Apache Kafka

* Kafka est un **broker de messages** métiers
* Broker signifie ici intermédiaire entre deux services (ou applications)
* Les services utilisent des technologies non homogènes or ils doivent communiquer entre eux
* Cela crée un couplage fort ce qui rend les choses très compliqués
* Kafka est un **bus de message** pour palier à ce problème
* Kafka est un système de messagerie distribuée en mode publish-subscribe



* Le producer (producteur) est un général le système opérationnel qui écrit le message sérialisé dans Kafka
* Le consumer (consommateur de messge) est souvent un système analytics
* Kafka est donc une zone qui stocke les messages sérialisés (sous forme de série binaire)
* Hautement scalable (distribué dans un cluster Hadoop)
* Mais peut être en dehors de Hadoop 🡪 distribution **Confluent**
* Durable car persistant (écriture sur disque + réplication…)
* **Evènementiel** (traite les messages **en temps réel**)
* Supporte une énorme charge en production et consommation
* Durée de persistance paramétrable

**Concepts clés**

* Un ou plusieurs brokers (une instance de Kafka) dans le cluster Kafta géré par le Zookeeper
* Le producer écrit le message sur un broker
* Le consumer lit les messages des brokers
* Mode distribué
* Les messages sont écrits dans des **topics** (lignes de message de même nature ~ notion abstraite assimilable à une table un SGBDR)
* Le message peut être de n’importe quel type (Kafka parle qu’en binaire)
* Les topics sont divisés en partitions
* Les partitions sont répliquées sur plusieurs brokers

**Topic**

* Le temps de rétention est configurable