



Institución
Universitaria
Reacreditada en Alta Calidad

INTRODUCCIÓN A LA CIENCIA DE DATOS

INTRODUCCIÓN A LA CIENCIA DE DATOS

530202018

www.itm.edu.co

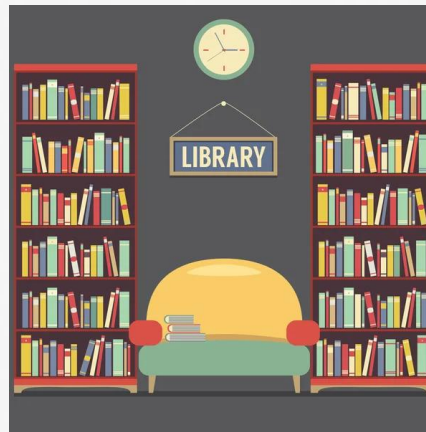
Bases de datos

¿Qué son las **bases de datos**?

Conjunto de datos organizados y relacionados que se almacena de manera sistemática para ser consultado, gestionado y actualizado de forma eficiente.



Colección de datos
organizada



Base de datos física

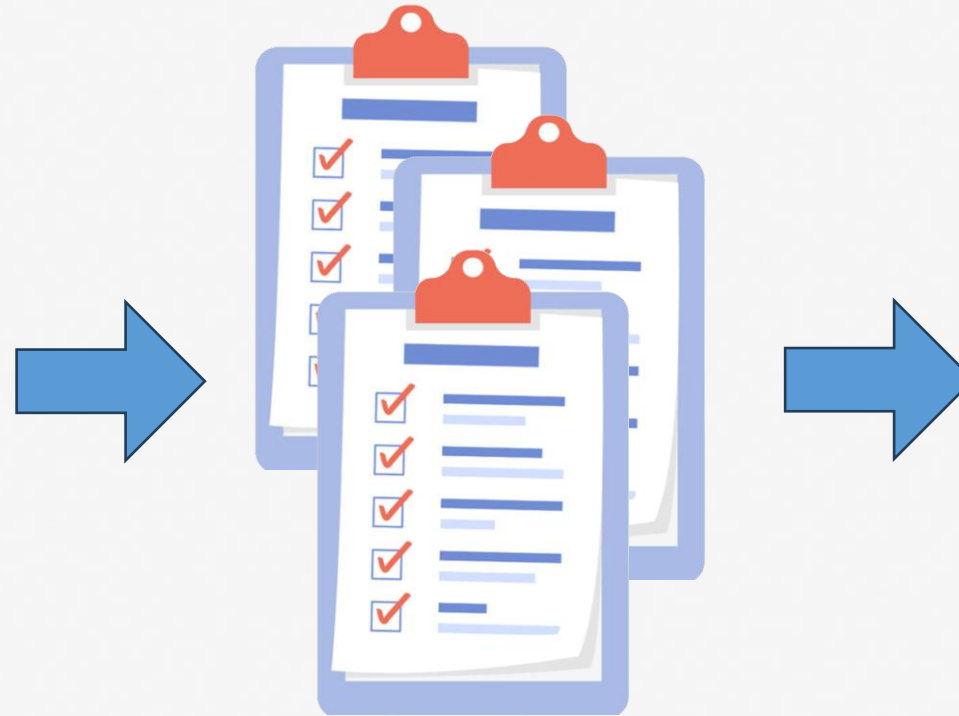
	A	B	C	D	E
1	No. Empleado	Departamento	Nombre	Apellido	
2	4000307	Recursos Humanos	Andrea	Mendoza	
3	4000308	Finanzas	Juan Carlos	Hernández	
4	4000309	Informática	Arturo	Becerril	
5	4000310	Marketing	Claudia	Camacho	
6	4000311	Dirección General	Jacob	Sarmiento	
7	4000312	Finanzas	Isidro	Jiménez	
8	4000313	Finanzas	Consuelo	Palacios	
9	4000314	Marketing	Fernanda	Fonseca	
10	4000315	Recursos Humanos	Jorge	Villarreal	
11					

Base de datos en
términos informáticos

Bases de datos



Datos



Etiquetar y numerar

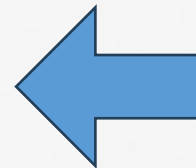


Base de datos
estructurada

Bases de datos

¿Qué es **SQL**?

Una cosa es la base de datos, que es la información en sí misma y otra cosa es el lenguaje estructurado para hacer consultas en esta base de datos.



Structured
Query
Language

Bases de datos

¿Qué es **SQL**?

SQL es un lenguaje, al igual que cualquier otro idioma tiene las siguientes características:

- Propia sintaxis (reglas)
- Propio vocabulario y significado de esas palabras
- Lenguaje estructurado (comando y reglas para comunicarse)
- Estándar internacional (diferentes gestiones de bases de datos, software: MySQL, PostgreSQL,...)

Bases de datos

Bases de datos relacionales



Bases de datos NO relacionales



fecha	cliente	idioma	nivel	suscripción	precio	descuento_%	precio final
25/06/2018	Pedro	Inglés	Intermedio	Mensual	7	0	7
25/06/2018	Pedro	Chino	Principiante	Mensual	9	0	9
01/07/2018	Aurelia	Francés	Avanzado	Anual	8	25	6
03/07/2018	Federico	Inglés	Intermedio	Trimestral	7	10	6.3

```
{
  "maquina_id":1,
  "timestamp":149992693000,
  "coordenadas":"75988823.567, 55375867.098, 12676444.311",
  "encontrado":[
    "roca",
    "agua",
    "roca",
    "roca",
    "algo que parece un animal",
    "roca"
  ]
}
```

Bases de datos

¿Qué son las bases de datos **Relacionales**?

- Se basa en la organización de la información en trozos pequeños, que se relacionan entre ellos mediante identificadores.
- Se organizan en tablas y se llaman relacionales porque las tablas se relacionan entre si.
- Usan SQL como su lenguaje principal de consulta.

Productos	
id	
producto	
cantidad	

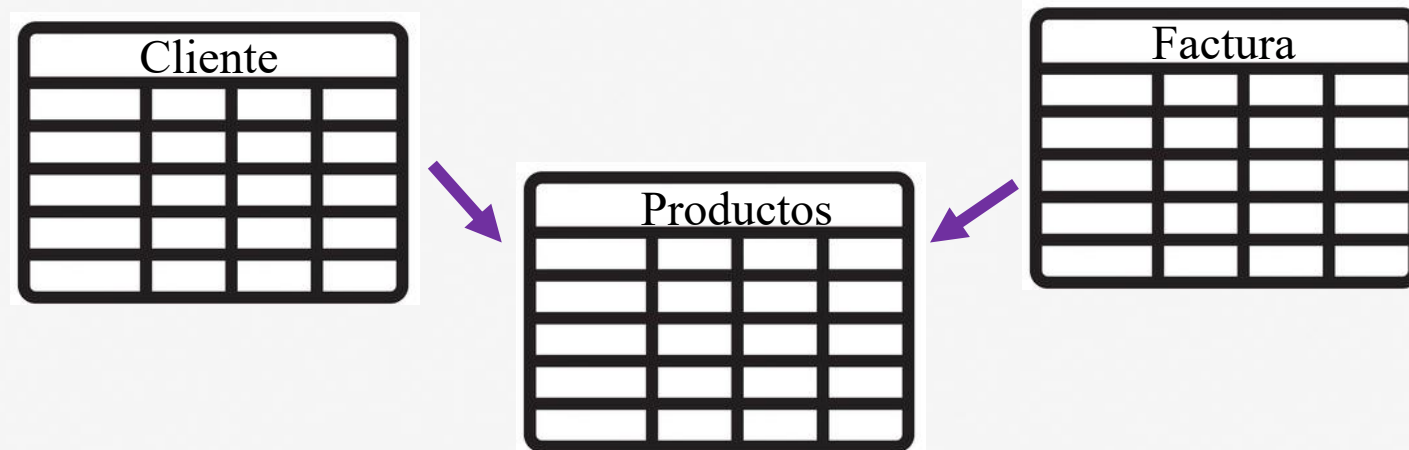
Factura	
Llave	id_producto
Llave	codigo_cliente

Cliente	
codigo	
nombre	
numeroTelefono	

Componentes de bases de datos **Relacionales**

Tablas

Cada tabla de una base de datos contiene datos sobre un tema específico



Ventajas de las bases de datos **Relacionales**

1. Eliminan de la redundancia y eficiencia del almacenamiento, ya que la información se divide en tablas, cada una con un propósito específico.
2. El uso de claves primarias y foráneas aseguran la **integridad y consistencia** de los datos.
3. Flexibilidad y facilidad para realizar consultas
4. Mantenimiento y escalabilidad, ya que cuando es necesario actualizar la información, solo es necesario hacerlo en un lugar.

Bases de datos

¿Qué son las bases de datos **NO Relacionales**?

A diferencia de las relacionales, no tienen un identificador que sirva de relación entre un conjunto de datos y otros.

La información se organiza normalmente mediante documentos y es muy útil cuando no tenemos un esquema exacto de lo que se va a almacenar

Se organizan en formatos muy distintos

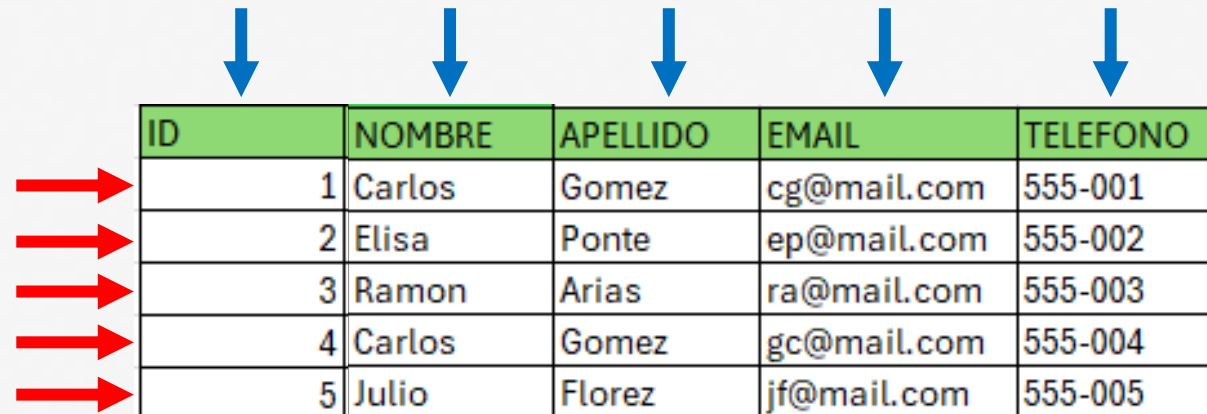
Necesitan una gestión distinta por lo tanto son NoSQL

Componentes de bases de datos Relacionales

Tablas

Las filas son los registros de cada caso que existe en una tabla.

Las columnas representan un atributo o propiedad de los datos, como: nombre, apellido, teléfono, etc.



ID	NOMBRE	APELLIDO	EMAIL	TELEFONO
1	Carlos	Gomez	cg@mail.com	555-001
2	Elisa	Ponte	ep@mail.com	555-002
3	Ramon	Arias	ra@mail.com	555-003
4	Carlos	Gomez	gc@mail.com	555-004
5	Julio	Florez	jf@mail.com	555-005

Bases de datos

¿Qué son las bases de datos **Relacionales**?

fecha	cliente	idioma	nivel	suscripción	precio	descuento_%	precio final
25/06/2018	Pedro	Inglés	Intermedio	Mensual	7	0	7
25/06/2018	Pedro	Chino	Principiante	Mensual	9	0	9
01/07/2018	Aurelia	Francés	Avanzado	Anual	8	25	6
03/07/2018	Federico	Inglés	Intermedio	Trimestral	7	10	6.3

Este modelo de datos tiene algunos problemas

Bases de datos

¿Qué son las bases de datos **Relacionales**?

fecha	cliente	idioma	nivel	suscripción	precio	descuento_%	precio final
25/06/2018	Pedro	Inglés	Intermedio	Mensual	7	0	7
25/06/2018	Pedro	Chino	Principiante	Mensual	9	0	9
01/07/2018	Aurelia	Francés	Avanzado	Anual	8	25	6
03/07/2018	Federico	Inglés	Intermedio	Trimestral	7	10	6.3

1) Los clientes no tienen un identificador único

Bases de datos

¿Qué son las bases de datos **Relacionales**?

fecha	cliente	idioma	nivel	suscripción	precio	descuento_%	precio final
25/06/2018	Pedro	Inglés	Intermedio	Mensual	7	0	7
25/06/2018	Pedro	Chino	Principiante	Mensual	9	0	9
01/07/2018	Aurelia	Francés	Avanzado	Anual	8	25	6
03/07/2018	Federico	Inglés	Intermedio	Trimestral	7	10	6.3

2) Si cambia un precio hay que actualizar filas en que aparece

Bases de datos

¿Qué son las bases de datos **Relacionales**?

fecha	cliente	idioma	nivel	suscripción	precio	descuento_%	precio final
25/06/2018	Pedro	Inglés	Intermedio	Mensual	7 8	0	7 8
25/06/2018	Pedro	Chino	Principiante	Mensual	9	0	9
01/07/2018	Aurelia	Francés	Avanzado	Anual	8	25	6
03/07/2018	Federico	Inglés	Intermedio	Trimestral	7 8	10	68 72

3) Si cambia un precio hay que actualizar las columnas precio y precio final

Bases de datos

¿Qué son las bases de datos **Relacionales**?

fecha	cliente	idioma	nivel	suscripción	precio	descuento_%	precio final
25/06/2018	Pedro	Inglés	Intermedio	Mensual	7	0	7
25/06/2018	Pedro	Chino	Principiante	Mensual	9	0	9
01/07/2018	Aurelia	Francés	Avanzado	Anual	8	25	6
03/07/2018	Federico	Inglés	Intermedio	Trimestral	7	10	6.3

4) La información en precio final es redundante

Bases de datos

¿Qué son las bases de datos **Relacionales**?

fecha	cliente	idioma	nivel	suscripción	precio	descuento_%	precio final
25/06/2018	Pedro	Inglés	Intermedio	Mensual	7	0	7
25/06/2018	Pedro	Chino	Principiante	Mensual	9	0	9
01/07/2018	Aurelia	Francés	Avanzado	Anual	8	25	6
03/07/2018	Federico	Inglés	Intermedio	Trimestral	7	10	6.3

4) No se ven todos los idiomas y niveles ofertados

Bases de datos

¿Qué son las bases de datos **Relacionales**?

Se recomienda un diseño de base de datos relacional, donde se recoja la información en varias tablas y no solo en una.

cliente_id	nombre_cliente

programa_id	idioma	nivel	precio

suscripcion_id	tipo	descuento_%

Bases de datos

¿Qué son las bases de datos **Relacionales**?

Se recomienda un diseño de base de datos relacional, donde se recoja la información en varias tablas y no solo en una.

cliente_id	nombre_cliente	programa_id	idioma	nivel	precio	suscripcion_id	tipo	descuento_%
1	Pedro	1	alemán	principiante	7	1	Mensual	0
2	Aurelia	2	chino	principiante	9	2	Trimestral	10
3	Federico	3	francés	avanzado	8	3	Anual	25
		4	inglés	intermedio	7			

Bases de datos

¿Qué son las bases de datos **Relacionales**?

Finalmente tenemos la tabla que relaciona todo:

id	cliente_id	programa_id	suscripcion_id	fecha
1	1	4	1	25/06/2018
2	1	2	1	25/06/2018
3	2	3	3	01/07/2018
4	3	4	2	03/07/2018

Componentes de bases de datos **Relacionales**

Claves primarias (Primary key)

La clave primaria es uno de los campos o columnas de la tabla que sirve para identificar de manera única cada fila en una tabla.

Es un identificador para asegurar de que no hayan duplicados y cada registro sea identificado de manera única.

Propiedades:

- Única: No se repite para diferentes filas.
- No nula: No puede contener valores NULL.
- Inmutable: En lo posible, su valor no debe cambiar con el tiempo.

Componentes de bases de datos Relacionales

Claves primarias (Primary key)

cliente_id	nombre_cliente	programa_id	idioma	nivel	precio
1	Pedro	1	alemán	principiante	7
2	Aurelia	2	chino	principiante	9
3	Federico	3	francés	avanzado	8
			inglés	intermedio	7

suscripcion_id	tipo	descuento_%
1	Mensual	0
2	Trimestral	10
3	Anual	25

id	cliente_id	programa_id	suscripcion_id
1	1	4	1
2	1	2	1
3	2	3	3
4	3	4	2

Componentes de bases de datos Relacionales

Claves primarias (Primary key)

cliente_id	nombre_cliente	programa_id	idioma	nivel	precio
1	Pedro	1	alemán	principiante	7
2	Aurelia	2	chino	principiante	9
3	Federico	3	francés	avanzado	8
			inglés	intermedio	7

suscripcion_id	tipo	descuento_%
1	Mensual	0
2	Trimestral	10
3	Anual	25

id	cliente_id	programa_id	suscripcion_id
1	1	4	1
2	1	2	1
3	2	3	3
4	3	4	2

Componentes de bases de datos **Relacionales**

Claves foráneas (Foreíng key)

Es uno o varios campos en una tabla que referencian la clave primaria de otra tabla.

Su propósito es establecer y reforzar la relación entre tablas.

Garantiza la integridad referencial, ya que no se pueden crear relaciones a registros que no existan en la tabla referenciada.

Componentes de bases de datos Relacionales

Claves foráneas (Foreíng key)

cliente_id	nombre_cliente	programa_id	idioma	nivel	precio
1	Pedro	1	alemán	principiante	7
2	Aurelia	2	chino	principiante	9
3	Federico	3	francés	avanzado	8
			inglés	intermedio	7

suscripcion_id	tipo	descuento_%
1	Mensual	0
2	Trimestral	10
3	Anual	25

id	cliente_id	programa_id	suscripcion_id
1	1	4	1
2	1	2	1
3	2	3	3
4	3	4	2

Componentes de bases de datos Relacionales

Claves foráneas (Foreíng key)

cliente_id	nombre_cliente	programa_id	idioma	nivel	precio
1	Pedro	1	alemán	principiante	7
2	Aurelia	2	chino	principiante	9
3	Federico	3	francés	avanzado	8
			inglés	intermedio	7

suscripcion_id	tipo	descuento_%
1	Mensual	0
2	Trimestral	10
3	Anual	25

id	cliente_id	programa_id	suscripcion_id
1	1	4	1
2	1	2	1
3	2	3	3
4	3	4	2

TAREA: Instalar MySQL para la próxima clase

MySQL es un software o un programa que actúa como intermedio entre los usuarios y las bases de datos.



MySQL:

Es el motor de gestión de bases de datos



MySQL Workbench:

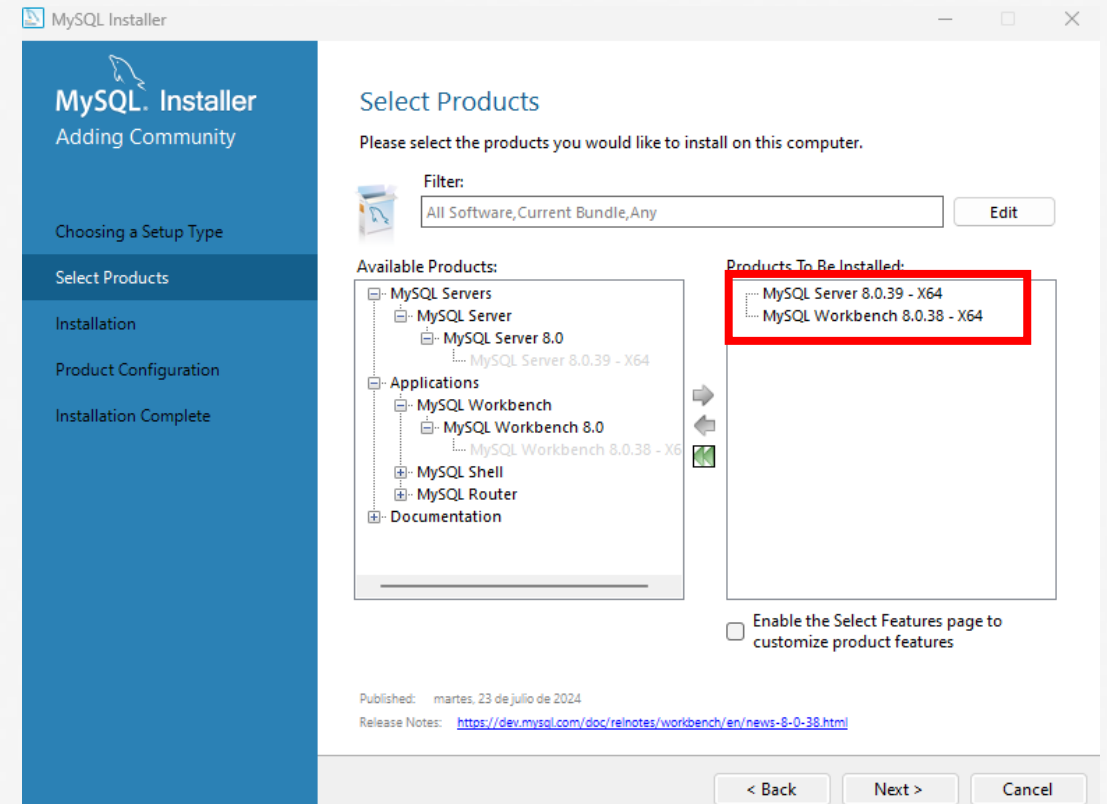
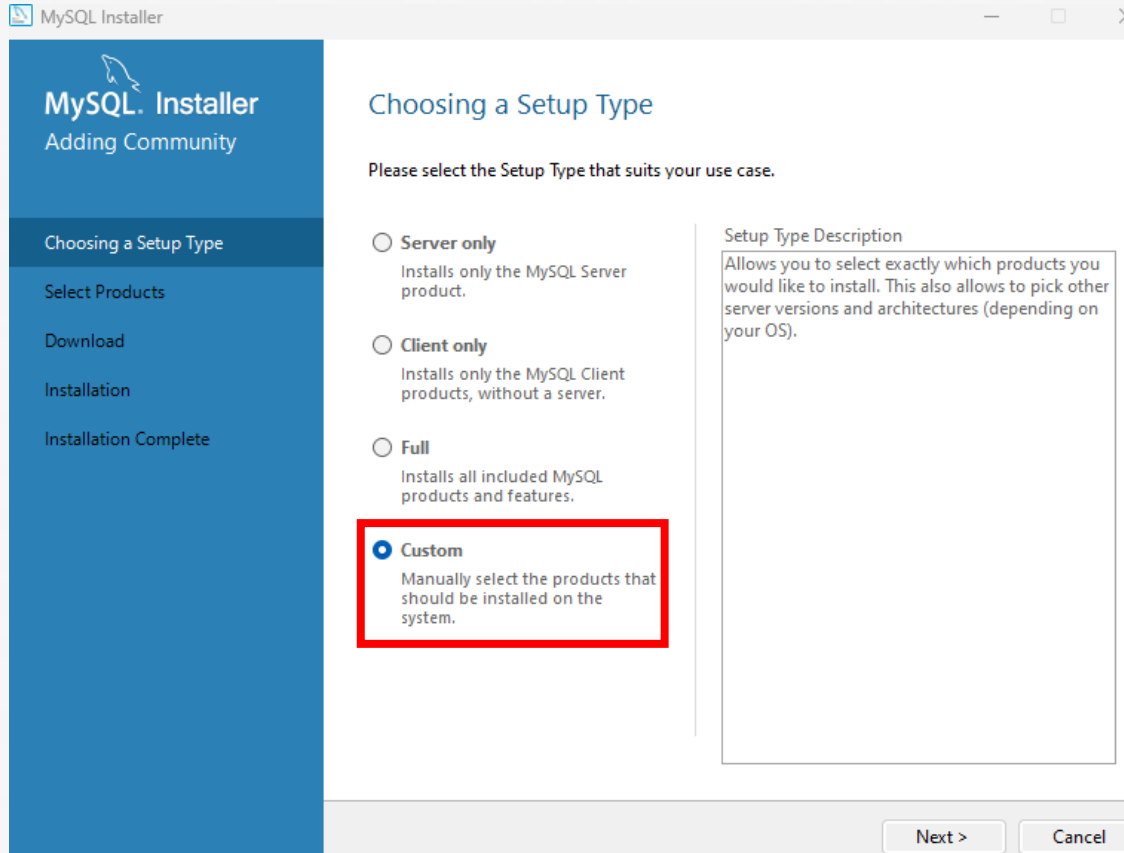
Es la herramienta visual y entorno integrado de desarrollo

Instalación de MySQL

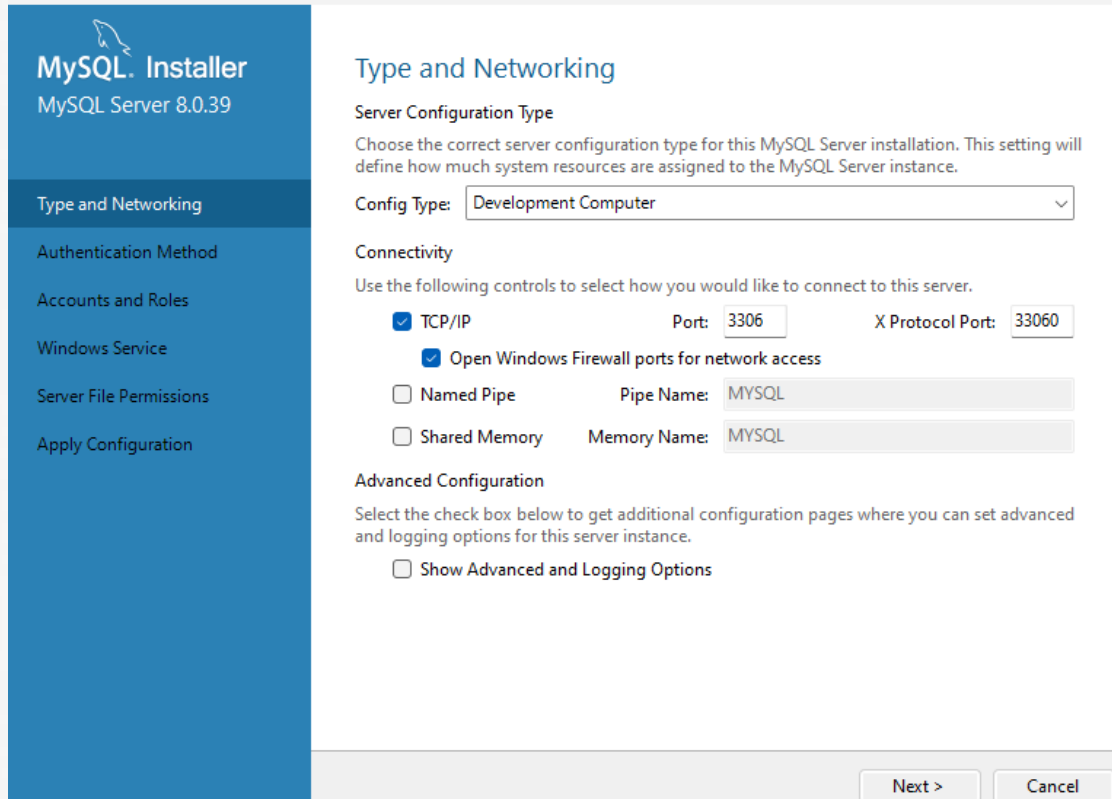
1. Ir a la dirección <https://www.mysql.com>
2. Ir a la sección “Downloads” y seleccionar “[MySQL Community Downloads](#)”

Windows (x86, 32-bit), MSI Installer (mysql-installer-web-community-8.0.39.0.msi)	8.0.39	2.1M	Download MD5: d8499da0b2c4b5dfa81a5c5185af9238 Signature
Windows (x86, 32-bit), MSI Installer (mysql-installer-community-8.0.39.0.msi)	8.0.39	303.6M	Download MD5: 353c5e5ab9350d0e9ddcb42264229b5d Signature

Instalación de MySQL



Instalación de MySQL



MySQL. Installer
MySQL Server 8.0.39

Type and Networking

Server Configuration Type

Choose the correct server configuration type for this MySQL Server installation. This setting will define how much system resources are assigned to the MySQL Server instance.

Config Type:

Connectivity

Use the following controls to select how you would like to connect to this server.

☒ TCP/IP Port: X Protocol Port:

☒ Open Windows Firewall ports for network access

☐ Named Pipe Pipe Name:

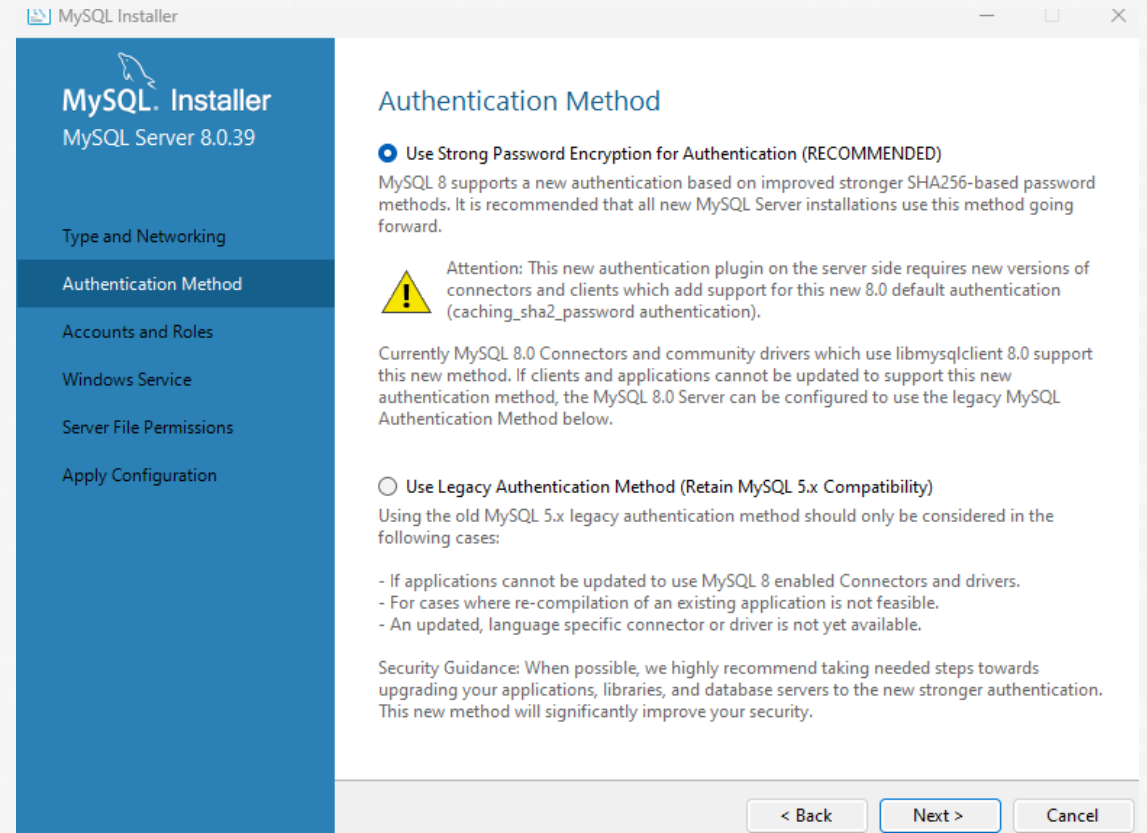
☐ Shared Memory Memory Name:

Advanced Configuration

Select the check box below to get additional configuration pages where you can set advanced and logging options for this server instance.

☐ Show Advanced and Logging Options

Next > Cancel




MySQL. Installer
MySQL Server 8.0.39

Authentication Method

☒ Use Strong Password Encryption for Authentication (RECOMMENDED)

MySQL 8 supports a new authentication based on improved stronger SHA256-based password methods. It is recommended that all new MySQL Server installations use this method going forward.

 Attention: This new authentication plugin on the server side requires new versions of connectors and clients which add support for this new 8.0 default authentication (caching_sha2_password authentication).

Currently MySQL 8.0 Connectors and community drivers which use libmysqlclient 8.0 support this new method. If clients and applications cannot be updated to support this new authentication method, the MySQL 8.0 Server can be configured to use the legacy MySQL Authentication Method below.

☐ Use Legacy Authentication Method (Retain MySQL 5.x Compatibility)

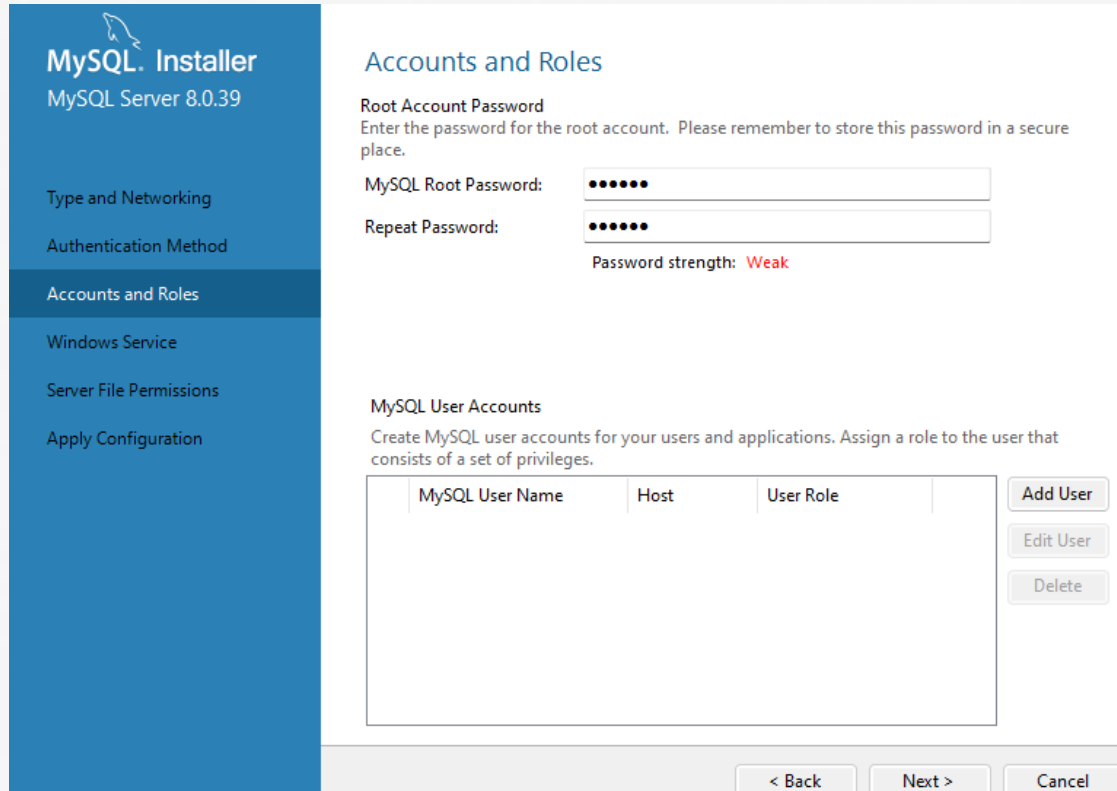
Using the old MySQL 5.x legacy authentication method should only be considered in the following cases:

- If applications cannot be updated to use MySQL 8 enabled Connectors and drivers.
- For cases where re-compilation of an existing application is not feasible.
- An updated, language specific connector or driver is not yet available.

Security Guidance: When possible, we highly recommend taking needed steps towards upgrading your applications, libraries, and database servers to the new stronger authentication. This new method will significantly improve your security.

< Back Next > Cancel

Instalación de MySQL



MySQL. Installer
MySQL Server 8.0.39

Type and Networking
Authentication Method
Accounts and Roles
Windows Service
Server File Permissions
Apply Configuration

Accounts and Roles

Root Account Password
Enter the password for the root account. Please remember to store this password in a secure place.

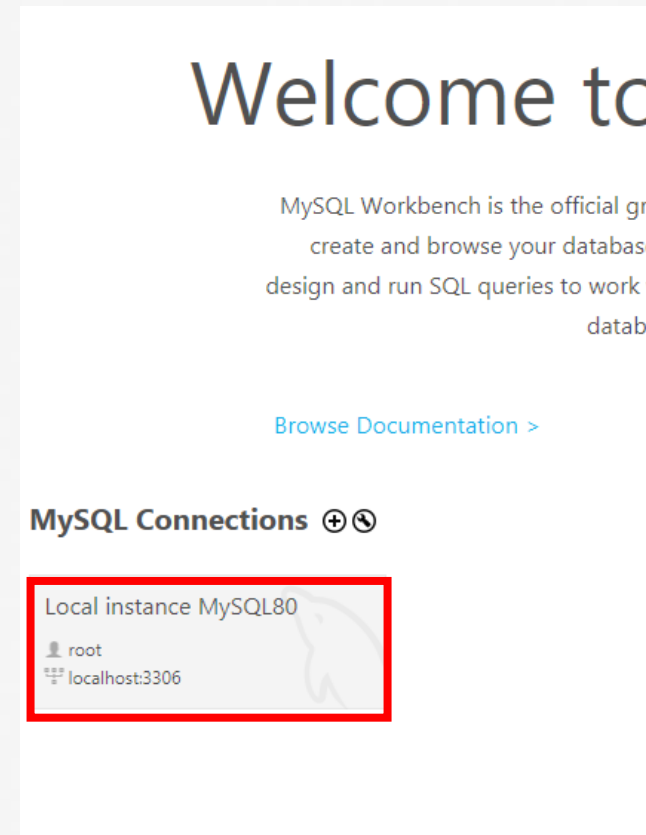
MySQL Root Password:

Repeat Password:

Password strength: **Weak**

MySQL User Accounts
Create MySQL user accounts for your users and applications. Assign a role to the user that consists of a set of privileges.

MySQL User Name	Host	User Role
-----------------	------	-----------



Bases de datos: **Ejercicios**

Teniendo en cuenta el ejemplo de los cursos de idiomas, organizar las siguientes tablas de datos en forma de tablas relacionales, y en cada una cuáles son las claves primarias y las claves foráneas.

Bases de datos: **Ejercicio 1**

VentaID	ClienteID	NombreCliente	CiudadCliente	LibroID	TituloLibro	Autor	PrecioVenta
1	C001	Ana Torres	Bogotá	L101	El Quijote	Miguel de Cervantes	45.000
2	C002	Juan Pérez	Medellín	L102	Cien Años de Soledad	Gabriel García Márquez	52.000
3	C001	Ana Torres	Bogotá	L103	La Odisea	Homero	38.000
4	C003	Laura Gómez	Cali	L101	El Quijote	Miguel de Cervantes	45.000
5	C004	Pedro Ruiz	Bogotá	L104	Don Juan Tenorio	José Zorrilla	40.000
6	C002	Juan Pérez	Medellín	L105	El Principito	Antoine de Saint-Exupéry	30.000
7	C005	María López	Barranquilla	L103	La Odisea	Homero	38.000
8	C003	Laura Gómez	Cali	L106	Hamlet	William Shakespeare	47.000
9	C006	Andrés Díaz	Medellín	L102	Cien Años de Soledad	Gabriel García Márquez	52.000
10	C001	Ana Torres	Bogotá	L105	El Principito	Antoine de Saint-Exupéry	30.000

Bases de datos: Ejercicio 2

Pedido ID	Cliente ID	Nombre Cliente	Ciudad Cliente	Producto ID	Nombre Producto	Sucursal	Cantidad	Precio Unitario
1	C01	Ana Torres	Medellín	P100	Café Caturra 340g	Laureles	2	28.000
2	C02	Juan Pérez	Bogotá	P101	Café Geisha 250g	Chapinero	1	95.000
2	C02	Juan Pérez	Bogotá	P102	Café Castillo 250g	Chapinero	1	42.000
3	C01	Ana Torres	Medellín	P102	Café Castillo 500g	Laureles	1	42.000
4	C03	María López	Cali	P100	Café Caturra 340g	San Antonio	3	28.000
4	C03	María López	Cali	P100	Café Geisha 250g	San Antonio	1	95.000
5	C02	Juan Pérez	Bogotá	P100	Café Caturra 340g	Chapinero	1	28.000

Ayuda: Es posible establecer una clave primaria compuesta, la cual es un tipo de clave primaria en una base de datos que está formada por **dos o más atributos de una misma tabla**. A diferencia de una clave simple (que utiliza un solo atributo), la combinación de los valores de estos atributos es lo que garantiza la **unicidad de cada registro**.

Bases de datos: Ejercicio 3

Venta ID	Fecha venta	Cliente ID	Nombre Cliente	email Cliente	Producto ID	Nombre producto	Categoria producto	Precio producto	Cantidad
101	2025-07-01	C001	Ana García	ana.garcia@email.com	P01	Laptop A	Electrónica	1200	1
102	2025-07-01	C002	Luis Pérez	luis.perez@email.com	P02	Teclado mecánico	Accesorios	150	2
103	2025-07-02	C001	Ana García	ana.garcia@email.com	P03	Mouse inalámbrico	Accesorios	50	1
104	2025-07-03	C003	Sofía López	sofia.lopez@email.com	P01	Laptop A	Electrónica	1200	1
105	2025-07-04	C002	Luis Pérez	luis.perez@email.com	P04	Monitor 4K	Electrónica	450	1
106	2025-07-05	C004	María Sol	maria.sol@email.com	P01	Laptop A	Electrónica	1200	1
106	2025-07-05	C004	María Sol	maria.sol@email.com	P02	Teclado mecánico	Accesorios	150	1

Respuesta: Ejercicio 1

VentaID	ClienteID	NombreCliente	CiudadCliente	LibroID	TituloLibro	Autor	PrecioVenta
---------	-----------	---------------	---------------	---------	-------------	-------	-------------

Clientes (ClienteID, NombreCliente, CiudadCliente)

Libros (LibroID, TituloLibro, Autor, PrecioVenta)

Ventas (VentaID, ClienteID, LibroID)

1. Tabla Clientes

Clave primaria: ClienteID (identifica de forma única a cada cliente)

2. Tabla Libros

Clave primaria: LibroID (identifica de forma única a cada libro)

3. Tabla Ventas

Clave primaria: VentaID (identifica de forma única cada transacción)

Claves foráneas:

ClienteID (referencia a Tabla Clientes)

LibroID (referencia a Tabla Libros)

Respuesta: Ejercicio 2

Pedido ID	Cliente ID	Nombre Cliente	Ciudad Cliente	Producto ID	Nombre Producto	Sucursal	Cantidad	Precio Unitario
-----------	------------	----------------	----------------	-------------	-----------------	----------	----------	-----------------

Clientes (ClienteID, NombreCliente, CiudadCliente)

Producto (ProductoID, NombreProducto, PrecioUnitario)

Pedido (PedidoID, ClienteID, Sucursal)

Detalle Pedido (PedidoID, ProductoID, Cantidad)

1. Tabla Clientes

Clave primaria: ClienteID (identifica de forma única a cada cliente)

2. Tabla Producto

Clave primaria: ProductoID (identifica de forma única a cada producto)

3. Tabla Pedidos

Clave primaria: PedidoID (identifica de forma única cada transacción)

Claves foráneas: ClienteID (referencia a Tabla Clientes)

4. Tabla Detalle Pedidos

Clave primaria compuesta: PedidoID + ProductoID

La combinación de los valores de estos atributos (PedidoID + ProductoID) garantiza la **unicidad de cada registro**.

Respuesta: **Ejercicio 3**

Venta ID	Fecha Venta	Cliente ID	Nombre Cliente	email Cliente	Producto ID	Nombre Producto	Categoría Producto	Precio Producto	Cantidad
----------	-------------	------------	----------------	---------------	-------------	-----------------	--------------------	-----------------	----------

Clientes (ClienteID, NombreCliente, emailCliente)

Producto (ProductoID, CategoríaProducto, PrecioProducto)

Venta (VentaID, FechaVenta, ClienteID)

Detalle Venta (VentaID, ProductoID, Cantidad)

1. Tabla Clientes

Clave primaria: ClienteID (identifica de forma única a cada cliente)

2. Tabla Producto

Clave primaria: ProductoID (identifica de forma única a cada producto)

3. Tabla Ventas

Clave primaria: VentaID (identifica de forma única cada transacción)

Claves foráneas: ClienteID (referencia a Tabla Clientes)

4. Tabla Detalle Ventas

Clave primaria compuesta: VentaID + ProductoID

La combinación de los valores de estos atributos (VentaID + ProductoID) garantiza la **unicidad de cada registro**.



Institución
Universitaria
Reacreditada en Alta Calidad

¡MUCHAS GRACIAS!

A decorative graphic consisting of several horizontal lines in various colors (yellow, orange, red, purple, blue, green) and a curved line on the right side, all in white.

Material de apoyo usado para las clases de Introducción a la Ciencia de Datos

Elaborado por:

Esteban Gonzalez Valencia

estebangonzalez@itm.edu.co

... Hacia una era de
Universidad y *Humanidad*