



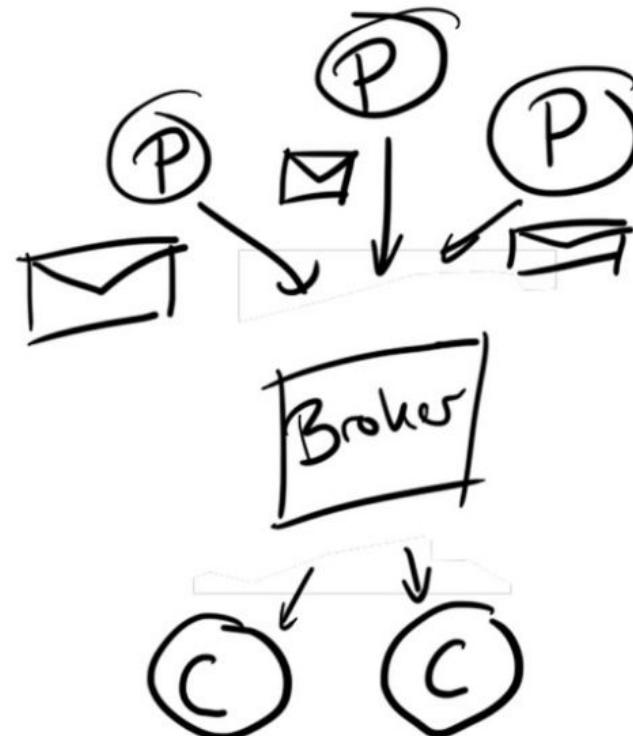
A distributed streaming platform

Timothée Depriester
VT 2018 - RICM5

Qu'est-ce que c'est ?

Une plateforme de streaming distribuée

- sauvegarder les messages
- résister aux pannes
- traiter les données



Introduction

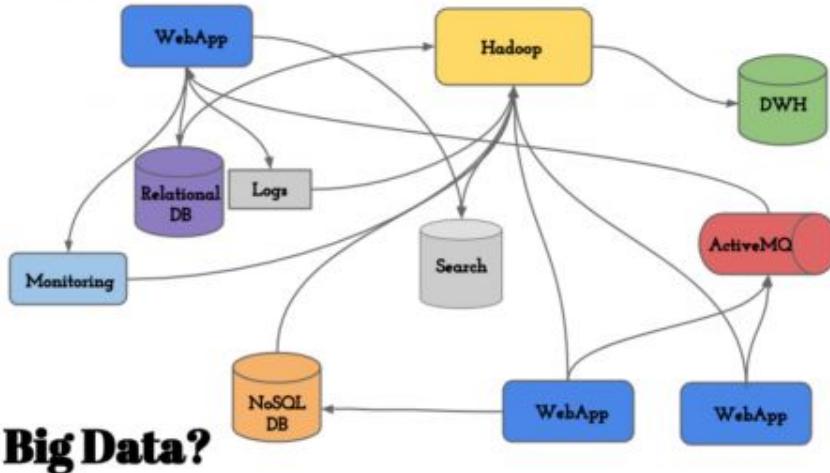
Facilite l'échange de données

Facilite l'évolution du système

Besoin spécialisé



Pourquoi ?



Pour qui ?



NETFLIX

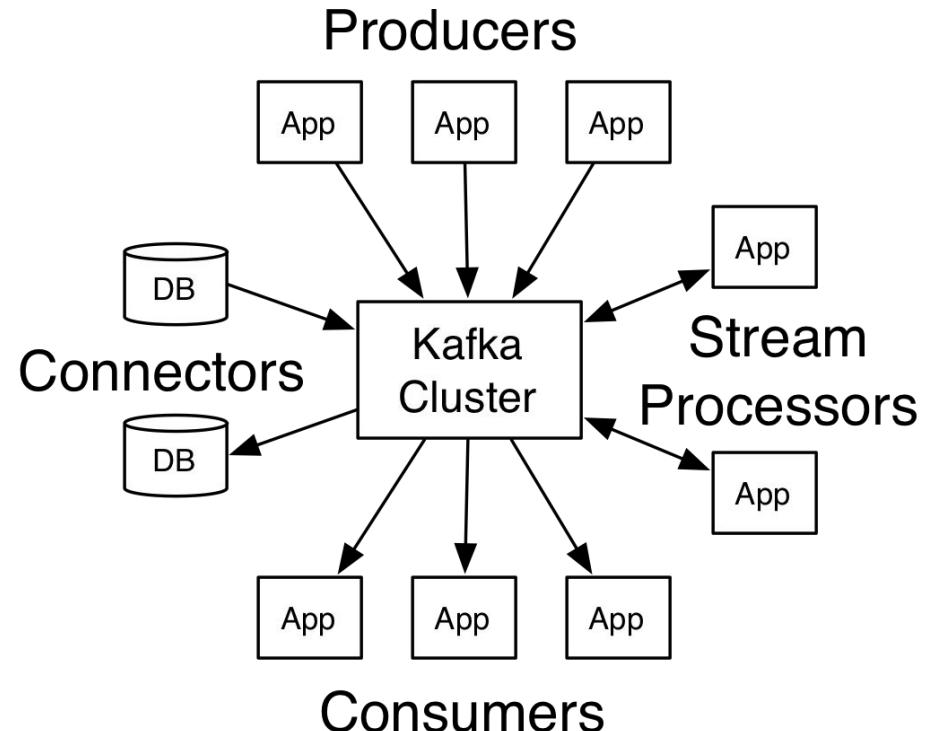


Et des milliers d'autres ...

Comment ça marche ?

4 APIs :

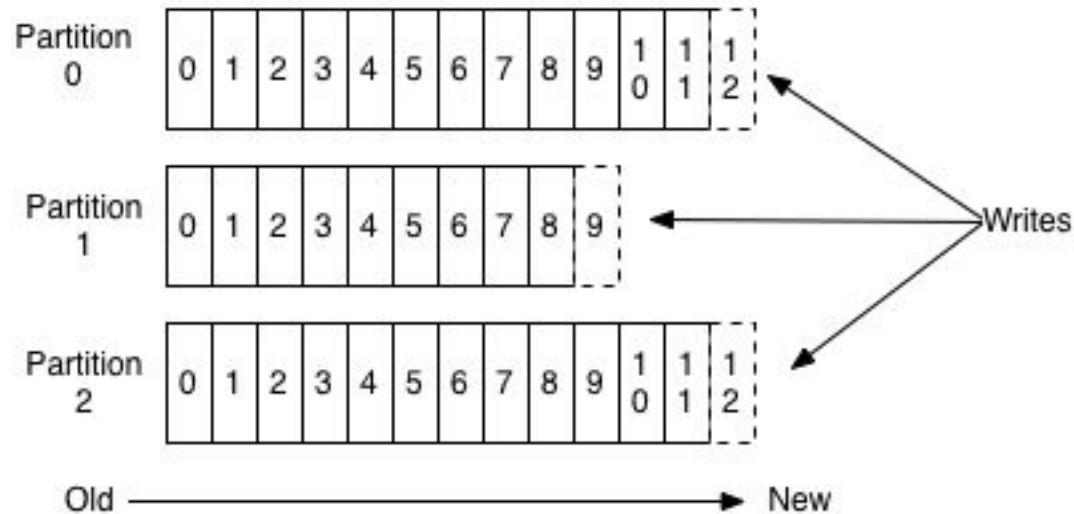
- [Producer API](#)
- [Consumer API](#)
- [Streams API](#)
- [Connector API](#)



Système de topic

- Distribution
- Réplication
- Taille & persistence
- MétaDonnées (offset, temps, clé, etc.)

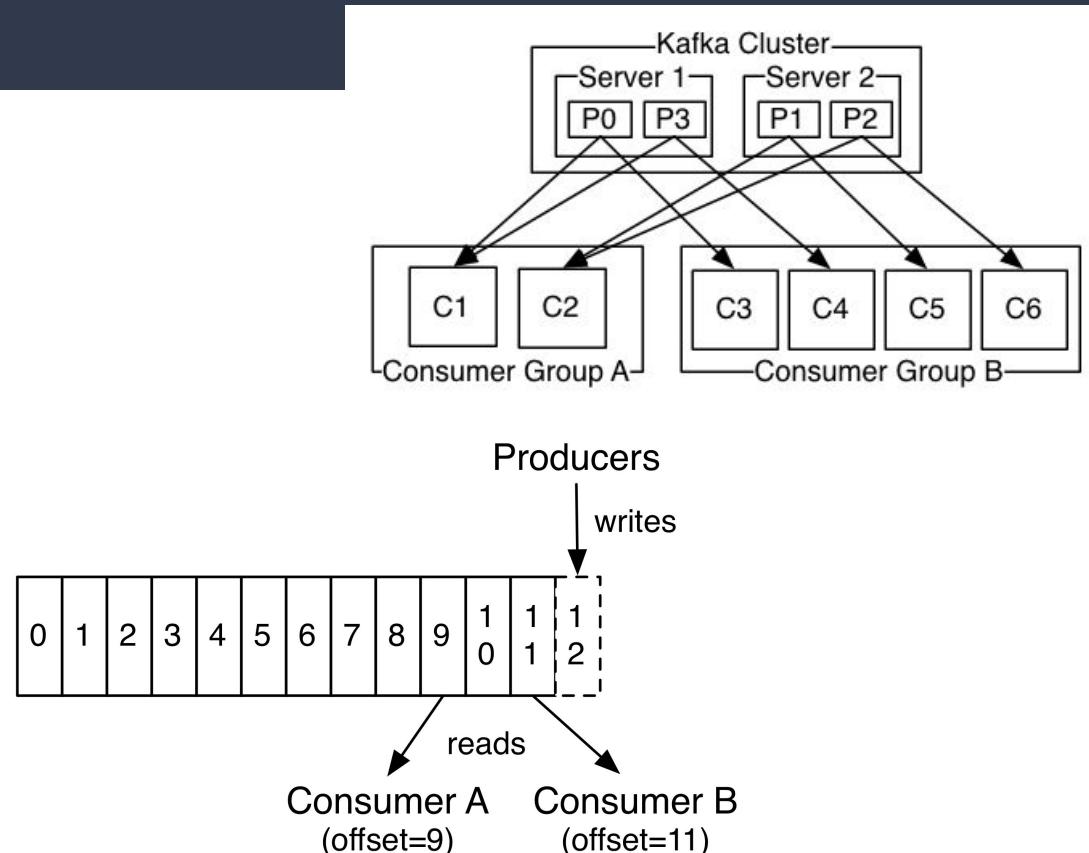
Anatomy of a Topic



Système de topic

→ Scalabilité horizontale

- Par topic
- Par partition
- Par consumer



Cas d'usage

- Traitement des logs
- Gestion des messages

Démonstration

- <https://cwiki.apache.org/confluence/display/KAFKA/Kafka+papers+and+presentations>
- <https://kafka.apache.org>
- <https://github.com/apache/kafka/blob/2.1/streams/examples/src/main/java/org/apache/kafka/streams/examples/wordcount/WordCountDemo.java>
- <http://confluent.io/product/>
- <https://hackernoon.com/distributed-log-analytics-using-apache-kafka-kafka-connect-and-fluentd-303330e478af>