:** Oracle Database puede tener otros servicios y componentes que utilizan puertos adicionales, como el Oracle Net Services, el servicio de notificación de cambios (Change Data Capture, CDC), entre otros.

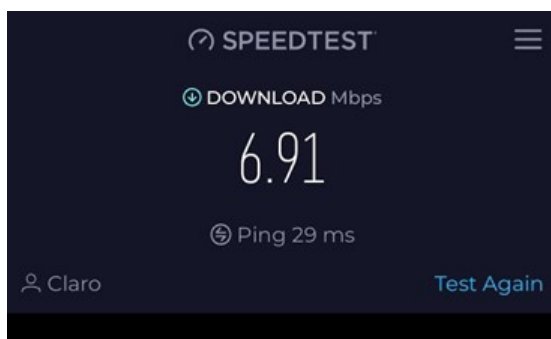
7 Resultados y análisis

En el primer laboratorio pudimos observar que entre más pestañas abiertas de diferentes páginas web, el esfuerzo en la máquina es mayor. Al inicio, cuando no había nada abierto, el performance era del 1% pero entre más pestañas se abrían, la arquitectura computacional empieza a exigir más recursos de la máquina, a usar más los procesadores y empieza a llenar la memoria Ram.



Con esto se puede ver la relación entre la capacidad de procesos que se ejecutan en el hardware y cómo afecta a un hardware específico de computación.

Para abordar el tema de la capacidad de información que se puede trasladar entre nodos, se realizó el laboratorio de medición de SpeedTest donde se comparó la información dada en los computadores que manejan fibra óptica frente a los celulares conectados a redes inalámbricas.



Al observar la comparación de los resultados obtenidos entre dispositivos, podemos entender que dependiendo de la conexión que tenga el servidor, van a tener un comportamiento diferente los servicios o aplicaciones que se estén ejecutando. Esto es lo que determina las capacidades que se tienen frente a un proceso.

El siguiente laboratorio que realizamos con el fin de evaluar cuánto se demora en conectar desde los servidores de la universidad a los servidores de diferentes partes del mundo mostró que tenemos que tener en cuenta esta condición de la transmisión física de los datos a través del mundo, que en su mayoría se centran en Estados Unidos y para llegar a distancias más lejanas requerirá mucho más

tiempo.

```
Paquetes: enviados = 4, recibidos = 0, perdidos = 4
(100% perdidos),
C:\Users\Sala_000>ping ub.co
La solicitud de ping no pudo encontrar el host ub.co. Compruebe el nombre y
resuelva a Internet.
C:\Users\Sala_000>tracert google.com
Tracert no se reconoce como un comando interno o externo,
programa o archivo por lotes ejecutable.
C:\Users\Sala_000>tracert google.com
Pasa a la dirección google.com [142.250.78.46]
sobre un máximo de 30 saltos:
  0  < 1 ms < 1 ms < 1 ms 172.20.35.182
  1  < 1 ms < 1 ms < 1 ms 172.20.35.182
  2  < 1 ms < 1 ms < 1 ms 172.17.174.185
  3  < 1 ms < 1 ms < 1 ms 172.17.174.185
  4  < 1 ms < 1 ms < 1 ms 172.17.174.185
  5  < 1 ms < 1 ms < 1 ms 172.17.174.185
  6  < 1 ms < 1 ms < 1 ms 172.17.174.185
  7  < 1 ms < 1 ms < 1 ms 172.17.174.185
  8  < 1 ms < 1 ms < 1 ms 172.17.174.185
  9  < 1 ms < 1 ms < 1 ms 172.17.174.185
 10  < 1 ms < 1 ms < 1 ms 172.17.174.185
Pasa completa.
```

```
Traza a la dirección google.com [142.250.78.46]
sobre un máximo de 30 saltos:
  0  < 1 ms < 1 ms < 1 ms 172.20.35.182
  1  < 1 ms < 1 ms < 1 ms 172.20.35.182
  2  < 1 ms < 1 ms < 1 ms 172.17.174.185
  3  < 1 ms < 1 ms < 1 ms 172.17.174.185
  4  < 1 ms < 1 ms < 1 ms 172.17.174.185
  5  < 1 ms < 1 ms < 1 ms 172.17.174.185
  6  < 1 ms < 1 ms < 1 ms 172.17.174.185
  7  < 1 ms < 1 ms < 1 ms 172.17.174.185
  8  < 1 ms < 1 ms < 1 ms 172.17.174.185
  9  < 1 ms < 1 ms < 1 ms 172.17.174.185
 10  < 1 ms < 1 ms < 1 ms 172.17.174.185
Pasa completa.
C:\Users\Sala_000>ipconfig
Configuración IP de Windows

Adaptador de Ethernet (Ethernet 2):
    Sufijo DNS específico para la conexión. . . :
    Dirección IPv4. . . . . : 172.20.112.1
    Máscara de subred. . . . . : 255.255.0.0
    Puerta de enlace predeterminada. . . . . : 172.20.35.182
```

```
C:\Users\Sala_000>nslookup google.com
Servidor: SERDAC.csi.ucentral.edu.co
Address: 172.20.129.250

Respuesta no autoritativa:
Nombre: google.com
Addresses: 2800:3f0:4005:40a::200e
142.250.78.46
```

8 Conclusiones

Los componentes de la computadora son los que permiten un mejor rendimiento del equipo, una buena relación entre cada uno nos permitirá un mejor performance.

Todos los sistemas funcionan de la misma forma. Un computador, el celular, los smartwatch. Para entender la capacidad de los sistemas y como los componentes de estos influyen debemos conocer cómo se mide el rendimiento de las máquinas.

Es importante conocer la arquitectura computacional en el contexto de automatización y la integración de datos debido a que estas son las herramientas que nos permiten transmitir la información y para realizar automatización es necesario conocer los límites a los que nos podemos enfrentar, el conocer esto nos permite generar soluciones compatibles y eficientes de acuerdo al contexto en el que se deba realizar la mejora.

El resultado del laboratorio de velocidad de internet nos permite entender que para que haya transporte la información se mide en cantidad de megabits por segundo que es capaz de transportar la red de acuerdo con el ancho de banda.

La actividad con el comando traceroute nos permite ver cómo se mueve la información por la red, cuántas IP debe atravesar para llegar a su destino y que los nombres comerciales de las páginas se traducen en direcciones IP para que se produzca el envío de información. A esto le llamamos protocolo de internet.

9 Referencias

- Andrew, S. Operacionais Modernos.
- Wetherall, D. J., & Tanenbaum, A. S. (2012). Redes de computadoras. Redes de Computadoras. Quinta ed. Mexico: Pearson Educación de México, SA de CV, 2.
- Andrew, S. (2000). Organización de computadoras.

10 Anexos

