APACHE KAFKA

By the Apache Software Foundation

Realisé par:

Chrif El Asri Hanane

El Gamous Khalid

APACHE KAFKA

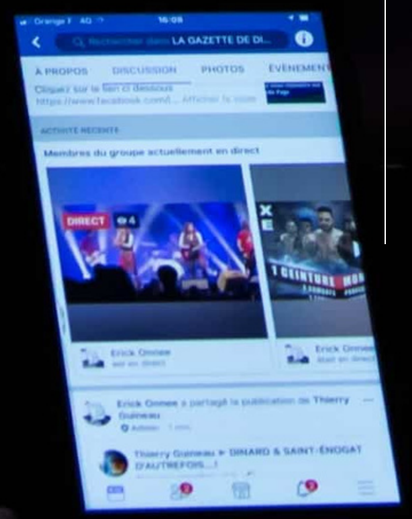
- Apache Kafka est un logiciel de traitement de flux (stream-processing).
- Créé par LinkedIn, actuellement open source et principalement entretenu par Apache Software Foundation.
- Il est écrit en Scala.
- Il sert à collecter, traiter, stocker et analyser des données à grande échelle.

Pourquoi ?

- Kafka est connu pour :
  - Excellentes performances.
  - Une faible latence.
  - La tolérance aux pannes.
  - Un débit élevé.
- Kafka a plus de 1000 cas d'utilisation, certains avantages courants sont :
  - La création de pipelines de données.
  - L'exploitation de flux de données en temps réel.
  - La mise en place de mesures opérationnelles.
  - L'intégration de données à travers d'innombrables sources.
- Kafka est capable de gérer des milliers de messages par seconde.

Le streaming d'événements

- Capturer des données en temps réel à partir de sources d'événements sous la forme de flux d'événements.
- Stocker durablement ces flux d'événements pour une récupération ultérieure.
- Manipuler, traiter et réagir aux flux d'événements en temps réel.
- Acheminer les flux d'événements vers différentes technologies de destination selon les besoins.



Cas d’utilisation de streaming d'événements

- Le traitement des paiements et les transactions financières en temps réel.
- La suivie et la surveillance des voitures, des camions, des flottes et des expéditions en temps réel.
- La capture et l’analyse en continu des données des capteurs à partir d'appareils IoT ou d'autres équipements.
- La surveillance des patients hospitalisés et la prévention des changements de condition.
- La connexion, le stockage et la mise à disposition des données produites par différentes divisions d'une entreprise.
- La création d’un système de messagerie.

CAS D'APPLICATIONS DE KAFKA

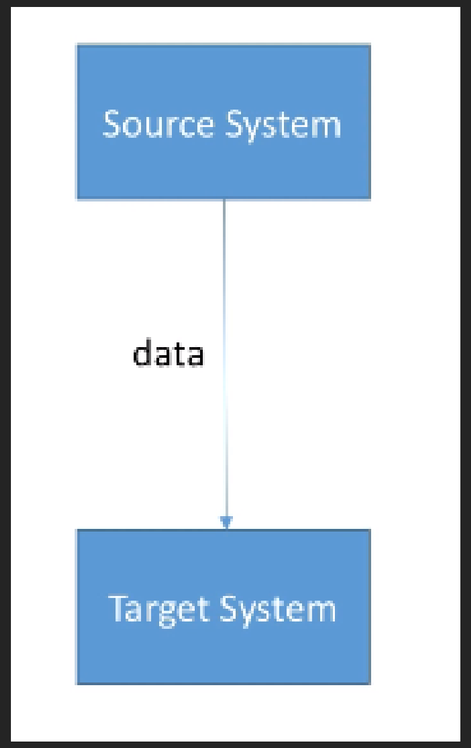
- Netflix utilise Kafka pour appliquer des recommandations en temps réel pendant que vous regardez des émissions de télévision.

- Uber utilise Kafka pour collecter en temps réel des données sur les utilisateurs, les taxis et les trajets pour calculer et prévoir la demande, et calculer la tarification des surtensions en temps réel.

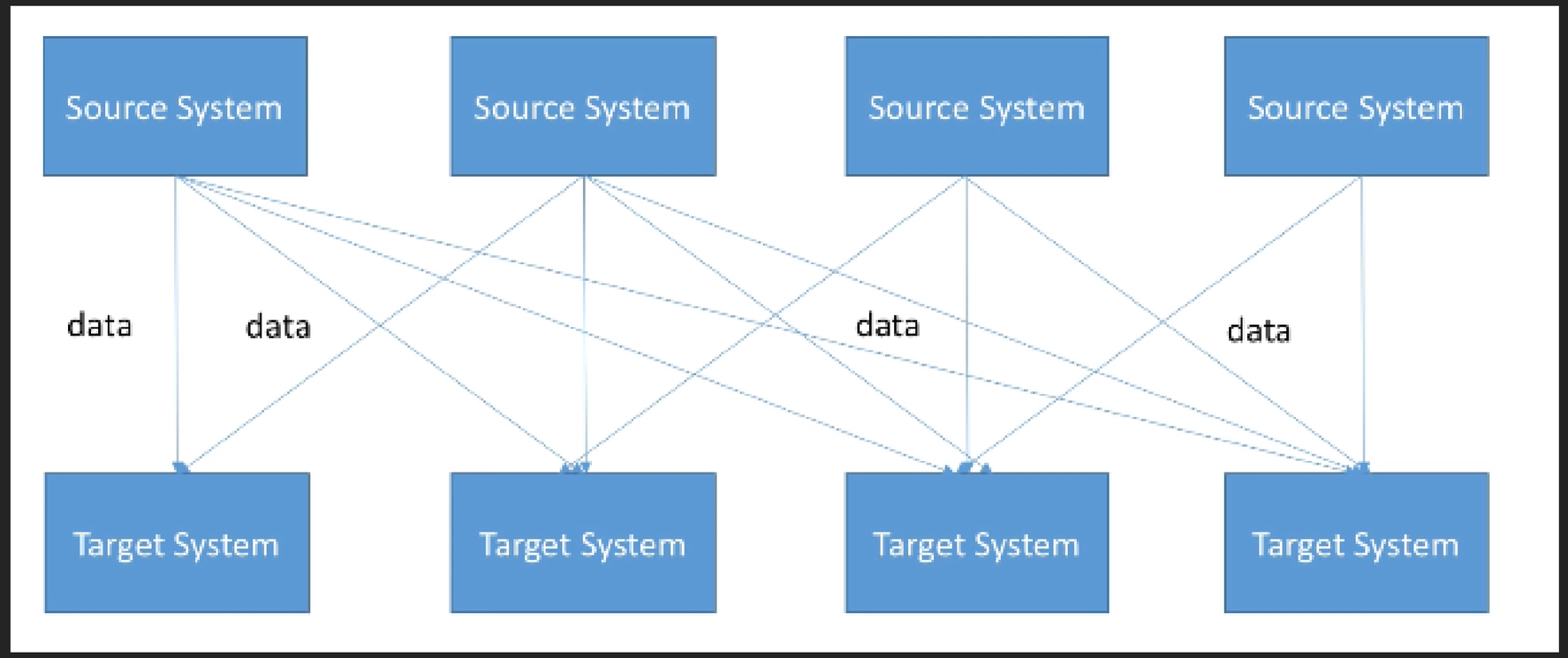
- Linkedin utilise Kafka pour empêcher le spam, collecter les interactions des utilisateurs pour faire de meilleures recommandations de connexion en temps réel.

FONCTIONNEMENT DE KAFKA

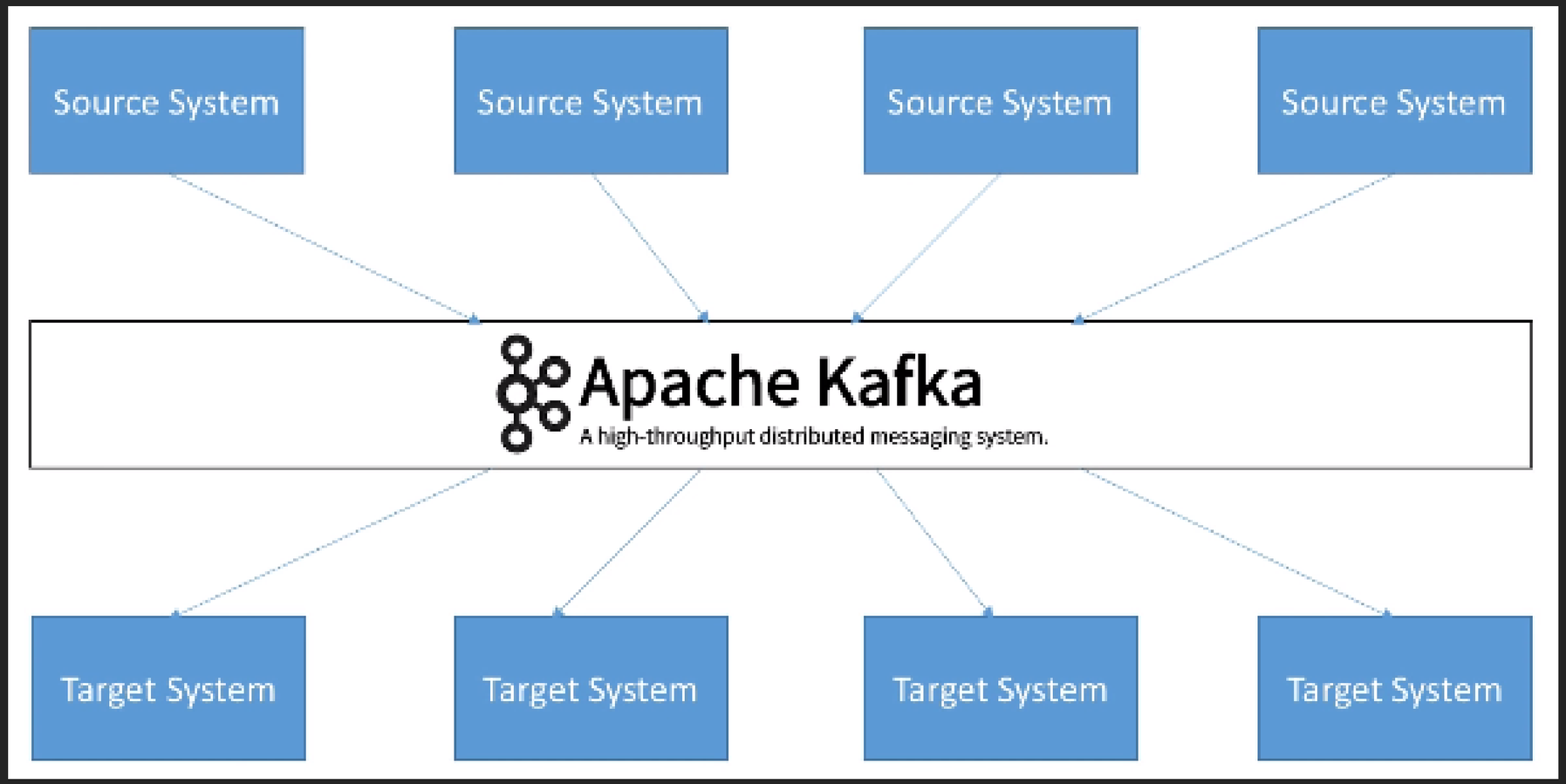
- Système distribué composé de serveurs et de clients.
- Communication via un protocole réseau TCP haute performance.
- Il peut être déployé sur du matériel nu, des machines virtuelles et des conteneurs dans des environnements sur site ou cloud.



Simple at first

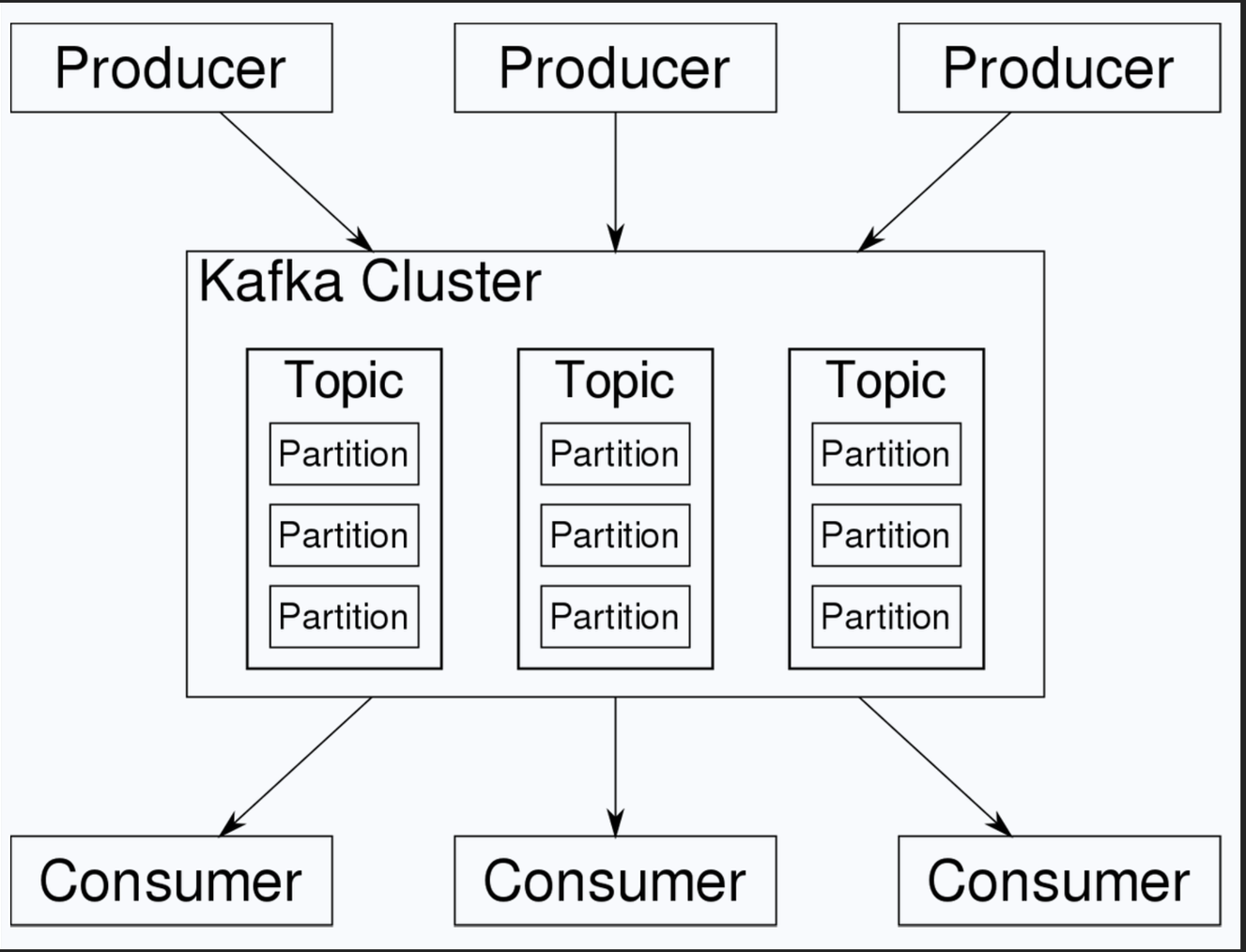


It becomes complicated !



Apache Kafka solution ?

- **Les producteurs (Producers)**
- **Les consommateurs (Consumers)**
- **Un événement (event)**
  Exemple:
  - Clé d'événement: «Alice»
  - Valeur de l'événement: «A effectué un paiement de 200 $ à Bob»
  - Horodatage de l'événement: «25 juin 2020 à 14 h 06».
- **Un sujet (topic)**
- **Partition**



Fonctionnement de Kafka

- **Les producteurs (Producers) :** Ceux qui publient (écrivent) des événements dans Kafka.
- **Les consommateurs (Consumers) :** Ceux qui s'abonnent à (lisent et traitent) ces événements.
- **Un événement (event) :** le fait que "quelque chose s'est passé" dans le monde ou dans votre entreprise. Exemple:
  - Clé d'événement: «Alice»
  - Valeur de l'événement: «A effectué un paiement de 200 $ à Bob»
  - Horodatage de l'événement: «25 juin 2020 à 14 h 06».
- **Un sujet (topic) :** similaire à un dossier contenant des fichiers représentés par des événements, Les sujets sont partitionnés, ce qui signifie qu'un sujet est réparti sur un certain nombre de «compartiments» situés sur différents Kafka brokers.
- Les sujets sont **partitionnés**, ce qui signifie qu'un sujet est réparti sur un certain nombre de «compartiments» situés sur différents Kafka brokers. Ce placement distribué de vos données est très important pour l'évolutivité car il permet aux applications clientes de lire et d'écrire les données de et vers de nombreux brokers en même temps.

Kafka APIs

- **Producer API :** permet aux applications d'envoyer des flux de données.
- **Consumer API :** permet aux applications de lire des flux de données.
- **Streams API :** permet de transformer des flux de données en topic de sortie.
- **Connect API :** permet d'implémenter des connecteurs qui récupèrent les données d'un système source ou d'une application vers Kafka ou qui poussent de Kafka vers une application.
- **AdminClient API :** permet de gérer et d'inspecter les topics, les brokers, et les autres objets Kafka .

Quelque pratiques ?

On y va

```
 ystem32\WindowsPowerShell\v1.0\;C:\Windows\System32\OpenSSH\;C
:\cmd;C:\ProgramData\ComposerSetup\bin;C:\wamp\bin\php\php7.4.8;C:
\Microsoft\WindowsApps;C:\Development\flutter\bin;C:\Users\Etudian
s\Etudiant\AppData\Local\GitHubDesktop\bin;C:\Users\Etudiant\AppDa
oaming\Composer\vendor\bin;C:\laragon\bin\;C:\laragon\bin\php\php7
Data\Roaming\npm;C:\Program Files\Git\usr\bin;C:\Program Files\Git
che.zookeeper.server.ZooKeeperServer)
[2021-02-15 23:30:47,799] INFO Server environment:java.io.tmpdir=C
eper.server.ZooKeeperServer)
[2021-02-15 23:30:47,799] INFO Server environment:java.compiler=<N
[2021-02-15 23:30:47,800] INFO Server environment:os.name=Windows
[2021-02-15 23:30:47,810] INFO Server environment:os.arch=amd64 (o
[2021-02-15 23:30:47,811] INFO Server environment:os.version=10.0
[2021-02-15 23:30:47,812] INFO Server environment:user.name=Etudia
[2021-02-15 23:30:47,813] INFO Server environment:user.home=C:\Use
er)
[2021-02-15 23:30:47,814] INFO Server environment:user.dir=C:\kafk
[2021-02-15 23:30:47,815] INFO Server environment:os.memory.free=4
[2021-02-15 23:30:47,816] INFO Server environment:os.memory.max=51
[2021-02-15 23:30:47,817] INFO Server environment:os.memory.total=
[2021-02-15 23:30:47,828] INFO minSessionTimeout set to 6000 (org.
[2021-02-15 23:30:47,829] INFO maxSessionTimeout set to 60000 (org
[2021-02-15 23:30:47,830] INFO Created server with tickTime 3000 m
\tmp\zookeeper\version-2 snapdir \tmp\zookeeper\version-2 (org.apa
[2021-02-15 23:30:47,965] INFO Using org.apache.zookeeper.server.N
apache.zookeeper.server.ServerCnxnFactory)
[2021-02-15 23:30:47,986] INFO Configuring NIO connection handler
read(s), 8 worker threads, and 64 kB direct buffers. (org.apache.z
[2021-02-15 23:30:47,997] INFO binding to port 0.0.0.0/0.0.0.0:218
[2021-02-15 23:30:48,073] INFO zookeeper.snapshotSizeFactor = 0.33
[2021-02-15 23:30:48,087] INFO Reading snapshot \tmp\zookeeper\ver
ence.FileSnap)
[2021-02-15 23:30:48,220] INFO Snapshotting: 0xb9 to \tmp\zookeepe
```

kafka - "C:\kafka" - .\bin\windows\kafka-console-producer.bat --broker-list localhost:9092 --topic Messaging

```
C:\kafka
λ .\bin\windows\kafka-console-producer.bat --broker-list localhost:9092 --topic Messaging
>Hello World
>
```

kafka - "C:\kafka" - .\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic Messaging --from-beginning

```
C:\kafka
λ .\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic Messaging --from-beginning
Hello World
```

Launch Zookeper → Launch Kafka → Create Producer & Consumer

Merci Pour votre attention

On va partager avec vous un fichier avec les étapes détaillées