

Apache kafka quick start

Site officiel : <https://kafka.apache.org>

STEP 1: DOWNLOAD AND INSTALL KAFKA

https://dlcdn.apache.org/kafka/3.7.0/kafka_2.13-3.7.0.tgz

STEP 2: START THE KAFKA ENVIRONMENT

Start the ZooKeeper service

```
C:\Users\RAMESH\Downloads\kafka>.bin\windows\zookeeper-server-start.bat  
.\config\zookeeper.properties
```

Start the Kafka broker service

```
C:\Users\RAMESH\Downloads\kafka>.bin\windows\kafka-server-start.bat .\config\server.properties
```

STEP 3: CREATE A TOPIC TO STORE YOUR EVENTS

```
C:\Users\RAMESH\Downloads\kafka>.bin\windows\kafka-topics.bat --create --topic topic_demo --  
bootstrap-server localhost:9092
```

STEP 4: WRITE SOME EVENTS INTO THE TOPIC

```
C:\Users\RAMESH\Downloads\kafka>.bin\windows\kafka-console-producer.bat --topic topic_demo  
--bootstrap-server localhost:9092
```

```
>hello world
```

```
>topic demo
```

STEP 5: READ THE EVENTS

```
C:\Users\RAMESH\Downloads\kafka>.bin\windows\kafka-console-consumer.bat --topic topic_demo  
--from-beginning --bootstrap-server localhost:9092
```

```
hello world
```

```
topic demo
```

Kafka-connect

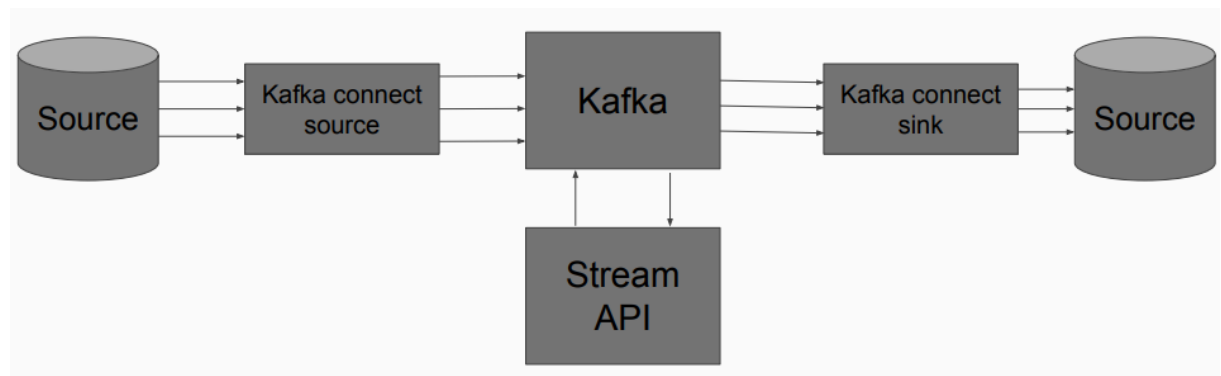
Connect kafka avec les systèmes externes (base de données, système de fichier,...)

2 catégories :

Connecteur source : prendre les données du monde extérieur et les connecter avec KAFKA.

Exemple : lire des messages depuis une BDD et les envoyer sur un topic kafka

Connecteur sink : prendre les données depuis un topic kafka et les envoyer à une destination par exemple HDFS (si on veut créer des fichiers sous HDFS)



- Test 1 : connect source file stream
- Test 2 : connect sink file stream
- Test 3 : connect sink désérialisation JSON

A. CONNECT SOURCE FILE STREAM

1. lancer le connecteur en mode standalone: (ou distributed)

```
C:\kafka\bin\windows>.\connect-standalone.bat ..\..\kafka-connect\connect-standalone.properties  
"..\..\kafka-connect\connect-file-source.properties"
```

2. lancer un consumer sur un topic de test pour tester :

```
.\bin\windows\kafka-console-consumer.bat --topic test-autre--from-beginning --bootstrap-server  
localhost:9092
```

produire : écrire sur test-source.txt -> aller à la ligne et sauvegarder

Le consumer va lire dans un fichier text

```
C:\kafka>.\bin\windows\kafka-console-consumer.bat --topic test-autre --bootstrap-server localhost:9092  
hi  
hello  
imene  
michelin  
capgemini
```

test-source.txt - Bloc-notes

```
Fichier Edition Format Affichage Aide  
hi  
hello  
imene  
michelin  
capgemini
```

B. Linverse : CONNECT SINK FILE STREAM

1. lancer le source et le sink en mode standalone (il faut arrêter celui d'avant car déjà utilisé)

```
.\connect-standalone.bat ..\..\kafka-connect\connect-standalone.properties ..\..\kafka-  
connect\connect-file-source.properties ..\..\kafka-connect\connect-file-sink.properties
```

2. écrire dans le source et voir le resultat dans le sink

C: > kafka > kafka-connect > resultat-sink.txt

```
1 hi
2 hello
3 imene
4 michelin
5 capgemini
6 hi
7 jjj
8 hhhhhhhhhhhhhhhhhhh
9
10 llo there
11
12
13
14 iiiiiiiiii
15 last one
16 le 09 avril 2024
17
```

test-source.txt - Bloc-notes

Fichier Edition Format Affichage Aide

```
hi
hello
imene
michelin
capgemini
hi
jjj
hhhhhhhhhhhhhhhhhh
hello there
cc
iiiiiiiiiii
last one
le 09 avril 2024
```

C:\Windows\System32\cmd.exe - .\bin\windows\kafka-console-consumer.bat --topic test-autre --bootstrap-server localhost:9092

```
C:\kafka>.\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic test-autre
[2024-04-09 11:07:41,931] WARN [Consumer clientId=console-consumer, groupId=console-consumer-71755] Error while fetching
metadata with correlation id 2 : {test-autre=LEADER_NOT_AVAILABLE} (org.apache.kafka.clients.NetworkClient)
[2024-04-09 11:07:42,092] WARN [Consumer clientId=console-consumer, groupId=console-consumer-71755] Error while fetching
metadata with correlation id 8 : {test-autre=LEADER_NOT_AVAILABLE} (org.apache.kafka.clients.NetworkClient)
Processed a total of 0 messages
Terminer le programme de commandes (O/N) ? o
```

```
C:\kafka>.\bin\windows\kafka-console-consumer.bat --topic test-autre --from-beginning --bootstrap-server localhost:9092
Processed a total of 0 messages
Terminer le programme de commandes (O/N) ? o
```

```
C:\kafka>.\bin\windows\kafka-console-consumer.bat --topic test-autre --bootstrap-server localhost:9092
```

```
hi
hello
imene
michelin
capgemini
hi
jjj
hhhhhhhhhhhhhhhhhh
llo there
```

```
iiiiiiiiiii
last one
le 09 avril 2024
```

C: > kafka > kafka-connect > resultat-sink.txt

```
1 hi
2 hello
3 imene
4 michelin
5 capgemini
6 hi
7 jjj
8 hhhhhhhhhhhhhhhhhhh
9
10 llo there
11
12
13
14 iiiiiiiiii
15 last one
16 le 09 avril 2024
17
```

test-source.txt - Bloc-notes

Fichier Edition Format Affichage Aide

```
hi
hello
imene
michelin
capgemini
hi
jjj
hhhhhhhhhhhhhhhhhh
hello there
cc
iiiiiiiiiii
last one
le 09 avril 2024
```

C. CONNECT SINK DESERIALISATION JSON

1. modifier standalone.properties

bootstrap.servers=localhost:9092

key.converter=org.apache.kafka.connect.json.JsonConverter

value.converter=org.apache.kafka.connect.json.JsonConverter

key.converter.schemas.enable=false

value.converter.schemas.enable=false

offset.storage.file.filename=C:\\kafka\\offset-storage.offsets

offset.flush.interval.ms=10000

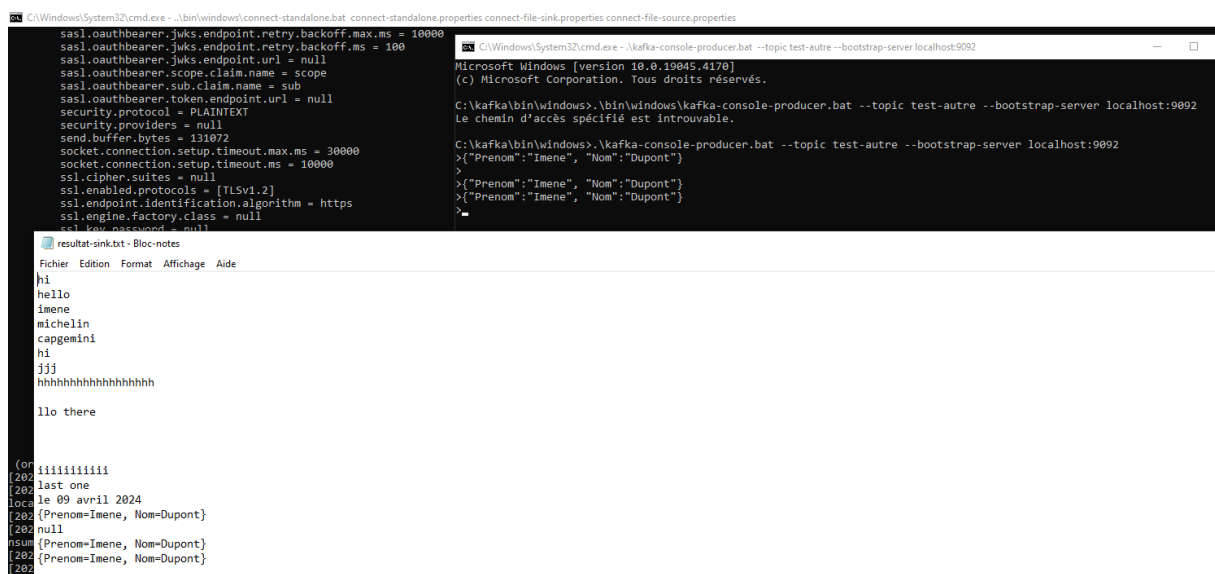
plugin.path=C:\\kafka\\libs

2. lancer le source et le sink en mode standalone (il faut arrêter celui d'avant car déjà utilisé)

C:\\kafka\\kafka-connect>..\\bin\\windows\\connect-standalone.bat connect-standalone.properties
connect-file-sink.properties connect-file-source.properties

3. Créer le producer : C:\\Users\\RAMESH\\Downloads\\kafka>..\\bin\\windows\\kafka-console-
producer.bat --topic test-autre --bootstrap-server localhost:9092 et tapez : {"Prenom":"Imene",
"Nom":"Dupond"}

4. lire sur le fichier resultat-sink.txt : {Prenom=Imene, Nom=Dupond}



```
C:\Windows\System32\cmd.exe - .\bin\windows\connect-standalone.bat connect-standalone.properties connect-file-sink.properties connect-file-source.properties
sasl.oauthbearer.jwks.endpoint.retry.backoff.max.ms = 10000
sasl.oauthbearer.jwks.endpoint.retry.backoff.ms = 100
sasl.oauthbearer.jwks.endpoint.url = null
sasl.oauthbearer.scope.claim.name = scope
sasl.oauthbearer.sub.claim.name = sub
sasl.oauthbearer.token.endpoint.url = null
security.protocol = PLAINTEXT
security.providers = null
send.buffer.bytes = 131072
socket.connection.setup.timeout.max.ms = 30000
socket.connection.setup.timeout.ms = 10000
ssl.cipher.suites = null
ssl.enabled.protocols = [TLSv1.2]
ssl.endpoint.identification.algorithm = https
ssl.engine.factory.class = null
ssl.key.password = null

C:\Windows\System32\cmd.exe - .\kafka-console-producer.bat --topic test-autre --bootstrap-server localhost:9092
Microsoft Windows [version 10.0.19045.4170]
(c) Microsoft Corporation. Tous droits réservés.

C:\kafka\bin\windows>..\\bin\windows\kafka-console-producer.bat --topic test-autre --bootstrap-server localhost:9092
Le chemin d'accès spécifié est introuvable.

C:\kafka\bin\windows>..\\bin\windows\kafka-console-producer.bat --topic test-autre --bootstrap-server localhost:9092
>{"Prenom":"Imene", "Nom":"Dupont"}
>
>{"Prenom":"Imene", "Nom":"Dupont"}
>{"Prenom":"Imene", "Nom":"Dupont"}
>{"Prenom":"Imene", "Nom":"Dupont"}
>

resultat-sink.txt - Bloc-notes
Fichier Edition Format Affichage Aide
hi
hello
imene
michelin
cappemini
hi
jjj
hhhhhhhhhhhhhhhhhhhh
llo there

{
  "Prenom": "Imene",
  "Nom": "Dupont"
}
null
{
  "Prenom": "Imene",
  "Nom": "Dupont"
}
{
  "Prenom": "Imene",
  "Nom": "Dupont"
}
```

1. Renommer à true

3. relancer le kafka-connect

4. tester avec :

```
{"schema":{"type": "struct", "fields": [ { "type": "string", "optional": "false", "field": "Prenom"},{  
"type": "string", "optional": "false", "field": "Nom" } ], "optional": "false", "name": "users" },  
"payload": {"Prenom": "Imene", "Nom": "OUAHRANI"}}
```

qui est:

```
{  
"schema":  
  {  
    "type": "struct",  
    "fields": [  
      {  
        "type": "string",  
        "optional": "false",  
        "field": "Prenom"  
      },  
      {  
        "type": "string",  
        "optional": "false",  
        "field": "Nom"  
      }  
    ],  
    "optional": "false",  
    "name": "users"  
  },  
  "payload": {"Prenom": "Imene", "Nom": "OUAHRANI"}  
}
```

5. lire sur le fichier resultat-sink.txt : Struct{Prenom=Imene,Nom=OUAHRANI}

```
h1
jjj
hhhhhhhhhhhhhhhhhhhh

llo there

iiiiiiiiiii
last one
le 09 avril 2024
{Prenom=Imene, Nom=Dupont}
null
{Prenom=Imene, Nom=Dupont}
{Prenom=Imene, Nom=Dupont}
Struct{Prenom=Imene,Nom=Dupont}

C:\Windows\System32\cmd.exe - .\kafka-console-producer.bat --topic test-autre --bootstrap-server localhost:9092
Microsoft Windows [version 10.0.19045.4170]
(c) Microsoft Corporation. Tous droits réservés.

C:\kafka\bin\windows>.bin\windows\kafka-console-producer.bat --topic test-autre --bootstrap-server localhost:9092
Le chemin d'accès spécifié est introuvable.

C:\kafka\bin\windows>.kafka-console-producer.bat --topic test-autre --bootstrap-server localhost:9092
>{"Prenom":"Imene", "Nom":"Dupont"}
>
>{"Prenom":"Imene", "Nom":"Dupont"}
>{"Prenom":"Imene", "Nom":"Dupont"}
>{"Prenom":"Imene", "Nom":"Dupont"}
>{"schema":{"type":"struct","fields":[{"type":"string","optional":"false","field":"Prenom"}, {"type":"string","optional":"false","field":"Nom"}]}, "optional":"false", "name":"users","payload":{"Prenom":"Imene", "Nom":"Dupont"}}
>
```

1.modier

name=local-file-sink

connector.class=FileStreamSink

tasks.max=1

file=C:\\kafka\\kafka-connect\\resultat-sink.txt

topics=test-autre

transforms=Udemy

transforms.Udemy.type=org.apache.kafka.connect.transforms.MaskField\$Value

transforms.Udemy.fields=nom

relancer

5. lire sur le fichier resultat-sink.txt : Struct{Prenom=Imene,Nom=}

KAFKA STREAM API

Remarque : la version de kafka installé c'est la 3.7.0 mais la majorité des tutoriels utilise la version 2.x

Basique

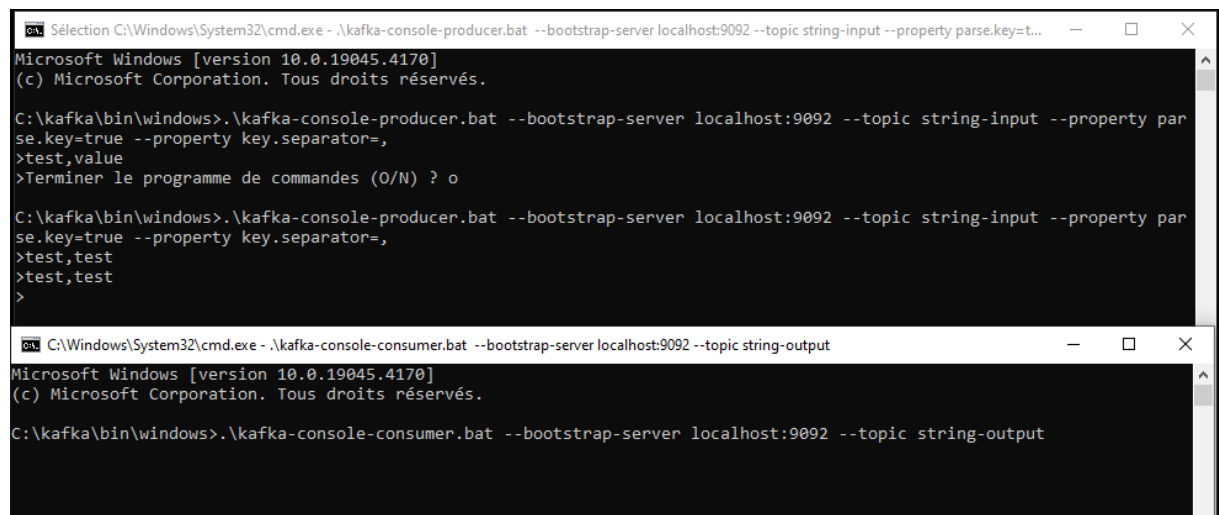
- Séparateur virgule clé,valeur <k,v> : écrire test, valeur sur un producer et lire valeur sur un consumer



The image shows two side-by-side Windows command prompt windows. The top window is titled 'C:\Windows\System32\cmd.exe - .\kafka-console-producer.bat --bootstrap-server localhost:9092 --topic string-input --property parse.key=true --prop...'. It shows the command prompt at 'C:\kafka\bin\windows>' where the command '.\kafka-console-producer.bat --bootstrap-server localhost:9092 --topic string-input --property parse.key=true --property key.separator=,' is entered. The user then enters 'test,value' and presses Enter. The bottom window is titled 'C:\Windows\System32\cmd.exe - .\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic string-output'. It shows the command prompt at 'C:\kafka\bin\windows>' where the command '.\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic string-output' is entered. The output 'value' is displayed on the screen.

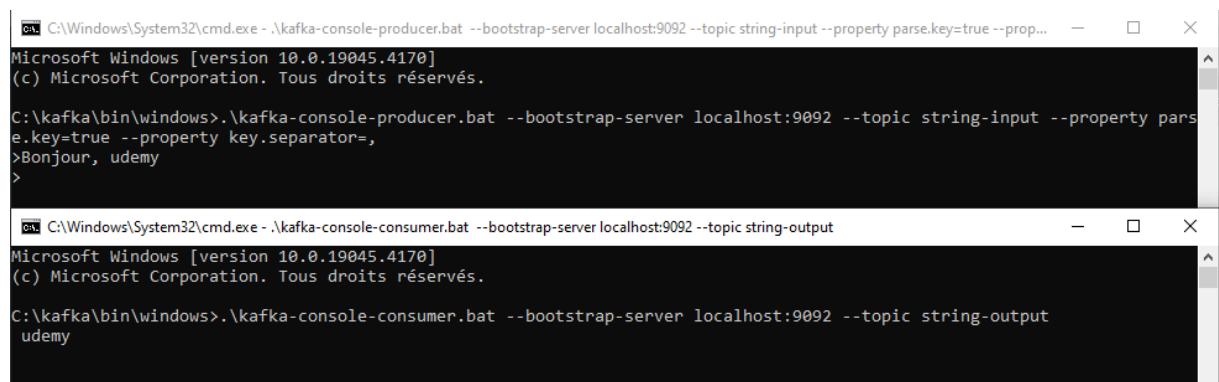
filter

- Ici, on ne peut pas recevoir car la taille est < 5



The image shows two side-by-side Windows command prompt windows. The top window is titled 'C:\Windows\System32\cmd.exe - .\kafka-console-producer.bat --bootstrap-server localhost:9092 --topic string-input --property parse.key=true...'. It shows the command prompt at 'C:\kafka\bin\windows>' where the command '.\kafka-console-producer.bat --bootstrap-server localhost:9092 --topic string-input --property parse.key=true --property key.separator=,' is entered. The user enters 'test,value', then 'Terminer le programme de commandes (O/N) ? o', then 'test,test', and finally 'test,test'. The bottom window is titled 'C:\Windows\System32\cmd.exe - .\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic string-output'. It shows the command prompt at 'C:\kafka\bin\windows>' where the command '.\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic string-output' is entered. No output is visible in the consumer window.

- Ici, on va bien recevoir car la taille > 5



The image shows two side-by-side Windows command prompt windows. The top window is titled 'C:\Windows\System32\cmd.exe - .\kafka-console-producer.bat --bootstrap-server localhost:9092 --topic string-input --property parse.key=true --prop...'. It shows the command prompt at 'C:\kafka\bin\windows>' where the command '.\kafka-console-producer.bat --bootstrap-server localhost:9092 --topic string-input --property parse.key=true --property key.separator=,' is entered. The user enters 'Bonjour, udey' and presses Enter. The bottom window is titled 'C:\Windows\System32\cmd.exe - .\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic string-output'. It shows the command prompt at 'C:\kafka\bin\windows>' where the command '.\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic string-output' is entered. The output 'udey' is displayed on the screen.

mapValue

- Renvoie une valeur en majuscule



```
C:\Windows\System32\cmd.exe - .\kafka-console-producer.bat --bootstrap-server localhost:9092 --topic string-input --property parse.key=true --property key.separator=,
Microsoft Windows [version 10.0.19045.4291]
(c) Microsoft Corporation. Tous droits réservés.

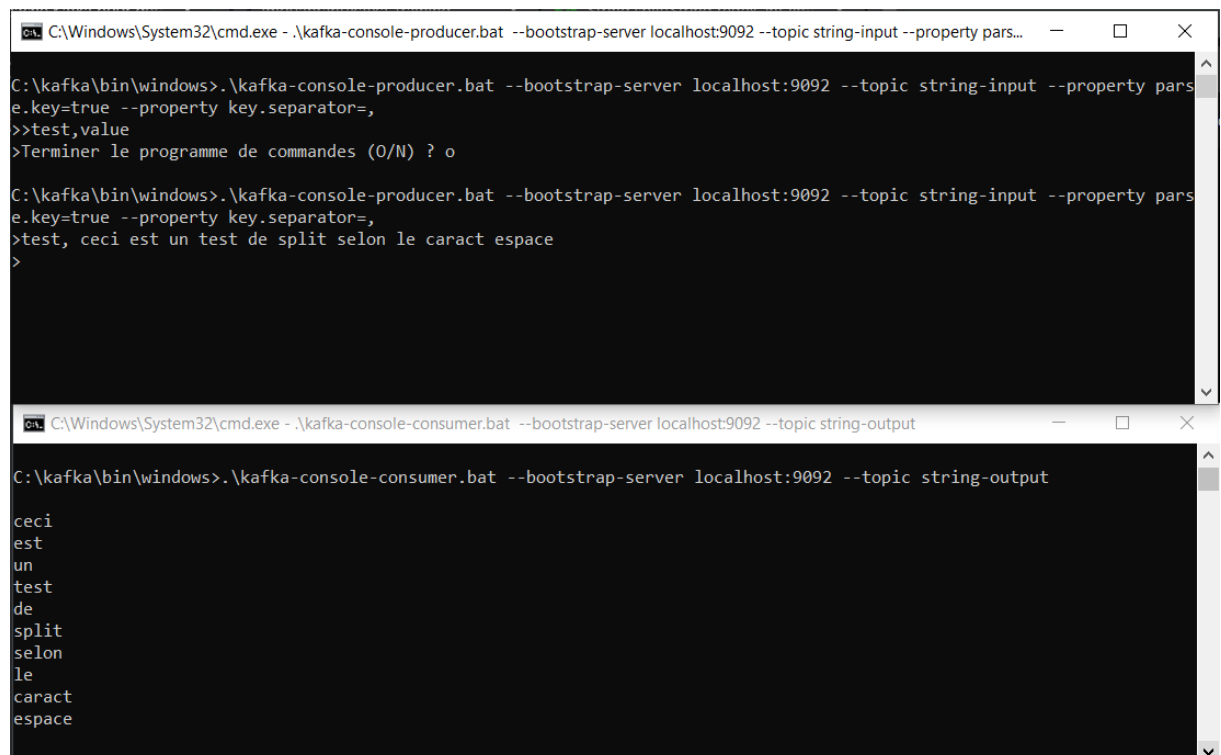
C:\kafka\bin\windows>.\kafka-console-producer.bat --bootstrap-server localhost:9092 --topic string-input --property parse.key=true --property key.separator=,
>>test,value
>

C:\Windows\System32\cmd.exe - .\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic string-output
Microsoft Windows [version 10.0.19045.4291]
(c) Microsoft Corporation. Tous droits réservés.

C:\kafka\bin\windows>.\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic string-output
VALUE
```

flatMap

- Split selon un caractere



```
C:\Windows\System32\cmd.exe - .\kafka-console-producer.bat --bootstrap-server localhost:9092 --topic string-input --property parse.key=true --property key.separator=,
Microsoft Windows [version 10.0.19045.4291]
(c) Microsoft Corporation. Tous droits réservés.

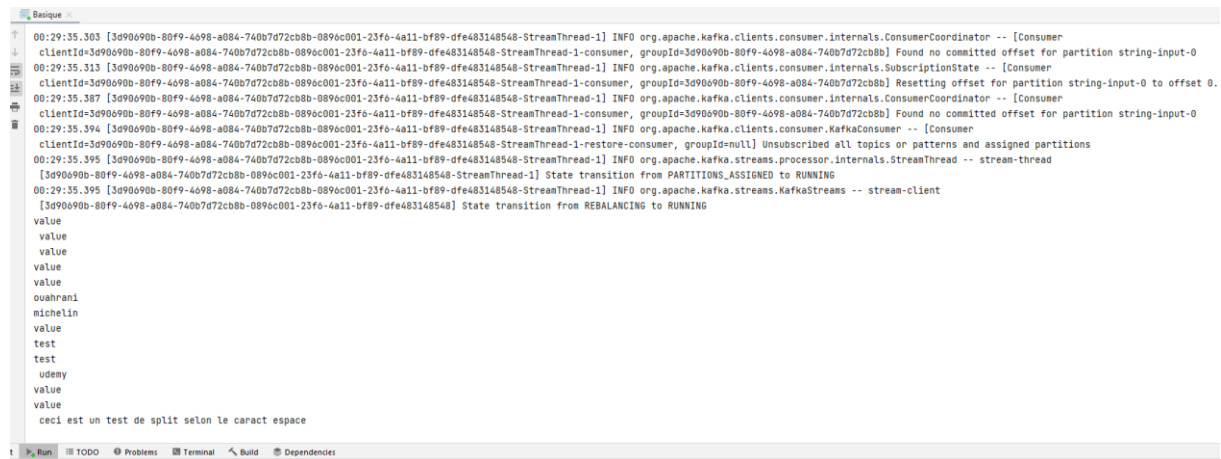
C:\kafka\bin\windows>.\kafka-console-producer.bat --bootstrap-server localhost:9092 --topic string-input --property parse.key=true --property key.separator=,
>>test,value
>Terminer le programme de commandes (O/N) ? o

C:\kafka\bin\windows>.\kafka-console-producer.bat --bootstrap-server localhost:9092 --topic string-input --property parse.key=true --property key.separator=,
>test, ceci est un test de split selon le caract espace
>

C:\Windows\System32\cmd.exe - .\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic string-output
Microsoft Windows [version 10.0.19045.4291]
(c) Microsoft Corporation. Tous droits réservés.

C:\kafka\bin\windows>.\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic string-output
ceci
est
un
test
de
split
selon
le
caract
espace
```

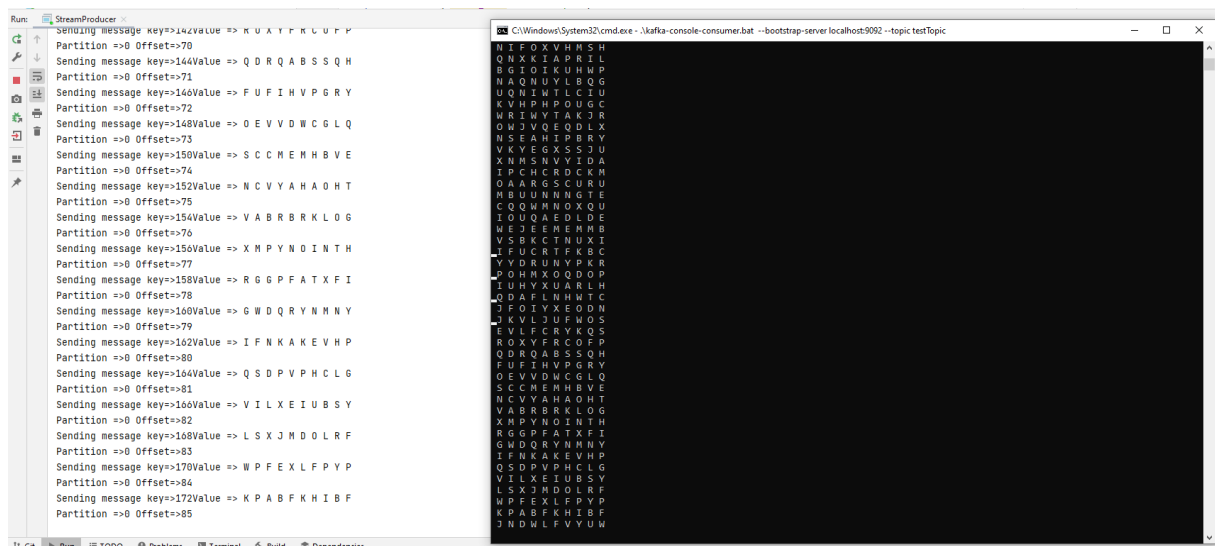

forEach



```
00:29:35.303 [3d90690b-80f9-4698-a084-740b7d72cb8b-0896c001-23f6-4a11-bf89-dfe483148548-StreamThread-1] INFO org.apache.kafka.clients.consumer.internals.ConsumerCoordinator -- [Consumer
clientId=3d90690b-80f9-4698-a084-740b7d72cb8b-0896c001-23f6-4a11-bf89-dfe483148548-StreamThread-1-consumer, groupId=3d90690b-80f9-4698-a084-740b7d72cb8b] Found no committed offset for partition string-input-0
00:29:35.313 [3d90690b-80f9-4698-a084-740b7d72cb8b-0896c001-23f6-4a11-bf89-dfe483148548-StreamThread-1] INFO org.apache.kafka.clients.consumer.internals.SubscriptionState -- [Consumer
clientId=3d90690b-80f9-4698-a084-740b7d72cb8b-0896c001-23f6-4a11-bf89-dfe483148548-StreamThread-1-consumer, groupId=3d90690b-80f9-4698-a084-740b7d72cb8b] Resetting offset for partition string-input-0 to offset 0.
00:29:35.387 [3d90690b-80f9-4698-a084-740b7d72cb8b-0896c001-23f6-4a11-bf89-dfe483148548-StreamThread-1] INFO org.apache.kafka.clients.consumer.internals.ConsumerCoordinator -- [Consumer
clientId=3d90690b-80f9-4698-a084-740b7d72cb8b-0896c001-23f6-4a11-bf89-dfe483148548-StreamThread-1-consumer, groupId=3d90690b-80f9-4698-a084-740b7d72cb8b] Found no committed offset for partition string-input-0
00:29:35.394 [3d90690b-80f9-4698-a084-740b7d72cb8b-0896c001-23f6-4a11-bf89-dfe483148548-StreamThread-1] INFO org.apache.kafka.clients.consumer.KafkaConsumer -- [Consumer
clientId=3d90690b-80f9-4698-a084-740b7d72cb8b-0896c001-23f6-4a11-bf89-dfe483148548-StreamThread-1-restore-consumer, groupId=null] Unsubscribed all topics or patterns and assigned partitions
00:29:35.395 [3d90690b-80f9-4698-a084-740b7d72cb8b-0896c001-23f6-4a11-bf89-dfe483148548-StreamThread-1] INFO org.apache.kafka.streams.processor.internals.StreamThread -- stream-thread
[3d90690b-80f9-4698-a084-740b7d72cb8b-0896c001-23f6-4a11-bf89-dfe483148548-StreamThread-1] State transition from PARTITIONS_ASSIGNED to RUNNING
00:29:35.395 [3d90690b-80f9-4698-a084-740b7d72cb8b-0896c001-23f6-4a11-bf89-dfe483148548-StreamThread-1] INFO org.apache.kafka.streams.KafkaStreams -- stream-client
[3d90690b-80f9-4698-a084-740b7d72cb8b-0896c001-23f6-4a11-bf89-dfe483148548] State transition from REBALANCING to RUNNING

value
value
value
value
value
value
ouahrani
michelin
value
test
test
udemy
value
value
value
ceci est un test de split selon le caract espace
```

- kafka stream producer



```
Run: StreamProducer
Sending message key=>142Value => K X Y F K C U P P
Partition =>0 Offset=>70
Sending message key=>144Value => Q D R Q A B S S Q H
Partition =>0 Offset=>71
Sending message key=>146Value => F U F I H V P 6 R Y
Partition =>0 Offset=>72
Sending message key=>148Value => O E V V D W C G L Q
Partition =>0 Offset=>73
Sending message key=>150Value => S C C M E M H B V E
Partition =>0 Offset=>74
Sending message key=>152Value => N C V Y A H A O H T
Partition =>0 Offset=>75
Sending message key=>154Value => V A B R B R K L O G
Partition =>0 Offset=>76
Sending message key=>156Value => X M P Y N O I N T H
Partition =>0 Offset=>77
Sending message key=>158Value => R G G P F A T X F I
Partition =>0 Offset=>78
Sending message key=>160Value => G W D Q R Y N M N Y
Partition =>0 Offset=>79
Sending message key=>162Value => I F N K A K E V H P
Partition =>0 Offset=>80
Sending message key=>164Value => Q S D P V P H C L G
Partition =>0 Offset=>81
Sending message key=>166Value => V I L X E I U B S Y
Partition =>0 Offset=>82
Sending message key=>168Value => L S X J M D O L R F
Partition =>0 Offset=>83
Sending message key=>170Value => W P F E X L F P Y P
Partition =>0 Offset=>84
Sending message key=>172Value => K P A B F K H I B F
Partition =>0 Offset=>85

C:\Windows\System32\cmd.exe - .\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic testTopic
N I F O X X V R H S H
Q N X X T A P R T I L
B G I O I K U H W P
N A Q N U Y L B Q G
U Q N Z M T L C I U
K V H P H P O U G C
W R I W Y T A K J R
O M J V O E Q D L X
U S E S H I P B R Y
V K Y E G X S S J U
X N M S N V V I D A
I P C H C R O C K N
O A A R G S C U R U
M B U U N N N G T E
C Q Q W M N D X Q U
I O U Q A E D L D E
W E J E E M E M M B
V S B K C T N U X I
I F U C R T F K B C
Y Y D R U N Y P K R
P O H N X Q Q D O P
I U H V X U A R L H
Q D A F L N H W T C
J F O I Y X E O D N
K V L J U F W O S
I V L F C R V K Q S
R O X V F R C O F P
Q D R Q A B S S Q H
F U F I H V P 6 R Y
O E V V D W C G L Q
S C C M E M H B V E
N C V Y A H A O H T
V A B R B R K L O G
X M P Y N O I N T H
R G G P F A T X F I
G W D Q R Y N M N Y
I F N K A K E V H P
Q S D P V P H C L G
V I L X E I U B S Y
L S X J M D O L R F
W P F E X L F P Y P
K P A B F K H I B F
J N D M L F V Y U H
```

Peek = foreach mais avec plus d'opérations

```
Run: Peek x Basique x
C:\Windows\System32\cmd.exe - .\bin\windows\kafka-console-producer.bat --topic string-input --bootstrap-server localhost:9092
Microsoft Windows [version 10.0.19045.4291]
(c) Microsoft Corporation. Tous droits réservés.

C:\kafka>.\bin\windows\kafka-console-producer.bat --topic string-input --bootstrap-server localhost:9092
>
>test
>de
>message
>

key null : value test
key null : value de
key null : value message
```

Word acount : compter combien de fois un mot a été écrit sur le producer

```
C:\Windows\System32\cmd.exe - .\kafka-console-producer.bat --bootstrap-server localhost:9092 --topic string-input
Microsoft Windows [version 10.0.19045.4291]
(c) Microsoft Corporation. Tous droits réservés.

C:\kafka\bin\windows>.\kafka-console-producer.bat --bootstrap-server localhost:9092 --topic string-input
>10
>50
>UU
>Terminer le programme de commandes (O/N) ?
^C
C:\kafka\bin\windows>.\kafka-console-producer.bat --bootstrap-server localhost:9092 --topic string-input
>hello kafka stream
>imene kafka stream level awesome
>ji ji
>hello kafka !!
>imene imene
>hi
>hi hi
>

C:\kafka\bin\windows>.\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic string-output --property print.key=true --property print.value=true --property key.deserializer=org.apache.kafka.common.serialization.StringDeserializer --property value.deserializer=org.apache.kafka.common.serialization.LongDeserializer
ji      2
hello   2
kafka   3
imene   4
hi       1
hi       3
```