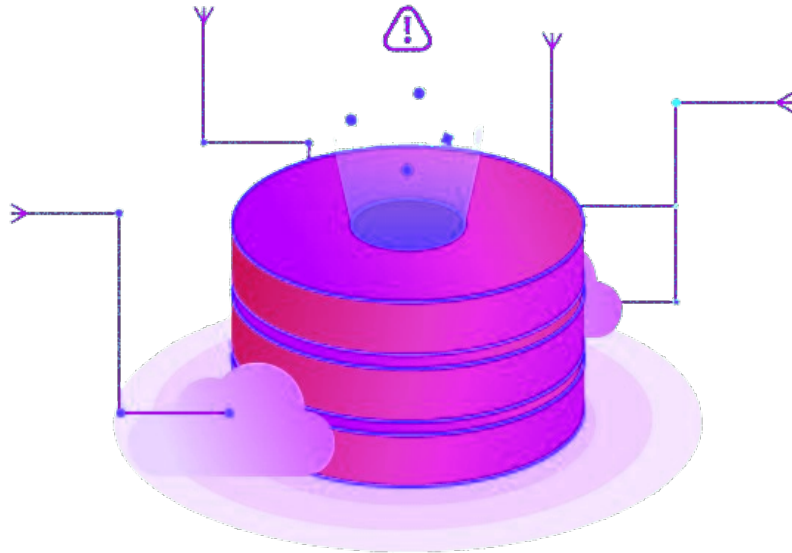


UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE CIENCIAS, 2024-I
ALMACENES & MINERÍA DE DATOS



PRÁCTICA 01:
Vistas Materializadas.

PROFESOR:
Gerardo Avilés Rosas

AYUDANTE DE TEORÍA:
Gerardo Uriel Soto Miranda

AYUDANTE DE LABORATORIO:
Ricardo Badillo Macías

Introducción.

Maquina Virtual

Las *Maquinas Virtuales* son **ordenadores de software que proporcionan la misma funcionalidad que los ordenadores físicos**. Como ocurre con los ordenadores físicos, ejecutan aplicaciones y un sistema operativo. Sin embargo, las máquinas virtuales son **archivos que se ejecutan en un ordenador físico, y se comportan como sistemas informáticos independientes**.

VirtualBox

VirtualBox es un potente producto que sirve para hacer máquinas virtuales con instalaciones de sistemas operativos para uso empresarial y doméstico. Es extremadamente rico en funciones, de alto rendimiento para clientes empresariales, también es la única solución profesional que está disponible gratuitamente como software de código abierto bajo los términos de la Licencia Pública General GNU.

Actualmente *VirtualBox* se ejecuta en hosts como **Windows, Linux, macOS y Solaris**, además admite una gran cantidad de sistemas operativos invitados, incluidos entre otros **Windows(NT 4.0, 2000, XP, Server 2003, Vista, Windows 7, Windows 8, Windows 10)**, **DOS/Windows 3.x**, **Linux (2.4, 2.6, 3.x y 4.x)**, **Solaris y OpenSolaris, OS/2 y OpenBSD**.



Figura 1: VirtualBox.

Microsoft SQL Server

Es un **sistema de manejador de bases de datos relacionales (RDBMS)** que admite una amplia variedad de aplicaciones de procesamiento de transacciones, inteligencia empresarial y análisis en entornos informativos corporativos.

SQL Server se construye principalmente en torno a una **estructura de tablas basada en filas que conecta los elementos de datos relacionados en diferentes tablas entre sí**, evitando la necesidad de almacenar datos de forma redundante en varios lugares dentro de una base de datos.



Figura 2: SQL Server.

Diferencias básicas entre SQL Server y PostgreSQL

SQL Server.	PostgreSQL.
Sistema manejador de bases de datos relacionales.	Sistema manejador de bases de datos relacionales de objetos
Producto comercial de Microsoft.	Código abierto (completamente gratuito)
Se ejecuta solo en Microsoft o Linux.	Se ejecuta en la mayoría de las máquinas y sistemas operativos.
Usa Transact-SQL o T-SQL (SQL estándar + funcionalidad adicional)	Usa SQL estándar.

Vistas Materializadas

Una **vista materializada** es una tabla de datos duplicada que se crea mediante la combinación de datos de varias tablas existentes para una recuperación de datos más rápida.

Una **vista materializada** se define como una **vista común, pero en lugar de almacenar la definición de la vista, almacena el resultado de la consulta**, es decir, la materializa, como un objeto persistente en la base de datos.

Los principales motivos para utilizar vistas materializadas son:

- Seguridad:** Evitar que se acceda a toda la información de la base de datos y solo mostrar un conjunto de datos específicos.
- Comodidad:** Si la estructura de nuestro modelo de datos es bastante complejo, las vistas nos permitiría organizar cierta información a modo que nos sea más comodo acceder a esta mediante consultas más sencillas.

Vistas vs Vistas Materializadas

Vista.	Vista materializada.
No se persisten datos, no ocupa espacio físico.	Se persisten los datos en la BD, por lo que ocupa espacio físico.
Mismo rendimiento que ejecutar query directamente.	Mejor rendimiento, los datos ya están cacheados
Cada vez que se ejecuta se obtiene una foto actualizada.	La foto de los datos es de cuando se generó o actualizó por última vez.

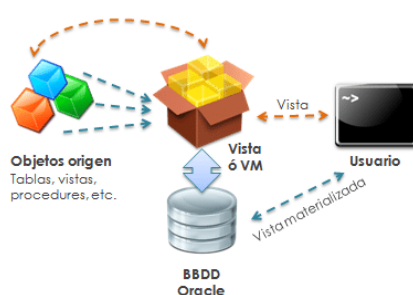


Figura 3: Vistas vs Vistas Ma.

Sintaxis

Para crear las vistas en SQL Server:

```
CREATE [OR REPLACE] VIEW nombre_vista AS  
SELECT select_query;
```

Para borrar vistas en SQL Server:

```
DROP VIEW nombre_vista;
```

Para crear las vistas materializadas en SQL Server:

```
CREATE MATERIALIZED VIEW nombreVista  
[REFRESH [FAST|COMPLETE|FORCE] [ON DEMAND|ON COMMIT]]  
[BUILD IMMEDIATE|BUILD DEFERRED]  
AS  
select_query;
```

Donde:

- **REFRESH FAST:** Utiliza un método de actualización incremental que utiliza los cambios realizados en las tablas subyacentes de un archivo de registro.
- **REFRESH COMPLETE:** Realiza una actualización completa volviendo a ejecutar la consulta en la vista materializada.
- **REFRESH FORCE:** Indica que se debe hacer una actualización FAST, si es posible, si no se realiza una actualización completa.
- **ON DEMAND:** Indica que se producirá una actualización manual, cada vez que se llamen a funciones específicas. (Esta viene predeterminado).
- **ON COMMIT:** Realiza una actualización rápida cada vez que se confirma una transacción que realiza cambios en cualquiera de las tablas (COMMIT).
- **BUILD IMMEDIATE:** La vista materializada se rellenará inmediatamente (predeterminado).
- **BUILD DEFERRED:** La vista materializada se rellenará en la siguiente operación de actualización.
- **AS select_query:** la consulta que se ejecuta, con los resultados almacenados en la vista materializada.

Para borrar las vistas materializadas en SQL Server:

```
DROP MATERIALIZED VIEW nombre_vista;.
```

Vistas Materializadas y Almacenamiento de Datos

Un **almacén de datos** contiene datos históricos detallados sobre la organización. Normalmente, los datos fluyen desde una o más bases de datos de procesamiento de transacciones en línea (*OLTP*) al almacén de datos mensual, semanal o diario. Los **almacenes de datos** suelen variar en tamaño desde decenas de gigabytes hasta unos pocos terabytes, **generalmente la mayoría de los datos son almacenados en tablas muy grandes**.

Una de las técnicas empleadas en los **almacenes de datos** para mejorar el rendimiento es la **creación de resúmenes o agregados**. Los cuales son un **tipo especial de vista agregada** que mejora los tiempos de ejecución de consultas mediante el cálculo previo de costosas uniones y operaciones de agregación antes de la ejecución, y el almacenamiento de los resultados en una tabla en la base de datos.

En los **almacenes de datos**, las **vistas materializadas** se pueden usar para **previamente calcular y almacenar los datos agregados**. Estas vistas en estos entornos suelen denominarse como **resúmenes**, ya que almacenan datos resumidos. También se pueden usar para **previamente calcular las uniones con o sin agregaciones**. Por lo tanto, **se utilizan para eliminar la sobrecarga asociada con uniones o agregaciones costosas para un gran cumulo de consultas**. El optimizador de consultas reconoce automáticamente cuándo una vista materializada existente puede y debe usarse para satisfacer una solicitud. Así que reescribir las consultas para usar vistas materializadas en lugar de tablas, puede ayudar a mejorar las respuestas.

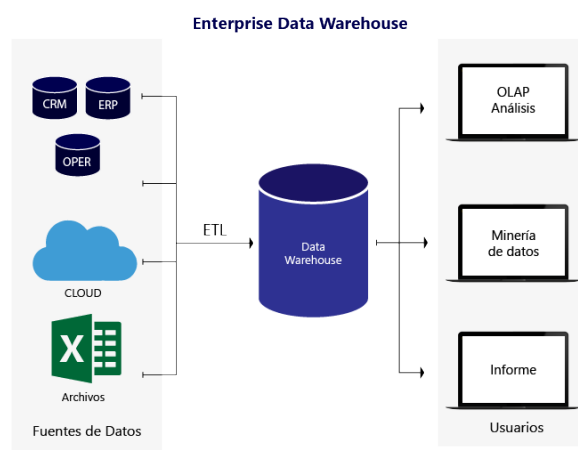


Figura 4: Almacenes de datos.

Instalación VirtualBox y SQL Server.

- i. Para instalar VirtualBox, nos dirigimos a la siguiente página: <https://www.virtualbox.org/wiki/Downloads>
- ii. Seleccionamos nuestro sistema operativo para descargar el instalador.



Figura 5: VirtualBox Downloads.

- iii. Una vez descargado el instalador, lo ejecutamos, y seguimos los pasos que se nos indica. Si todo salió bien debería salirnos lo siguiente:



Figura 6: VirtualBox.

- iv. Antes de crear la máquina virtual descargamos el siguiente archivo que contiene a SQL Server: <https://www.mediafire.com/file/svwjnm1xkjm4z57/SQLServerDWH.zip/file>

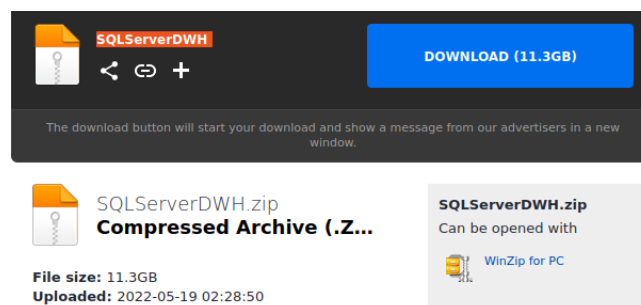


Figura 7: SQLServerDWH.

- v. Una vez que fue descargado el archivo y descomprimido, volvemos a VirtualBox y seleccionamos el botón "Nueva", el cual nos abrirá una ventana para realizar nuestra máquina virtual. Deberá quedarnos algo parecido a lo siguiente:

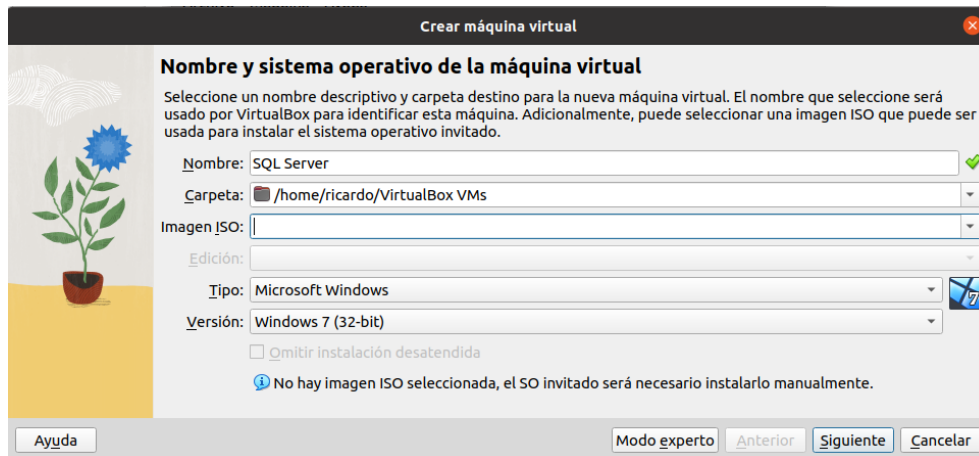


Figura 8: Creación máquina Virtual.

- vi. Al seleccionar siguiente, nos saldrá una ventana la cual nos indicará el tamaño de memoria y procesadores usados, para la máquina virtual, lo dejamos como viene y damos en siguiente.

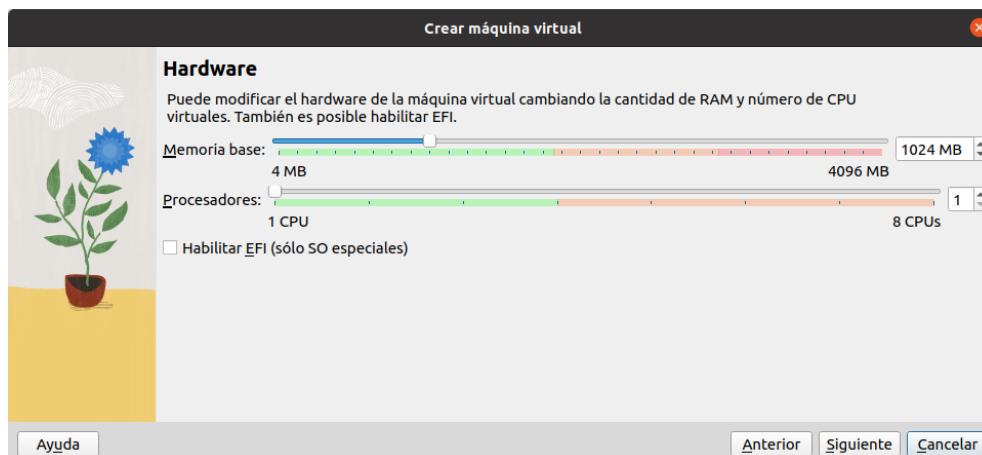


Figura 9: Memoria.

- vii. En la siguiente pantalla seleccionamos la opción "Usar un archivo de disco duro virtual existente". El cual aquí será donde tendremos que buscar el archivo que descomprimimos.

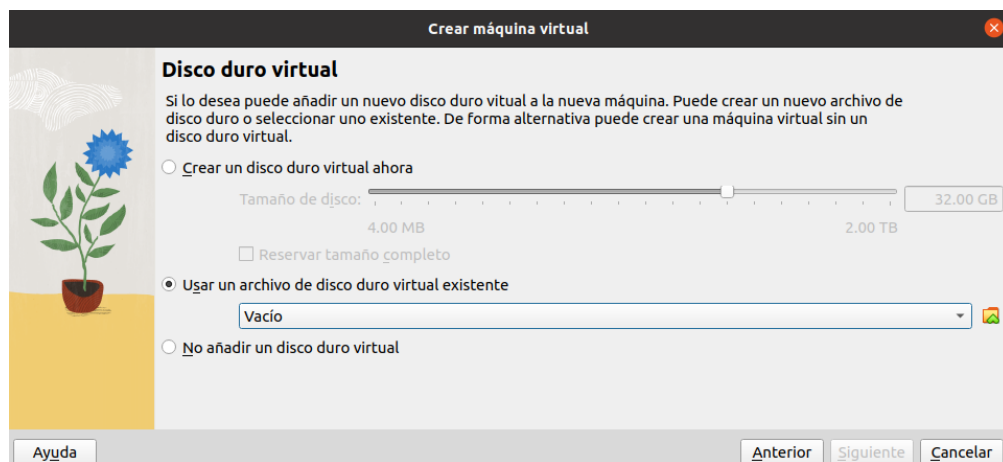


Figura 10: Disco duro virtual 1.

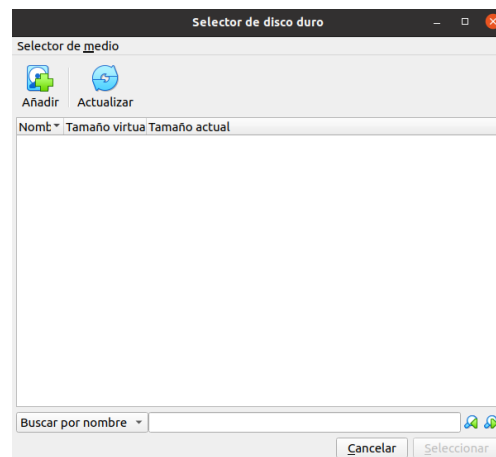


Figura 11: Disco duro virtual 2.

Nombre	Tamaño
Win7Starter.vdi	25.4 GB

Figura 12: Disco duro virtual 3.

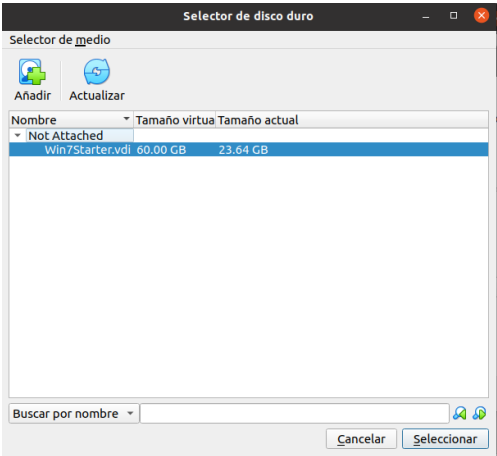


Figura 13: Disco duro virtual 4.

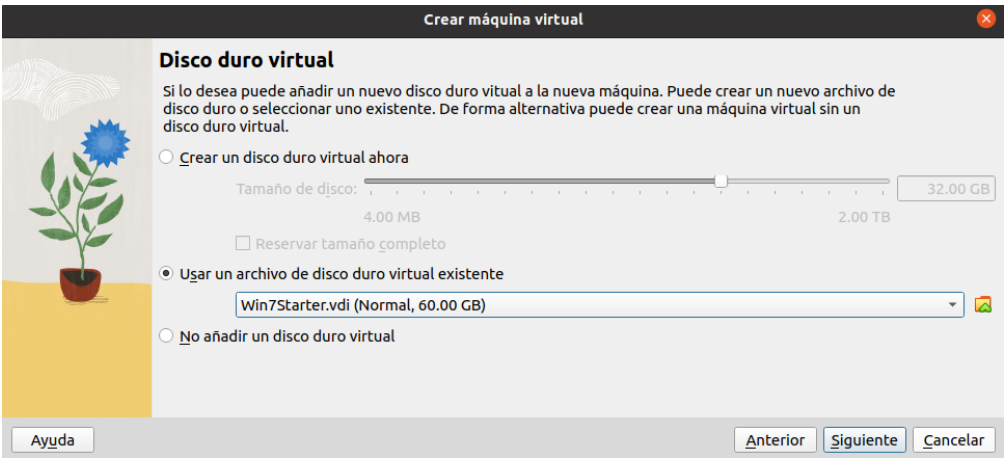


Figura 14: Disco duro virtual 5.

viii. Una vez que seleccionemos el archivo al presionar siguiente, debería salir algo parecido. Y terminamos.

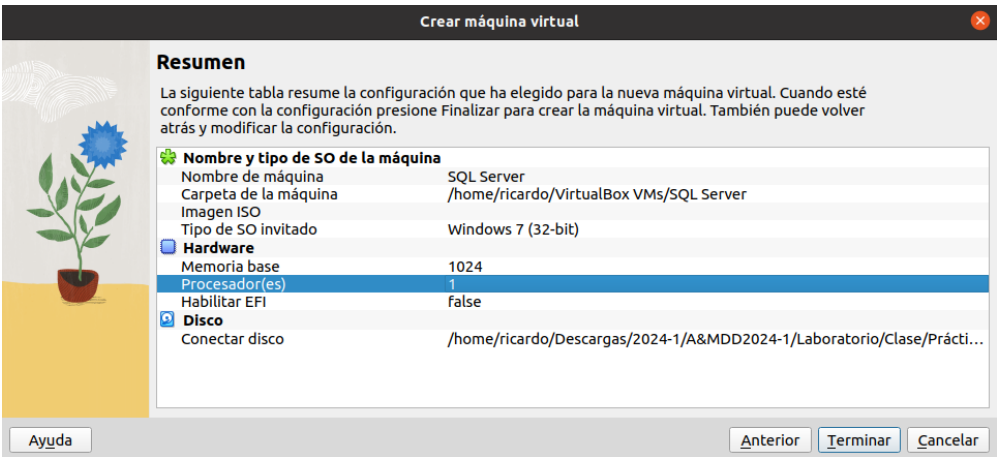


Figura 15: Máquina Virtual Completa.

- ix. Si todo salio bien, debemos poder iniciar la maquina virtual.

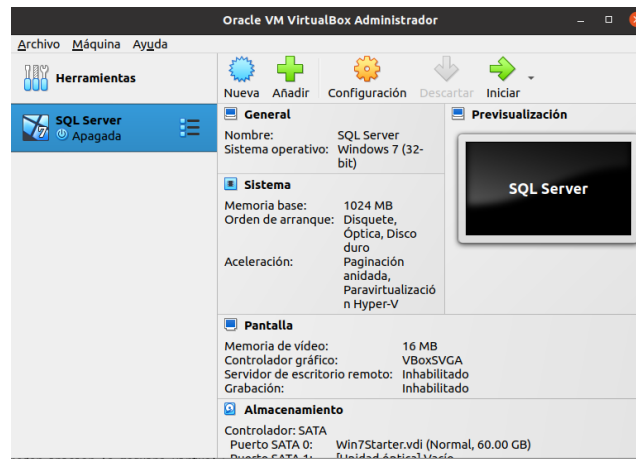


Figura 16: Maquina Virtual.

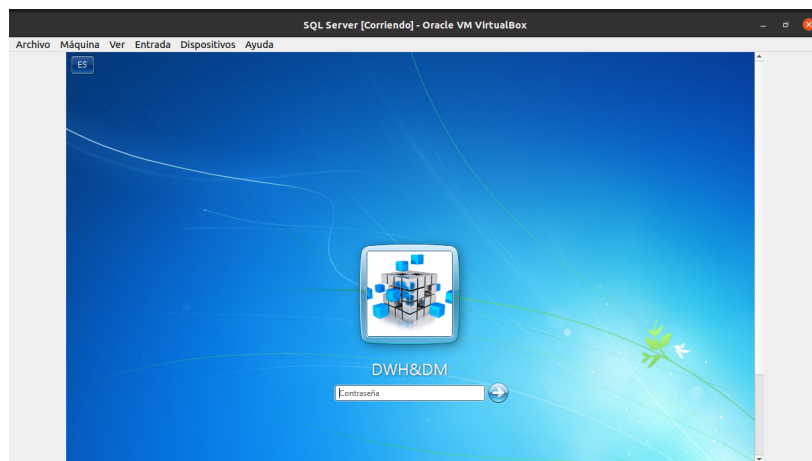


Figura 17: Maquina Virtual Windows7.

- x. Para arreglar problemas en la resolución, una vez abierta la maquina virtual, en la barra herramientas seleccionamos dispositivos, y seleccionamos 'Insertar imagen de CD de las Guest Additions '. El cual empezara a descargar archivos necesarios y nosotros seguiremos los pasos.

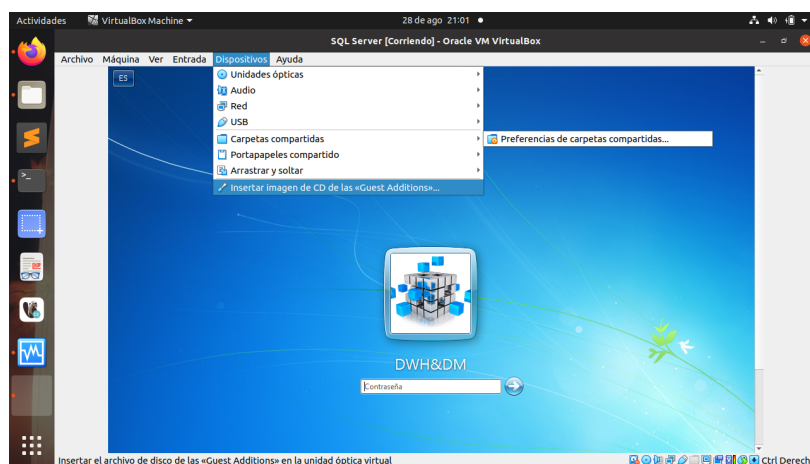


Figura 18: Resolución.

- xi. La contraseña para ingresar es: **dwhdm** y una vez adentro podremos iniciar SQL Server.

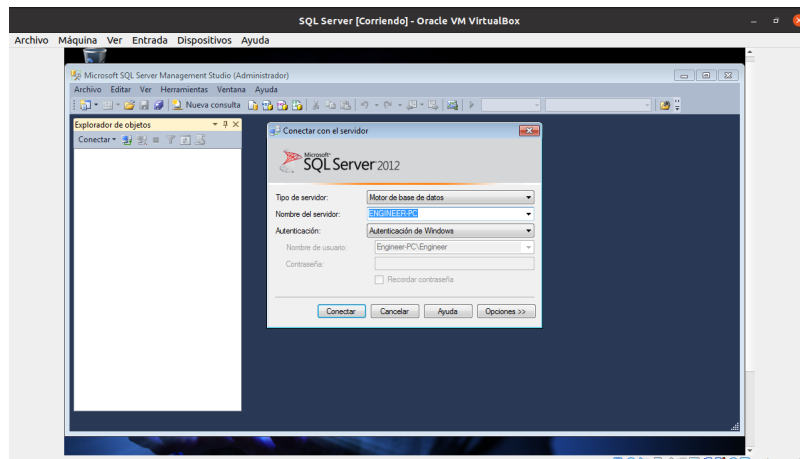


Figura 19: SQL Server.

Agregar carpeta compartida

Al utilizar Maquinas Virtuales, el mayor inconveniente es que no reconocera memorias usb asi que necesitaremos crear una **carpeta compartida**. El cual nos permitira mover archivos desde el sistema operativo host y el sistema operativo creado por la maquina Virtual.

- Creamos una carpeta cualquiera en cualquier ubicación.
- Abrimos VirtualBox, y seleccionando la maquina virtual creada, presionamos configuración, y en la pantalla que sale nos vamos a carpeta compartida.

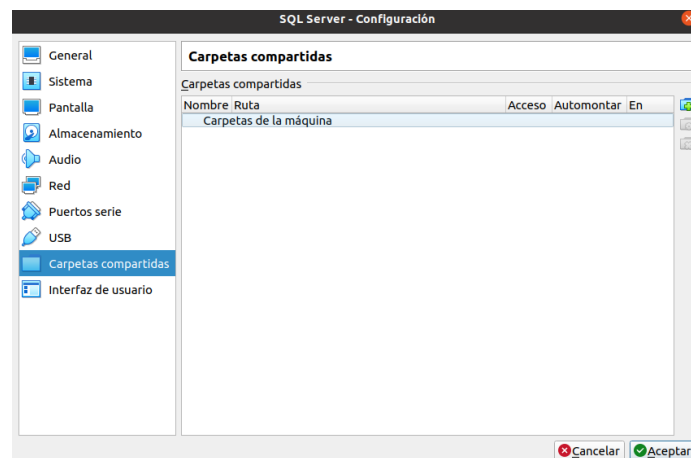


Figura 20: Carpeta Compartida1.

- iii. Agregamos la carpeta que acabamos de crear, seleccionando el símbolo de carpeta con cruz verde, y en ruta de carpeta buscamos nuestra carpeta recién hecha. Y marcamos automontar. Y aceptamos.

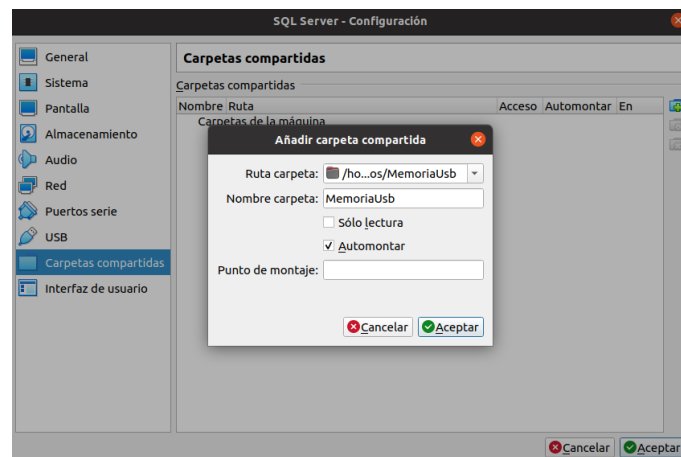


Figura 21: Carpeta Compartida2.

- iv. Iniciamos nuestra Máquina Virtual, iniciamos sesión y nos vamos al apartado Red.

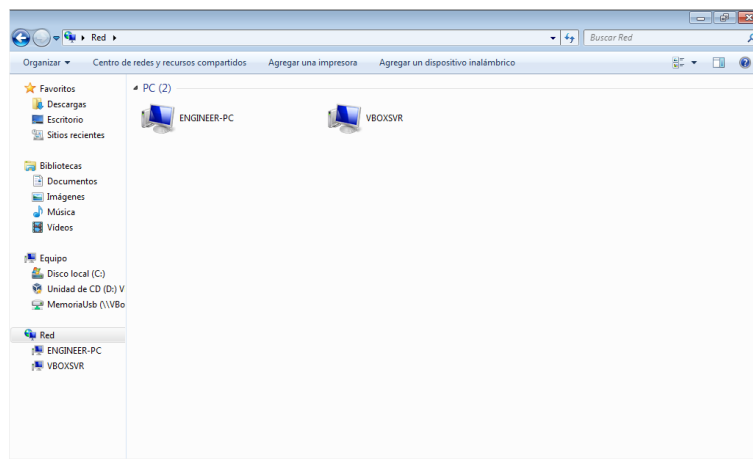


Figura 22: Red1 .

- v. Al meternos a VBOXSVR debería aparecernos nuestra carpeta que hemos creado.

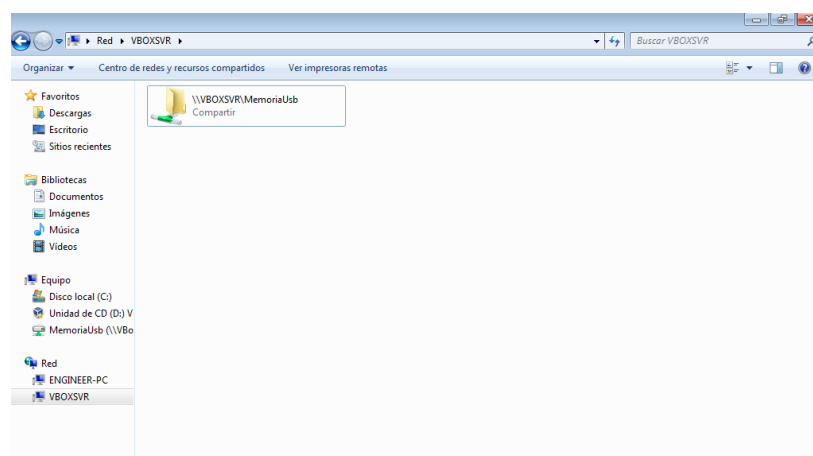


Figura 23: Red2 .

Actividades.

Debido a que es necesario que todos los integrantes del equipo tengan ya instalado VirtualBox, con el SMBD SQL Server.

- i. **Cada integrante** deberá realizar un reporte en donde indiquen los pasos que realizaron para instalar Virtual-Box, crear la maquina virtual con SQL Server, y crear la carpeta compartida dentro de la misma. Todos los reportes iran en el mismo documento PDF, el cual sera llamado **Reporte.pdf**.

IMPORTANTE: Si en dado caso ya realizaron la instalación de la maquina virtual antes de esta práctica, solo sera necesario poner el como crearon la carpeta compartida.

- ii. Dado la siguiente base de datos deberan hacer lo siguiente:

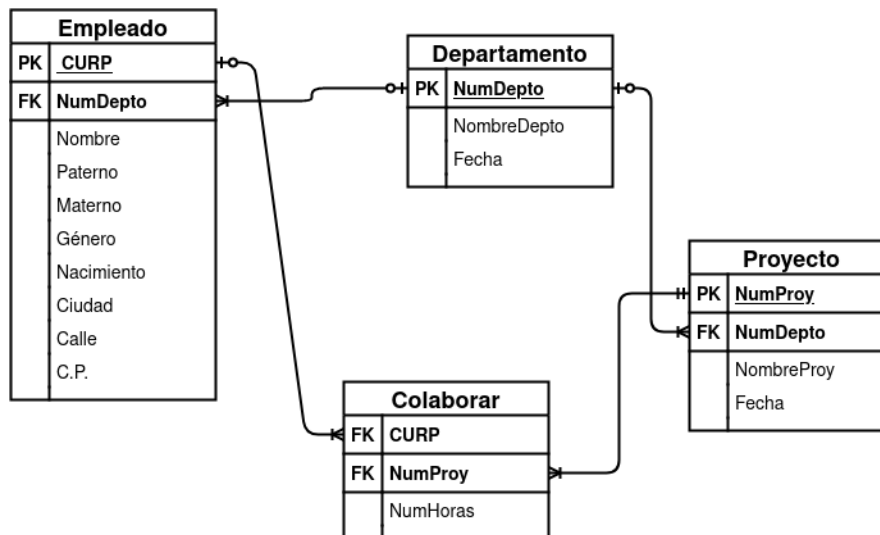


Figura 24: Base De Datos.

- Deberan crear una vista para la tabla empleado, en donde solo aparezca el curp, y su nombre completo (nombre,paterno,materno), y utilizando esta vista, deberan obtener a todos los empleados cuyo nombre comience con la letra A.
- Deberan crear una vista que muestre toda la información de los empleados y departamentos en una sola vista.
- Deberan crear una vista que muestre el numProyecto, NombreProyecto, , el numero de empleados y el numero de horas colaboradas que tiene cada proyecto.
- De las 3 vistas que crearon, deberan realizar operaciones de mantenimiento (insert, delete, update), y deberan contestar lo siguiente:¿En cuales vistas se pudieron ver los insert, delete y update?, ¿Por qué?.
- Contesta: ¿Cuales son las limitaciones de las vistas normales?.
- Realiza las mismas 3 consultas, pero ahora utilizando vistas materializadas.Y aplica las mismas operaciones de mantenimiento. Y contesta ¿Hubo algun cambio?, ¿Por qué?.

Su archivo se llamara **script.sql**, y la respuesta a las preguntas lo escribieran en un documento PDF llamado **Practica01.pdf**.

- iii. Deberan leer el siguiente articulo **Materialized Views and Data Warehouses**, y deberan realizar un resumen del documento, destacando los puntos más relevantes. El cual llamaran **Resumen.pdf**



Figura 25: Actividades.

Entregables.

Deberán subir un archivo con formato *zip* a *Google Classroom*, de acuerdo a lo indicado en los lineamientos de entrega. Debe de estar organizado de la siguiente manera, (suponiendo que el nombre del equipo que está entregando es *Dream Team* y los integrantes del equipo son los profesores del curso).

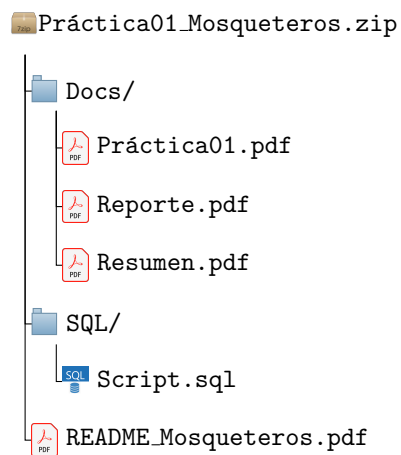


Figura 26: Entregables.

Nota.

Para cualquier duda o comentario que pudiera surgirles al hacer este trabajo, recuerden que cuentan con la asignación de este entregable en el grupo de *Classroom*, en donde seguramente encontrarás las respuestas que necesites.



Figura 27: Nota.