



Módulo Profesional: Big Data Aplicado

Change Data Capture
Debezium

¿QUÉ ES DEBEZIUM?

Debezium: código abierto para la captura de cambios de datos, permitiendo detectar las inserts, updates y deletes de manera muy rápida.



Haga más con sus datos



Simplifica tus aplicaciones

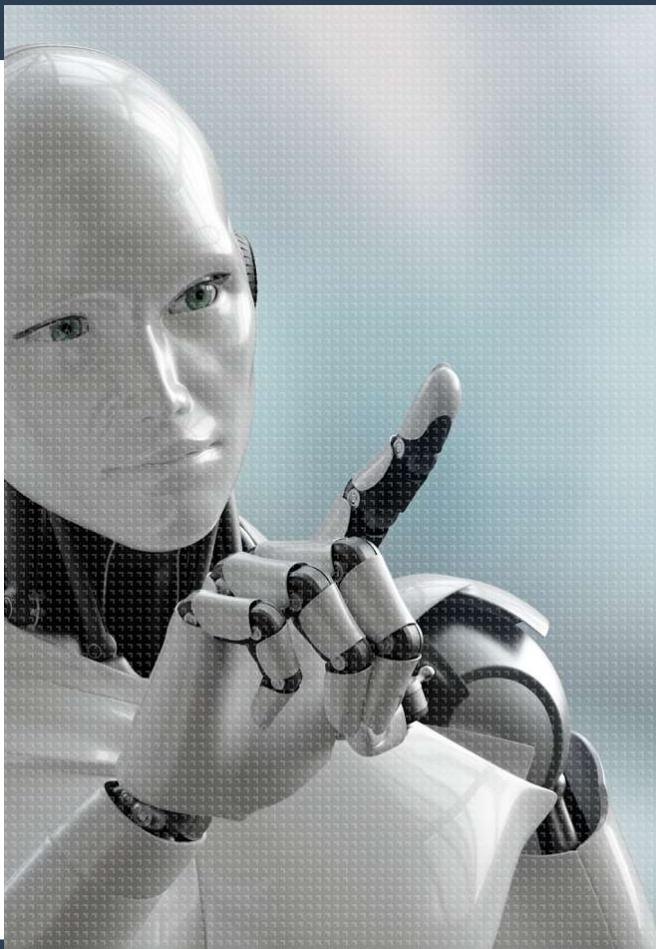


Reaccionar con rapidez



No perder ningún dato

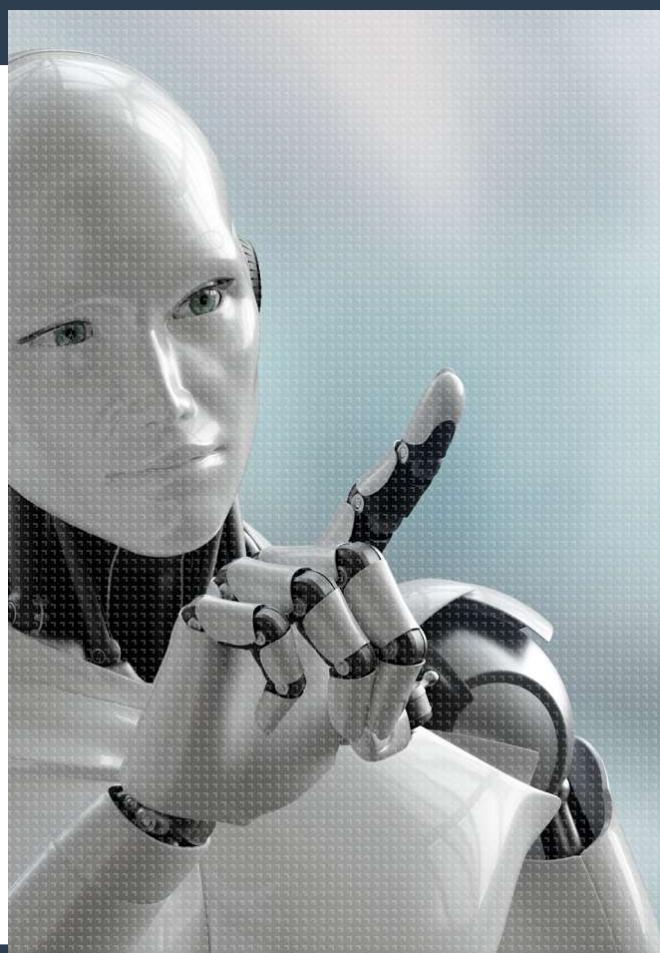
CARACTERÍSTICAS



Change Data Capture:

- Se capturan todos los cambios de datos (fiabilidad)
- Retraso muy bajo (rango ms.) → no polling (no llamadas a la base de datos)
- Sin cambios en el modelo de datos.
- Captura los deletes
- Captura old record y metadatos

CARACTERÍSTICAS



Otras capacidades:

- Permite un snapshots inicial, donde te envía todos los datos.
- Permite filtrar esquemas, tablas y columnas de la base de datos.
- Enmascaramiento de datos sensibles
- Monitorización
- Permite realizar ciertas transformaciones de mensajes para enrutamiento, filtrado, etc.

DATABASE SUPPORT



INICIAR DEBEZIUM - PASOS

1.- Descarga de Aules el ZIP

2.- Accede desde comandos al directorio local donde has descomprimido el ZIP.

```
C:\Users\SONIA>cd "\CLASES\1 clases 2024-2025\BigData\Material_Clase_Big Data\2_Ingesta_Datos_Debezium\ING-002-Debezium"
```

3.- Una vez has accedido a tu carpeta local, ejecuta el siguiente comando:

- # In PowerShell → \$env:DEBEZIUM_VERSION="2.5"

- # In Windows Command Line → set DEBEZIUM_VERSION="2.5"

- # In Linux → export DEBEZIUM_VERSION="2.5"

```
C:\CLASES\1 clases 2024-2025\BigData\Material_Clase_Big Data\2_Ingesta_Datos_Debezium\ING-002-Debezium>set DEBEZIUM_VERSION=2.5
```

INICIAR DEBEZIUM - PASOS

4.- Linea de comandos ejecutar → docker compose up -d

```
C:\CLASES\1_clases_2024-2025\BigData\Material_Clase_Big_Data\2_Ingesta_Datos_Debezium\ING-002-Debezium>docker compose up -d
time="2024-10-27T12:11:03+01:00" level=warning msg="C:\\CLASES\\1_clases_2024-2025\\BigData\\\\Material_Clase_Big_Data\\\\2_Ingesta_Datos_Debezium\\\\ING-002-Debe
zium\\docker-compose.yml: the attribute 'version' is obsolete, it will be ignored, please remove it to avoid potential confusion"
[+] Running 12/27
- kafka [####] 235.7MB / 434.6MB Pulling                                         616.4s
- adminer [#] Pulling                                                       616.4s
- connect [          ] Pulling                                              616.4s
- mysql [          ] Pulling                                              616.4s
- redpanda-console [        ] Pulling                                         616.4s
- zookeeper [      ] Pulling                                              616.4s
```

INICIAR DEBEZIUM - PASOS

5.- Instalar Debezium el conector Kafka (lo tenéis en el fichero conector_Kafka)

```
curl -i -X POST -H "Accept:application/json" -H "Content-Type:application/json"  
http://localhost:8083/connectors/ -d @register-mysql.json
```

```
C:\CLASES\1 clases 2024-2025\BigData\Material_Clase_Big Data\2_Ingesta_Datos_Dbezahlum\ING-002-Debezahlum>curl -i -X POST -H "Accept:application/json" -H "Content-Type:application/json" http://localhost:8083/connectors/ -d @register-mysql.json  
HTTP/1.1 201 Created  
Date: Sun, 27 Oct 2024 18:49:15 GMT  
Location: http://localhost:8083/connectors/inventory-connector  
Content-Type: application/json  
Content-Length: 501  
Server: Jetty(9.4.52.v20230823)  
  
{"name": "inventory-connector", "config": {"connector.class": "io.debezahlum.connector.mysql.MySqlConnector", "tasks.max": "1", "database.hostname": "mysql", "database.port": "3306", "database.user": "debezahlum", "database.password": "dbz", "database.server.id": "184054", "topic.prefix": "dbserver1", "database.include.list": "inventory", "schema.history.internal.kafka.bootstrap.servers": "kafka:9092", "schema.history.internal.kafka.topic": "schema-changes.inventory", "name": "inventory-connector"}, "tasks": [], "type": "source"}  
C:\CLASES\1 clases 2024-2025\BigData\Material_Clase_Big Data\2_Ingesta_Datos_Dbezahlum\ING-002-Debezahlum>
```

COMPONENTES

En esta instalación tenemos los siguientes componentes:

- **Kafka:** Message broker, utilizado para crear canalizaciones de datos en tiempo real y aplicaciones de streaming. Es escalable horizontalmente, tolerante a fallos y rápido.
- **Zookeeper:** Open-source server que permite una coordinación distribuida altamente fiable
- **Kafka Connect: MySQL Debezium connector,** que lee los registros de MySQL y envía datos a un tema Kafka (topic)
- **Kafka UI (Repanda Console):** UI para visualizar cluster, topics, mensajes y conectores.
- **MySQL: Open-source database,** fuente de datos.
- **Adminer (MySQL UI):** Front-end de base de datos web para navegar e interactuar con la base de datos

CONFIGURACIÓN DOCKER COMPOSE

Componentes	Servicios	Puertos	URL
Kafka	kafka	9092	N/A
Zookeeper	zookeeper	2181, 2888 y 3888	N/A
Kafka Connect	connect	8083	N/A
Kafka UI	redpanda-console	9000	http://localhost:9000/
MySQL	mysql	3306	Server: mysql Credentials: mysqluser/mysqlpw
Adminer (MySQL UI)	adminer	8090	http://localhost:8090/

REGISTER-MYSQL.JSON

```
{  
  "name": "inventory-connector",  
  "config": {  
    "connector.class": "io.debezium.connector.mysql.MySqlConnector",  
    "tasks.max": "1",  
    "database.hostname": "mysql",  
    "database.port": "3306",  
    "database.user": "debezium",  
    "database.password": "dbz",  
    "database.server.id": "184054",  
    "topic.prefix": "dbserver1",  
    "database.include.list": "inventory",  
    "schema.history.internal.kafka.bootstrap.servers": "kafka:9092",  
    "schema.history.internal.kafka.topic": "schema-changes.inventory"  
  }  
}
```

DOCKER DESKTOP

<input type="checkbox"/>	▼	 ing-002-debezium		Running (6/6)		448.99%	15 seconds ago	  	
<input type="checkbox"/>		 connect-1 bbd289b41906 	debezium/connect:2.5	Running	8083:8083 ↗	215.82%	19 seconds ago	  	
<input type="checkbox"/>		 redpanda-console-1 3c5a4f0fcc21 	redpandadata/console:v2.3.1	Running	9000:8080 ↗	0.04%	17 seconds ago	  	
<input type="checkbox"/>		 kafka-1 9e5b4c9a697e 	debezium/kafka:2.5	Running	9092:9092 ↗	219.89%	15 seconds ago	  	
<input type="checkbox"/>		 mysql-1 4665dd244781 	debezium/example-mysql:2.5	Running	3306:3306 ↗	1.07%	2 minutes ago	  	
<input type="checkbox"/>		 zookeeper-1 685953a5d872 	debezium/zookeeper:2.5	Running	2181:2181 ↗ Show all ports (3)	12.17%	2 minutes ago	  	
<input type="checkbox"/>		 adminer-1 559bcd1c2d9 	adminer:<none>	Running	8090:8080 ↗	0%	3 minutes ago	  	

KAFKA UI http://localhost:9000

Redpanda

Overview

Topics

Schema Registry

Consumer Groups

Security

Quotas

Connectors

Reassign Partitions

Create Topic

Show internal topics

Name	Partitions	Replicas	CleanupPolicy	Size
dbserver1	1	1	delete	66.3 kB
dbserver1.inventory.addresses	1	1	delete	22.4 kB
dbserver1.inventory.customers	1	1	delete	16.3 kB
dbserver1.inventory.geom	1	1	delete	9.8 kB
dbserver1.inventory.orders	1	1	delete	11.6 kB
dbserver1.inventory.products	1	1	delete	23.9 kB
dbserver1.inventory.products_on_hand	1	1	delete	22.2 kB
my_connect_configs	1	1	compact	2.18 kB
my_connect_offsets	25	1	compact	636 B
my_connect_statuses	5	1	compact	4.77 kB
schema-changes.inventory	1	1	delete	18.8 kB

Total 11 items < 1 > 50 / page

13

Adminer (MySQL UI) <http://localhost:8090>

Idioma: Español ▾

Adminer 4.8.1

Server: mysql
Credentials:
mysqluser
mysqlpw

Login

Motor de base de datos	MySQL
Servidor	db
Usuario	
Contraseña	
Base de datos	

Guardar contraseña

Adminer (MySQL UI) Schema Inventory

Idioma: Español ▾ MySQL » mysql

Adminer 4.8.1

Seleccionar Base de datos

DB:

Crear Base de datos Privilegios Lista de procesos Variables Estado

Versión MySQL: 8.2.0 a través de la extensión de PHP MySQLi

Logueado como: mysqluser@172.19.0.2

	Base de datos - Refrescar	Colación	Tablas	Size - Compute
<input type="checkbox"/>	information_schema		?	?
<input type="checkbox"/>	inventory	utf8mb4_0900_ai_ci	?	?
<input type="checkbox"/>	performance_schema		?	?

Selected (0)

Adminer (MySQL UI) Tablas Schema Inventory

Idioma: Español [Cerrar sesión](#)

MySQL » mysql » Base de datos: inventory

Adminer 4.8.1

DB: inventory

Comando SQL Importar
Exportar Crear tabla

registros addresses
registros customers
registros geom
registros orders
registros products
registros products_on_hand

Base de datos: inventory

Modificar Base de datos Esquema de base de datos Privilegios

Tablas y vistas

Buscar datos en tablas (6)

<input type="checkbox"/>	Tabla	Motor?	Colación?	Longitud de datos?	Longitud de índice?	Espacio libre?	Incremento automático?	Registros?	Comentario?
<input type="checkbox"/>	addresses	InnoDB	utf8mb4_0900_ai_ci	16 384	16 384	0	17	~ 7	
<input type="checkbox"/>	customers	InnoDB	utf8mb4_0900_ai_ci	16 384	16 384	0	1 005	~ 4	
<input type="checkbox"/>	geom	InnoDB	utf8mb4_0900_ai_ci	16 384	0	0	4	0	
<input type="checkbox"/>	orders	InnoDB	utf8mb4_0900_ai_ci	16 384	32 768	0	10 005	0	
<input type="checkbox"/>	products	InnoDB	utf8mb4_0900_ai_ci	16 384	0	0	110	~ 9	
<input type="checkbox"/>	products_on_hand	InnoDB	utf8mb4_0900_ai_ci	16 384	0	0		~ 9	
6 en total		InnoDB	utf8mb4_0900_ai_ci	98 304	65 536	0			

Selected (0)

Analizar Optimizar Comprobar Reparar Vaciar Eliminar

Mover a otra base de datos: Mover Copiar

[Crear tabla](#) [Crear vista](#)

Procedimientos

[Crear procedimiento](#) [Crear función](#)

Adminer (MySQL UI). Tabla Customer

Idioma: Español ▾ MySQL » mysql » inventory » Tabla: customers

Adminer 4.8.1

DB: inventory ▾

Comando SQL Importar
Exportar Crear tabla

registros addresses
registros **customers**
registros geom
registros orders
registros products
registros products_on_hand

Tabla: customers

Visualizar contenido **Mostrar estructura** Modificar tabla Nuevo Registro

Columna	Tipo	Comentario
id	int <i>Incremento automático</i>	
first_name	varchar(255)	
last_name	varchar(255)	
email	varchar(255)	

Índices

PRIMARY	<i>id</i>
UNIQUE	<i>email</i>

Modificar índices

Claves externas

Agregar clave externa

Disparadores

Agregar disparador

Adminer (MySQL UI). Tabla Customer

SELECT * FROM `customers` LIMIT 50 (0.001 s) Modificar

<input type="checkbox"/> Modify	id	first_name	last_name	email
<input checked="" type="checkbox"/> modificar	1001	Sally 2223333	Thomas	sally.thomas@acme.com
<input type="checkbox"/> modificar	1002	George	Bailey	nbailey@foobar.com
<input type="checkbox"/> modificar	1003	Edward	Wal	
<input type="checkbox"/> modificar	1004	Anne	Kret	

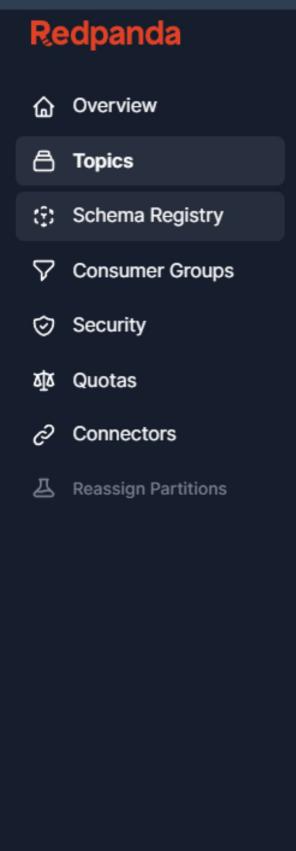
Modificar: customers

Modificamos un dato por
ejemplo el nombre
first_name

id	<input type="text"/> 1001
first_name	<input type="text"/> Sally 4444
last_name	<input type="text"/> Thomas
email	<input type="text"/> sally.thomas@acme.com

Guardar Guardar y continuar editando Eliminar

Kafka UI (Redpanda-console)



Create Topic

Name

- dbserver1
- dbserver1.inventory.addresses
- dbserver1.inventory.customers** (highlighted with a red box)
- dbserver1.inventory.geom
- dbserver1.inventory.orders
- dbserver1.inventory.products
- dbserver1.inventory.products_on_hand
- my_connect_configs
- my_connect_offsets
- my_connect_statuses
- schema-changes.inventory

Key	Value	Headers	Compression	Transactional
Json (168 B)	Json (2.57 kB)	No headers set	uncompressed	false

Key

Value

Headers

▼ { 2 items

 ▼ "schema" : { 5 items

 "type" : "struct"

 fields : [...] 6 items

 "optional" : false

 "name" : "dbserver1.inventory.customers.Envelope"

 "version" : 1

 }

 ▼ "payload" : { 6 items

 ▼ "before" : { 4 items

 "id" : 1001

 "first_name" : "Sally 2223333"

 "last_name" : "Thomas"

 "email" : "sally.thomas@acme.com"

 }

 ▼ "after" : { 4 items

 "id" : 1001

 "first_name" : "Sally 4444"

 "last_name" : "Thomas"

 "email" : "sally.thomas@acme.com"

 }

 }

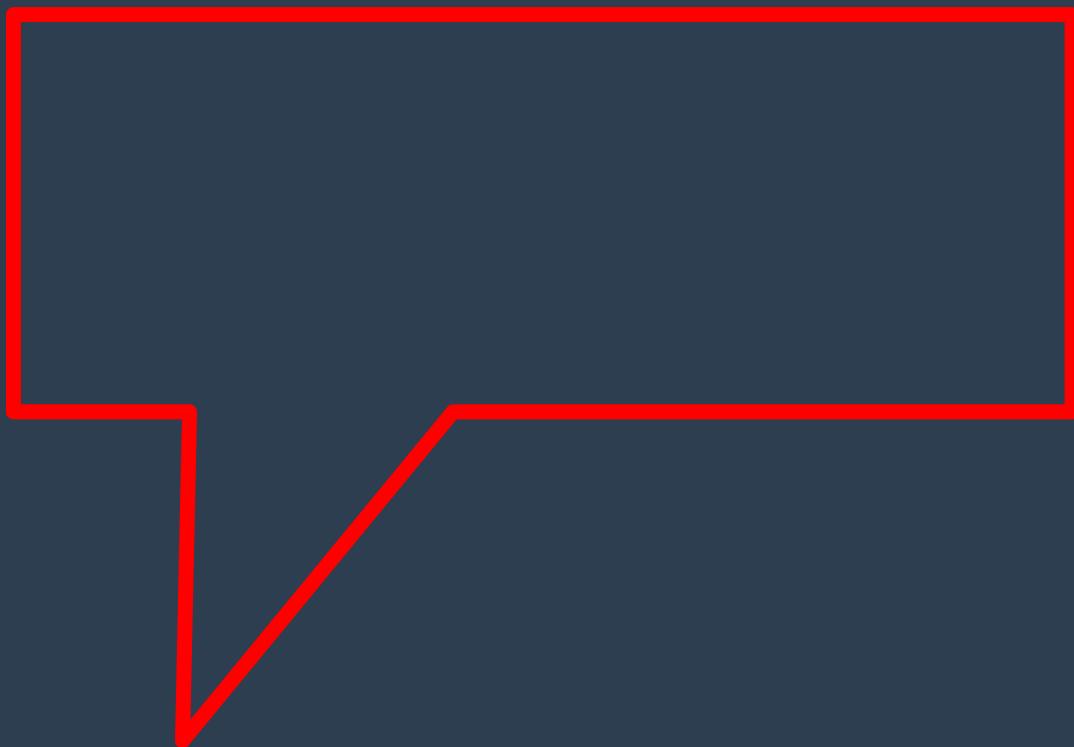
 ▼ "source" : { ... } 15 items

 "op" : "u"

 "ts_ms" : 1730107081772

 "transaction" : NULL

}



20