

5075. Big Data aplicado - 1ª Evaluación (RA 1 – CE b)

Unidad Didáctica 1. Estrategias de ingestión y almacenamiento de datos en Big Data

Práctica 6. Practicando la ingesta de datos en Batch.

Paso 1. Instalamos y configuramos MySQL.

```
hadoop@hadoop-virtualbox:~$ sudo apt update
Obj:1 http://security.ubuntu.com/ubuntu noble-security InRelease
Obj:2 http://archive.ubuntu.com/ubuntu noble InRelease
Obj:3 http://archive.ubuntu.com/ubuntu noble-updates InRelease
Obj:4 http://archive.ubuntu.com/ubuntu noble-backports InRelease
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias... Hecho
Leyendo la información de estado... Hecho
Se pueden actualizar 2 paquetes. Ejecute «apt list --upgradable» para verlos.
hadoop@hadoop-virtualbox:~$
hadoop@hadoop-virtualbox:~$ sudo apt install mysql-server
Leyendo lista de paquetes... Hecho
Creando árbol de dependencias... Hecho
Leyendo la información de estado... Hecho
```

```
hadoop@hadoop-virtualbox:~$ sudo mysql_secure_installation utility

Securing the MySQL server deployment.

Connecting to MySQL using a blank password.

VALIDATE PASSWORD COMPONENT can be used to test passwords
and improve security. It checks the strength of password
and allows the users to set only those passwords which are
secure enough. Would you like to setup VALIDATE PASSWORD component?

Press y|Y for Yes, any other key for No: y

There are three levels of password validation policy:

LOW      Length >= 8
MEDIUM  Length >= 8, numeric, mixed case, and special characters
STRONG Length >= 8, numeric, mixed case, special characters and dictionary file

Please enter 0 = LOW, 1 = MEDIUM and 2 = STRONG: 0

Skipping password set for root as authentication with auth_socket is used by default.
If you would like to use password authentication instead, this can be done with the "ALTER_USER" command.
See https://dev.mysql.com/doc/refman/8.0/en/alter-user.html#alter-user-password-management for more information.

By default, a MySQL installation has an anonymous user,
allowing anyone to log into MySQL without having to have
a user account created for them. This is intended only for
testing, and to make the installation go a bit smoother.
You should remove them before moving into a production
environment.

Remove anonymous users? (Press y|Y for Yes, any other key for No) : y
Success.
```

```
Normally, root should only be allowed to connect from
'localhost'. This ensures that someone cannot guess at
the root password from the network.

Disallow root login remotely? (Press y|Y for Yes, any other key for No) : y
Success.

By default, MySQL comes with a database named 'test' that
anyone can access. This is also intended only for testing,
and should be removed before moving into a production
environment.

Remove test database and access to it? (Press y|Y for Yes, any other key for No) : y
- Dropping test database...
Success.

- Removing privileges on test database...
Success.

Reloading the privilege tables will ensure that all changes
made so far will take effect immediately.

Reload privilege tables now? (Press y|Y for Yes, any other key for No) : y
Success.

All done!
hadoop@hadoop-virtualbox:~$
```

Actualmente podemos seguir accediendo al sistema como root utilizando el comando `sudo`. Si queremos definir una contraseña de root para localhost haremos:

```
hadoop@hadoop-virtualbox:~$ sudo mysql -u root
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 24
Server version: 8.0.39-0ubuntu0.24.04.2 (Ubuntu)

Copyright (c) 2000, 2024, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> ALTER USER 'root'@'localhost' IDENTIFIED WITH mysql_native_password BY 'hadoopHadoop';
Query OK, 0 rows affected (0,00 sec)

mysql> FLUSH PRIVILEGES;
Query OK, 0 rows affected (0,01 sec)

mysql> EXIT;
Bye
```

Hemos indicado que queremos utilizar el plugin de autenticación “mysql_native_password” (antiguas versiones). Esto lo hemos hecho dado que está habilitado “auth_socket”, utilizando el usuario de sistema para autenticarse.


```
hadoop@hadoop-virtualbox:~$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 25
Server version: 8.0.39-0ubuntu0.24.04.2 (Ubuntu)

Copyright (c) 2000, 2024, Oracle and/or its affiliates.
```

Además de lo anterior, si queremos añadir el acceso, desde cualquier ip, como root al sistema, introduciremos los comandos:

```
mysql> CREATE USER 'root'@'%' IDENTIFIED BY 'hadoopHadoop';
Query OK, 0 rows affected (0,02 sec)

mysql> GRANT ALL PRIVILEGES ON *.* TO 'root'@'%' WITH GRANT OPTION;
Query OK, 0 rows affected (0,01 sec)

mysql> FLUSH PRIVILEGES;
Query OK, 0 rows affected (0,00 sec)
```

Con el usuario creado, vamos a ver la dirección de enlace usando “netstat” para comprobar que puede conectar correctamente:

```
hadoop@hadoop-virtualbox:~$ netstat -atunpo
(No todos los procesos pueden ser identificados, no hay información de propiedad, no se mostrarán, necesita ser superusuario para verlos todos.)
Conexiones activas de Internet (servidores y establecidos)
Proto Recib Envíad Dirección local Dirección remota Estado
tcp 0 0 127.0.0.54:53 0.0.0.0:* ESCUCHAR
tcp 0 0 127.0.0.1:3306 0.0.0.0:* ESCUCHAR
tcp 0 0 127.0.0.53:53 0.0.0.0:* ESCUCHAR
```

```
hadoop@hadoop-virtualbox:~$
hadoop@hadoop-virtualbox:~$ sudo nano /etc/mysql/mysql.conf.d/mysqld.cnf
hadoop@hadoop-virtualbox:~$
```

```
# Instead of skip-networking the default
# localhost which is more compatible and
bind-address            = 0.0.0.0
mysqlx-bind-address     = 0.0.0.0
```

```
hadoop@hadoop-virtualbox:~$ sudo systemctl restart mysql
hadoop@hadoop-virtualbox:~$
hadoop@hadoop-virtualbox:~$ sudo systemctl status mysql
● mysql.service - MySQL Community Server
   Loaded: loaded (/usr/lib/systemd/system/mysql.service; enabled; preset: enabled)
   Active: active (running) since Fri 2024-08-30 21:18:41 CEST; 6s ago
     Process: 3924 ExecStartPre=/usr/share/mysql/mysql-systemd-start pre (code=exited, status=0/SUCCESS)
    Main PID: 3933 (mysqld)
      Status: "Server is operational"
        Tasks: 38 (limit: 6838)
       Memory: 365.5M (peak: 379.7M)
          CPU: 541ms
      CGroup: /system.slice/mysql.service
              └─3933 /usr/sbin/mysqld

ago 30 21:18:40 hadoop-virtualbox systemd[1]: Starting mysql.service - MySQL Community Server...
ago 30 21:18:41 hadoop-virtualbox systemd[1]: Started mysql.service - MySQL Community Server.
hadoop@hadoop-virtualbox:~$
```

```
ricardo@ricardo-virtualbox:~$
ricardo@ricardo-virtualbox:~$ mysql -u root -h 192.168.0.20 -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 8
Server version: 8.0.39-0ubuntu0.24.04.2 (Ubuntu)

Copyright (c) 2000, 2024, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql>
mysql>
mysql> exit
Bye
ricardo@ricardo-virtualbox:~$
```

Paso 2. Creamos la base de datos.

Descargaremos el fichero SQL de la dirección

<https://raw.githubusercontent.com/ricardomariscalquintero/bda/main/OnlineRetail.sql>

```
hadoop@hadoop-virtualbox:~$ wget https://raw.githubusercontent.com/ricardomariscalquintero/bda/main/OnlineRetail.sql
--2024-08-30 21:23:13-- https://raw.githubusercontent.com/ricardomariscalquintero/bda/main/OnlineRetail.sql
Resolviendo raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.108.133, 185.199.111.133, 185.199.109.133, ...
Conectando con raw.githubusercontent.com (raw.githubusercontent.com)[185.199.108.133]:443... conectado.
Petición HTTP enviada, esperando respuesta... 200 OK
Longitud: 1325071 (1.3M) [text/plain]
Guardando como: 'OnlineRetail.sql'

OnlineRetail.sql 100%[=====] 1,26M --.-KB/s en 0,09s
2024-08-30 21:23:14 (13,4 MB/s) - 'OnlineRetail.sql' guardado [1325071/1325071]
```

```
wget https://raw.githubusercontent.com/ricardomariscalquintero/bda/main/OnlineRetail.sql
```

```
DROP DATABASE IF EXISTS onlineretail;

CREATE DATABASE IF NOT EXISTS onlineretail;

USE onlineretail;

CREATE TABLE ventas (
    `id` int AUTO_INCREMENT NOT NULL,
    `pedido` varchar(7),
    `codigoStock` varchar(12),
    `descripcion` varchar(35),
    `cantidad` integer,
    `fecha` date,
    `hora` time,
    `precioUnitario` numeric(9,3),
    `idCliente` integer,
    `pais` varchar(20),
    primary key(`id`)
);

CREATE USER 'retail'@'%' IDENTIFIED BY 'retailRetail123';

GRANT ALL ON onlineretail.* TO 'retail'@'%';

FLUSH PRIVILEGES;
```

Comprobamos las variables globales de validación de las contraseñas:

```
mysql> SHOW VARIABLES LIKE 'validate_password%';
+-----+-----+
| Variable_name | Value |
+-----+-----+
| validate_password.changed_characters_percentage | 0 |
| validate_password.check_user_name | ON |
| validate_password.dictionary_file | |
| validate_password.length | 8 |
| validate_password.mixed_case_count | 1 |
| validate_password.number_count | 1 |
| validate_password.policy | MEDIUM |
| validate_password.special_char_count | 1 |
+-----+-----+
8 rows in set (0,00 sec)
```

```
mysql> SET GLOBAL validate_password.policy=LOW;
Query OK, 0 rows affected (0,00 sec)

mysql> SHOW VARIABLES LIKE 'validate_password%';
+-----+-----+
| Variable_name | Value |
+-----+-----+
| validate_password.changed_characters_percentage | 0 |
| validate_password.check_user_name | ON |
| validate_password.dictionary_file | |
| validate_password.length | 8 |
| validate_password.mixed_case_count | 1 |
| validate_password.number_count | 1 |
| validate_password.policy | LOW |
| validate_password.special_char_count | 1 |
+-----+-----+
8 rows in set (0,01 sec)
```

La explicación de lo anterior es que, aunque tras la instalación seleccionamos un nivel bajo de seguridad, como hemos reiniciado se volvió a configurar a nivel medio. Para evitar esto deberíamos de añadir, directamente a la configuración (/etc/mysql/mysql.conf.d/mysql.cnf) la directiva asociada.

```
hadoop@hadoop-virtualbox:~$ mysql -u root -p < OnlineRetail.sql
Enter password:
hadoop@hadoop-virtualbox:~$
```

```
mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| mysql |
| onlineretail |
| performance_schema |
| sys |
+-----+
5 rows in set (0,00 sec)

mysql> use onlineretail
Database changed
```



```
mysql> show tables;
+-----+
| Tables_in_onlineretail |
+-----+
| ventas                  |
+-----+
1 row in set (0,00 sec)

mysql> select * from ventas limit 5;
+-----+-----+-----+-----+-----+
| id | pedido | codigoStock | descripcion | cantidad |
+-----+-----+-----+-----+-----+
| 1 | 536365 | 85123A      | WHITE HANGING HEART T-LIGHT HOLDER | 6 |
| 2 | 536365 | 71053       | WHITE METAL LANTERN | 6 |
| 3 | 536365 | 84406B      | CREAM CUPID HEARTS COAT HANGER | 8 |
| 4 | 536365 | 84029G      | KNITTED UNION FLAG HOT WATER BOTTLE | 6 |
| 5 | 536365 | 84029E      | RED WOOLLY HOTTIE WHITE HEART. | 6 |
+-----+-----+-----+-----+-----+
5 rows in set (0,00 sec)
```

```
mysql> select * from ventas limit 5;
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| id | pedido | codigoStock | descripcion | cantidad | fecha | hora | precioUnitario | idCliente | pais |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 1 | 536365 | 85123A      | WHITE HANGING HEART T-LIGHT HOLDER | 6 | 2010-12-01 | 08:26:00 | 2.550 | 17850 | United Kingdom |
| 2 | 536365 | 71053       | WHITE METAL LANTERN | 6 | 2010-12-01 | 08:26:00 | 3.390 | 17850 | United Kingdom |
| 3 | 536365 | 84406B      | CREAM CUPID HEARTS COAT HANGER | 8 | 2010-12-01 | 08:26:00 | 2.750 | 17850 | United Kingdom |
| 4 | 536365 | 84029G      | KNITTED UNION FLAG HOT WATER BOTTLE | 6 | 2010-12-01 | 08:26:00 | 3.390 | 17850 | United Kingdom |
| 5 | 536365 | 84029E      | RED WOOLLY HOTTIE WHITE HEART. | 6 | 2010-12-01 | 08:26:00 | 3.390 | 17850 | United Kingdom |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
5 rows in set (0,00 sec)
```

Paso 3. Descargamos e instalamos la librería necesaria.

Para descargar el conector de Java nos iremos a la dirección

<https://ftp.ntu.edu.tw/MySQL/Downloads/Connector-J/>

mysql-connector-java-8.0.28.tar.gz.md5	16-Dec-2021 00:49	69
mysql-connector-java-8.0.28.zip	15-Dec-2021 23:25	5020078
mysql-connector-java-8.0.28.zip.asc	16-Dec-2021 01:18	833
mysql-connector-java-8.0.28.zip.md5	16-Dec-2021 00:49	66
mysql-connector-java-8.0.29-1.el6.noarch.rpm	11-Mar-2022 09:53	2491736
mysql-connector-java-8.0.29-1.el6.src.rpm	11-Mar-2022 09:53	4248045
mysql-connector-java-8.0.29-1.el7.noarch.rpm	11-Mar-2022 09:53	2482212
mysql-connector-java-8.0.29-1.el7.src.rpm	11-Mar-2022 09:53	4248060
mysql-connector-java-8.0.29-1.el8.noarch.rpm	11-Mar-2022 09:53	2492604
mysql-connector-java-8.0.29-1.el8.src.rpm	11-Mar-2022 09:53	4251213
mysql-connector-java-8.0.29-1.fc34.noarch.rpm	11-Mar-2022 09:53	2466050
mysql-connector-java-8.0.29-1.fc34.src.rpm	11-Mar-2022 09:53	4251289
mysql-connector-java-8.0.29-1.fc35.noarch.rpm	11-Mar-2022 09:53	2465983
mysql-connector-java-8.0.29-1.fc35.src.rpm	11-Mar-2022 09:53	4251276
mysql-connector-java-8.0.29-1.sl15.noarch.rpm	11-Mar-2022 09:53	2476240
mysql-connector-java-8.0.29-1.sl15.src.rpm	11-Mar-2022 09:53	4251143
mysql-connector-java-8.0.29-1.sles12.noarch.rpm	11-Mar-2022 09:53	2472092
mysql-connector-java-8.0.29-1.sles12.src.rpm	11-Mar-2022 09:53	4247941
mysql-connector-java-8.0.29.tar.gz	08-Mar-2022 18:21	4273713
mysql-connector-java-8.0.29.tar.gz.asc	11-Mar-2022 10:29	833
mysql-connector-java-8.0.29.tar.gz.md5	11-Mar-2022 09:48	69
mysql-connector-java-8.0.29.zip	08-Mar-2022 18:21	5100429
mysql-connector-java-8.0.29.zip.asc	11-Mar-2022 10:29	833
mysql-connector-java-8.0.29.zip.md5	11-Mar-2022 09:48	66


```
hadoop@hadoop-virtualbox:~$ wget https://ftp.ntu.edu.tw/MySQL/Downloads/Connector-J/mysql-connector-java-8.0.29.tar.gz
--2024-08-30 21:57:05-- https://ftp.ntu.edu.tw/MySQL/Downloads/Connector-J/mysql-connector-java-8.0.29.tar.gz
Resolviendo ftp.ntu.edu.tw (ftp.ntu.edu.tw)... 140.112.36.185
Conectando con ftp.ntu.edu.tw (ftp.ntu.edu.tw)[140.112.36.185]:443... conectado.
Petición HTTP enviada, esperando respuesta... 200 OK
Longitud: 4273713 (4,1M) [application/octet-stream]
Guardando como: 'mysql-connector-java-8.0.29.tar.gz'

mysql-connector-java-8.0.29.tar.gz 100%[=====]
2024-08-30 21:57:18 (383 KB/s) - 'mysql-connector-java-8.0.29.tar.gz' guardado [4273713/4273713]
```

Copiamos el driver dentro del directorio de NiFi.

```
hadoop@hadoop-virtualbox:~$
hadoop@hadoop-virtualbox:~$ mkdir /opt/nifi/drivers
hadoop@hadoop-virtualbox:~$
```

```
hadoop@hadoop-virtualbox:~$ tar -xvzf mysql-connector-java-8.0.29.tar.gz
mysql-connector-java-8.0.29/
mysql-connector-java-8.0.29/src/
mysql-connector-java-8.0.29/src/build/
mysql-connector-java-8.0.29/src/build/java/
mysql-connector-java-8.0.29/src/build/java/documentation/
```

```
hadoop@hadoop-virtualbox:~$ cp mysql-connector-java-8.0.29/mysql-connector-java-8.0.29.jar /opt/nifi/drivers/
hadoop@hadoop-virtualbox:~$
hadoop@hadoop-virtualbox:~$ ls -lart /opt/nifi/drivers/
total 2472
drwxrwxr-x 16 hadoop hadoop 4096 ago 30 21:59 ..
drwxrwxr-x 2 hadoop hadoop 4096 ago 30 22:01 .
-rw-r--r-- 1 hadoop hadoop 2519547 ago 30 22:01 mysql-connector-java-8.0.29.jar
hadoop@hadoop-virtualbox:~$
```

Paso 4. Accedemos a la interfaz de NiFi y damos de alta el proceso para consultar la base de datos creada.

Displaying 2 of 360



Type ^	Version	Tags
QueryDatabaseTable	1.27.0	database, select, query, jdbc, sql
QueryDatabaseTableRecord	1.27.0	database, select, query, record, j...


Configure Processor | QueryDatabaseTable 1.27.0

 Invalid

SETTINGS	SCHEDULING	PROPERTIES	RELATIONSHIPS	COMMENTS
----------	------------	------------	---------------	----------

Required field

Property	Value
Database Connection Pooling Service	 No value set
Database Type	 Generic

Reference parameter... 

Reference parameter...

Create new service...

CANCEL


OK

Add Controller Service

Requires Controller Service

DBCPService 1.27.0 from org.apache.nifi - nifi-standard-services-api-nar

Compatible Controller Services

DBCPConnectionPool 1.27.0 

Controller Service Name

MySQL

Bundle

org.apache.nifi - nifi-dbcp-service-nar

Tags

database, pooling, dbcp, restricted, jdbc, connection, store

Description

CANCEL

CREATE

Configure Processor | QueryDatabaseTable 1.27.0

 Invalid

SETTINGS

SCHEDULING

PROPERTIES

RELATIONSHIPS

COMMENTS

Required field



Property	Value
Database Connection Pooling Service	MySQL
Database Type	Generic

Processor Configuration

Save changes before going to this Controller Service?

NO

YES

NiFi Flow Configuration



GENERAL

CONTROLLER SERVICES

Name	Type	Bundle	State	Scope
MySQL	DBCPConnectionPool 1.27.0	org.apache.nifi-nifi-dbg-service-nar	Invalid	NiFi Flow

Scope

NiFi Flow



Configure Controller Service | DBCPConnectionPool 1.27.0

INVALID

SETTINGS PROPERTIES COMMENTS

Required field

Property	Value
Database Connection URL	
Database Driver Class Name	
Database Driver Location(s)	
Kerberos User Service	1 jdbc:mysql://localhost:3306/onlineRetail?serverTimezone=UTC
Kerberos Credentials Service	

Configure Controller Service | DBCPConnectionPool 1.27.0

INVALID

SETTINGS PROPERTIES COMMENTS

Required field

Property	Value
Database Connection URL	jdbc:mysql://localhost:3306/onlineRetail?serv...
Database Driver Class Name	com.mysql.jdbc.Driver
Database Driver Location(s)	/opt/nifi/drivers/mysql-connector-java-8.0.29.jar
Kerberos User Service	No value set
Kerberos Credentials Service	No value set
Kerberos Principal	No value set
Kerberos Password	No value set
Database User	retail
Password	Sensitive value set
Max Wait Time	500 millis
Max Total Connections	8
Validation query	No value set

Scope
NiFi Flow

Enable Controller Service

Service
MySQL

Scope

Service only

Referencing Components ?

▼ Processors (1)

▲ QueryDatabaseTable QueryDatabaseTable

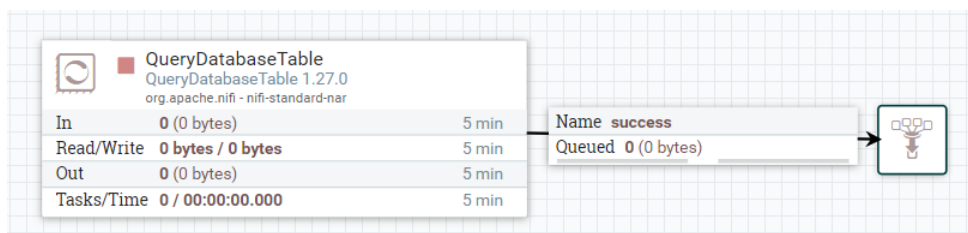
Configure Processor | QueryDatabaseTable 1.27.0

Invalid

SETTINGS	SCHEDULING	PROPERTIES	RELATIONSHIPS	COMMENTS																										
<p>Required field</p>																														
<table><thead><tr><th>Property</th><th>Value</th></tr></thead><tbody><tr><td>Database Connection Pooling Service</td><td>MySQL</td></tr><tr><td>Database Type</td><td>MySQL</td></tr><tr><td>Table Name</td><td>ventas</td></tr><tr><td>Columns to Return</td><td>No value set</td></tr><tr><td>Additional WHERE clause</td><td>No value set</td></tr><tr><td>Custom Query</td><td>No value set</td></tr><tr><td>Maximum-value Columns</td><td>No value set</td></tr><tr><td>Initial Load Strategy</td><td>Start at Beginning</td></tr><tr><td>Max Wait Time</td><td>0 seconds</td></tr><tr><td>Fetch Size</td><td>0</td></tr><tr><td>Set Auto Commit</td><td>No value set</td></tr><tr><td>Max Rows Per Flow File</td><td>0</td></tr></tbody></table>					Property	Value	Database Connection Pooling Service	MySQL	Database Type	MySQL	Table Name	ventas	Columns to Return	No value set	Additional WHERE clause	No value set	Custom Query	No value set	Maximum-value Columns	No value set	Initial Load Strategy	Start at Beginning	Max Wait Time	0 seconds	Fetch Size	0	Set Auto Commit	No value set	Max Rows Per Flow File	0
Property	Value																													
Database Connection Pooling Service	MySQL																													
Database Type	MySQL																													
Table Name	ventas																													
Columns to Return	No value set																													
Additional WHERE clause	No value set																													
Custom Query	No value set																													
Maximum-value Columns	No value set																													
Initial Load Strategy	Start at Beginning																													
Max Wait Time	0 seconds																													
Fetch Size	0																													
Set Auto Commit	No value set																													
Max Rows Per Flow File	0																													
<p>CANCEL APPLY</p>																														

El campo “Maximum-value Columns” es muy importante. Esto es debido a que es la columna que tomará de referencia como delta. De esta manera, cuando realice una ingesta, la siguiente vez que lo haga la realizará a partir del máximo valor guardado.

El motivo por el que lo hemos dejado en blanco es para que realice la ingesta infinitas veces.



Paso 5. Comprobamos el acceso a la tabla de ventas.

QueryDatabaseTable
QueryDatabaseTable 1.27.0
org.apache.nifi - nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 945,34 KB	5 min
Out	1 (945,34 KB)	5 min
Tasks/Time	1 / 00:00:00.640	5 min

Name success
Queued 1 (945,34 KB)

- Configure
- View status history
- List queue
- Go to source
- Go to destination
- Bring to front
- Create template
- Empty queue
- Delete

SUCCESS

Displaying 1 of 1 (945.34 KB)

Position	UUID	Filename	File Size
1	62f1edd8-0f0f-4b6d-a84d-f806b091dacb	62f1edd8-0f0f-4b6d-a84d-f806b091dacb	945.34 KB

The destination of this queue is currently running. This listing may no longer be accurate.

Lineage Duration	Penalized
00:01:04.375	No

Al visualizar los datos, teniendo en cuenta que el formato de salida es Avro, vamos a poder ver, además de los datos, el esquema de los mismos.

View as: original

```

1  Obj{...avro.schema*{"type":"record","name":"ventas","namespace":"any.data","fields":[{"name":"id","type":["null","int"]}, {"name":"pedido".
2  5e...536365...85123A...WHITE HANGING HEART T-LIGHT HOLDER...2010-12-01...08:26:00
3  2.550...5...United Kingdom...536365
4  71053...WHITE METAL LANTERN...2010-12-01...08:26:00
5  3.390...5...United Kingdom...536365...844068...CREAM CUPID HEARTS COAT HANGER...2010-12-01...08:26:00
6  2.750...5...United Kingdom...536365...84029G...FKNITTED UNION FLAG HOT WATER BOTTLE...2010-12-01...08:26:00
7  3.390...5...United Kingdom
8  ...536365...84029E...RED WOOLLY HOTTIE WHITE HEART...2010-12-01...08:26:00
9  3.390...5...United Kingdom...536365
10  22752...8SET 7 BABUSHKA NESTING BOXES...2010-12-01...08:26:00
11  7.650...5...United Kingdom...536365
12  21730...8GLASS STAR FROSTED T-LIGHT HOLDER...2010-12-01...08:26:00
13  4.250...5...United Kingdom...536366
14  77623...HAPPY WHISKY BOTTLE...2010-12-01...08:26:00

```


Paso 6. Convertimos la salida en JSON y la escribimos en disco.

Displaying 7 of 360

Convert

Type ^	Version	Tags
ConvertAvroToJson	1.27.0	json, convert, avro
ConvertAvroToParquet	1.27.0	convert, avro, parquet
ConvertCharacterSet	1.27.0	character set, character set, text...
ConvertExcelToCSVProcessor	1.27.0	excel, csv, poi
ConvertJSONToSQL	1.27.0	database, rdbms, flat, json, inse...
ConvertRecord	1.27.0	schema, log, record, csv, freefor...
LookupRecord	1.27.0	filter, lookup, enrichment, datab...

Configure Processor | ConvertAvroToJson 1.27.0

Invalid

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Required field

Property	Value
JSON container options	array
Wrap Single Record	false
Avro schema	No value set

Configure Processor | UpdateAttribute 1.27.0

Invalid

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

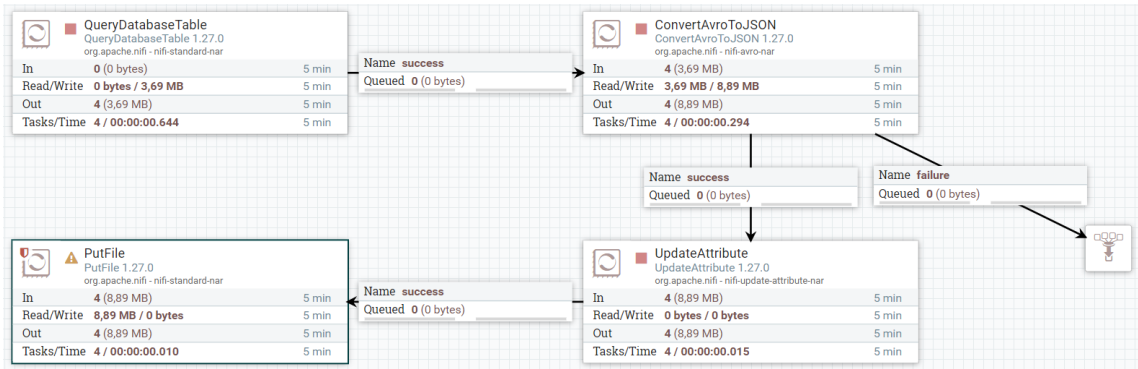
Required field

Property	Value
Delete Attributes Expression	No value set
Store State	Do not store state
Stateful Variables Initial Value	No value set
Cache Value Lookup Cache Size	100
filename	\${now().format('yyyy-MM-dd_HH-mm-ss')}

Configure Processor | PutFile 1.27.0

Invalid

SETTINGS	SCHEDULING	PROPERTIES	RELATIONSHIPS	COMMENTS
Required field				
Property		Value		
Directory		workspace/salidaNifi/		
Conflict Resolution Strategy		replace		
Create Missing Directories		true		
Maximum File Count		No value set		
Last Modified Time		No value set		
Permissions		No value set		
Owner		No value set		
Group		No value set		



NiFi Data Provenance

Displaying 20 of 20
Oldest event available: 08/29/2024 03:07:23 CEST

Filter	by component name				
	Date/Time	Type	FlowFile Uuid	Size	Component Name
	08/31/2024 00:52:56.509 CEST	DROP	b72c6409-9a89-4742-9fd2-c712fdaa45ad	222 MB	PutFile
	08/31/2024 00:52:56.508 CEST	SEND	b72c6409-9a89-4742-9fd2-c712fdaa45ad	222 MB	PutFile
	08/31/2024 00:52:36.416 CEST	DROP	fd7abab8-ea84-4645-b35d-b3dc78e6ca5e	222 MB	PutFile
	08/31/2024 00:52:36.415 CEST	SEND	fd7abab8-ea84-4645-b35d-b3dc78e6ca5e	222 MB	PutFile
	08/31/2024 00:52:16.308 CEST	DROP	7902dace-a362-4f22-9da0-ac3e1ee08a3	222 MB	PutFile

```
hadoop@hadoop-virtualbox:~$ ls -latrh workspace/salidaNifi/
total 9,0M
drwxrwxr-x 4 hadoop hadoop 4,0K ago 29 02:38 ..
-rw-rw-r-- 1 hadoop hadoop 2,3M ago 31 00:51 2024-08-31_00-51-36
-rw-rw-r-- 1 hadoop hadoop 2,3M ago 31 00:51 2024-08-31_00-51-56
-rw-rw-r-- 1 hadoop hadoop 2,3M ago 31 00:52 2024-08-31_00-52-16
-rw-rw-r-- 1 hadoop hadoop 2,3M ago 31 00:52 2024-08-31_00-52-36
drwxrwxr-x 2 hadoop hadoop 4,0K ago 31 00:52 .
hadoop@hadoop-virtualbox:~$
```

```

{"id": 1, "pedido": "536365", "codigoStock": "85123A", "descripcion": "WHITE HANGING HEART T-LIGHT HOLDER", "cantidad": 1, "precioUnitario": "2.550", "idCliente": 17850, "pais": "United Kingdom"}, {"id": 2, "pedido": "536365", "codigoStock": "71053", "descripcion": "2010-12-01", "cantidad": 1, "precioUnitario": "3.390", "idCliente": 17850, "pais": "United Kingdom"}, {"id": 3, "pedido": "536365", "codigoStock": "84029C", "descripcion": "CREAM CUPID HEARTS COAT HANGER", "cantidad": 8, "fecha": "2010-12-01", "hora": "08:26:00", "precioUnitario": "3.390", "idCliente": 17850, "pais": "United Kingdom"}, {"id": 4, "pedido": "536365", "codigoStock": "84029C", "descripcion": "KNITTED UNION FLAG HOT WATER BOTTLE", "cantidad": 1, "precioUnitario": "3.390", "idCliente": 17850, "pais": "United Kingdom"}, {"id": 5, "pedido": "536365", "codigoStock": "84029C", "descripcion": "HEART", "cantidad": 6, "fecha": "2010-12-01", "hora": "08:26:00", "precioUnitario": "3.390", "idCliente": 17850, "pais": "United Kingdom"}, {"id": 6, "pedido": "536365", "codigoStock": "22752", "descripcion": "SET 7 BABUSHKA NESTING BOXES", "cantidad": 2, "fecha": "2010-12-01", "hora": "08:26:00", "precioUnitario": "4.250", "idCliente": 17850, "pais": "United Kingdom"}, {"id": 7, "pedido": "536365", "codigoStock": "21730", "descripcion": "GLASS STAR FROSTED T-MUG", "cantidad": 6, "fecha": "2010-12-01", "hora": "08:28:00", "precioUnitario": "1.850", "idCliente": 17850, "pais": "United Kingdom"}, {"id": 8, "pedido": "536365", "codigoStock": "22632", "descripcion": "HAND WARMER RED POLKA DOT", "cantidad": 6, "fecha": "2010-12-01", "hora": "08:34:00", "precioUnitario": "1.690", "idCliente": 13047, "pais": "United Kingdom"}, {"id": 9, "pedido": "536367", "codigoStock": "84879", "descripcion": "ASSORTED COLOUR PLAYHOUSE BEDROOM", "cantidad": 6, "fecha": "2010-12-01", "hora": "08:34:00", "precioUnitario": "2.100", "idCliente": 17850, "pais": "United Kingdom"}, {"id": 10, "pedido": "536367", "codigoStock": "22748", "descripcion": "POPPY'S PLAYHOUSE KITCHEN", "cantidad": 6, "fecha": "2010-12-01", "hora": "08:34:00", "precioUnitario": "2.100", "idCliente": 17850, "pais": "United Kingdom"}

```

Displaying 1 of 360		ConvertRe
Type ^	Version	Tags
ConvertRecord	1.27.0	schema, log, record, csv, freefor...

Configure Processor

ConvertRecord 1.27.0

Invalid

SETTINGS

SCHEDULING

PROPERTIES

RELATIONSHIPS

COMMENTS

Required field

Property	Value
Record Reader	No value set
Record Writer	No value set
Include Zero Record FlowFiles	true

SETTINGS

SCHEDULING

PROPERTIES

RELATIONSHIPS

COMMENTS

Required field

Record Reader	
Record Writer	
Include Zero Record FlowFiles	

Reference parameter...

Reference parameter...

Create new service...

Create new service... OK

Add Controller Service

Requires Controller Service

RecordReaderFactory 1.27.0 from org.apache.nifi - nifi-standard-services-api-nar

Compatible Controller Services

JsonTreeReader 1.27.0

Controller Service Name

JsonTreeReader

Property	Value
Record Reader	
Record Writer	<div>Reference parameter... Reference parameter... Create new service...</div>
Include Zero Record FlowFiles	

Create new service...

Add Controller Service

Requires Controller Service

RecordSetWriterFactory 1.27.0 from org.apache.nifi - nifi-standard-services-api-nar

Compatible Controller Services




CSVRecordSetWriter 1.27.0












Controller Service Name

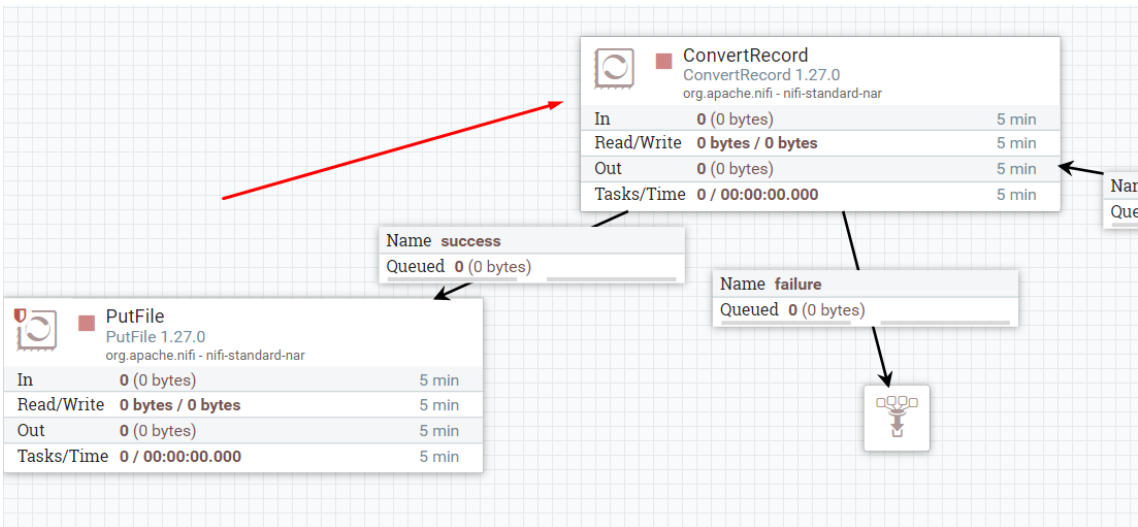
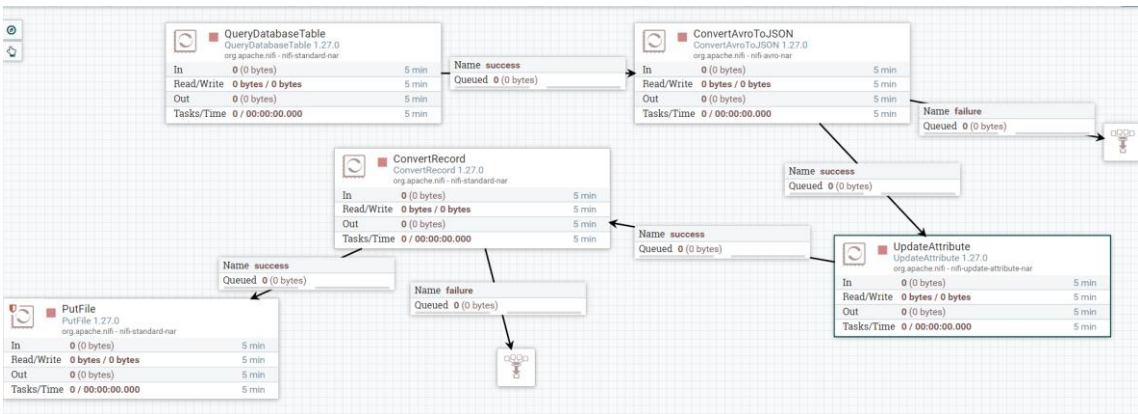
CSVRecordSetWriter

NIFI Flow Configuration

GENERAL						
CONTROLLER SERVICES						
Name	Type	Bundle	State	Scope		
CSVRecordSetWriter	CSVRecordSetWriter 1.27.0	org.apache.nifi-nifi-record-serialization-services-nar	Enabled	NIFI Flow	1	1
JsonTreeReader	JsonTreeReader 1.27.0	org.apache.nifi-nifi-record-serialization-services-nar	Enabled	NIFI Flow	1	1
MySQL	DBCPConnectionPool 1.27.0	org.apache.nifi-nifi-dbcp-service-nar	Enabled	NIFI Flow	1	1

	Name ^	Type
	CSVRecordSetWriter	CSVRecordSetWriter 1.27.0
	JsonTreeReader	JsonTreeReader 1.27.0
	MySQL	DBCPConnectionPool 1.27.0

State	Scope	
 Disabled	NiFi Flow	  
 Disabled	NiFi Flow	  
 Enabled	NiFi Flow	 



```
hadoop@hadoop-virtualbox:~$ ls -lart workspace/salidaNifi/  
total 3656  
drwxrwxr-x 4 hadoop hadoop 4096 ago 29 02:38 ..  
-rw-rw-r-- 1 hadoop hadoop 932582 ago 31 01:06 2024-08-31_01-06-58.csv  
-rw-rw-r-- 1 hadoop hadoop 932582 ago 31 01:07 2024-08-31_01-07-11.csv  
-rw-rw-r-- 1 hadoop hadoop 932582 ago 31 01:07 2024-08-31_01-07-31.csv  
-rw-rw-r-- 1 hadoop hadoop 932582 ago 31 01:07 2024-08-31_01-07-51.csv  
drwxrwxr-x 2 hadoop hadoop 4096 ago 31 01:07 .  
hadoop@hadoop-virtualbox:~$
```

```
id,pedido,codigoStock,descripcion,cantidad,fecha,hora,precioUnitario,idCliente,pais  
1,536365,85123A,WHITE HANGING HEART T-LIGHT HOLDER,6,2010-12-01,08:26:00,2.550,17850,United Kingdom  
2,536365,71053,WHITE METAL LANTERN,6,2010-12-01,08:26:00,3.390,17850,United Kingdom  
3,536365,84406B,CREAM CUPID HEARTS COAT HANGER,8,2010-12-01,08:26:00,2.750,17850,United Kingdom  
4,536365,84029G,KNITTED UNION FLAG HOT WATER BOTTLE,6,2010-12-01,08:26:00,3.390,17850,United Kingdom  
5,536365,84029E,RED WOOLLY HOTTIE WHITE HEART.,6,2010-12-01,08:26:00,3.390,17850,United Kingdom  
6,536365,22752,SET 7 BABUSHKA NESTING BOXES,2,2010-12-01,08:26:00,7.650,17850,United Kingdom  
7,536365,21730,GLASS STAR FROSTED T-LIGHT HOLDER,6,2010-12-01,08:26:00,4.250,17850,United Kingdom  
8,536366,22633,HAND WARMER UNION JACK,6,2010-12-01,08:28:00,1.850,17850,United Kingdom  
9,536366,22632,HAND WARMER RED POLKA DOT,6,2010-12-01,08:28:00,1.850,17850,United Kingdom  
10,536367,84879,ASSORTED COLOUR BIRD ORNAMENT,32,2010-12-01,08:34:00,1.690,13047,United Kingdom  
11,536367,22745,POPPY'S PLAYHOUSE BEDROOM,6,2010-12-01,08:34:00,2.100,13047,United Kingdom  
12,536367,22748,POPPY'S PLAYHOUSE KITCHEN,6,2010-12-01,08:34:00,2.100,13047,United Kingdom  
13,536367,22749,FELTCRAFT PRINCESS CHARLOTTE DOLL,8,2010-12-01,08:34:00,3.750,13047,United Kingdom  
14,536367,22310,IVORY KNITTED MUG COSY,6,2010-12-01,08:34:00,1.650,13047,United Kingdom  
15,536367,84969,BOX OF 6 ASSORTED COLOUR TEASPOONS,6,2010-12-01,08:34:00,4.250,13047,United Kingdom  
16,536367,22623,BOX OF VINTAGE JIGSAW BLOCKS,3,2010-12-01,08:34:00,4.950,13047,United Kingdom  
17,536367,22622,BOX OF VINTAGE ALPHABET BLOCKS,2,2010-12-01,08:34:00,9.950,13047,United Kingdom  
18,536367,21754,HOME BUILDING BLOCK WORD,3,2010-12-01,08:34:00,5.950,13047,United Kingdom  
19,536367,21755,LOVE BUILDING BLOCK WORD,3,2010-12-01,08:34:00,5.950,13047,United Kingdom  
20,536367,21777,RECIPE BOX WITH METAL HEART,4,2010-12-01,08:34:00,7.950,13047,United Kingdom  
21,536367,48187,DOORMAT NEW ENGLAND,4,2010-12-01,08:34:00,7.950,13047,United Kingdom  
22,536368,22960,JAM MAKING SET WITH JARS,6,2010-12-01,08:34:00,4.250,13047,United Kingdom  
23,536368,22913,RED COAT RACK PARIS FASHION,3,2010-12-01,08:34:00,4.950,13047,United Kingdom  
24,536368,22912,YELLOW COAT RACK PARIS FASHION,3,2010-12-01,08:34:00,4.950,13047,United Kingdom  
25,536368,22914,BLUE COAT RACK PARIS FASHION,3,2010-12-01,08:34:00,4.950,13047,United Kingdom  
26,536369,21756,BATH BUILDING BLOCK WORD,3,2010-12-01,08:35:00,5.950,13047,United Kingdom
```