

# Internet of Things (IoT)

Traitement de flux

Mickaël Bettinelli  
([mickael.bettinelli@univ-smb.fr](mailto:mickael.bettinelli@univ-smb.fr))



# Objectifs

Présentation générale des principaux frameworks pour le traitement des flux de données.

Objectifs :

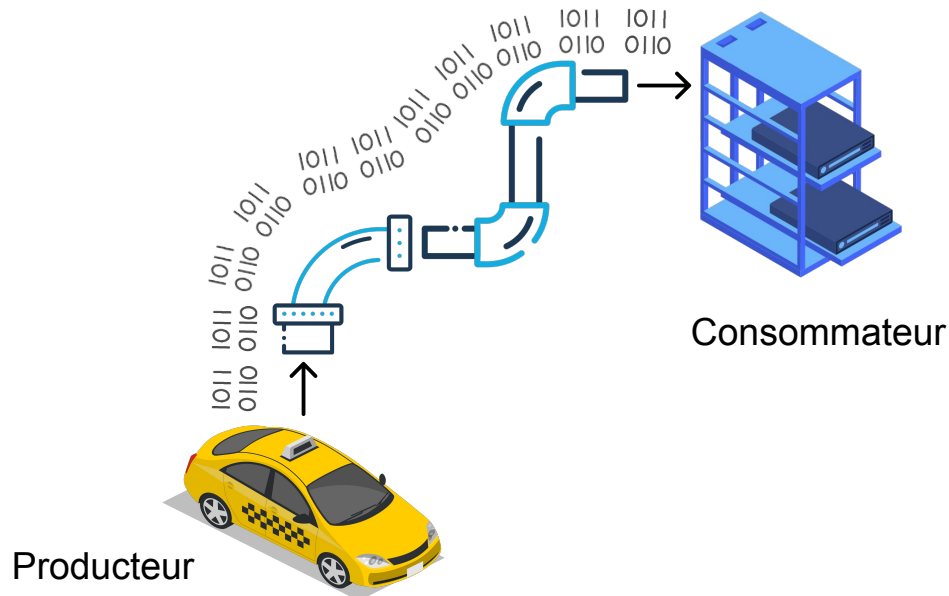
- Comprendre leur intérêt et fonctionnement
- Savoir quand les utiliser



# Les flux de données

Qu'est-ce que c'est ?

- Un flux continu de données entre un producteur et un consommateur

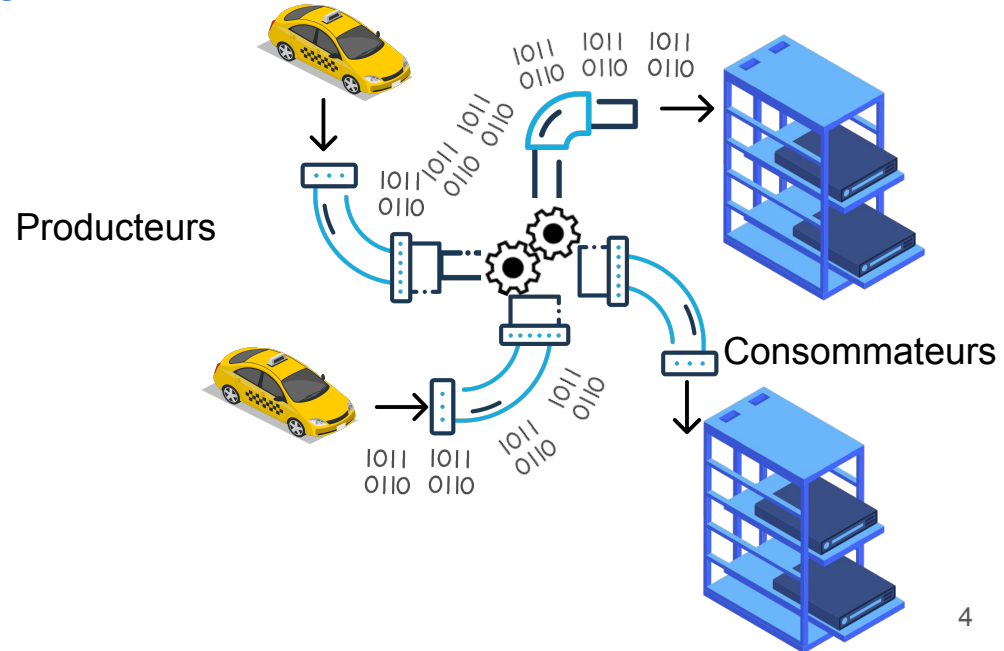




# Le traitement de flux de données

Comment ?

- Traitement temps réel des données
- Analyse de gros volumes de données
- Traitement distribué de données

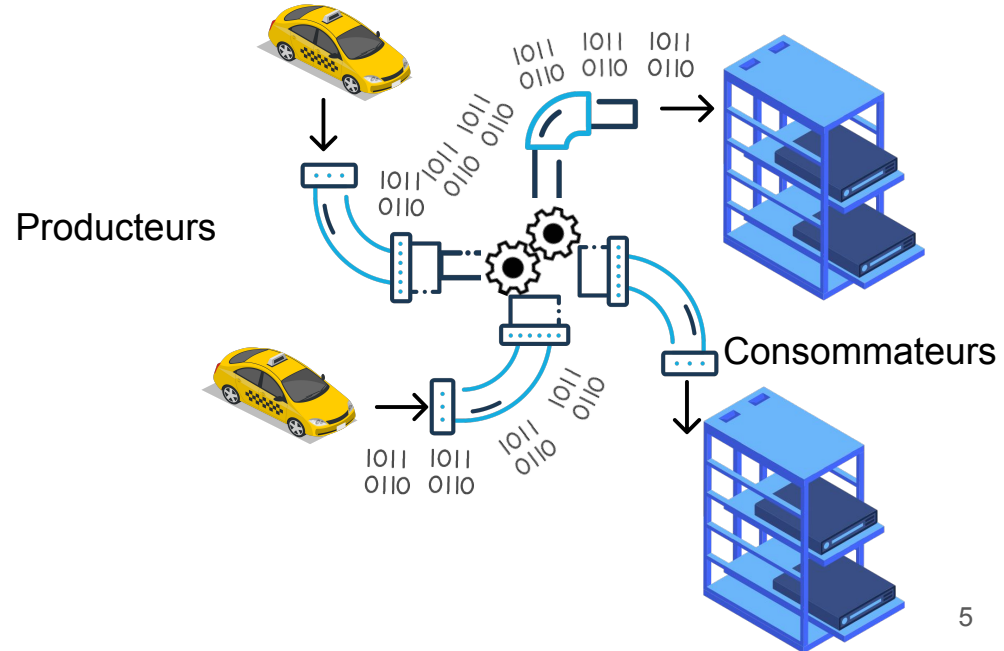




# Le traitement de flux de données

Quelques frameworks pour le traitement de flux :

- Kafka
- Hadoop MapReduce
- Apache Spark





# La communication

Les problématiques :

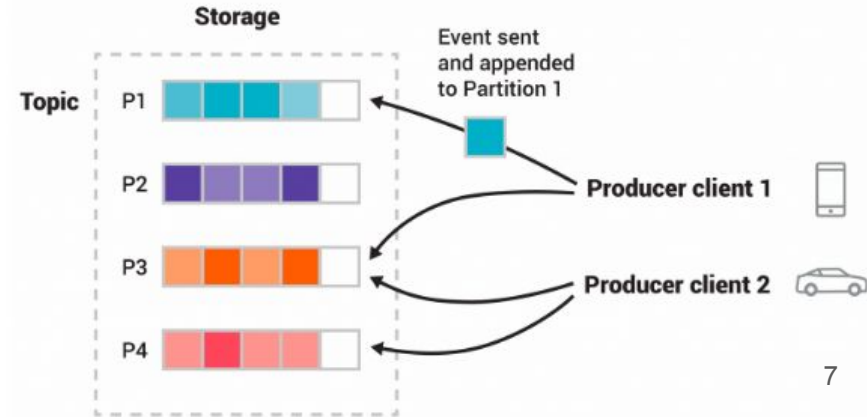
- Communiquer en temps réel dans des systèmes distribués
  - Exemple : problématiques industrielles
- Communiquer dans un système avec de nombreux noeuds
- Gérer un volume de données important



# La communication : Apache Kafka

Vue d'ensemble :

- Créer par LinkedIn en 2011 pour gérer les messages sur leur plateforme
- Fonctionne par événement
- Capable de gérer des millions d'événements par minute
- Framework temps réel / faible latence





# La communication : Apache Kafka

Les API de Kafka :

- Producer : pour envoyer des données
- Consumer : pour recevoir les données
- Streams : pour créer des pipelines de transformations des données
- Connect : pour connecter Kafka à des sources de données ou des BDD

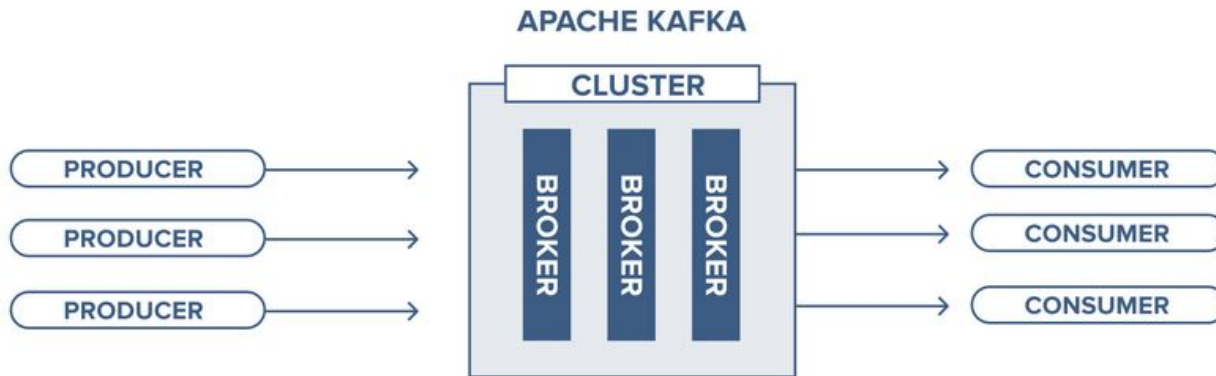




# La communication : Apache Kafka

Les concepts de Kafka :

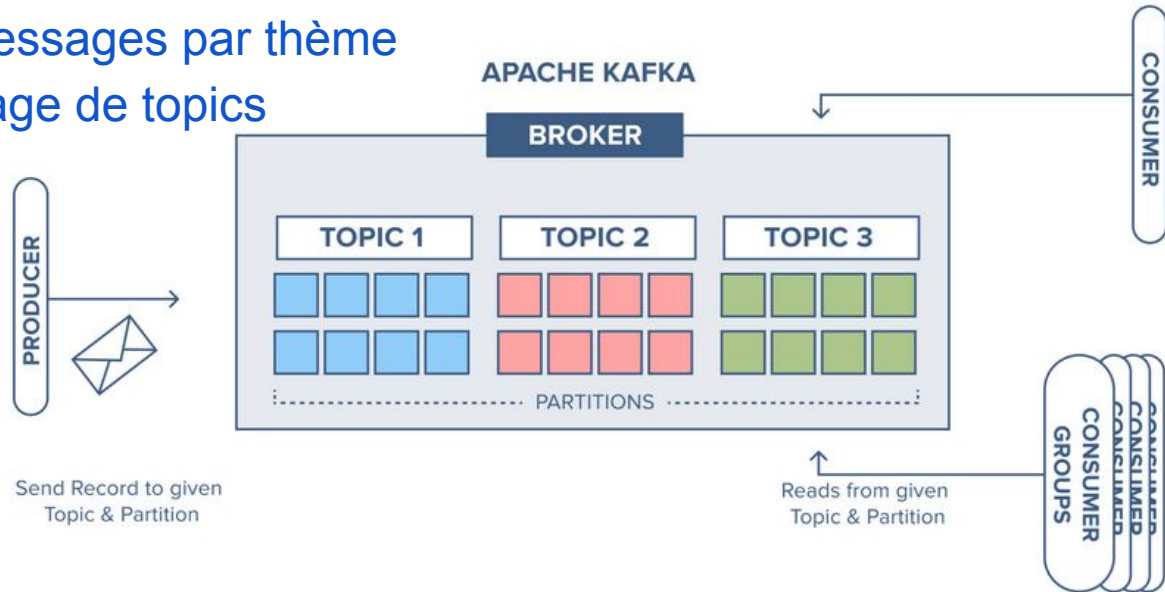
- Broker : traite et distribue les messages
  - 1 cluster = N brokers
- Topics : pour trier les messages par thème
- Partitions : pour stocker les messages dans chaque broker



# La communication : Apache Kafka

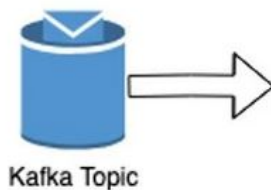
Les concepts de Kafka :

- Broker : traite et distribue les messages
  - 1 cluster = N brokers
- Topics : pour trier les messages par thème
- Partition : pour le stockage de topics





# Apache Kafka



Kafka Broker 0

Partition 0

0 3 6 9 12 15

Partition 1 Replica

1 4 7 10 13 16

Partition 2 Replica

2 5 8 11 14 17

Kafka Broker 1

Partition 1

1 4 7 10 13 16

Partition 0 Replica

0 3 6 9 12 15

Partition 2 Replica

2 5 8 11 14 17

Kafka Broker 2

Partition 2

2 5 8 11 14 17

Partition 0 Replica

0 3 6 9 12 15

Partition 1 Replica

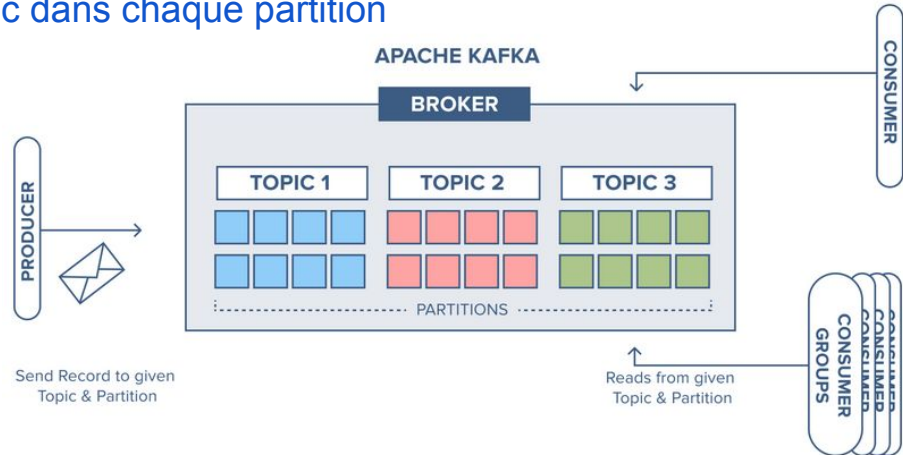
1 4 7 10 13 16



# Apache Kafka

Partitionnement des messages, 3 méthodes :

- Round Robin :
  - Les messages sont rangés de manière à équilibrer les partitions
  - Ne permet pas de garder l'ordre d'arrivé des messages
- Message key partitioning :
  - Les messages sont triés par topic dans chaque partition
  - Maintien l'ordre des messages
- Personnalisée !





# Affichage des données

## Grafana : InfluxDB : séries temporelles

