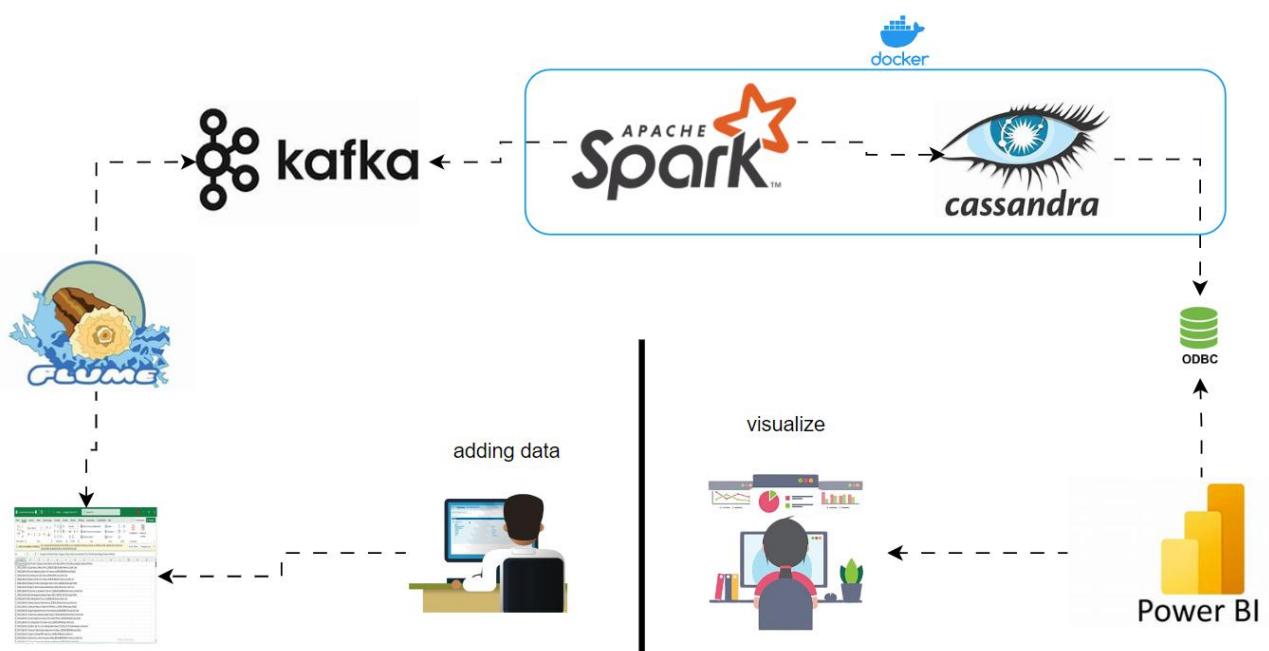


# GUIDE COMPLET DE MISE EN PLACE D'UN

## PIPELINE DE DONNEES TEMPS REEL



Edson KANOU TAYOUTSOP



Edson KANOU



Edsonkanou55@gmail.com



# Sommaire

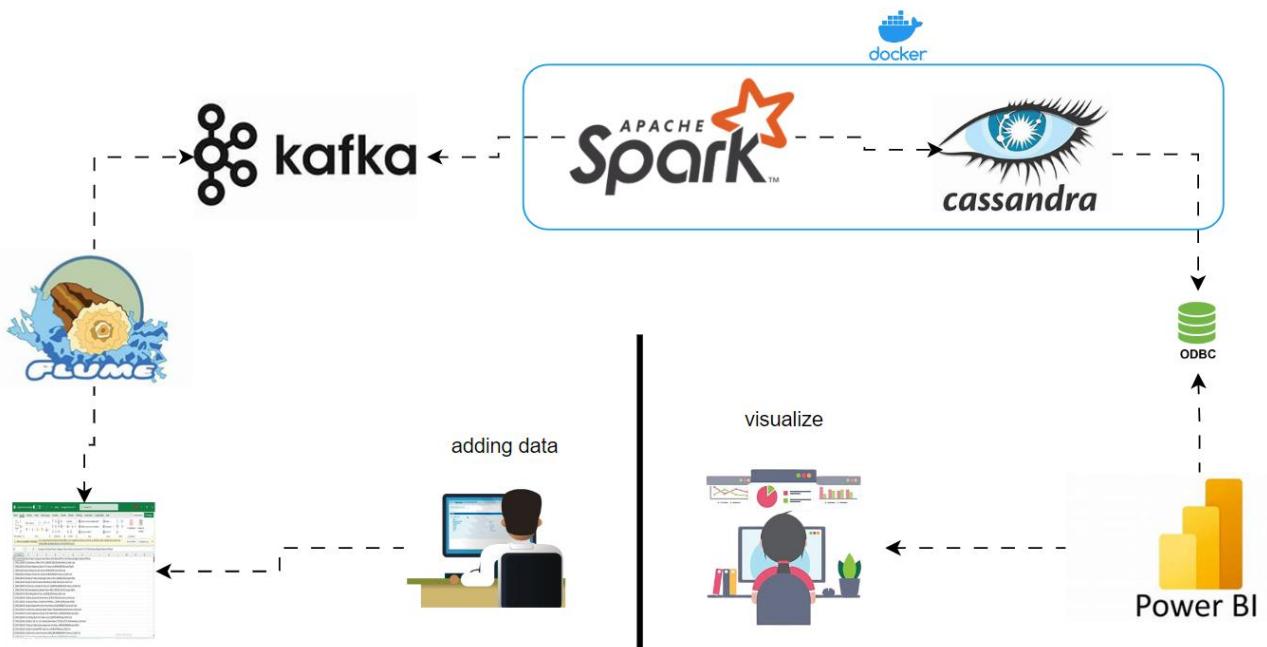
INTRODUCTION .....	3
PARTIE I : Etapes contractuelles.....	4
PARTIE II : Présentation de la source de données.....	8
Etape I : Apache FLUME .....	9
Etape II : Apache KAFKA.....	14
Etape III : Apache CASSANDRA .....	17
Etape IV : Apache Spark.....	22
Etape V : Configuration de la source de donnees ODBC .....	24
Etape VI : Visualisation POWER BI .....	26

# INTRODUCTION

Le présent document détaille un pipeline de traitement de données en temps réel conçu pour ingérer, transformer et stocker des flux de données volumineux. Ce pipeline s'appuie sur une architecture distribuée composée

## Pipeline :

1. **Apache Flume** collecte les données de la source et les envoie à **Apache Kafka**.
2. **Spark Streaming (pyspark)** lit les données de Kafka, les transforme et les écrit dans **Apache Cassandra**.
3. **Power BI** se connecte à Cassandra pour créer des visualisations interactives.



# PARTIE I : Etapes contractuelles

## 1. Exécuter le fichier flume

```
flume-ng agent -n al -c conf -f C:/flume/conf/q.conf -property flume.root.logger=DEBUG,console
```



```
C:\flume>flume-ng agent -n al -c conf -f C:/flume/conf/flume-kafka.conf -property flume.root.logger=DEBUG,console
C:\flume>powershell.exe -NoProfile -InputFormat none -ExecutionPolicy unrestricted -File C:\flume\bin\flume-ng.ps1 agent -n al -c con
f -f C:/flume/conf/flume-kafka.conf -property flume.root.logger=DEBUG,console

Avertissement de sécurité
N'exécutez que des scripts que vous approuvez. Bien que les
scripts en provenance d'Internet puissent être utiles, ce
script est susceptible d'endommager votre ordinateur. Si vous
approuvez ce script, utilisez l'applet de commande Unblock-File
pour autoriser le script à s'exécuter sans ce message
d'avertissement. Voulez-vous exécuter
C:\flume\bin\flume-ng.ps1 ?
[N] Ne pas exécuter [O] Exécuter une fois [S] Suspendre
[?] Aide(Cla valeur par défaut est « N ») : 0

WARN: Config directory not set. Defaulting to C:\flume\conf
Sourcing environment configuration script C:\flume\conf\flume-env.ps1
Including Hadoop libraries found in (C:\hadoop) for DFS access
Le chemin d'accès spécifié est introuvable.
Le chemin d'accès spécifié est introuvable.
'-Dhadoop.security.logger' n'est pas reconnu en tant que commande interne
ou externe, un programme exécutable ou un fichier de commandes.
WARN: HBASE_HOME not found
WARN: HIVE_HOME not found

Running FLUME agent :
  class: org.apache.flume.node.Application
  arguments: -n al -f "C:/flume/conf/flume-kafka.conf"

2024-11-09 09:43:05,412 (main) [DEBUG - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95)] No global SSL keystore
path specified.
2024-11-09 09:43:05,417 (main) [DEBUG - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95)] No global SSL keystore
```

## 2. Vérifier le consommateur kafka

```
kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic flume-topic2 --from-beginning
```



```
C:\Software\kafka\bin\windows>kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic flume-topic1 --from-beginning
Processed a total of 0 messages
Terminer le programme de commandes (O/N) ?
^C
C:\Software\kafka\bin\windows>kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic flume-topic1 --from-beginning
10002, 2024-01-02, Home Appliances, Dyson V11 Vacuum, 1,499.99, 499.99, Europe, PayPal
10005, 2024-01-05, Beauty Products, Neutrogena Skincare Set, 1,89.99, 89.99, Europe, PayPal
Transaction ID, Date, Product Category, Product Name, Units Sold, Unit Price, Total Revenue, Region, Payment Method
10003, 2024-01-03, Clothing, Levi's 501 Jeans, 3,69.99, 209.97, Asia, Debit Card
10006, 2024-01-06, Sports, Wilson Evolution Basketball, 5,29.99, 149.95, Asia, Credit Card
10001, 2024-01-01, Electronics, iPhone 14 Pro, 2,999.99, 1999.98, North America, Credit Card
10004, 2024-01-04, Books, The Da Vinci Code, 4,15.99, 63.96, North America, Credit Card
10007, 2024-01-07, Electronics, MacBook Pro 16-inch, 1,2499.99, 2499.99, North America, Credit Card
10008, 2024-01-08, Home Appliances, Blueair Classic 480i, 2,599.99, 1199.98, Europe, PayPal
10011, 2024-01-11, Beauty Products, Chanel No. 5 Perfume, 1,129.99, 129.99, Europe, PayPal
10014, 2024-01-14, Home Appliances, Keurig K-Elite Coffee Maker, 1,189.99, 189.99, Europe, PayPal
10009, 2024-01-09, Clothing, Nike Air Force 1, 6,89.99, 539.94, Asia, Debit Card
10012, 2024-01-12, Sports, Babolat Pure Drive Tennis Racket, 3,199.99, 599.97, Asia, Credit Card
10015, 2024-01-15, Clothing, North Face Down Jacket, 2,249.99, 499.98, Asia, Debit Card
10010, 2024-01-10, Books, Dune by Frank Herbert, 2,25.99, 51.98, North America, Credit Card
10013, 2024-01-13, Electronics, Samsung Galaxy Tab S8, 2,749.99, 1499.98, North America, Credit Card
10017, 2024-01-17, Beauty Products, Dyson Supersonic Hair Dryer, 1,399.99, 399.99, Europe, PayPal
10020, 2024-01-20, Home Appliances, Ninja Professional Blender, 1,99.99, 99.99, Europe, PayPal
10023, 2024-01-23, Beauty Products, Olay Regenerist Face Cream, 1,49.99, 49.99, Europe, PayPal
10018, 2024-01-18, Sports, Manduka PRO Yoga Mat, 4,119.99, 479.96, Asia, Credit Card
10021, 2024-01-21, Clothing, Zara Summer Dress, 3,59.99, 179.97, Asia, Debit Card
10024, 2024-01-24, Sports, Adidas FIFA World Cup Football, 3,29.99, 89.97, Asia, Credit Card
10016, 2024-01-16, Books, "Salt, Fat, Acid, Heat by Samin Nosrat", 3,35.99, 107.97, North America, Credit Card
10019, 2024-01-19, Electronics, Garmin Forerunner 945, 2,499.99, 999.98, North America, Credit Card
10022, 2024-01-22, Books, Gone Girl by Gillian Flynn, 2,22.99, 45.98, North America, Credit Card
10025, 2024-01-25, Electronics, Bose QuietComfort 35 Headphones, 1,299.99, 299.99, North America, Credit Card
10026, 2024-01-26, Home Appliances, Panasonic NN-SN9665 Microwave, 1,179.99, 179.99, Europe, PayPal
10029, 2024-01-29, Beauty Products, MAC Ruby Woo Lipstick, 1,29.99, 29.99, Europe, PayPal
10027, 2024-01-27, Clothing, Adidas Ultraboost Shoes, 2,179.99, 359.98, Asia, Debit Card
10030, 2024-01-30, Sports, Nike Air Zoom Pegasus 37, 2,129.99, 259.98, Asia, Credit Card
```

### 3. Créer un Keyspace (my\_keyspace)

```
kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic flume-topic2 --from-beginning
```

```
cqlsh> CREATE KEYSPACE IF NOT EXISTS my_keyspace WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication_factor' : 1 };
cqlsh> CREATE TABLE IF NOT EXISTS my_keyspace.transactions (
...   transaction_id text PRIMARY KEY,
...   date text,
...   product_category text,
...   product_name text,
...   units_sold int,
...   unit_price double,
...   total_revenue double,
...   region text,
...   payment_method text
... );
cqlsh> DESCRIBE KEYSPACES;
my_keyspace  system_auth      system_schema  system_views
system       system_distributed system_traces  system_virtual_schema
```

### 4. Créer une Table (

```
CREATE TABLE IF NOT EXISTS my_keyspace.transactions (
    transaction_id text PRIMARY KEY,
    date text,
    product_category text,
    product_name text,
    units_sold int,
    unit_price double,
    total_revenue double,
    region text,
    payment_method text);
```

```
cqlsh> CREATE KEYSPACE IF NOT EXISTS my_keyspace WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication_factor' : 1 };
cqlsh> CREATE TABLE IF NOT EXISTS my_keyspace.transactions (
...   transaction_id text PRIMARY KEY,
...   date text,
...   product_category text,
...   product_name text,
...   units_sold int,
...   unit_price double,
...   total_revenue double,
...   region text,
...   payment_method text
... );
cqlsh> DESCRIBE KEYSPACES;
my_keyspace  system_auth      system_schema  system_views
system       system_distributed system_traces  system_virtual_schema
```

### 5. Exécuter le code pyspark

```
spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.0.1,com.datastax.spark:spark-
cassandra-connector_2.12:3.0.0 --conf "spark.driver.extraClassPath=/path/to/jnr-posix.jar"
"C:\Users\hp\Desktop\ECE\M2\Cours\Concevoir un pipeline de
donnees\TP\rendu\Spark_work\test\exemple1.py"
```

```

C:\>spark-submit --packages com.datastax.spark:spark-cassandra-connector_2.12:3.0.0,com.github.jnr:jnr-posix:3.1.6 "C:\Users\hp\Desktop\CEC\2\Cours\Concevoir un pipeline de donnees\TP\rendu\Spark_work\test\exemple1.py"
:: loading settings :: url = jar:file:/C:/Spark/jars/ivy-2.5.1.jar!/_org/apache/ivy/core/settings/ivysettings.xml
Ivy Default Cache set to: C:\Users\hp\.ivy2\cache
The jars for the packages stored in: C:\Users\hp\.ivy2\jars
com.datastax.spark#spark-cassandra-connector_2.12 added as a dependency
com.github.jnr#jnr-posix added as a dependency
:: resolving dependencies :: org.apache.spark#spark-submit-parent-3dc3f5c-a215-460e-82d2-b0ff55267eb0;1.0
  confs: [default]
    found com.datastax.spark#spark-cassandra-connector_2.12;3.0.0 in central
    found com.datastax.spark#spark-cassandra-connector-driver_2.12;3.0.0 in central
    found com.datastax.oss#java-driver-core-shaded;4.7.2 in central
    found com.datastax.oss#native-protocol;1.4.10 in central
    found com.datastax.oss#java-driver-shaded-guava;25.1-jre-graal-sub-1 in central
    found com.typesafe#config;1.3.4 in central
    found org.slf4j#slf4j-api;1.7.26 in central
    found com.dropwizard.metrics#metrics-core;4.0.5 in central
    found org.hdrhistogram#HdrHistogram;2.1.11 in central
    found org.javatuples#javatuples;1.2 in central
    found org.reactivestreams#reactive-streams;1.0.2 in central
    found com.github.stephenc.jcip#jcip-annotations;1.0-1 in local-m2-cache
    found com.github.spotbugs#spotbugs-annotations;3.1.12 in central
    found com.google.code.findbugs#jsr305;3.0.2 in local-m2-cache
    found com.datastax.oss#java-driver-mapper-runtime;4.7.2 in central
    found com.datastax.oss#java-driver-query-builder;4.7.2 in central
    found com.google.common#commons-lang3;3.9 in central
    found com.thoughtworks.parancamer#parancamer;2.8 in local-m2-cache
    found org.scala-lang#scala-reflect;2.12.11 in central
    found com.github.jnr#jnr-posix;3.1.6 in central
    found com.github.jnr#jnr-ffi;2.2.3 in central
    found com.github.jnr#jffi;1.3.3 in central
    found org.ow2.asm#asm;9.1 in local-m2-cache
    found org.ow2.asm#asm-commons;9.1 in central
    found org.ow2.asm#asm-tree;9.1 in central

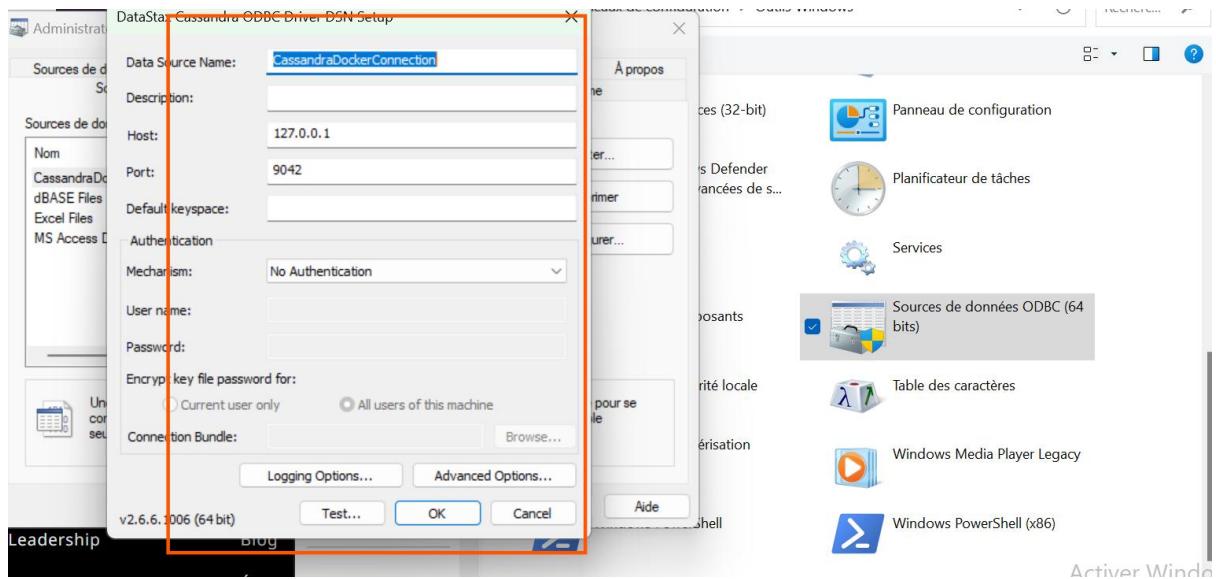
```

Activer Windows  
Accédez aux paramètres pour activer Windows.

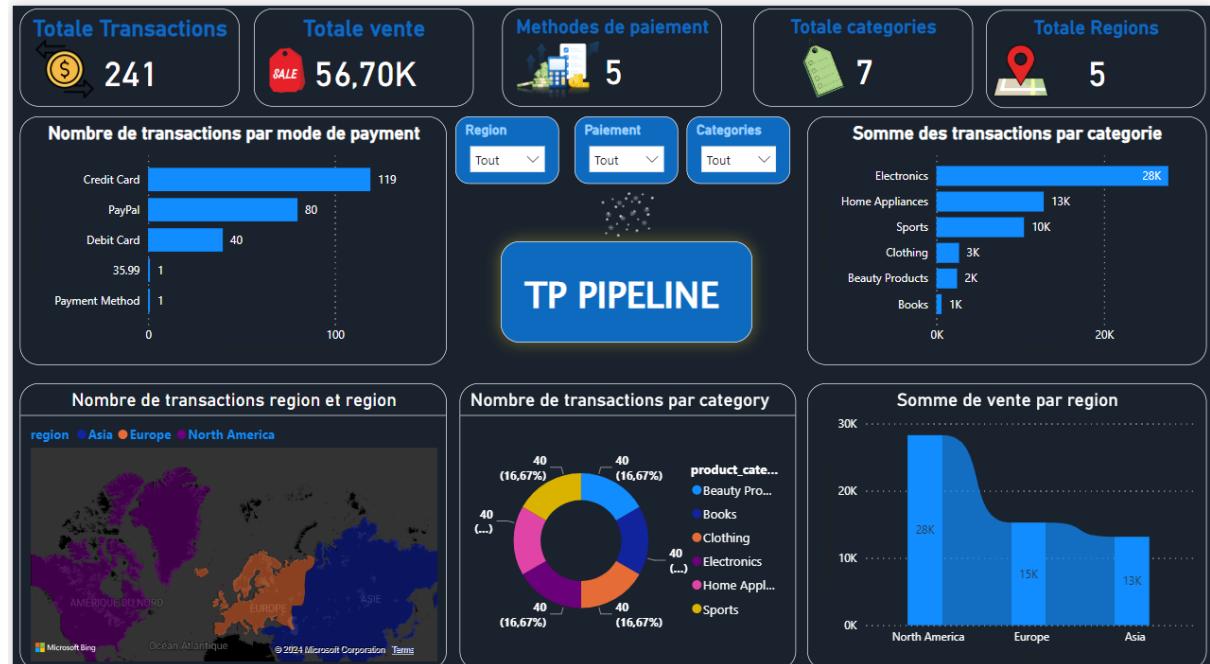
## 6. Vérifier la bd cassandra

(0 rows)	bd												
transaction_id	date	payment_method	product_category	product_name	region	total_revenue	unit_price	units_sold					
10056	2024-02-25	PayPal	Home Appliances	KitchenAid Artisan Stand Mixer	Europe	499.99	499.99	1					
10116	2024-04-25	PayPal	Home Appliances	Breville Smart Grill	Europe	599.9	299.95	2					
10179	2024-06-27	PayPal	Beauty Products	First Aid Beauty Ultra Repair Cream	Europe	68	30	2					
10180	2024-06-09	PayPal	Beauty Products	Youth to the People Superfood Antioxidant Cleanser	Europe	68	30	1					
100658	2024-01-01	PayPal	Home Appliances	Philips DiamondClean SmartPro Max Vacuum	Europe	659.99	219.99	3					
10011	2024-01-11	PayPal	Beauty Products	Chanel No. 5 Perfume	Europe	129.99	129.99	1					
10143	2024-05-22	PayPal	Beauty Products	Caudalie Vinoperfect Radiance Serum	Europe	79	79	1					
10209	2024-07-27	PayPal	Beauty Products	De'Longhi Magnifica Espresso Machine	Europe	899.99	899.99	1					
100874	2024-03-14	PayPal	Home Appliances	Dyson V11 Vacuum	Europe	499.99	499.99	1					
10002	2024-01-02	PayPal	Home Appliances	Clinique Moisture Surge	Europe	52	52	1					
10087	2024-02-16	PayPal	Beauty Products	Breville Smart Coffee Grinder Pro	Europe	199.95	199.95	1					
10194	2024-07-12	PayPal	Home Appliances	Breville K-Mini Compact Kettle	Europe	159.99	159.99	1					
10181	2024-01-14	PayPal	Home Appliances	Keurig K-Elite Coffee Maker	Europe	189.99	189.99	1					
10020	2024-01-20	PayPal	Home Appliances	Romka 174	Europe	1599.98	1599.98	2					
10033	2024-02-07	PayPal	Home Appliances	Ninja Professional Blender	Europe	99.99	99.99	1					
10227	2024-08-14	PayPal	Beauty Products	The Ordinary Caffeine Solution 5% + EGCG	Europe	6.7	6.7	1					
10083	2024-03-23	PayPal	Beauty Products	Charlotte Tilbury Magic Cream	Europe	100	100	1					
10236	2024-08-23	PayPal	Home Appliances	Nespresso Vertuo Next Coffee and Espresso Maker	Europe	159.99	159.99	1					
10158	2024-06-06	PayPal	Home Appliances	Breville Nespresso Creatista Plus	Europe	499.95	499.95	1					
10064	2024-02-10	PayPal	Beauty Products	Estee Lauder Advanced Night Repair	Europe	185	185	1					
10173	2024-01-15	PayPal	Beauty Products	Paula's Choice Skin Perfecting 2% BHA Liquid Exfoliant	Europe	29.99	29.99	1					
10017	2024-01-17	PayPal	Beauty Products	Dyson Supersonic Hair Dryer	Europe	399.99	399.99	1					
10203	2024-07-21	PayPal	Beauty Products	Dr. Jart+ Cicapair Tiger Grass Correcting Treatment	Europe	52	52	1					
10137	2024-05-16	PayPal	Beauty Products	The Ordinary Hyaluronic Acid Serum	Europe	6.8	6.8	1					
10077	2024-03-17	PayPal	Beauty Products	Glossier Boy Brow	Europe	32	16	2					
10116	2024-05-25	PayPal	Home Appliances	Cuisinart Coffee Center	Europe	399.9	199.95	2					
10035	2024-02-04	PayPal	Beauty Products	Sunday Riley Luminous Sleeping Oil	Europe	79.98	39.99	2					
10239	2024-09-13	PayPal	Home Appliances	Philips Norelco Shaver IQ	Europe	699.99	349.99	1					
10122	2024-05-01	PayPal	Home Appliances	Hamilton Beach FlexBrew Coffee Maker	Europe	105.99	105.99	1					
10113	2024-04-22	PayPal	Beauty Products	Philips Sonicare DiamondClean Toothbrush	Europe	129.99	129.99	2					
10176	2024-06-24	PayPal	Home Appliances	Crock-Pot 6-Quart Slow Cooker	Europe	99.98	49.99	2					
10095	2024-04-04	PayPal	Beauty Products	Tatcha The Water Cream	Europe	68	68	1					
10149	2024-05-28	PayPal	Beauty Products	Laneige Water Sleeping Mask	Europe	25	25	1					
10089	2024-03-29	PayPal	Beauty Products	Sunday Riley Good Genes	Europe	105	105	1					
10023	2024-01-23	PayPal	Beauty Products	Olay Regenerist Face Cream	Europe	49.99	49.99	1					
10106	2024-07-10	PayPal	Home Appliances	Instant Pot Ultra	Europe	139.99	139.99	1					
10062	2024-03-27	PayPal	Home Appliances	Philips Air Fryer	Europe	399.99	199.99	2					
10122	2024-05-01	PayPal	Home Appliances	Hamilton Beach FlexBrew Coffee Maker	Europe	89.99	89.99	1					
10113	2024-04-22	PayPal	Beauty Products	Biore UV Aqua Rich Watery Essence Sunscreen	Europe	15	15	1					
10152	2024-05-31	PayPal	Home Appliances	Ninja Foodi Pressure Cooker	Europe	399.98	199.99	2					
10185	2024-07-03	PayPal	Beauty Products	La Mer Crème de la Mer Moisturizer	Europe	198	198	1					
10170	2024-06-18	PayPal	Home Appliances	Instant Pot Duo Eva Plus	Europe	279.98	139.99	2					
10032	2024-02-01	PayPal	Home Appliances	Instant Pot Duo	Europe	269.97	89.99	3					
10200	2024-07-18	PayPal	Home Appliances	Instant Pot Duo Nova	Europe	99.95	99.95	1					
10162	2024-06-30	PayPal	Home Appliances	Dyson V8 Absolute	Europe	399.99	399.99	1					
10191	2024-07-09	PayPal	Beauty Products	Glossier Cloud Paint	Europe	18	18	1					

## 7. Configurer une source de données



## 8. Visualisation



## PARTIE II : Présentation de la source de données

Transaction ID	Date	Product Category	Product Name	Units Sold	Unit Price	Total Revenue	Region	Payment Method
10001	2024-01-01	Electronics	iPhone 14 Pro	2,999.99	1,999.98	North America	Credit Card	
10002	2024-01-02	Home Appliances	Dyson V11 Vacuum	1,499.99	499.99	Europe	PayPal	
10003	2024-01-03	Clothing	Levi's 501 Jeans	3,69.99	209.97	Asia	Debit Card	
10004	2024-01-04	Books	The Da Vinci Code	4,15.99	63.96	North America	Credit Card	
10005	2024-01-05	Beauty Products	Neutrogena Skincare Set	1,89.99	89.99	Europe	PayPal	
10006	2024-01-06	Sports	Wilson Evolution Basketball	5,29.99	149.95	Asia	Credit Card	
10007	2024-01-07	Electronics	MacBook Pro 16-inch	1,2499.99	2499.99	North America	Credit Card	
10008	2024-01-08	Home Appliances	Blueair Classic 480i	2,599.99	1199.98	Europe	PayPal	
10009	2024-01-09	Clothing	Nike Air Force 1	6,89.99	539.94	Asia	Debit Card	
10010	2024-01-10	Books	Dune by Frank Herbert	2,25.99	51.98	North America	Credit Card	
10011	2024-01-11	Beauty Products	Chanel No. 5 Perfume	1,129.99	129.99	Europe	PayPal	
10012	2024-01-12	Sports	Babolat Pure Drive Tennis Racket	3,199.99	599.97	Asia	Credit Card	
10013	2024-01-13	Electronics	Samsung Galaxy Tab S8	2,749.99	1499.98	North America	Credit Card	
10014	2024-01-14	Home Appliances	Keurig K-Elite Coffee Maker	1,189.99	189.99	Europe	PayPal	
10015	2024-01-15	Clothing	North Face Down Jacket	2,249.99	499.98	Asia	Debit Card	
10016	2024-01-16	Books	"Salt, Fat, Acid, Heat by Samin Nosrat"	3,35.99	107.97	North America	Credit Card	
10017	2024-01-17	Beauty Products	Dyson Supersonic Hair Dryer	1,399.99	399.99	Europe	PayPal	
10018	2024-01-18	Sports	Manduka PRO Yoga Mat	4,119.99	479.96	Asia	Credit Card	
10019	2024-01-19	Electronics	Garmin Forerunner 945	2,499.99	999.98	North America	Credit Card	
10020	2024-01-20	Books	Atlas of the Human Brain	1,000.00	100.00	Europe	PayPal	

# Etape I : Apache FLUME

## 1. Installation

### 1. Aller sur le site et le telecharger le zip pour windows

Apache Flume binary (tar.gz)	apache-flume-1.11.0-bin.tar.gz	apache-flume-1.11.0-bin.tar.gz.sha512	apache-flume-1.11.0-bin.tar.gz.asc
Apache Flume source (tar.gz)	apache-flume-1.11.0-src.tar.gz	apache-flume-1.11.0-src.tar.gz.sha512	apache-flume-1.11.0-src.tar.gz.asc

### 2. Fichier bin.tar.gz

We suggest the following location for your download:

<https://dlcdn.apache.org/flume/1.11.0/apache-flume-1.11.0-bin.tar.gz>

Alternate download locations are suggested below.

It is essential that you [verify the integrity](#) of the downloaded file using the PGP signature ([.asc](#) file) or a hash ([.md5](#) or [.sha\\*](#) file).

### HTTP

<https://dlcdn.apache.org/flume/1.11.0/apache-flume-1.11.0-bin.tar.gz>

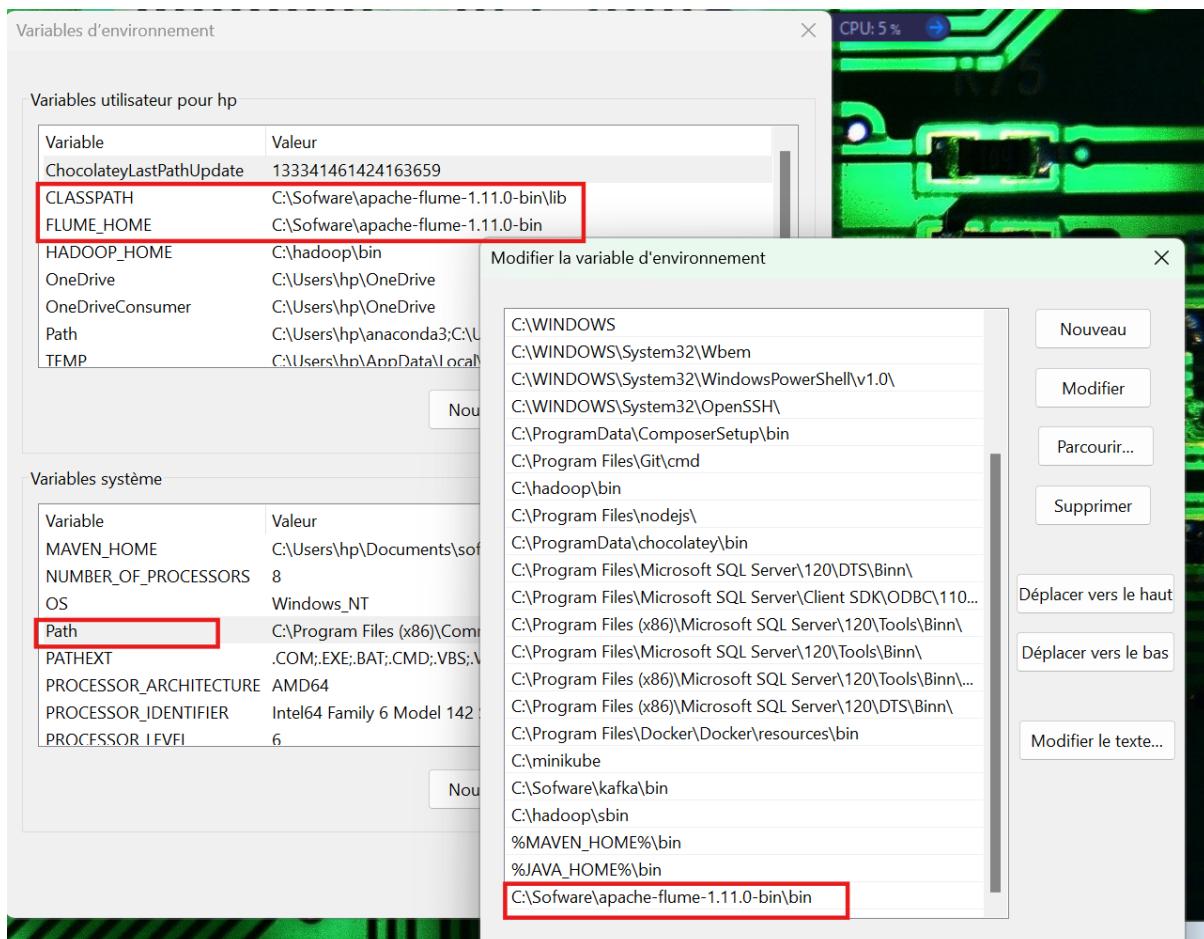
### BACKUP SITES

<https://dlcdn.apache.org/flume/1.11.0/apache-flume-1.11.0-bin.tar.gz>

### VERIFY THE INTEGRITY OF THE FILES

It is essential that you verify the integrity of the downloaded file using the PGP signature ([.asc](#) file) or a hash ([.md5](#) or [.sha\\*](#) file). Please read

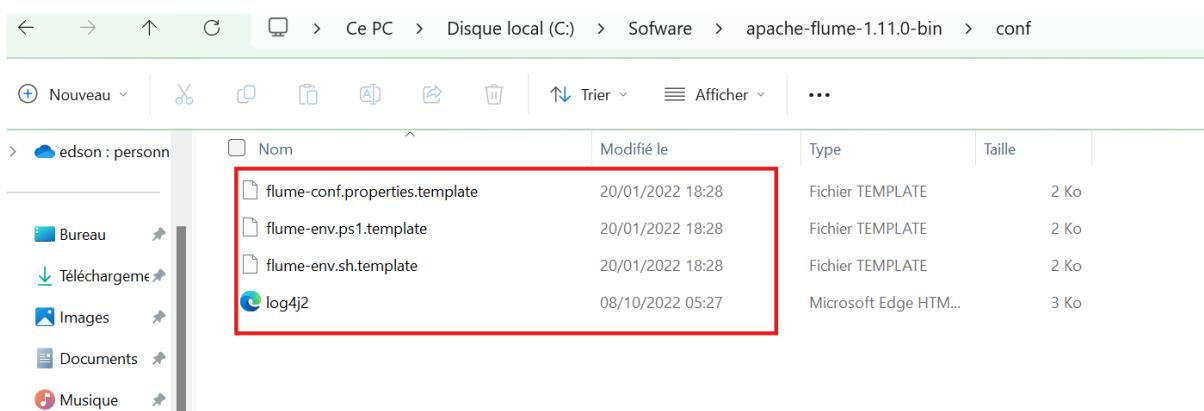
### 3. Configurer la variable d'environnement



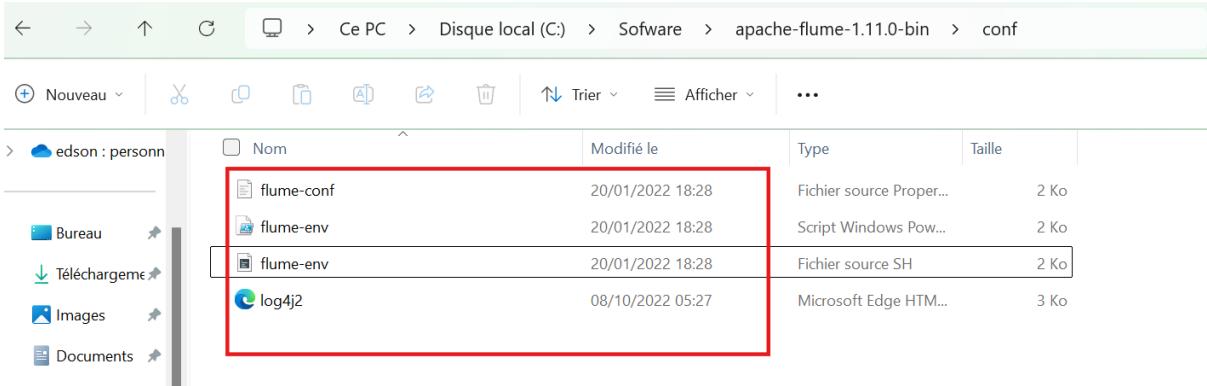
#### 4. Modifier les fichiers du dossier conf

i. Ayant

Retirer les extensions .template



## ii. Apres



## 4. Modifier le chemin du jdk et classpath dans le fichier flume-env.sh

```
$ flume-env.sh X
C:> Software > apache-flume-1.11.0-bin > conf > $ flume-env.sh
21
22 export JAVA_HOME=C:\Program Files\Java\jdk-1.8
23
24 # Give Flume more memory and pre-allocate, enable remote monitoring via JMX
25 # export JAVA_OPTS="-Xms100m -Xmx2000m -Dcom.sun.management.jmxremote"
26
27 # Let Flume write raw event data and configuration information to its log files for debugging
28 # purposes. Enabling these flags is not recommended in production,
29 # as it may result in logging sensitive user information or encryption secrets.
30 # export JAVA_OPTS="$JAVA_OPTS -Dorg.apache.flume.log.rawdata=true -Dorg.apache.flume.log.printconfig=true "
31
32 # Note that the Flume conf directory is always included in the classpath.
33 #FLUME_CLASSPATH=""
34
35 export CLASSPATH = C:\Software\apache-flume-1.11.0-bin\lib
36
37
```

## 5. Vérifier si c'est bien installé

```
C:\Software\apache-flume-1.11.0-bin\bin\flume-ng.ps1 ?
[N] Ne pas exécuter [O] Exécuter une fois [S] Suspendre [?] Aide (la valeur par défaut est « N ») : o
WARN: Config directory not set. Defaulting to C:\Software\apache-flume-1.11.0-bin\conf
Sourcing environment configuration script C:\Software\apache-flume-1.11.0-bin\conf\flume-env.ps1
Avertissement de sécurité
N'exécutez que des scripts que vous approuvez. Bien que les scripts en provenance d'Internet puissent être ut
script est susceptible d'endommager votre ordinateur. Si vous approuvez ce script, utilisez l'applet de comma
Unblock-File pour autoriser le script à s'exécuter sans ce message d'avertissement. Voulez-vous exécuter
C:\Software\apache-flume-1.11.0-bin\conf\flume-env.ps1 ?
[N] Ne pas exécuter [O] Exécuter une fois [S] Suspendre [?] Aide (la valeur par défaut est « N ») : O
Including Hadoop libraries found in (C:\hadoop) for DFS access
Le chemin d'accès spécifié est introuvable.
Le chemin d'accès spécifié est introuvable.
'Dhadoop.security.logger' n'est pas reconnu en tant que commande interne
ou externe, un programme exécutable ou un fichier de commandes.
WARN: HBASE_HOME not found
WARN: HIVE_HOME not found

Running FLUME version :
  class: org.apache.flume.tools.VersionInfo
  arguments:

Flume 1.11.0
Source code repository: https://git.apache.org/repos/asf/flume.git
Revision: la15927e594fd0d05a59d804b90a9c31ec93f5e1
Compiled by rgoers on Sun Oct 16 14:44:15 MST 2022
From source with checksum bbbca682177262aac3a89defde369a37
C:\Software\apache-flume-1.11.0-bin\bin>
```

## 6. Le fichier de configuration

```
$ flume-env.sh      q1.conf      last.conf      q.conf      flume-kafka.conf X      flume-env.ps1
EXPLORER      ...
OPEN EDITORS
CONF
  flume-conf.properties
  flume-env.ps1
  flume-env.sh
  flume-kafka.conf
  last.conf
  log4j.properties
  q.conf
  q1.conf

flume-kafka.conf
1  # Defining the agent 'a1'
2  a1.sources = r1
3  a1.sinks = k1
4  a1.channels = c1
5
6  # Source: Spooldir (monitore un répertoire pour de nouveaux fichiers CSV)
7  a1.sources.r1.type = spooldir
8  a1.sources.r1.spoolDir = C:/CSV
9  a1.sources.r1.fileHeader = true
10 a1.sources.r1.basenameHeader = true
11 a1.sources.r1.ignorePattern = ^$    # Ignore les fichiers vides
12
13 # Sink: Kafka
14 a1.sinks.k1.type = org.apache.flume.sink.kafka.KafkaSink
15 a1.sinks.k1.kafka.bootstrap.servers = localhost:9092
16 a1.sinks.k1.kafka.topic = flume-topic1
17
18 # Channel: Memory channel
19 a1.channels.c1.type = memory
20 a1.channels.c1.capacity = 1000
21 a1.channels.c1.transactionCapacity = 100
22
23 # Liaison source et sink au channel
24 a1.sources.r1.channels = c1
25 a1.sinks.k1.channel = c1
26
27
```

## 7. Lancer flume

**NOTE (1) :** Avant d'exécuter ceci il faut que le topic « flume-kafka » existe déjà. Donc allez d'abord à la prochaine étape (Kafka). Quand il va falloir revenir sur ce point il sera signalé dans la suite du document.

```
flume-ng agent -n a1 -c conf -f C:/flume/conf/q.conf -property flume.root.logger=DEBUG,console
```

Si ça dérange taper cette requête dans le powershell en tant que administrateur

```
powershell.exe -NoProfile -InputFormat none -ExecutionPolicy Bypass -File C:\flume\bin\flume-ng.ps1
agent -n a1 -c conf -f C:/flume/conf/q.conf -property flume.root.logger=DEBUG,console
```

```

C:\flume>flume-ng agent -n al -c conf -f C:/flume/conf/flume-kafka.conf -property flume.root.logger=DEBUG,console
C:\flume>powershell.exe -NoProfile -InputFormat none -ExecutionPolicy unrestricted -File C:\flume\bin\flume-ng.ps1 agent -n al -c con
f -f C:/flume/conf/flume-kafka.conf -property flume.root.logger=DEBUG,console

Avertissement de sécurité
N'exécutez que des scripts que vous approuvez. Bien que les
scripts en provenance d'Internet puissent être utiles, ce
script est susceptible d'endommager votre ordinateur. Si vous
approuvez ce script, utilisez l'outil de commande Unblock-File
pour autoriser le script à s'exécuter sans ce message
d'avertissement. Voulez-vous exécuter
C:\flume\bin\flume-ng.ps1 ?
[N] Ne pas exécuter [O] Exécuter une fois [S] Suspendre
[?] Aide(la valeur par défaut est « N ») : O

WARN: Config directory not set. Defaulting to C:\flume\conf
Sourcing environment configuration script C:\flume\conf\flume-env.ps1
Including Hadoop libraries found in (C:\hadoop) for DFS access
Le chemin d'accès spécifié est introuvable.
Le chemin d'accès spécifié est introuvable.
'-Dhadoop.security.logger' n'est pas reconnu en tant que commande interne
ou externe, un programme exécutable ou un fichier de commandes.
WARN: HBASE_HOME not found
WARN: HIVE_HOME not found

Running FLUME agent :
  class: org.apache.flume.node.Application
  arguments: -n al -f "C:/flume/conf/flume-kafka.conf"

2024-11-09 09:43:05,412 (main) [DEBUG - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95)] No global SSL keystore
path specified.
2024-11-09 09:43:05,417 (main) [DEBUG - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95)] No global SSL keystore

C:\flume>flume-ng agent -n al -c conf -f C:/flume/conf/flume-kafka.conf -property flume.root.logger=DEBUG,console
C:\flume>powershell.exe -NoProfile -InputFormat none -ExecutionPolicy unrestricted -File C:\flume\bin\flume-ng.ps1 agent -n al -c conf -f C:/flume/conf/flume-kafka.conf -property flume.root.logger=DEBUG,console

WARN: Config directory not set. Defaulting to C:\flume\conf
Sourcing environment configuration script C:\flume\conf\flume-env.ps1
WARN: Did not find C:\flume\conf\flume-env.ps1
Including Hadoop libraries found in (C:\hadoop) for DFS access
Le chemin d'accès spécifié est introuvable.
Le chemin d'accès spécifié est introuvable.
'-Dhadoop.security.logger' n'est pas reconnu en tant que commande interne
ou externe, un programme exécutable ou un fichier de commandes.
WARN: HBASE_HOME not found
WARN: HIVE_HOME not found

Running FLUME agent :
  class: org.apache.flume.node.Application
  arguments: -n al -f "C:/flume/conf/flume-kafka.conf"

2024-10-12 15:21:18,987 (main) [DEBUG - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95)] No global SSL keystore path specified.
2024-10-12 15:21:18,992 (main) [DEBUG - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95)] No global SSL keystore password specified.
2024-10-12 15:21:18,992 (main) [DEBUG - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95)] No global SSL keystore type specified.
2024-10-12 15:21:18,993 (main) [DEBUG - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95)] No global SSL truststore path specified.
2024-10-12 15:21:18,998 (main) [DEBUG - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95)] No global SSL truststore password specified.
2024-10-12 15:21:18,998 (main) [DEBUG - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95)] No global SSL truststore type specified.
2024-10-12 15:21:18,999 (main) [DEBUG - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95)] No global SSL include protocols specified.
2024-10-12 15:21:18,999 (main) [DEBUG - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95)] No global SSL exclude protocols specified.
2024-10-12 15:21:18,999 (main) [DEBUG - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95)] No global SSL include cipher suites specified.
2024-10-12 15:21:18,999 (main) [DEBUG - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95)] No global SSL exclude cipher suites specified.
2024-10-12 15:21:19,177 (LifecycleSupervisor-1-0) [INFO - org.apache.flume.node.PollingPropertiesFileConfigurationProvider.start(PollingPropertiesFileConfigurationProvider.java:62)] Configuration provider starting
2024-10-12 15:21:19,179 (LifecycleSupervisor-1-0) [DEBUG - org.apache.flume.node.PollingPropertiesFileConfigurationProvider.start(PollingPropertiesFileConfigurationProvider.java:79)] Configuration provider started
2024-10-12 15:21:19,183 (conf-file-poller-0) [DEBUG - org.apache.flume.node.PollingPropertiesFileConfigurationProvider$FileWatcherRunnable.run(PollingPropertiesFileConfigurationProvider.java:131)] Checking file:C:\flume\conf\flume-kafka.conf for changes
2024-10-12 15:21:19,184 (conf-file-poller-0) [INFO - org.apache.flume.node.PollingPropertiesFileConfigurationProvider$FileWatcherRunnable.run(PollingPropertiesFileConfigurationProvider.java:138)] Reloading configuration file:C:\flume\conf\flume-kafka.conf
2024-10-12 15:21:19,210 (conf-file-poller-0) [INFO - org.apache.flume.conf.FlumeConfiguration$AgentConfiguration.addComponentConfig(FlumeConfiguration.java:1203)] Processing:k1

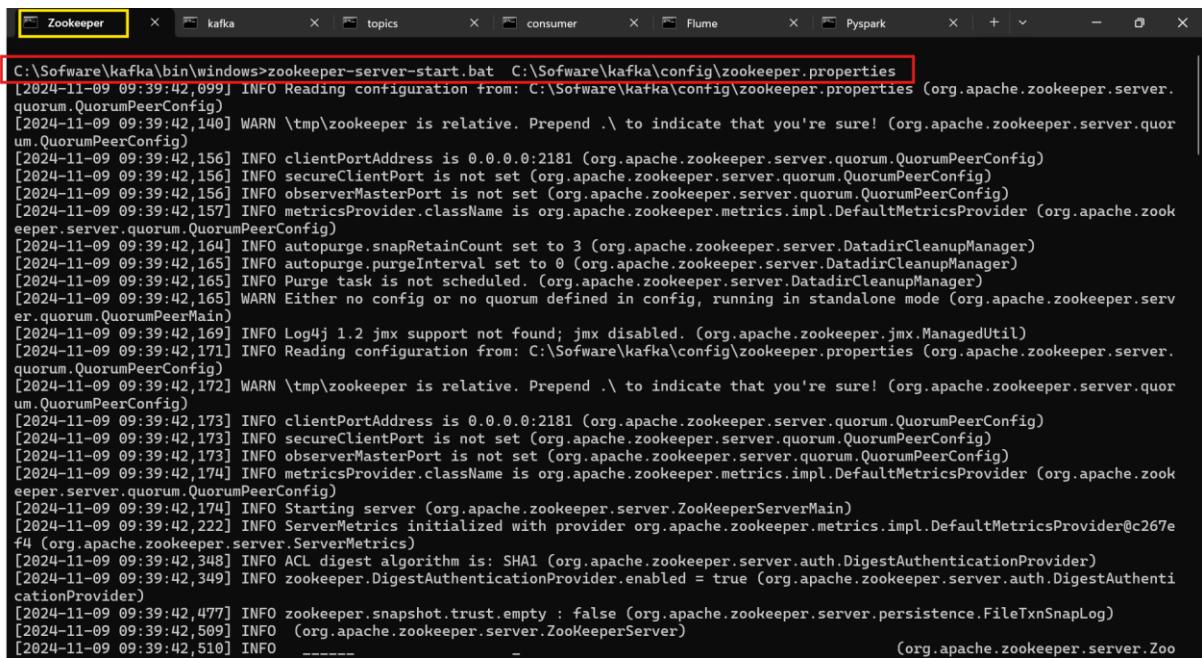
```

## Etape II : Apache KAFKA

1. Installation et configuration (voir Tp précédent )
2. Utilisation

### 1. Lancer Zookeeper

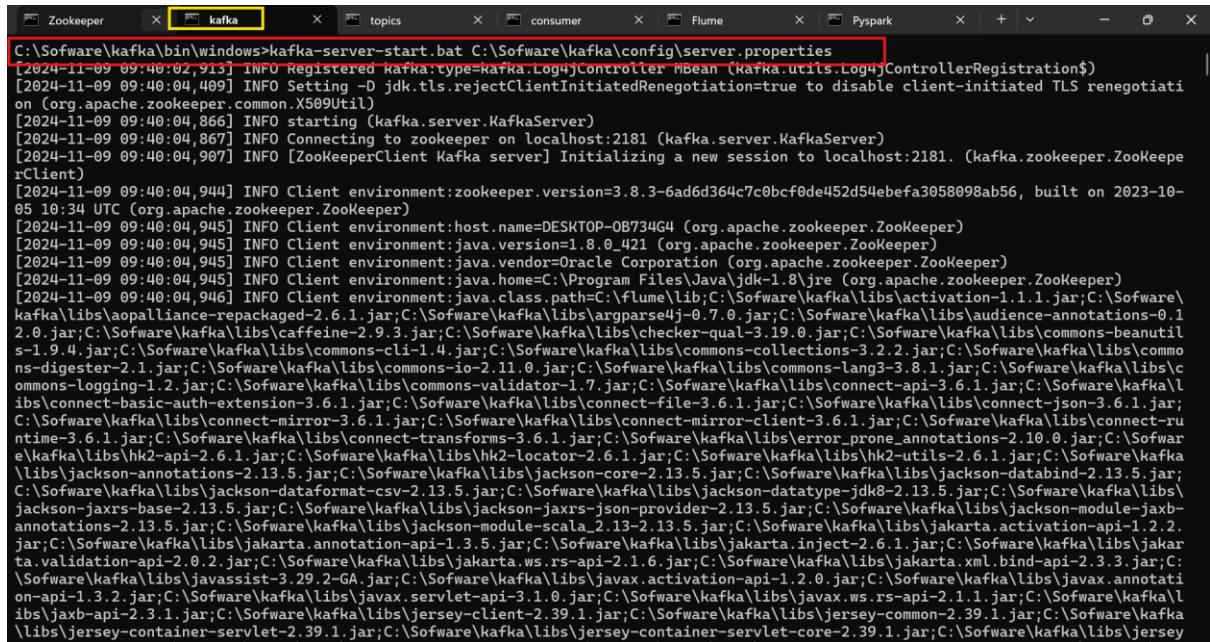
```
zookeeper-server-start.bat C:\Software\kafka\config\zookeeper.properties
```



```
C:\Software\kafka\bin\windows>zookeeper-server-start.bat C:\Software\kafka\config\zookeeper.properties
[2024-11-09 09:39:42,099] INFO Reading configuration from: C:\Software\kafka\config\zookeeper.properties (org.apache.zookeeper.server.QuorumPeerConfig)
[2024-11-09 09:39:42,140] WARN \tmp\zookeeper is relative. Prepend .\ to indicate that you're sure! (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-11-09 09:39:42,156] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-11-09 09:39:42,156] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-11-09 09:39:42,156] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-11-09 09:39:42,157] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-11-09 09:39:42,164] INFO autopurge.snapRetainCount set to 3 (org.apache.zookeeper.server.DatadirCleanupManager)
[2024-11-09 09:39:42,165] INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DatadirCleanupManager)
[2024-11-09 09:39:42,165] INFO Purge task is not scheduled. (org.apache.zookeeper.server.DatadirCleanupManager)
[2024-11-09 09:39:42,165] WARN Either no config or no quorum defined in config, running in standalone mode (org.apache.zookeeper.server.quorum.QuorumPeerMain)
[2024-11-09 09:39:42,169] INFO Log4j 1.2 jmx support not found; jmx disabled. (org.apache.zookeeper.jmx.ManagedUtil)
[2024-11-09 09:39:42,171] INFO Reading configuration from: C:\Software\kafka\config\zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-11-09 09:39:42,172] WARN \tmp\zookeeper is relative. Prepend .\ to indicate that you're sure! (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-11-09 09:39:42,173] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-11-09 09:39:42,173] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-11-09 09:39:42,173] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-11-09 09:39:42,174] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-11-09 09:39:42,174] INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
[2024-11-09 09:39:42,222] INFO ServerMetrics initialized with provider org.apache.zookeeper.metrics.impl.DefaultMetricsProvider@ (org.apache.zookeeper.server.ServerMetrics)
[2024-11-09 09:39:42,348] INFO ACL digest algorithm is: SHA1 (org.apache.zookeeper.server.auth.DigestAuthenticationProvider)
[2024-11-09 09:39:42,349] INFO zookeeper.DigestAuthenticationProvider.enabled = true (org.apache.zookeeper.server.auth.DigestAuthenticationProvider)
[2024-11-09 09:39:42,477] INFO zookeeper.snapshot.trust.empty : false (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
[2024-11-09 09:39:42,509] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2024-11-09 09:39:42,510] INFO ----- (org.apache.zookeeper.server.Zoo
```

## 2. Lancer kafka

```
kafka-server-start.bat C:\Software\kafka\config\server.properties
```

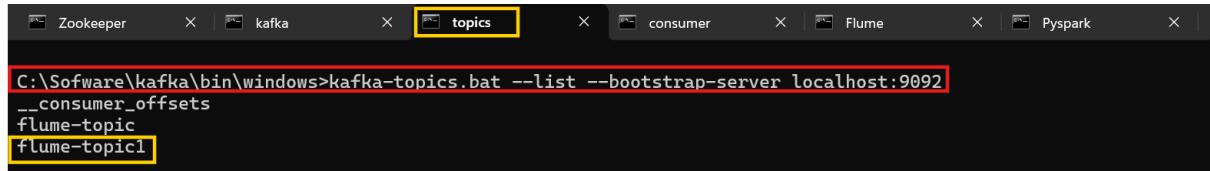


A screenshot of a terminal window titled "kafka". The window contains several tabs: Zookeeper, kafka (which is active), topics, consumer, Flume, and Pyspark. The kafka tab displays log output from the command "C:\Software\kafka\bin\windows>kafka-server-start.bat C:\Software\kafka\config\server.properties". The log includes various INFO messages related to Kafka's internal components like Log4jController, ZooKeeperClient, and KafkaServer.

```
C:\Software\kafka\bin\windows>kafka-server-start.bat C:\Software\kafka\config\server.properties
[2024-11-09 09:40:02,915] INFO Registered kafka:type=kafka.Log4jController Mbean (kafka.utils.Log4jControllerRegistration$)
[2024-11-09 09:40:04,409] INFO Setting -D jdk.tls.rejectClientInitiatedRenegotiation=true to disable client-initiated TLS renegotiation (org.apache.zookeeper.common.X509Util)
[2024-11-09 09:40:04,866] INFO starting (kafka.server.KafkaServer)
[2024-11-09 09:40:04,867] INFO Connecting to zookeeper on localhost:2181 (kafka.server.KafkaServer)
[2024-11-09 09:40:04,907] INFO [ZooKeeperClient Kafka server] Initializing a new session to localhost:2181. (kafka.zookeeper.ZooKeeperClient)
[2024-11-09 09:40:04,944] INFO Client environment:zookeeper.version=3.8.3-6ad6d364c7c0bcf0de452d54befa3058098ab56, built on 2023-10-05 10:34 UTC (org.apache.zookeeper.ZooKeeper)
[2024-11-09 09:40:04,945] INFO Client environment:host.name=DESKTOP-0B734G4 (org.apache.zookeeper.ZooKeeper)
[2024-11-09 09:40:04,945] INFO Client environment:java.version=1.8.0_421 (org.apache.zookeeper.ZooKeeper)
[2024-11-09 09:40:04,945] INFO Client environment:java.vendor=Oracle Corporation (org.apache.zookeeper.ZooKeeper)
[2024-11-09 09:40:04,945] INFO Client environment:java.home=C:\Program Files\Java\jdk-1.8.jre (org.apache.zookeeper.ZooKeeper)
[2024-11-09 09:40:04,946] INFO Client environment:java.class.path=C:\flume\lib;C:\Software\kafka\libs\activation-1.1.1.jar;C:\Software\kafka\libs\apolliance-repackaged-2.6.1.jar;C:\Software\kafka\libs\argparse4j-0.7.0.jar;C:\Software\kafka\libs\audience-annotations-0.1.2.0.jar;C:\Software\kafka\libs\caffeine-2.9.3.jar;C:\Software\kafka\libs\checker-qual-3.19.0.jar;C:\Software\kafka\libs\commons-beanutils-1.9.4.jar;C:\Software\kafka\libs\commons-cli-1.4.jar;C:\Software\kafka\libs\commons-collections-3.2.2.jar;C:\Software\kafka\libs\commons-digester-2.1.jar;C:\Software\kafka\libs\commons-io-2.11.0.jar;C:\Software\kafka\libs\commons-lang3-3.8.1.jar;C:\Software\kafka\libs\commons-logging-1.2.jar;C:\Software\kafka\libs\commons-validator-1.7.jar;C:\Software\kafka\libs\connect-api-3.6.1.jar;C:\Software\kafka\libs\connect-basic-auth-extension-3.6.1.jar;C:\Software\kafka\libs\connect-file-3.6.1.jar;C:\Software\kafka\libs\connect-json-3.6.1.jar;C:\Software\kafka\libs\connect-mirror-3.6.1.jar;C:\Software\kafka\libs\connect-transforms-3.6.1.jar;C:\Software\kafka\libs\error_prone_annotations-2.10.0.jar;C:\Software\kafka\libs\h2-api-2.6.1.jar;C:\Software\kafka\libs\h2-locator-2.6.1.jar;C:\Software\kafka\libs\h2-utils-2.6.1.jar;C:\Software\kafka\libs\jackson-annotations-2.13.5.jar;C:\Software\kafka\libs\jackson-core-2.13.5.jar;C:\Software\kafka\libs\jackson-databind-2.13.5.jar;C:\Software\kafka\libs\jackson-dataformat-csv-2.13.5.jar;C:\Software\kafka\libs\jacksondatatype-jdk8-2.13.5.jar;C:\Software\kafka\libs\jackson-jaxrs-base-2.13.5.jar;C:\Software\kafka\libs\jackson-jaxrs-json-provider-2.13.5.jar;C:\Software\kafka\libs\jackson-module-scala-2.13-2.13.5.jar;C:\Software\kafka\libs\jakarta.activation-api-1.2.2.jar;C:\Software\kafka\libs\jakarta.annotation-api-1.3.5.jar;C:\Software\kafka\libs\jakarta.inject-2.6.1.jar;C:\Software\kafka\libs\jakarta.validation-api-2.0.2.jar;C:\Software\kafka\libs\jakarta.ws.rs-api-2.1.6.jar;C:\Software\kafka\libs\jaxws.activation-api-1.2.0.jar;C:\Software\kafka\libs\javax.annotation-api-1.3.2.jar;C:\Software\kafka\libs\javassist-3.29.2-GA.jar;C:\Software\kafka\libs\jaxws.activation-api-1.2.0.jar;C:\Software\kafka\libs\jaxws.ws.rs-api-2.1.1.jar;C:\Software\kafka\libs\jaxb-api-2.3.1.jar;C:\Software\kafka\libs\jersey-client-2.39.1.jar;C:\Software\kafka\libs\jersey-common-2.39.1.jar;C:\Software\kafka\libs\jersey-container-servlet-2.39.1.jar;C:\Software\kafka\libs\jersey-container-servlet-core-2.39.1.jar;C:\Software\kafka\libs\jersey
```

## 3. Afficher la liste des topics (flume-topic1)

```
kafka-topics.bat --list --bootstrap-server localhost:9092
```



A screenshot of a terminal window titled "topics". The window contains several tabs: Zookeeper, kafka, topics (which is active), consumer, Flume, and Pyspark. The topics tab displays the output of the command "C:\Software\kafka\bin\windows>kafka-topics.bat --list --bootstrap-server localhost:9092". It shows three topics: \_\_consumer\_offsets, flume-topic, and flume-topic1.

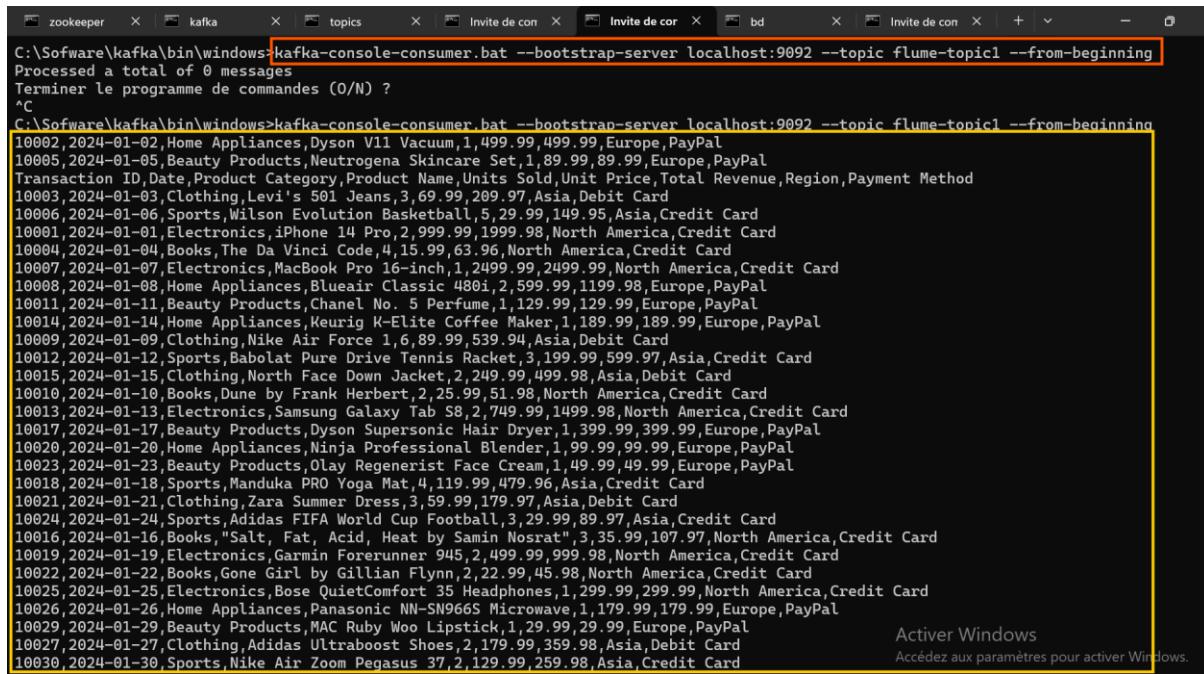
```
C:\Software\kafka\bin\windows>kafka-topics.bat --list --bootstrap-server localhost:9092
__consumer_offsets
flume-topic
flume-topic1
```

**NOTE (2) :** Le topic crée vous pouvez donc aller lancer exécuter votre fichier de configuration d'apache flume (note ) et ensuite continuer avec l'étape suivante :

#### 4. Lancer le consommateur kafka

Rappel : Apache flume déjà lancer et qui a collecté et envoyé à kafka (Note1)

```
kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic flume-topic2 --from-beginning
```



```
C:\Software\kafka\bin\windows>kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic flume-topic1 --from-beginning
Processed a total of 0 messages
Terminer le programme de commandes (O/N) ?
^C
C:\Software\kafka\bin\windows>kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic flume-topic1 --from-beginning
10002, 2024-01-02, Home Appliances, Dyson V11 Vacuum, 1, 499.99, 499.99, Europe, PayPal
10005, 2024-01-05, Beauty Products, Neutrogena Skincare Set, 1, 89.99, 89.99, Europe, PayPal
Transaction ID, Date, Product Category, Product Name, Units Sold, Unit Price, Total Revenue, Region, Payment Method
10003, 2024-01-03, Clothing, Levi's 501 Jeans, 3, 69.99, 209.97, Asia, Debit Card
10006, 2024-01-06, Sports, Wilson Evolution Basketball, 5, 29.99, 149.95, Asia, Credit Card
10001, 2024-01-01, Electronics, iPhone 14 Pro, 2, 999.99, 1999.98, North America, Credit Card
10004, 2024-01-04, Books, The Da Vinci Code, 4, 15.99, 63.96, North America, Credit Card
10007, 2024-01-07, Electronics, MacBook Pro 16-inch, 1, 2499.99, 2499.99, North America, Credit Card
10008, 2024-01-08, Home Appliances, Blueair Classic 480i, 2, 599.99, 1199.98, Europe, PayPal
10011, 2024-01-11, Beauty Products, Chanel No. 5 Perfume, 1, 129.99, 129.99, Europe, PayPal
10014, 2024-01-14, Home Appliances, Keurig K-Elite Coffee Maker, 1, 189.99, 189.99, Europe, PayPal
10009, 2024-01-09, Clothing, Nike Air Force 1, 6, 89.99, 539.94, Asia, Debit Card
10012, 2024-01-12, Sports, Babolat Pure Drive Tennis Racket, 3, 199.99, 599.97, Asia, Credit Card
10015, 2024-01-15, Clothing, North Face Down Jacket, 2, 249.99, 499.98, Asia, Debit Card
10010, 2024-01-10, Books, Dune by Frank Herbert, 2, 25.99, 51.98, North America, Credit Card
10013, 2024-01-13, Electronics, Samsung Galaxy Tab S8, 2, 749.99, 1499.98, North America, Credit Card
10017, 2024-01-17, Beauty Products, Dyson Supersonic Hair Dryer, 1, 399.99, 399.99, Europe, PayPal
10020, 2024-01-20, Home Appliances, Ninja Professional Blender, 1, 99.99, 99.99, Europe, PayPal
10023, 2024-01-23, Beauty Products, Olay Regenerist Face Cream, 1, 49.99, 49.99, Europe, PayPal
10018, 2024-01-18, Sports, Manduka PRO Yoga Mat, 4, 119.99, 479.96, Asia, Credit Card
10021, 2024-01-21, Clothing, Zara Summer Dress, 3, 59.99, 179.97, Asia, Debit Card
10024, 2024-01-24, Sports, Adidas FIFA World Cup Football, 3, 29.99, 89.97, Asia, Credit Card
10016, 2024-01-16, Books, "Salt, Fat, Acid, Heat" by Samin Nosrat*, 3, 35.99, 107.97, North America, Credit Card
10019, 2024-01-19, Electronics, Garmin Forerunner 945, 2, 499.99, 999.98, North America, Credit Card
10022, 2024-01-22, Books, Gone Girl by Gillian Flynn, 2, 22.99, 45.98, North America, Credit Card
10025, 2024-01-25, Electronics, Bose QuietComfort 35 Headphones, 1, 299.99, 299.99, North America, Credit Card
10026, 2024-01-26, Home Appliances, Panasonic NN-SN9665 Microwave, 1, 179.99, 179.99, Europe, PayPal
10029, 2024-01-29, Beauty Products, MAC Ruby Woo Lipstick, 1, 29.99, 29.99, Europe, PayPal
10027, 2024-01-27, Clothing, Adidas Ultraboost Shoes, 2, 179.99, 359.98, Asia, Debit Card
10030, 2024-01-30, Sports, Nike Air Zoom Pegasus 37, 2, 129.99, 259.98, Asia, Credit Card
```

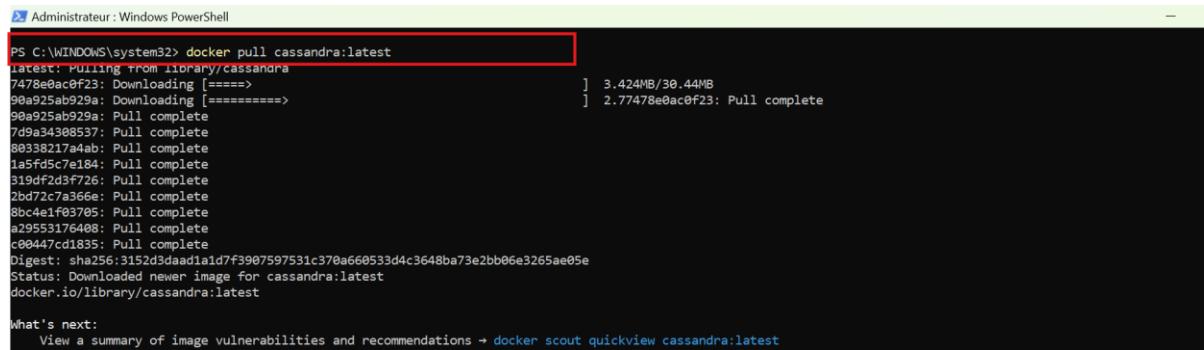
# Etape III : Apache CASSANDRA

## Installation et configurations (docker)

### 9. Extrayez l'image Docker

Cette docker pull commande obtiendra la dernière version de l'image officielle Docker Apache Cassandra disponible sur [Dockerhub](#).

```
kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic flume-topic2 --from-beginning
```

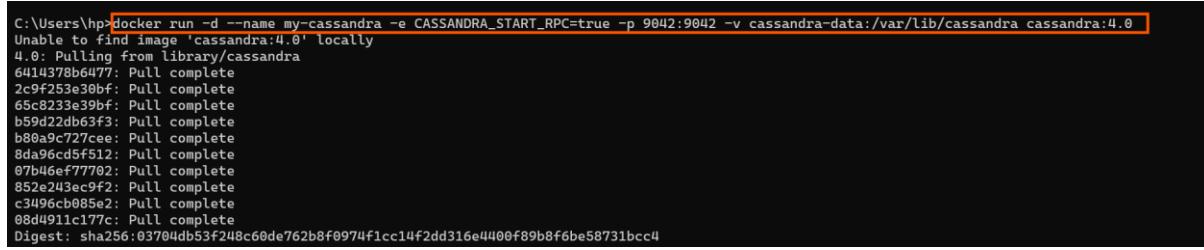


```
PS C:\WINDOWS\system32> docker pull cassandra:latest
latest: Pulling from library/cassandra
7478e0ac0f23: Downloading [====>] 3.424MB/30.44MB
9ea925ab929a: Downloading [=====] 2.77478e0ac0f23: Pull complete
9ea925ab929a: Pull complete
7d9a34308537: Pull complete
80338217a4ab: Pull complete
1a5fd5c7e184: Pull complete
319df2d3f726: Pull complete
2bd72c7a366e: Pull complete
8bc4e1f693705: Pull complete
a29553176408: Pull complete
c00447cd1835: Pull complete
Digest: sha256:3152d3daad1a1d7f3907597531c370a660533d4c3648ba73e2bb06e3265ae05e
Status: Downloaded newer image for cassandra:latest
docker.io/library/cassandra:latest

What's next:
  View a summary of image vulnerabilities and recommendations -> docker scout quickview cassandra:latest
```

### 10. Démarrez Cassandra

```
docker run -d --name my-cassandra -e CASSANDRA_START_RPC=true -p 9042:9042 -v cassandra-data:/var/lib/cassandra cassandra:4.0
```



```
C:\Users\hp>docker run -d --name my-cassandra -e CASSANDRA_START_RPC=true -p 9042:9042 -v cassandra-data:/var/lib/cassandra cassandra:4.0
Unable to find image 'cassandra:4.0' locally
4.0: Pulling from library/cassandra
6414378b6477: Pull complete
2c9f253e30bf: Pull complete
65c8233e39bf: Pull complete
b59d22db63f3: Pull complete
b8ea9c727cee: Pull complete
8da96cd5f512: Pull complete
07b46ef77702: Pull complete
852e243ec9f2: Pull complete
c3496cb085e2: Pull complete
08d4911c177c: Pull complete
Digest: sha256:03704db53f248c69de762b8f0974f1cc14f2dd316e4400f89b8f6be58731bcc4
```

## 11. Démarrez le shell CQL cqlsh

```
docker exec -it my-cassandra cqlsh
```

```
C:\Users\hp>docker run -d --name my-cassandra -e CASSANDRA_START_RPC=true -p 9042:9042 -v cassandra-data:/var/lib/cassandra cassandra:latest
ec1c743ee23e474c836b47f23dbe5709279043ed4194d803ffa57bd53b49fae
C:\Users\hp>docker exec -it my-cassandra cqlsh
Connected to Test Cluster at 127.0.0.1:9042
[cqlsh 6.2.0 | Cassandra 5.0.1 | CQL spec 3.4.7 | Native protocol v5]
Use HELP for help.
```

## 12. Créer un Keyspace (my\_keyspace)

```
CREATE KEYSPACE IF NOT EXISTS my_keyspace WITH REPLICATION = { 'class' : 'SimpleStrategy',
'replication_factor' : 1 };
```

```
cqlsh> CREATE KEYSPACE IF NOT EXISTS my_keyspace WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication_factor' : 1 };
cqlsh> CREATE TABLE IF NOT EXISTS my_keyspace.transactions (
...     transaction_id text PRIMARY KEY,
...     date text,
...     product_category text,
...     product_name text,
...     units_sold int,
...     unit_price double,
...     total_revenue double,
...     region text,
...     payment_method text
... );
cqlsh> DESCRIBE KEYSPACES;
my_keyspace  system_auth      system_schema  system_views
system      system_distributed system_traces   system_virtual_schema
```

## 13. Créer une Table

```
CREATE TABLE IF NOT EXISTS my_keyspace.transactions (
    transaction_id text PRIMARY KEY,
    date text,
    product_category text,
    product_name text,
    units_sold int,
    unit_price double,
    total_revenue double,
    region text,
    payment_method text
); "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"
```

```
cqlsh> CREATE KEYSPACE IF NOT EXISTS my_keyspace WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication_factor' : 1 };
cqlsh> CREATE TABLE IF NOT EXISTS my_keyspace.transactions (
...     transaction_id text PRIMARY KEY,
...     date text,
...     product_category text,
...     product_name text,
...     units_sold int,
...     unit_price double,
...     total_revenue double,
...     region text,
...     payment_method text
... );
cqlsh> DESCRIBE KEYSPACES;
my_keyspace  system_auth      system_schema  system_views
system       system_distributed system_traces   system_virtual_schema
```

#### 14. Vérifier les keyspace créée

```
DESCRIBE KEYSPACES;
```

```
cqlsh> CREATE KEYSPACE IF NOT EXISTS my_keyspace WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication_factor' : 1 };
cqlsh> CREATE TABLE IF NOT EXISTS my_keyspace.transactions (
...     transaction_id text PRIMARY KEY,
...     date text,
...     product_category text,
...     product_name text,
...     units_sold int,
...     unit_price double,
...     total_revenue double,
...     region text,
...     payment_method text
... );
cqlsh> DESCRIBE KEYSPACES;
my_keyspace  system_auth      system_schema  system_views
system       system_distributed system_traces   system_virtual_schema
```

#### 15. Vérifier la création de la table

```
USE my_keyspace;
```

```
cqlsh> USE my_keyspace;
cqlsh:my_keyspace> DESCRIBE TABLES;
transactions
```

## 16. Vérifier la structure de la table

DESCRIBE transactions:

```
cqlsh:my_keyspace> DESCRIBE transactions;
```

```
CREATE TABLE my_keyspace.transactions (
    transaction_id text PRIMARY KEY,
    date text,
    payment_method text,
    product_category text,
    product_name text,
    region text,
    total_revenue double,
    unit_price double,
    units_sold int
) WITH additional_write_policy = '99p'
    AND allow_auto_snapshot = true
    AND bloom_filter_fp_chance = 0.01
    AND caching = {'keys': 'ALL', 'rows_per_partition': 'NONE'}
    AND cdc = false
    AND comment = ''
    AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy', 'max_threshold': '32', 'min_threshold': '4'}
    AND compression = {'chunk_length_in_kb': '16', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor'}
    AND memtable = 'default'
    AND crc_check_chance = 1.0
    AND default_time_to_live = 0
    AND extensions = {}
    AND gc_grace_seconds = 864000
    AND incremental_backups = true
    AND max_index_interval = 2048
    AND memtable_flush_period_in_ms = 0
    AND min_index_interval = 128
    AND read_repair = 'BLOCKING'
    AND speculative_retry = '99n'.
```

## 17. Vérifier le contenu de la BD

select \* from transactions;

```
cqlsh:my_keyspace> select * from transactions ;
transaction_id | date | payment_method | product_category | product_name | region | total_revenue | unit_price | units_sold
-----+-----+-----+-----+-----+-----+-----+-----+-----+
(0 rows)
```

## 18. Verification après l'envoi des données de spark a cassandra

**NOTE (3) :** Avant d'exécuter cette étape il faut d'abord exécuter le code pyspark pour la transformation de données . Donc allez d'abord à la prochaine étape (pyspark). Quand il va falloir revenir sur ce point il sera signalé dans la suite du document.

select \* from transactions;

(0 rows)

```
cqlsh:my_keyspace> select * from transactions;
```

transaction_id	date	payment_method	product_category	product_name	region	total_revenue	unit_price	units_sold
10856	2024-02-25	PayPal	Home Appliances	KitchenAid Artisan Stand Mixer	Europe	499.99	499.99	1
10857	2024-02-25	Credit Card	Beauty Products	First Aid Beauty Ultra Repair Cream	North America	599.99	299.99	2
10179	2024-02-27	PayPal	Beauty Products	First Aid Beauty Superfood Antioxidant Cleanser	Europe	68.99	68.99	1
10161	2024-02-29	PayPal	Beauty Products	Youth to the People Superfood Antioxidant Cleanser	Europe	36	36	1
10968	2024-03-08	PayPal	Home Appliances	Eufy RoboVac 11S	Europe	659.97	219.99	3
10011	2024-01-11	PayPal	Beauty Products	Chanel No. 5 Perfume	Europe	129.99	129.99	1
10143	2024-05-22	PayPal	Beauty Products	Caudalie Vinoperfect Radiance Serum	Europe	79	79	1
10289	2024-07-27	PayPal	Beauty Products	De'Longhi Magnifica Espresso Machine	Europe	899.99	899.99	1
10066	2024-01-09	PayPal	Home Appliances	Dyson V11 Vacuum	Europe	499.99	499.99	1
10007	2024-01-16	PayPal	Beauty Products	Clinique Dramatically Different Moisturizing Lotion	Europe	52	52	1
10194	2024-07-12	PayPal	Home Appliances	Breville Smart Coffee Grinder Pro	Europe	199.95	199.95	1
10134	2024-05-13	PayPal	Home Appliances	Koerig K-Mini Coffee Maker	Europe	159.98	79.99	2
10014	2024-01-14	PayPal	Home Appliances	Keurig K-Elite Coffee Maker	Europe	189.99	189.99	1
10038	2024-02-07	PayPal	Home Appliances	Roomba i7+	Europe	1599.98	799.99	2
10020	2024-01-20	PayPal	Home Appliances	Ninja Professional Blender	Europe	99.99	99.99	1
10227	2024-08-14	PayPal	Beauty Products	The Ordinary Caffeine Solution 5% + EGCG	Europe	6.7	6.7	1
10065	2024-03-23	PayPal	Beauty Products	Charlotte Tilbury Magic Cream	Europe	100	100	1
10226	2024-02-26	PayPal	Home Appliances	Nespresso Vertuo Next Coffee and Espresso Maker	Europe	159.99	159.99	1
10158	2024-06-06	PayPal	Beauty Products	Breville Nespresso Capsule Plus	Europe	899.95	899.95	1
10801	2024-02-19	PayPal	Beauty Products	Estee Lauder Advanced Night Repair	Europe	185	185	1
10173	2024-06-21	PayPal	Beauty Products	Paula's Choice Skin Perfecting 2% BHA Liquid Exfoliant	Europe	29.5	29.5	1
10017	2024-01-17	PayPal	Beauty Products	Dyson Supersonic Hair Dryer	Europe	399.99	399.99	1
10203	2024-07-21	PayPal	Beauty Products	Dr. Jart+ Cicapair Tiger Grass Color Correcting Treatment	Europe	52	52	1
10137	2024-05-16	PayPal	Beauty Products	The Ordinary Hyaluronic Acid Serum	Europe	6.8	6.8	1
10077	2024-03-17	PayPal	Beauty Products	Glossier Boy Brow	Europe	32	16	2
10146	2024-02-19	PayPal	Home Appliances	Cuisinart Coffee Maker	Europe	399.95	199.95	2
10035	2024-02-04	PayPal	Beauty Products	L'Oréal Revitalift Serum	Europe	79.98	39.99	2
10239	2024-08-26	PayPal	Beauty Products	Sunday Riley Luna Sleeping Night Oil	Europe	55	55	1
10104	2024-04-13	PayPal	Home Appliances	Shark IQ Robot Vacuum	Europe	699.98	349.99	2
10224	2024-08-11	PayPal	Home Appliances	Philips Sonicare DiamondClean Toothbrush	Europe	459.98	229.99	2
10176	2024-06-24	PayPal	Home Appliances	Crock-Pot 6-Quart Slow Cooker	Europe	99.98	49.99	2
10095	2024-04-04	PayPal	Beauty Products	Tatché The Water Cream	Europe	68	68	1
10109	2024-05-28	PayPal	Beauty Products	Laneige Water Sleeping Mask	Europe	25	25	1
10106	2024-02-17	PayPal	Home Appliances	Sunday Riley Advanced Night Repair	Europe	105	30.99	1
10140	2024-05-19	PayPal	Beauty Products	Olay Regenerist Face Cream	Europe	49.99	49.99	1
10062	2024-03-02	PayPal	Home Appliances	Instant Pot Ultra	Europe	139.99	139.99	1
10122	2024-05-01	PayPal	Home Appliances	Philips Airfryer XXL	Europe	399.98	199.99	2
10113	2024-04-22	PayPal	Beauty Products	Hamilton Beach FlexBrew Coffee Maker	Europe	89.99	89.99	1
10152	2024-05-31	PayPal	Home Appliances	Biore UV Aqua Rich Watery Essence Sunscreen	Europe	15	15	1
10185	2024-07-03	PayPal	Beauty Products	Ninja Foodi Pressure Cooker	Europe	399.98	199.99	2
10109	2024-01-23	PayPal	Beauty Products	La Mer Crème de la Mer Moisturizer	Europe	198	198	1
10132	2024-02-01	PayPal	Home Appliances	Instant Pot Duo	Europe	279.99	139.99	2
10289	2024-07-18	PayPal	Home Appliances	Instant Pot Duo Nova	Europe	99.95	99.95	1
10182	2024-06-30	PayPal	Home Appliances	Instant Pot Duo Nova	Europe	399.99	399.99	1
10191	2024-07-09	PayPal	Beauty Products	Glossier Cloud Paint	Europe	18	18	1

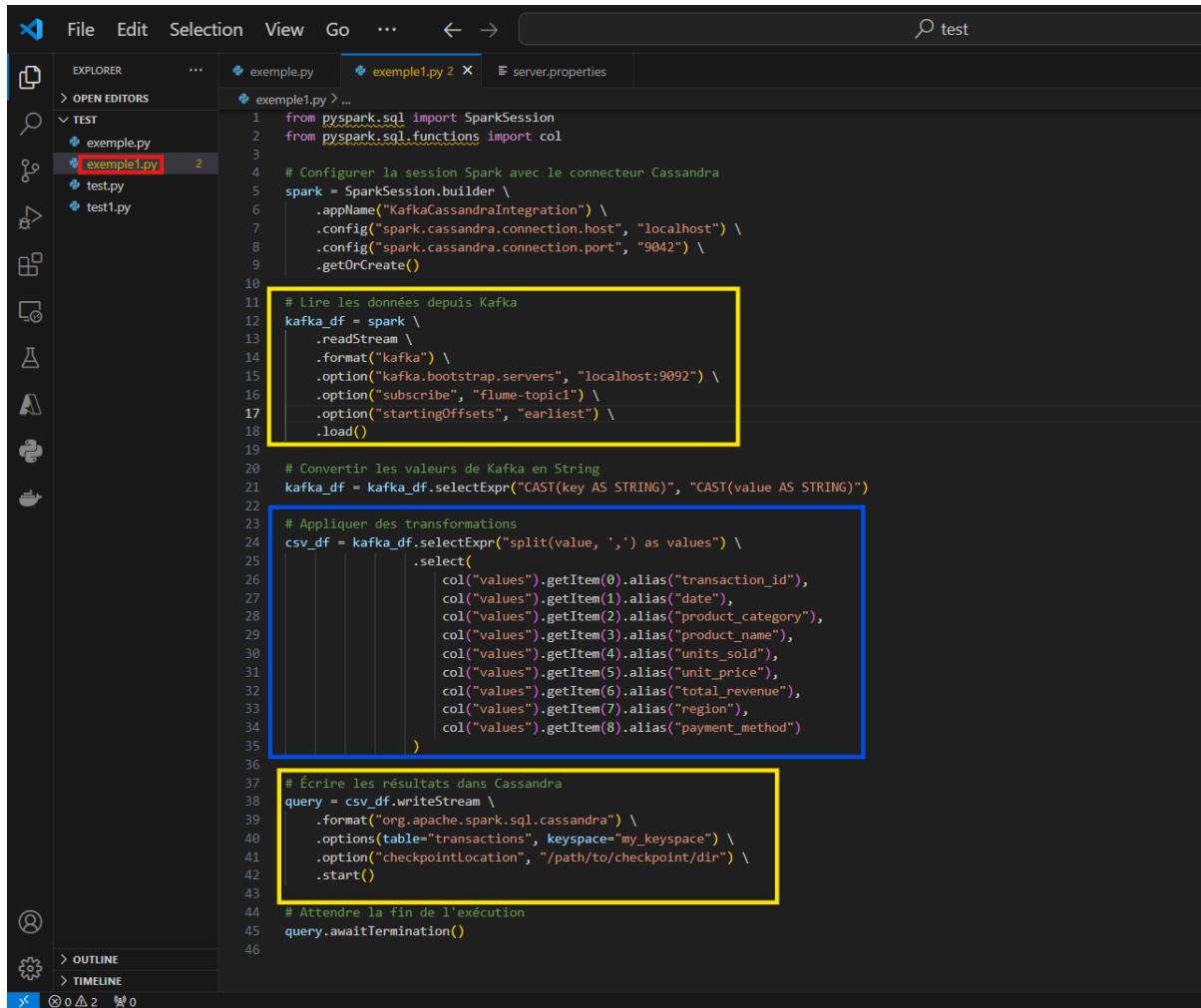
```
cqlsh:moviesdata> use my_keyspace ;
cqlsh:my_keyspace> select * from transactions ;
```

transaction_id	date	payment_method	product_category	product_name	region	total_revenue	unit_price	units_sold
10866	2024-02-29	Credit Card	Sports	Hyperice Hypervolt Massager	Asia	349	349	1
10867	2024-02-29	Credit Card	Sports	Thinx Period Underwear	North America	25.99	11.99	2
10113	2024-02-21	Credit Card	Books	Sapiens: A Brief History of Humankind by Yuval Noah Harari	North America	57.98	10.99	5
10856	2024-02-25	PayPal	Home Appliances	KitchenAid Artisan Stand Mixer	Europe	499.99	499.99	1
10034	2024-01-19	Credit Card	Sports	The Girl with the Dragon Tattoo by Stieg Larsson	North America	29.99	29.99	1
10248	2024-08-27	Credit Card	Sports	Yves Saint Laurent Ysl Radiance Lipstick	Asia	59.98	29.99	2
10201	2024-07-19	Debit Card	Clothing	Gap 1969 Original Fit Jeans	Asia	179.97	59.99	3
10041	2024-01-20	Credit Card	Sports	Yeezy Season 10 Track Top	North America	229.99	79.99	3
10118	2024-06-27	Credit Card	Books	Atomic Habits by James Clear	North America	47.96	16.99	4
10116	2024-06-26	PayPal	Home Appliances	Breville Smart Grill	Europe	599.9	299.95	2
10030	2024-02-05	Credit Card	Clothing	Ray-Ban Aviator Sunglasses	Asia	469.99	199.95	3
10036	2024-02-05	Credit Card	Sports	Peloton Bike	Asia	1995	1995	1
10171	2024-06-19	Debit Card	Clothing	Nike Tempo Running Shorts	Asia	184.97	34.99	3
10038	2024-01-20	Credit Card	Sports	Vilebrequin Men's Tiki Short	Asia	109.99	40.99	3
10021	2024-02-21	Debit Card	Clothing	Zara Summer Dress	Asia	179.97	59.99	3
10207	2024-07-25	Credit Card	Clothing	Lululemon Under High-Rise Leggings	Asia	19	98	2
10030	2024-01-13	Credit Card	Beauty Products	First Aid Beauty Hydrating Eye Cream	Europe	24.99	24.99	1
10216	2024-06-03	Credit Card	Sports	YETI Tundra 65 Cooler	Asia	349.99	349.99	1
10042	2024-09-13	Credit Card	Sports	Fitbit Charge 5	Asia	259.98	129.99	2
10031	2024-09-13	Credit Card	Beauty Products	Yves Saint Laurent Ysl Radiance Lipstick	Europe	89.99	89.99	1
10087	2024-03-27	Debit Card	Clothing	Uniqlo Light Down Jacket	Asia	239.97	79.99	3
10025	2024-01-25	Credit Card	Electronics	Bose QuietComfort 35 Headphones	North America	299.99	299.99	1
10032	2024-01-25	Credit Card	Clothing	Puma Suede Sneakers	Asia	109.99	34.99	3
10068	2024-03-08	PayPal	Home Appliances	Eufy RoboVac 11S	Europe	659.97	219.99	3
10235	2024-08-22	Credit Card	Sports	YETI Tundra 45 Cooler	Asia	299.99	299.99	1
10033	2024-01-11	Credit Card	Beauty Products	Bose QuietComfort 35 II Wireless Headphones	North America	299	499	1
10041	2024-08-22	Credit Card	Electronics	CafePress Custom T-Shirts	Europe	29.99	29.99	1
10072	2024-09-12	Credit Card	Sports	CLINIQUE Dramatically Different Moisturizing Lotion	Europe	29	29.99	1
10049	2024-07-29	PayPal	Beauty Products	NO 7 Specialist Skincare Range	North America	1599.99	1599.99	1
10022	2024-08-13	Credit Card	Electronics	Nike Metcon 6	Asia	389.97	129.99	3
10038	2024-01-14	Credit Card	Beauty Products	Caudalie Vinosource Rosé Hydrating Cream	Europe	899.99	899.99	1
10013	2024-01-13	Credit Card	Electronics	Samsung Galaxy Tab S8	North America	1499.98	799.99	2
10001	2024-01-01	Credit Card	Books	The Four Agreements by Don Miguel Ruiz	North America	17.99	8.99	2
10082	2024-03-21	Credit Card	Electronics	The Silent Patient by Alex Michaelides	North America	51.98	26.99	2
10083	2024-03-22	Credit Card	Books	The Hobbit: Desolation of Smaug by J.R.R. Tolkien	Europe	49.99	49.99	1
10190	2024-07-08	Credit Card	Books	The Catcher in the Rye by J.D. Salinger	North America	32.97	16.99	3
10194	2024-07-12	PayPal	Beauty Products	Urban Decay Naked Heat Eyeshadow Palette	Europe	199.95	199.95	1
10032	2024-08-19	Credit Card	Books	Dune by Frank Herbert	North America	19.98	9.99	2
10014	2024-01-19	Credit Card	Home Appliances	Keurig K-Elite Coffee Maker	Europe	189.99	189.99	1
10038	2024-02-07	PayPal	Home Appliances	Rooms i7+	Europe	1599.98	799.99	2
10024	2024-06-03	Credit Card	Electronics	Bose SoundLink Revolve Bluetooth Speaker	North America	899.99	499.99	2
10217	2024-08-04	Credit Card	Electronics	Product Name: Region: null null	Europe	2399	2399	1
10030	2024-01-20	Credit Card	Clothing	Apple MacBook Pro 16-inch	North America	2399	2399	1
10027	2024-08-27	Debit Card	Clothing	Hydro Flask Standard Mouth Water Bottle	Asia	42.99	32.99	3
10028	2024-01-28	PayPal	Home Appliances	Adidas Ultraboost Shoes	Asia	389.98	179.99	3
10033	2024-01-23	Credit Card	Electronics	Nirvana Professional Blender	Europe	99.99	99.99	1
10227	2024-08-10	Credit Card	Beauty Products	The Ordinary Caffeine Solution 5% + EGCG	Europe	6.7	6.7	1
10045	2024-02-14	Debit Card	Clothing	Patagonia Better Sweater	Asia	279.98	139.99	2
10031	2024-01-14	Credit Card	Books	One Hundred Years of Solitude by Gabriel García Márquez	Asia	14.99	14.99	3
10083	2024-03-23	PayPal	Beauty Products	Charlotte Tilbury Magic Cream	Europe	180	180	1
10236	2024-08-23	PayPal	Home Appliances	Nespresso Vertuo Next Coffee and Espresso Maker	Europe	159.99	159.99	1
10034	2024-01-14	Credit Card	Electronics	Microsoft Surface Laptop 4	North America	1299.99	1299.99	1
10133	2024-08-12	Credit Card	Home Appliances	Rowenta Waffle Maker	Europe	200.99	200.99	1

# Etape IV : Apache Spark

## Installation et configurations SPARK (pyspark)

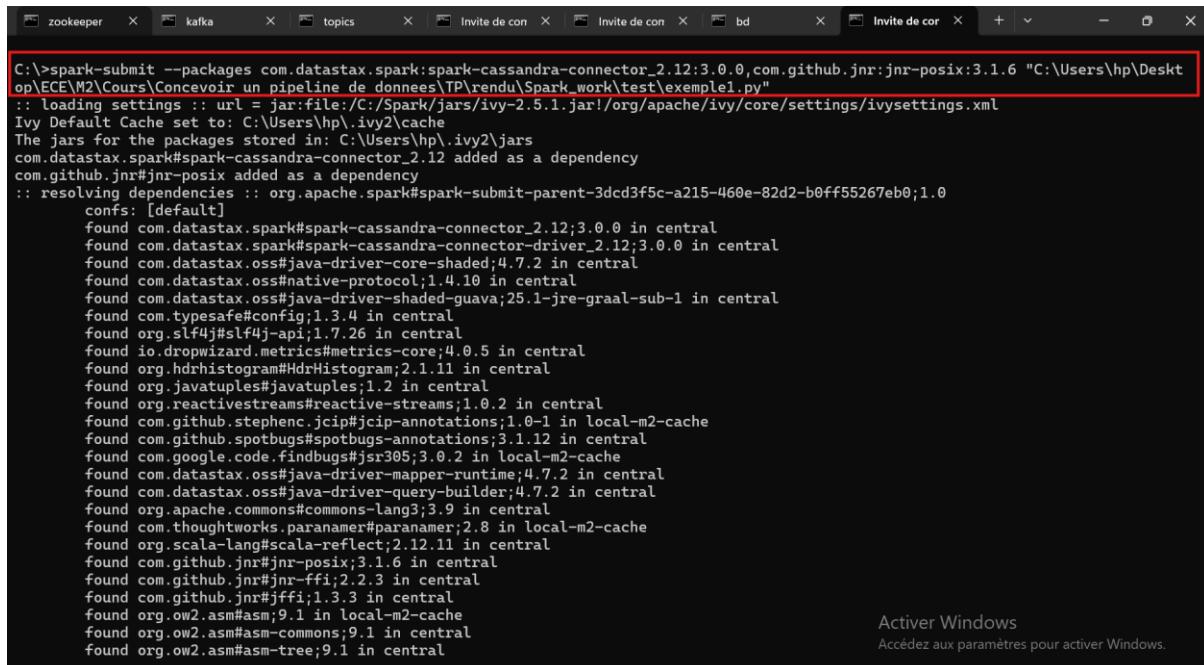
### 1. Code python



```
File Edit Selection View Go ... ← → ⌂ test
EXPLORER ... exemple.py exemple1.py 2 server.properties
> OPEN EDITORS
TEST
exemple.py
exemple1.py 2
test.py
test1.py
1  from pyspark.sql import SparkSession
2  from pyspark.sql.functions import col
3
4  # Configurer la session Spark avec le connecteur Cassandra
5  spark = SparkSession.builder \
6      .appName("KafkaCassandraIntegration") \
7      .config("spark.cassandra.connection.host", "localhost") \
8      .config("spark.cassandra.connection.port", "9042") \
9      .getOrCreate()
10
11 # Lire les données depuis Kafka
12 kafka_df = spark \
13     .readStream \
14     .format("kafka") \
15     .option("kafka.bootstrap.servers", "localhost:9092") \
16     .option("subscribe", "flume-topic1") \
17     .option("startingOffsets", "earliest") \
18     .load()
19
20 # Convertir les valeurs de Kafka en String
21 kafka_df = kafka_df.selectExpr("CAST(key AS STRING)", "CAST(value AS STRING)")
22
23 # Appliquer des transformations
24 csv_df = kafka_df.selectExpr("split(value, ',') as values") \
25     .select(
26         col("values").getItem(0).alias("transaction_id"),
27         col("values").getItem(1).alias("date"),
28         col("values").getItem(2).alias("product_category"),
29         col("values").getItem(3).alias("product_name"),
30         col("values").getItem(4).alias("units_sold"),
31         col("values").getItem(5).alias("unit_price"),
32         col("values").getItem(6).alias("total_revenue"),
33         col("values").getItem(7).alias("region"),
34         col("values").getItem(8).alias("payment_method")
35     )
36
37 # Écrire les résultats dans Cassandra
38 query = csv_df.writeStream \
39     .format("org.apache.spark.sql.cassandra") \
40     .options(table="transactions", keyspace="my_keyspace") \
41     .option("checkpointLocation", "/path/to/checkpoint/dir") \
42     .start()
43
44 # Attendre la fin de l'exécution
45 query.awaitTermination()
46
```

## 2. Exécution

```
spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.0.1,com.datastax.spark:spark-cassandra-connector_2.12:3.0.0 --conf "spark.driver.extraClassPath=/path/to/jnr-posix.jar"  
"C:\Users\hp\Desktop\ECE\M2\Cours\Concevoir un pipeline de donnees\TP\rendu\Spark_work\test\exemple1.py"
```



```
C:\>spark-submit --packages com.datastax.spark:spark-cassandra-connector_2.12:3.0.0,com.github.jnr:jnr-posix:3.1.6 "C:\Users\hp\Desktop\ECE\M2\Cours\Concevoir un pipeline de donnees\TP\rendu\Spark_work\test\exemple1.py"  
:: loading settings :: url = jar:file:/C:/Spark/jars/ivy-2.5.1.jar!/org/apache/ivy/core/settings/ivysettings.xml  
Ivy Default Cache set to: C:\Users\hp\.ivy2\cache  
The jars for the packages stored in: C:\Users\hp\.ivy2\jars  
com.datastax.spark#spark-cassandra-connector_2.12 added as a dependency  
com.github.jnr#jnr-posix added as a dependency  
:: resolving dependencies :: org.apache.spark#spark-submit-parent-3cd3f5c-a215-460e-82d2-b0ff55267eb0;1.0  
  confs: [default]  
    found com.datastax.spark#spark-cassandra-connector_2.12;3.0.0 in central  
    found com.datastax.spark#spark-cassandra-connector-driver_2.12;3.0.0 in central  
    found com.datastax.oss#java-driver-core-shaded;4.7.2 in central  
    found com.datastax.oss#native-protocol;1.4.10 in central  
    found com.datastax.oss#java-driver-shaded-guava;25.1-jre-graal-sub-1 in central  
    found com.typesafe#config;1.3.4 in central  
    found org.slf4j#slf4j-api;1.7.26 in central  
    found io.dropwizard.metrics#metrics-core;4.0.5 in central  
    found org.hdrhistogram#HdrHistogram;2.1.11 in central  
    found org.javatuples#javatuples;1.2 in central  
    found org.reactivestreams#reactive-streams;1.0.2 in central  
    found com.github.stephenc.jcip#jcip-annotations;1.0-1 in local-m2-cache  
    found com.github.spotbugs#spotbugs-annotations;3.1.12 in central  
    found com.google.code.findbugs#jsr305;3.0.2 in local-m2-cache  
    found com.datastax.oss#java-driver-mapper-runtime;4.7.2 in central  
    found com.datastax.oss#java-driver-query-builder;4.7.2 in central  
    found org.apache.commons#commons-lang3;3.9 in central  
    found com.thoughtworks.parancamer#parancamer;2.8 in local-m2-cache  
    found org.scala-lang#scala-reflect;2.12.11 in central  
    found com.github.jnr#jnr-posix;3.1.6 in central  
    found com.github.jnr#jffi;2.2.3 in central  
    found com.github.jnr#jffi;1.3.3 in central  
    found org.ow2.asm#asm;9.1 in local-m2-cache  
    found org.ow2.asm#asm-commons;9.1 in central  
    found org.ow2.asm#asm-tree;9.1 in central
```

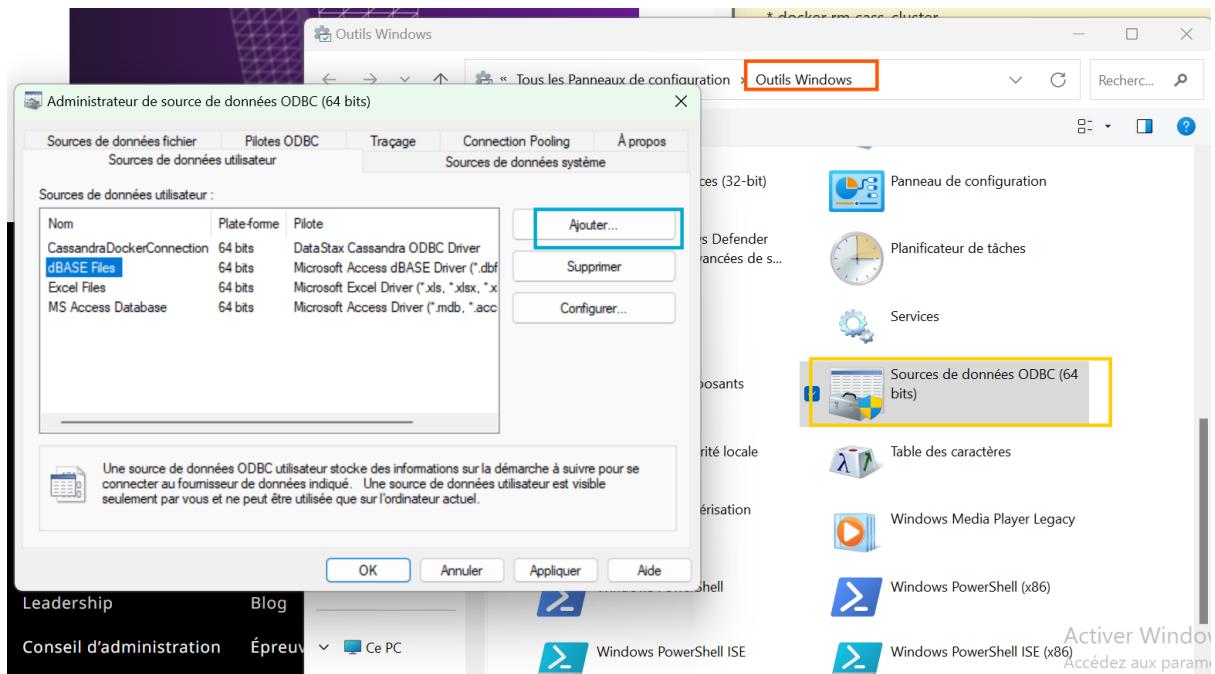
Activer Windows  
Accédez aux paramètres pour activer Windows.

**NOTE (4) :** Apres cette etape vous pouvez revenir a la NOTE (3) Pour vérifier le contenu de la keyspace cassandra .

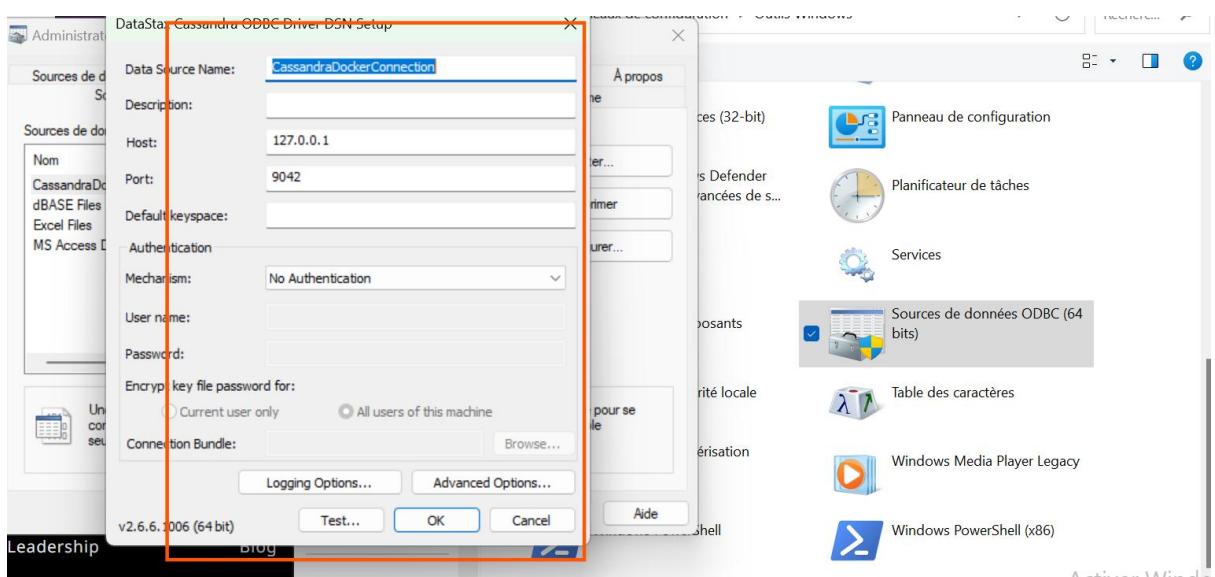
# Etape V : Configuration de la source de données ODBC

Cette étape va permettre la connexion de cassandra et power bi

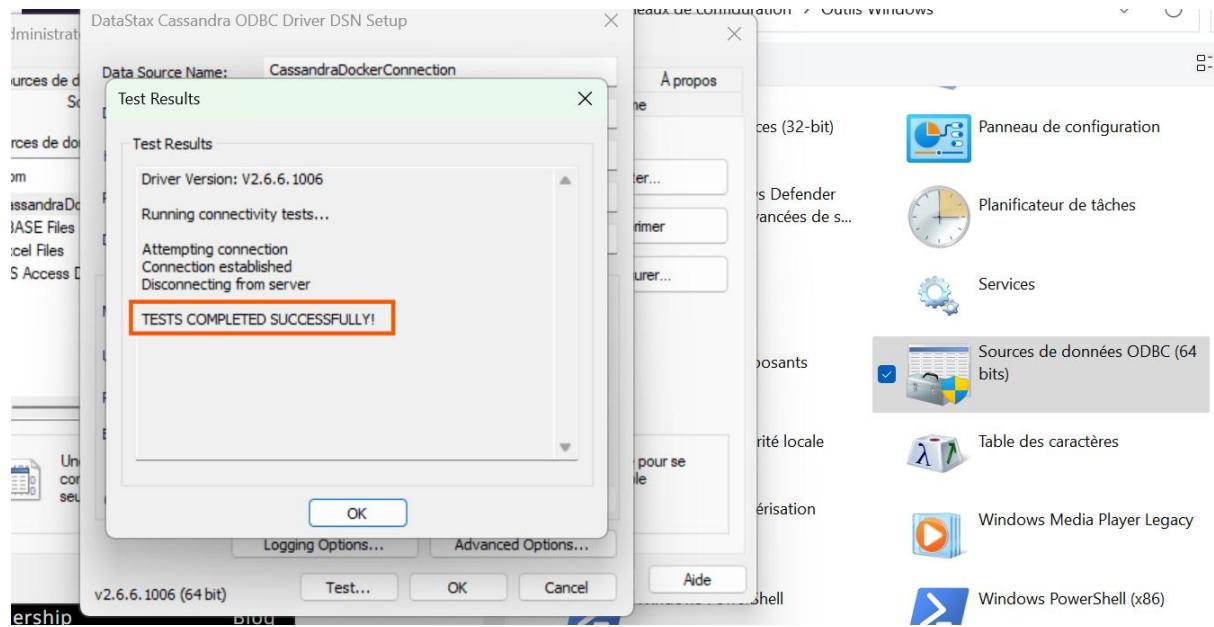
## 1. Ajouter une source de donnée



## 2. Remplir les informations de la source

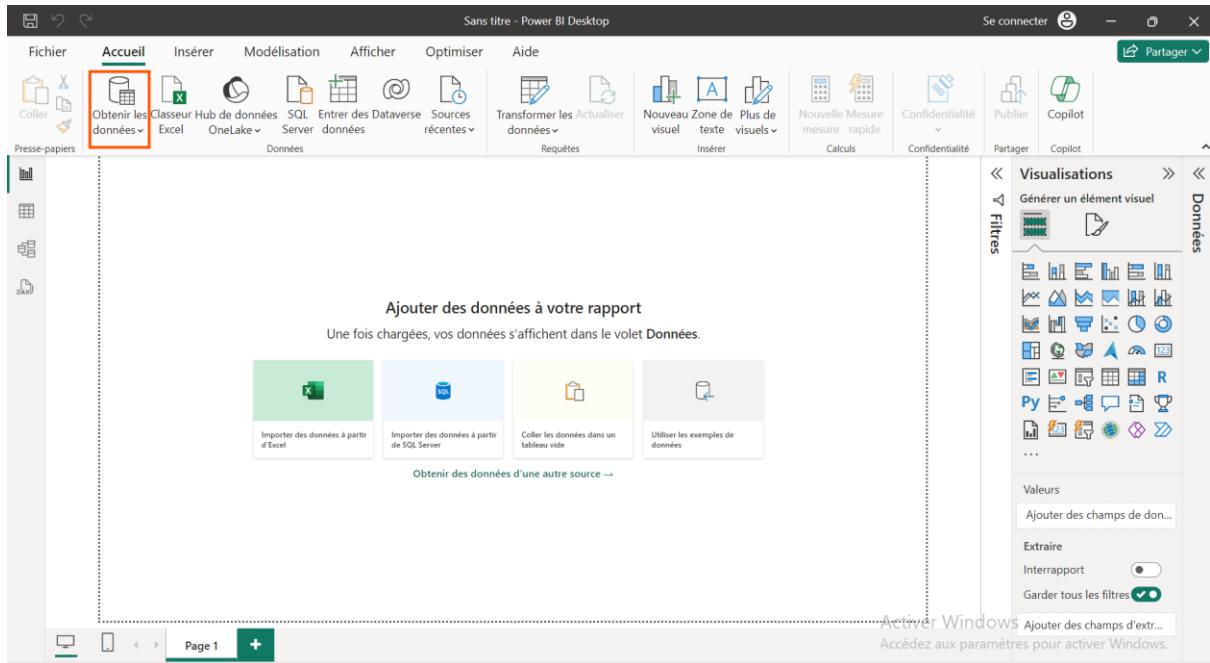


### 3. Tester la connexion avec la nouvelle source

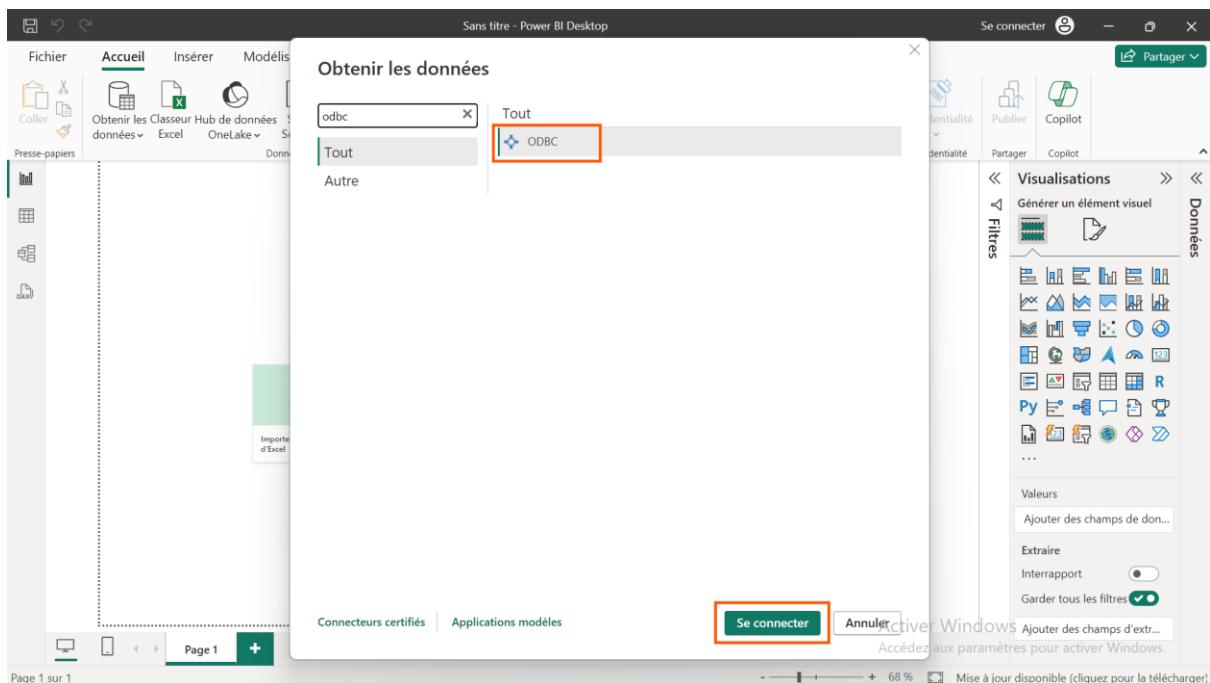


# Etape VI : Visualisation POWER BI

## 1. Obtenir la source de donnée



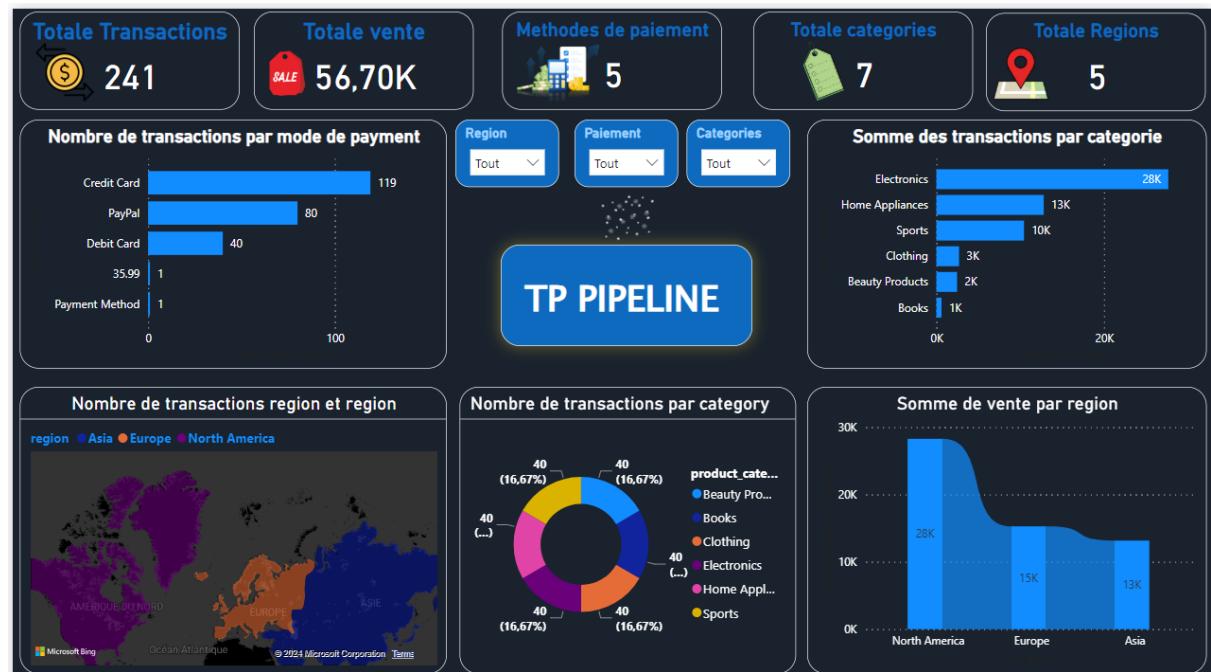
## 2. Choisir la source ODBC créer précédemment



### 3. Choisir la table

The screenshot shows the Microsoft Power BI Data Explorer interface. On the left, there's a navigation pane titled "Navigateur" (Explorer) with a tree view of databases and tables. The "transactions" table is selected under the "my\_keyspace [!]" database. To the right is a preview of the "transactions" table data, which contains columns: transaction\_id, date, payment\_method, product\_category, and product\_name. The table has 100 rows of data. At the bottom right of the preview area, there are buttons for "Charger" (Load), "Transformer les données" (Transform data), and "Annuler" (Cancel). A green box highlights the "Charger" button.

### 4. Quelques visualisations



Sans titre - Power BI Desktop

Se connecter

Fichier Accueil Insérer Modélisation Afficher Optimiser Aide

Nouvelle page Nouveau visuel Plus de visuels Eléments visuels Pages Visuels IA Questions et Influenceurs Arborescence de Narratif réponses clés décomposition Rapport Power paginé Apps Automate Power Platform Zone de Boutons Formes Image texte Ajouter un graphique sparkline Éléments Sparklines Partager

**Total Transactions** 241 **Total vente** 56.70K **Méthodes de paiement** 5 **Total categories** 7 **Total Regions** 5

**Nombre de transactions par mode de payment**

Payment Method	Count
Credit Card	119
PayPal	80
Debit Card	40
35.99	1
Payment Method	1
0	0

**TP PIPELINE**

**Nombre de transactions par region et region**



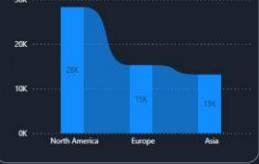
**Nombre de transactions par category**



**Somme des transactions par catégorie**

Category	Sum
Electronics	13K
Home Appliances	10K
Sports	3K
Clothing	2K
Beauty Products	1K
Books	0K

**Somme de vente par region**



Activer Windows  
Accédez aux paramètres pour activer Windows.

Je vous remercie pour la relecture approfondie de  
mon document. N'hésitez pas à me contacter si  
vous constatez des problèmes

**PARTAGEZ SVP**



Merci!



Edson KANOU TAYOUTSOP



Edson KANOU



Edsonkanou55@gmail.com

