

Informe Técnico N°3: Máquinas virtuales y Cloud Computing.

Alexander Victoria Garcia¹, Fernando Andrés Roa Martín²

Facultad de Ingeniería y Ciencias Básicas
Universidad Central
Maestria Analitica de Datos
Automatizacion e Integracion datos para IA
Bogotá, Colombia
`{1avictoriag,2froam1}@ucentral.edu.co`

September 9, 2023

1 Resumen o resumen ejecutivo

Se llevaron a cabo 3 exposiciones por parte de estudiantes cuyos temas fueron: Apache Spark, MariaDB y RStudio, siendo estos un framework para el procesamiento de datos masivos, un sistema de gestión de bases de datos y un entorno de desarrollo integrado para el lenguaje R respectivamente.

Utilizando Oracle VM Virtualbox se crearon dos máquinas virtuales, asignando recursos como la RAM, almacenamiento secundario, además de la descarga y suministro del respectivo disco de arranque del sistema operativo elegido para cada máquina virtual. Se realizó un esquema gráfico donde se mostraron los componentes que puede tener un proyecto final de la asignatura, haciendo hincapié en el uso de recursos como máquinas virtuales y bases de datos en la nube. La sesión de clase finalizó con una revisión general de la interfaz de Oracle Cloud.

Índice o tabla de contenidos

Contents

1	Resumen o resumen ejecutivo	1
2	Introducción	2
3	Contexto y antecedentes	2
4	Metodología	4
5	Desarrollo	5
6	Resultados y análisis	10
7	Conclusiones	10
8	Referencias	12

2 Introducción

Este informe técnico pretende explicar el concepto de máquinas virtuales (VM) y cloud computing, dos tecnologías con bastantes ventajas para la optimización de recursos. Una VM permite tener más de un entorno de software compartiendo el mismo hardware y el cloud computing brinda la posibilidad de acceder a recursos de software y hardware a través de internet pagando una cuota de acuerdo al uso.

Se describe además el procedimiento para la creación de máquinas virtuales mediante el software Oracle VM virtualbox, a la vez que se monitorea la variación del rendimiento y uso de recursos del ordenador a través del administrador de tareas al encender una máquina virtual. Finalmente, se realiza la inspección visual de la interfaz gráfica de una alternativa de cloud computing bastante popular, Oracle Cloud, a la vez que se enuncian algunos de los recursos gratuitos disponibles que pueden ser aprovechados en proyectos académicos y/o personales.

3 Contexto y antecedentes

Las máquinas virtuales (VM) representan versiones digitales de un ordenador físico, de acuerdo con la definición de Google Cloud. En esencia, las máquinas

virtuales actúan como interfaces de virtualización, lo cual permite ejecutar programas, sistemas operativos, aplicaciones, almacenar datos, establecer conexiones de red y llevar a cabo una amplia variedad de funciones informáticas.

Sin embargo, las máquinas virtuales no son sólo entornos de visualización, sino que también de ejecución, esto significa que además de simular sistemas operativos y aplicaciones, también pueden ejecutarlos de manera efectiva, a continuación se detallan algunas capacidades clave que ofrecen estas herramientas:

- **Gestión de Recursos:** Administran eficientemente los recursos de hardware, como CPU, memoria y almacenamiento, garantizando que los sistemas operativos y aplicaciones tengan acceso a los recursos necesarios y no se malgaste capacidad operativa.
- **Programación de Tareas:** Permiten programar tareas y procesos, lo que es fundamental para la optimización de recursos. Esto incluye la capacidad de programar la ejecución de scripts, tareas programadas y procesos de integración de datos en momentos específicos o en respuesta a eventos.
- **Programación de Tareas:** Permiten programar tareas y procesos, lo que es fundamental para la optimización de recursos. Esto incluye la capacidad de programar la ejecución de scripts, tareas programadas y procesos de integración de datos en momentos específicos o en respuesta a eventos.
- **Gestión de Archivos y Datos:** Proporcionan herramientas para la gestión de archivos y datos, esenciales para procesos que implican la lectura y escritura dentro del entorno virtual.
- **Escalabilidad y Administración de Recursos:** Permiten la gestión eficiente de múltiples máquinas virtuales, procesos y cargas de trabajo distribuidos según la asignación de recursos inicial a cada máquina virtual.
- **Compatibilidad y Estándares:** Trabajan con estándares y protocolos que facilitan la interoperabilidad entre aplicaciones y sistemas.

- Seguridad: Brindan aislamiento gracias a su ejecución en entornos virtuales separados, lo que disminuye la propagación de virus entre ellas y hacia el sistema anfitrión. También permiten la captura de snapshots o copias de seguridad que pueden ser utilizados para revertir infecciones.

De otra parte, la computación en la nube o Cloud Computing se originó en 1960, cuando John McCarthy propuso que algún día la informática sería realizada por empresas de servicios públicos a nivel nacional, dando origen a un paradigma tecnológico que revolucionará la forma en que almacenamos, procesamos y accedemos a datos y aplicaciones. En lugar de depender de recursos locales, la computación en la nube aprovecha servidores remotos a través de internet para ofrecer servicios flexibles y escalables según la capacidad requerida por el usuario, algunas de las ventajas clave de esta herramienta son las siguientes:

- Escalabilidad: Permite aumentar o disminuir recursos según las necesidades sin grandes inversiones de capital.
- Acceso Remoto: Facilita el acceso a datos y aplicaciones desde cualquier lugar con conexión a internet.
- Eficiencia y Costo: Reduce los costos asociados con la infraestructura física y el mantenimiento de servidores locales.
- Adopción Generalizada: Cloud computing se ha convertido en la columna vertebral de muchas industrias, desde startups hasta grandes corporaciones, proporcionando flexibilidad y agilidad para impulsar la innovación y el crecimiento empresarial.

4 Metodología

La metodología desarrollada durante la clase se muestra a continuación:

- Socialización del banco de noticias sobre inteligencia artificial.
- Exposiciones individuales.
- Conceptualización de máquinas virtuales.
- Creación de máquinas virtuales.

- Boceto de los componentes necesarios para los proyectos finales de automatización.
- Introducción al Cloud Computing mostrando la interfaz de Oracle Cloud.

5 Desarrollo

La sesión comenzó con una socialización sobre las últimas novedades en tecnología, con un enfoque particular en la inteligencia artificial, cuya duración fue de 10 minutos aproximadamente.

Posteriormente, se llevaron a cabo tres exposiciones individuales cuyos temas fueron: Apache Spark, MariaDB Server y R-Studio. Cada exposición tuvo una duración aproximada de entre 10 y 15 minutos. A modo de recordatorio se describe brevemente en qué consiste cada uno de los temas:

- Apache Spark. Es un framework de procesamiento de datos masivos de forma distribuida y en paralelo.



Figure 1: Apache Spark.

- MariaDB. Sistema de gestión de bases de datos de código abierto compatible con MySQL.



Figure 2: MariaDB.

- R Studio. Es un entorno de desarrollo integrado para el lenguaje R, con un buen repertorio de paquetes para estadística.



Figure 3: RSTUDIO.

Seguidamente, se inició la explicación detallada del funcionamiento de las máquinas virtuales, destacando cómo se integran en la infraestructura de la máquina principal, además de las ventajas de éstas como el uso de software y aplicaciones que ya no son compatibles con sistemas operativos actuales. Se enfatizó que las máquinas virtuales operan mediante la compartición de recursos de hardware del equipo anfitrión o principal.

Para comprender mejor el funcionamiento de las máquinas virtuales, se llevó a cabo un ejercicio práctico utilizando el software Oracle VM VirtualBox, permitiendo a los estudiantes experimentar de primera mano la creación y ejecución de una máquina virtual.

En el transcurso de la actividad, se crearon dos máquinas virtuales: una con el sistema operativo Windows y otra con un sistema operativo (SO) basado en Linux, donde los estudiantes podían escoger el de su preferencia. Durante este proceso, se seleccionaron y asignaron recursos específicos a cada máquina virtual, como lo era la capacidad de procesamiento, almacenamiento, además del suministro de un disco de arranque de SO, el cual para aquellos basados en linux no tiene costo. Como evidencia del proceso se muestran las siguientes imágenes.

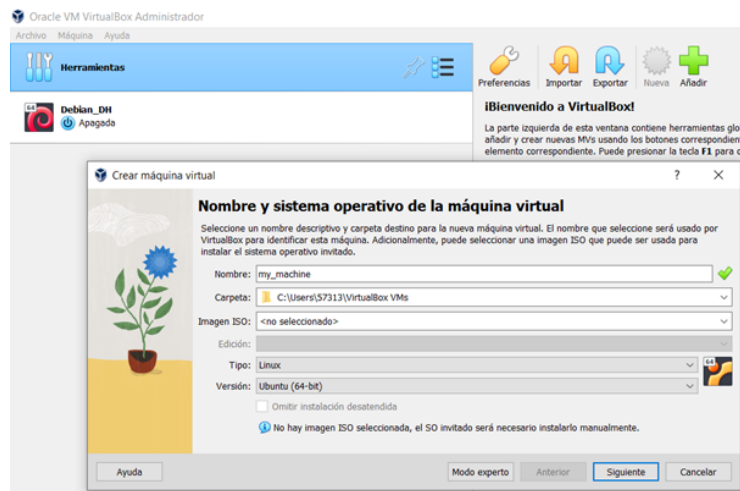


Figure 4: Creación de máquina virtual

Asignación del nombre de la máquina virtual, SO y versión o distribución del mismo.

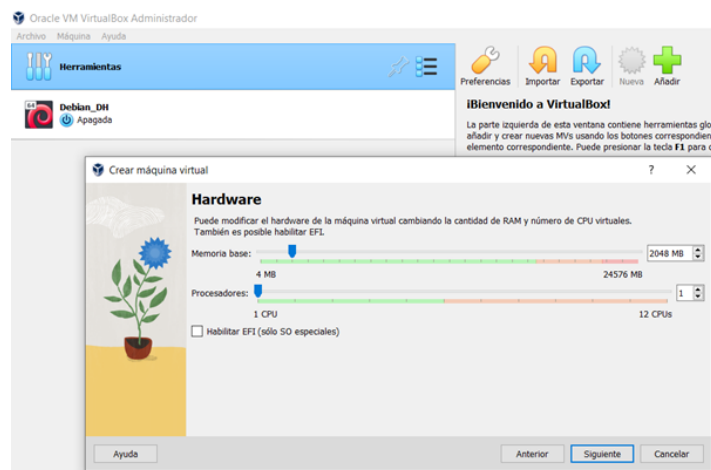


Figure 5: Asignación de RAM y procesador

Asignación de memoria ram de 2 GB y de procesador.

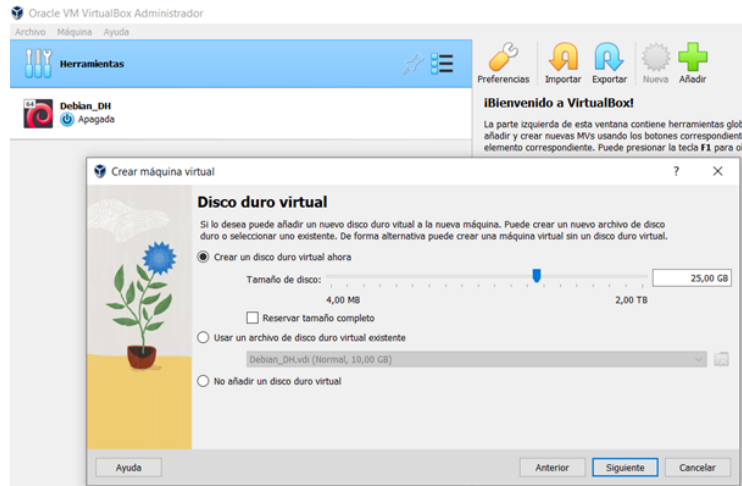


Figure 6: Asignación de recursos.

Asignación de un disco duro virtual de 25GB, sin marcar la opción de reservar el tamaño completo, lo que permite que el espacio se vaya modificando de forma dinámica.

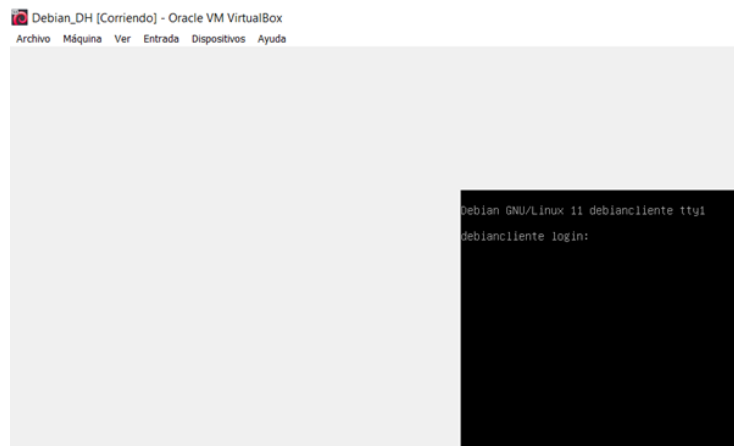


Figure 7: Máquina virtual encendida.

Máquina virtual de linux cuya distribución es Debian encendida.

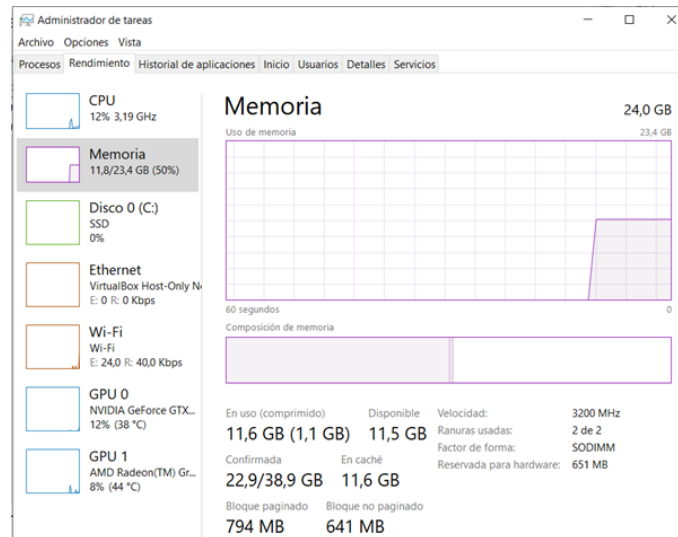


Figure 8: Monitoreo de recursos del ordenador

Monitoreo del uso de los recursos del ordenador, teniendo en cuenta que se ha iniciado la máquina virtual.

Esta experiencia proporcionó una comprensión más profunda sobre la operación del software Oracle VM VirtualBox para la creación de máquinas virtuales, además de las nociones vitales sobre criterios de asignación de recursos de hardware del ordenador principal.

Luego de crear las máquinas virtuales, se realizó un bosquejo sobre la estructura de los proyectos finales de automatización en la asignatura, resaltando componentes como las bases de datos, máquinas virtuales y containers necesarios para el end to end del proyecto.

La sesión finalizó con los conceptos básicos de cloud computing o computación en la nube, en donde en resumen, podemos acceder a recursos de hardware y software mediante el uso de internet, suministrados por compañías especializadas, pagando solamente aquellos recursos que se utilicen, cuyo costo puede ser gratis o tener un valor de acuerdo a las características de los mismos. El servicio en la nube sugerido para la clase es Oracle Cloud, que de forma gratuita permite la creación de dos bases de datos de gran capacidad y dos máquinas virtuales. En caso de que el proyecto requiera el uso de data no estructurada, el docente recomienda el uso de MongoDB Atlas, el cual cuenta con los recursos requeridos de forma gratuita.

6 Resultados y análisis

Se realizó la creación de las dos máquinas virtuales sin errores, apalancados de las claras indicaciones del docente, la intuitiva interfaz de la Herramienta Oracle VM Virtualbox y por supuesto, la holgada capacidad de hardware de los ordenadores utilizados en el laboratorio.

En el aspecto relacionado con el hardware del ordenador principal, es clave comprender la disponibilidad de recursos con un uso habitual al momento de crear una máquina virtual, en combinación con el propósito de uso de la misma y los recursos necesarios para SO, aplicaciones, procesamiento y almacenamiento. En este sentido, una excelente herramienta utilizada en la clase fue el monitoreo de recursos del ordenador a través del administrador de tareas, en donde a pesar de evidenciarse un incremento en el uso de los mismos, el componente más exigido fue la memoria ram, alcanzando apenas cerca de un 50% de la capacidad total.

Otro aspecto a destacar es que el sistema de arranque necesario para la correcta creación de la máquina virtual debe coincidir con el sistema y versión elegidos en el primer paso de creación de la misma, garantizando el éxito del proceso.

La contextualización de máquinas virtuales, brinda a los estudiantes herramientas para diversos usos, siendo uno básico la experimentación con nuevos sistemas operativos, además de que se introduce la idea de la existencia de las mismas en entornos de la nube.

Finalmente, al explicar de forma general lo que se espera del proyecto final de automatización, se introduce la necesidad del uso de recursos en la nube para la consecución de dicho trabajo y se fomenta a la creación de cuentas de usuario para que los estudiantes se familiaricen con este tipo de plataformas como Oracle Cloud, servicio en la nube con gran popularidad en la actualidad.

7 Conclusiones

El uso de máquinas virtuales y la implementación de computación en la nube son pasos significativos hacia la optimización y eficiencia en el ámbito de la tecnología de la información, las máquinas virtuales permiten gestionar de manera eficiente los recursos de hardware debido a que se pueden gestionar la capacidad de estos. De otro modo, el cloud computing proporciona acceso a recursos de software y hardware a través de internet, ofreciendo escalabilidad y flexibilidad sin grandes inversiones de capital.

La correcta configuración del sistema de arranque y la elección del sistema operativo son pasos críticos en el proceso de creación de máquinas virtuales, debido a que diferentes sistemas operativos tendrán costos adicionales.

Un factor esencial que se debe tener en cuenta es la capacidad de la máquina principal al crear máquinas virtuales. Si no se asignan correctamente los recursos, puede resultar en una subutilización de los mismos, lo que limitaría la eficacia y eficiencia del entorno virtual.

La adopción de Cloud Computing ofrece ventajas notables, como la capacidad de ajustar los recursos según las necesidades sin comprometer la estabilidad operativa, esto proporciona a los usuarios contar con una capacidad de procesamiento según sus requerimientos.

8 Referencias

- Oracle VM VirtualBox. (s/f). Oracle.com. Recuperado el 9 de septiembre de 2023, de <https://www.oracle.com/co/virtualization/virtualbox/>
- ¿Qué es cloud computing? (s/f). Google Cloud. Recuperado el 9 de septiembre de 2023, de <https://cloud.google.com/learn/what-is-cloud-computing?hl=es>
- Exposición Apache Spark, de https://github.com/faroamroa/automatizacion_de_datos/blob/main/apache_spark_exposicion.pdf
- Exposición MariaDB, de <https://docs.google.com/presentation/d/1mgSzxE40nxBaPL38y8GzS4ddk9DfqQzQudWdgA79X0k/edit?usp=sharing>
- Exposición RStudio, de https://github.com/diquinteroc/HV_juan_DiegoQuintero/blob/main/R