

Message Broker

Kafka metrics

Elective Master 2

Data Engineer & Data Scientist

Amine Benmansour
Philippe ZOZOR

Intervenant : Alex LIMA

Sommaire

- Introduction
- Sources de données
- Architecture
- Metrics Kafka
- Conclusion



Introduction

Présentation du sujet

- Connexion et streaming de données à partir d'une source (api, base de données...)
 - utilisation de kafka
 - Stockage des données dans une bdd
 - Récupération des Metrics de Kafka
-
- Recherche api et sujet
 - Evaluations de 2 architectures
 - slack ELK
 - Prometheus / Grafana



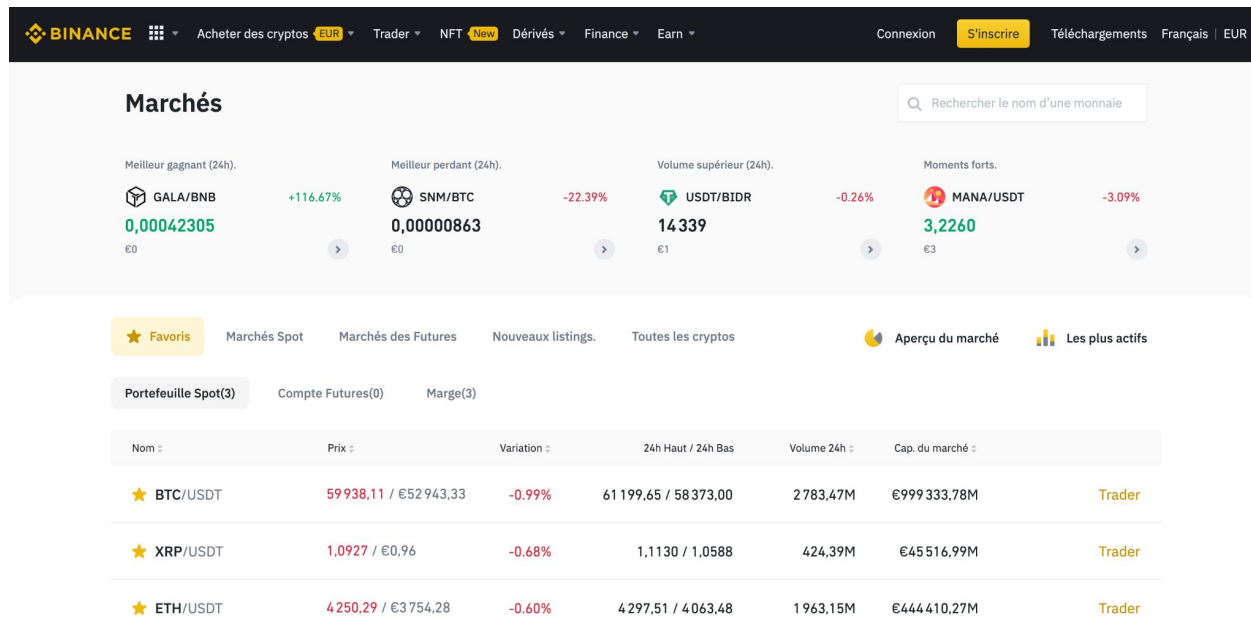


Sources de données



Sources de données

Binance



Plateforme d'échange de crypto monnaies

Sources de données

CryptoPanic

- Outil qui absorbe tout type de données présentent sur la toile internet ayant pour sujet la crypto monnaie et les listes.
- Similaire à Bloomberg pour la bourse classique
- Ces informations sont essentielles pour les professionnels et les amateurs car il recense chaque news pouvant impacter le prix des actifs numériques.
- De plus, les utilisateurs de la plateforme peuvent voter pour marquer leurs ressentiment sur une news : bonne, mauvaise, troll
- Ainsi que le ressentiment sur la tendance de l'actif : importants, haussiers ou baissiers.

The screenshot displays the CryptoPanic website, which serves as a news aggregator for the cryptocurrency market. The interface is divided into several sections:

- Top News:** A list of recent news items with headlines, timestamps, and source links.
- Trending:** A section highlighting popular news stories, such as "Dogecoin Killer? Shiba Inu Integrated by One of Leading Crypto Payment Processors" and "Fidelity Approved to Become Canada's First Bitcoin Custodian".
- Market Data:** A table on the right side showing the prices and percentage changes for various cryptocurrencies, including Bitcoin (BTC), Ethereum (ETH), and others.

[CryptoPanic](https://cryptopanic.com)

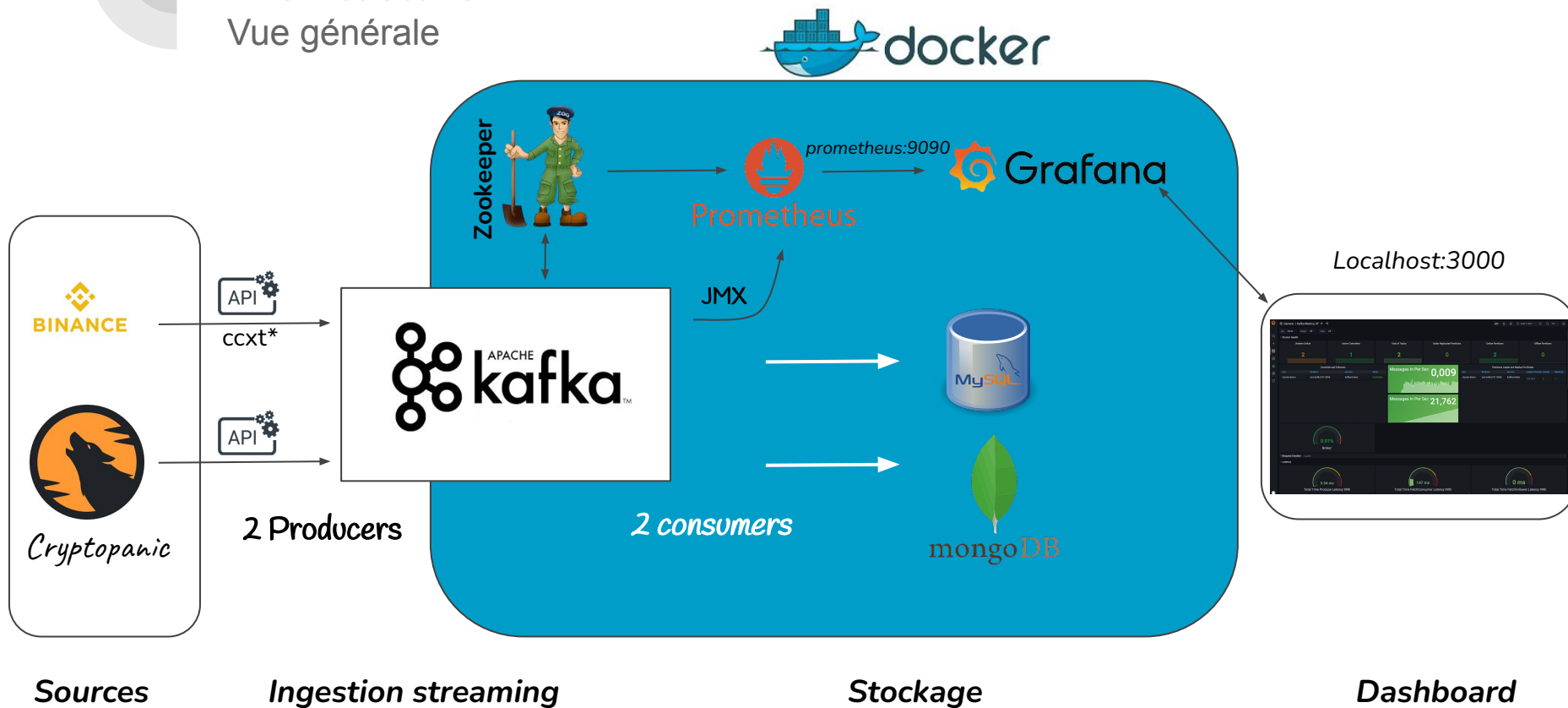


Architecture



Architecture

Vue générale



Sources

Ingestion streaming

Stockage

Dashboard






















*Library Java/ Python / PHP pour le commerce des crypto-monnaies avec un support pour de nombreux marchés d'échange de bitcoin/ether/altcoin et des API de commerçants.



Architecture

Composition du docker

- Kafka
- Zookeeper
- JMX-Kafka Exporter
- Prometheus
- Grafana
- Mongo
- MySql

<input type="checkbox"/>		NAME
<input type="checkbox"/>	▼ 	docker
<input type="checkbox"/>		jmx-kafka101 f1e24f9aa484 
<input type="checkbox"/>		grafana e7b2294c92ab 
<input type="checkbox"/>		kafka101 e55b3a86bd49 
<input type="checkbox"/>		mongo-1 e7eb288e9ab5 
<input type="checkbox"/>		mysql_db 84ee73ea2e47 
<input type="checkbox"/>		zookeeper1 26f7e17f3971 
<input type="checkbox"/>		mysql_adminer 06c9474c7f6c 
<input type="checkbox"/>		jupyter_notebook 742d2868701c 
<input type="checkbox"/>		mongo-express-1 8e80a14ea5ef 
<input type="checkbox"/>		prometheus 72baf642dbf3 



Architecture

Producer CryptoPanic

- Cryptos suivies : **BTC, XRP, ETH**
- Récupération des données “**Hot**” de nos 3 cryptos
- Fréquence de streaming : **1 heure**
- Topic du producer : **CryptoPanic**
- Données recueillies :

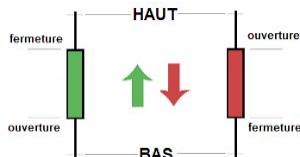
```
{'id': 13384320,
'currencies': [{'code': 'BTC',
'title': 'Bitcoin',
'slug': 'bitcoin',
'url': 'https://cryptopanic.com/news/bitcoin/'},
{'code': 'ETH',
'title': 'Ethereum',
'slug': 'ethereum',
'url': 'https://cryptopanic.com/news/ethereum/'}],
'title': 'Binance Smart Chain sets record after 14.7 million daily
transactions in one day',
'published_at': '2021-11-18T20:00:43Z',
'votes': {'negative': 1,
'positive': 6,
'important': 3,
'liked': 5,
'disliked': 1,
'lol': 0,
'toxic': 0,
'saved': 0,
'comments': 1},
'url':
'https://cryptopanic.com/news/13384320/Binance-Smart-Chain
-sets-record-after-147-million-daily-transactions-in-one-day'}
```



Architecture

Producer Binance

- Cryptos suivies : **BTC, XRP, ETH**
- Récupération des données **ohlcv** de nos cryptos, grâce à la méthode `fetch_ohlcv` de ccxt
- Fréquence de streaming : **1 min**
- Topic du producer : **Binance**



id	BTC	Timestamp	Open	Low	High	Close	Volume
	ETH	Timestamp	Open	Low	High	Close	Volume
	XRP	Timestamp	Open	Low	High	Close	Volume

```
{'id': 0, 'BTC': [[1637235720000, 59591.21, 59601.8, 59515.01, 59523.35, 19.73556]], 'ETH':  
[[1637235720000, 4219.75, 4219.99, 4215.15, 4215.61, 84.5654]], 'XRP': [[1637235720000, 1.093, 1.0931,  
1.0904, 1.0909, 269886.0]]}
```



Architecture

Consumers

- Réception des données des **2 producteurs** (Binance et cryptoPanic)
- Enregistrement des données de CryptoPanic dans MongoDB (3 DataBases):
 - Hot spot CryptoPanic
- Enregistrement des données de Binance dans MySQL :
 - Bitcoin
 - Ethereum
 - Ripple

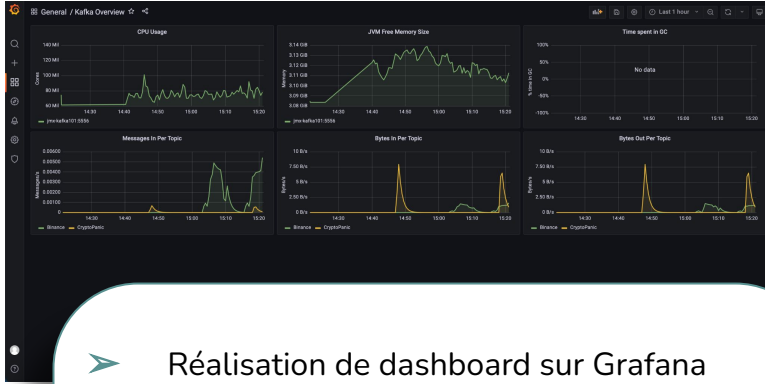
The screenshot shows a database management tool interface. On the left, a sidebar lists the database structure: 'Binance_db' containing 'Tables', 'Bitcoin', 'Ethereum', 'Ripple', 'Views', 'Stored Procedures', 'Functions', and 'sys'. The 'Bitcoin' table is selected. The main area displays a 'Result Grid' with columns: 'datetime', 'Open', 'High', 'Low', 'Close', and 'Volume'. The data represents hourly price movements for Bitcoin, starting from 2022-12-01 09:44:00. The interface includes a search bar, a 'Filter Rows' button, and an 'Export' button.

datetime	Open	High	Low	Close	Volume
2022-12-01 09:44:00	17105.85	17110.91	17102.23	17105.79	144.20953
2022-12-01 09:44:00	17105.88	17110.91	17102.23	17104.66	167.10376
2022-12-01 09:44:00	17105.88	17110.91	17100.91	17101.88	212.30488
2022-12-01 09:45:00	17102.47	17106.7	17101.72	17105.12	29.19088
2022-12-01 09:45:00	17102.47	17106.7	17101.72	17103	94.71203
2022-12-01 09:45:00	17102.47	17106.7	17096.96	17102.36	208.87817
2022-12-01 09:45:00	17102.47	17106.7	17096.96	17103.78	232.21782
2022-12-01 09:46:00	17103.54	17104.33	17101.85	17102.7	9.53902
2022-12-01 09:46:00	17103.54	17110.38	17101.64	17106.67	82.35328
2022-12-01 09:46:00	17103.54	17110.38	17101.64	17106.53	114.44939
2022-12-01 09:46:00	17103.54	17110.38	17101.64	17105.43	143.57557
2022-12-01 09:46:00	17103.54	17110.38	17101.64	17106.02	174.95901
2022-12-01 09:47:00	17107.47	17107.55	17106.32	17107.21	3.80482
2022-12-01 09:47:00	17107.47	17107.62	17105.01	17107.1	23.27415
2022-12-01 09:47:00	17107.47	17107.62	17104.49	17104.88	35.36208
2022-12-01 09:47:00	17107.47	17111.75	17103.73	17109.6	98.42731
2022-12-01 09:47:00	17107.47	17111.75	17103.73	17108.76	119.55288
2022-12-01 09:48:00	17108.95	17111.9	17107.84	17111.18	18.3338
2022-12-01 09:48:00	17108.95	17111.9	17107.84	17109.2	50.13246
2022-12-01 09:48:00	17108.95	17111.9	17107.26	17107.63	77.01573
2022-12-01 09:48:00	17108.95	17111.9	17106.97	17107.78	95.77889

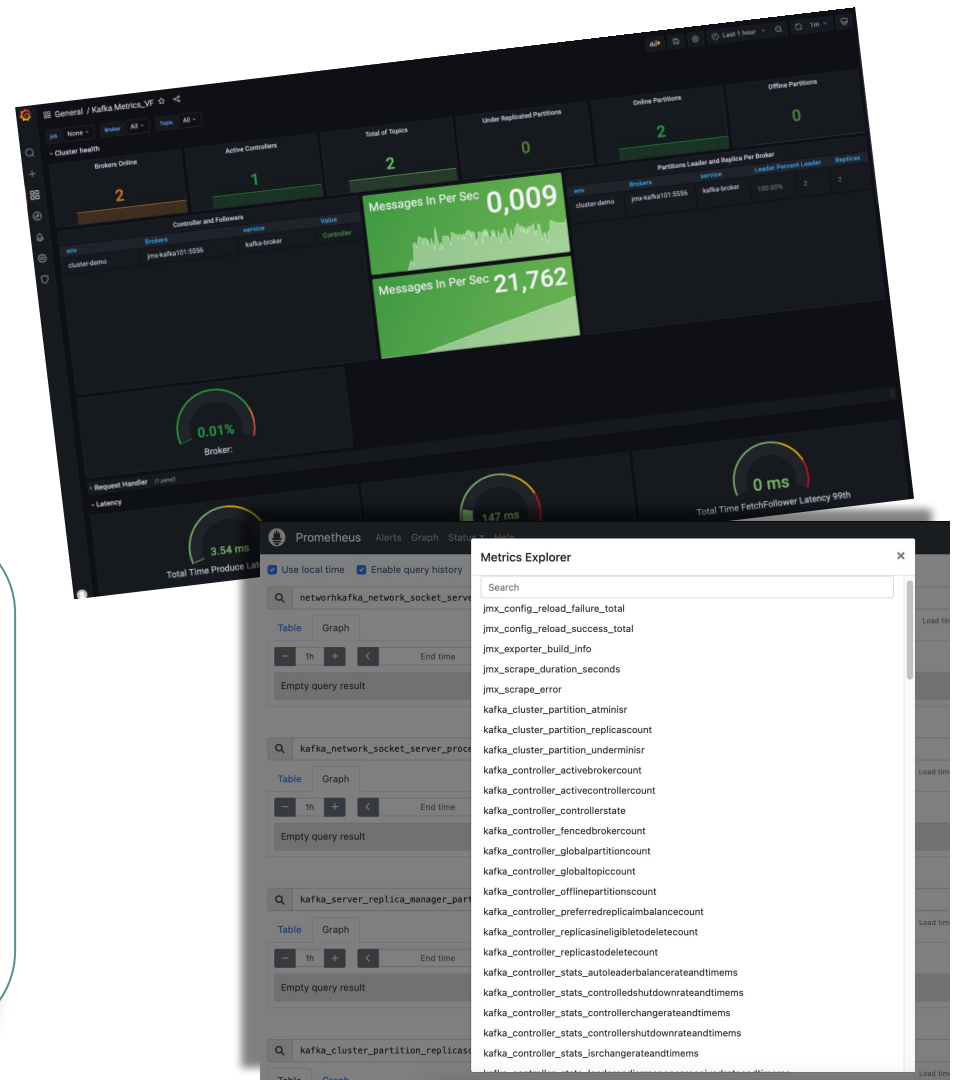


Kafka Metrics

Kafka Metrics



- Réalisation de dashboard sur Grafana
- Utilisation des modèles de Dashboard proposé par Grafana pour le suivi des metrics de Kafka.
Pb de nom des Metrics → pas de remonté des valeurs
- Re-affectation des Metrics pour obtenir un Dashboard fonctionnel





Kafka Metrics

Démonstration



Conclusion & perspectives





Conclusion & Perspectives

- des révisions...
 - Fonctionnement et connection à une API
 - Streaming de données provenant de sources web.
 - Utilisation de MongoDB et MySQL
 - Déploiement d'un environnement complet sur Docker et interaction entre ses différentes composantes

- ... et des apprentissages
 - Stack Kafka, Prometheus, Grafana
 - Réalisation de dashboard sur Grafana
 - Découverte des diverses Metrics disponibles

- Difficultés rencontrées :
 - Choix de l'api / sujet
 - Choix de l'infrastructure à mettre en place
 - Réalisation du dashboard sur Grafana

- Perspectives:
 - Meilleure compréhension des différentes Metrics



Merci pour votre attention

