Documentation Technique

Projet Big Data / Dataviz



Version 1.0 – 11/18/2021

Écris par: Jérémy RICHARD, Clément RAMOS LAGE, Louis ARDILLY

Table of contents

[I. Introduction 3](#_Toc88133163)

[II. Contexte 4](#_Toc88133164)

[III. Solution technique 5](#_Toc88133165)

[IV. Outils utilisés 6](#_Toc88133166)

[V. Données 6](#_Toc88133167)

[VI. Traitement des données 7](#_Toc88133168)

[VII. Interface 8](#_Toc88133169)

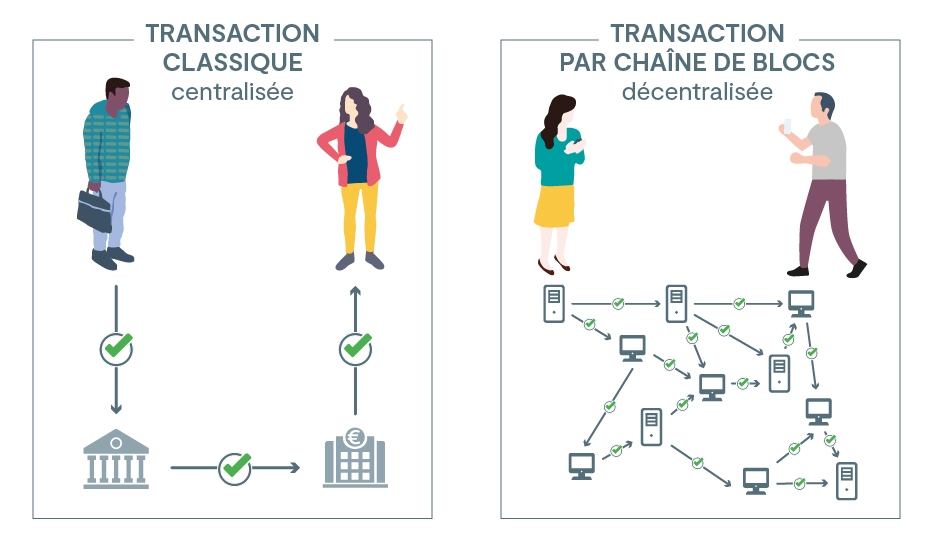
­­­

1. Introduction

Ces dernières années les cryptomonnaies sont en pleine expansion et démocratisation à l’international. Ce sujet aussi passionnant soit-il est complexe et ne doit pas être pris à la légère. Avant toute chose il est important de définir ce qu’est une cryptomonnaie ainsi que le fonctionnement de celle-ci.

Qu’est-ce qu’une cryptomonnaie ?

C'est un actif qui s'échange de pair-à-pair (P2P) sans tiers de confiance comme les banques. Elles n'ont pas de support physique comme des pièces ou des billets, ne sont pas régulées par un organe central et ne sont pas indexées sur le dollar ou l'or par exemple. Ces nouvelles **monnaies électroniques utilisent la technologie blockchain (grand livre de compte) pour transférer la propriété des cryptos entre leurs différents propriétaires**.

Qu’est-ce que la blockchain ?

Une blockchain est un registre, une grande base de données qui a la particularité d’être partagée simultanément avec tous ses utilisateurs, tous également détenteurs de ce registre, et qui ont également tous la capacité d’y inscrire des données, selon des règles spécifiques fixées par un protocole informatique très bien sécurisé grâce à la cryptographie.

Une image contenant texte

Description générée automatiquement

Les principales crypto monnaies :

1. Bitcoin
2. Ethereum
3. Binance Coin
4. Tether
5. Solana
6. Cardano
7. XRP
8. Polkadot
9. USD Coin
10. Dogecoin
11. Contexte

Le sujet étant introduit et les principaux acteurs définis nous pouvons présenter le contexte de notre projet.

Dans le cadre d’un projet de Big-Data et Dataviz nous avons donc réfléchis à la mise en place d’une solution répondant à un problématique liant l’intitulé du projet et le sujet de la crypto monnaie. Suite à une réflexion en groupe nous avons choisi de récupérer de la donnée provenant de plusieurs sources, une source numérique (prix de la crypto, prix le plus haut, prix le plus bas, moyenne etc…) et une source textuelle (news concernant la crypto monnaie).

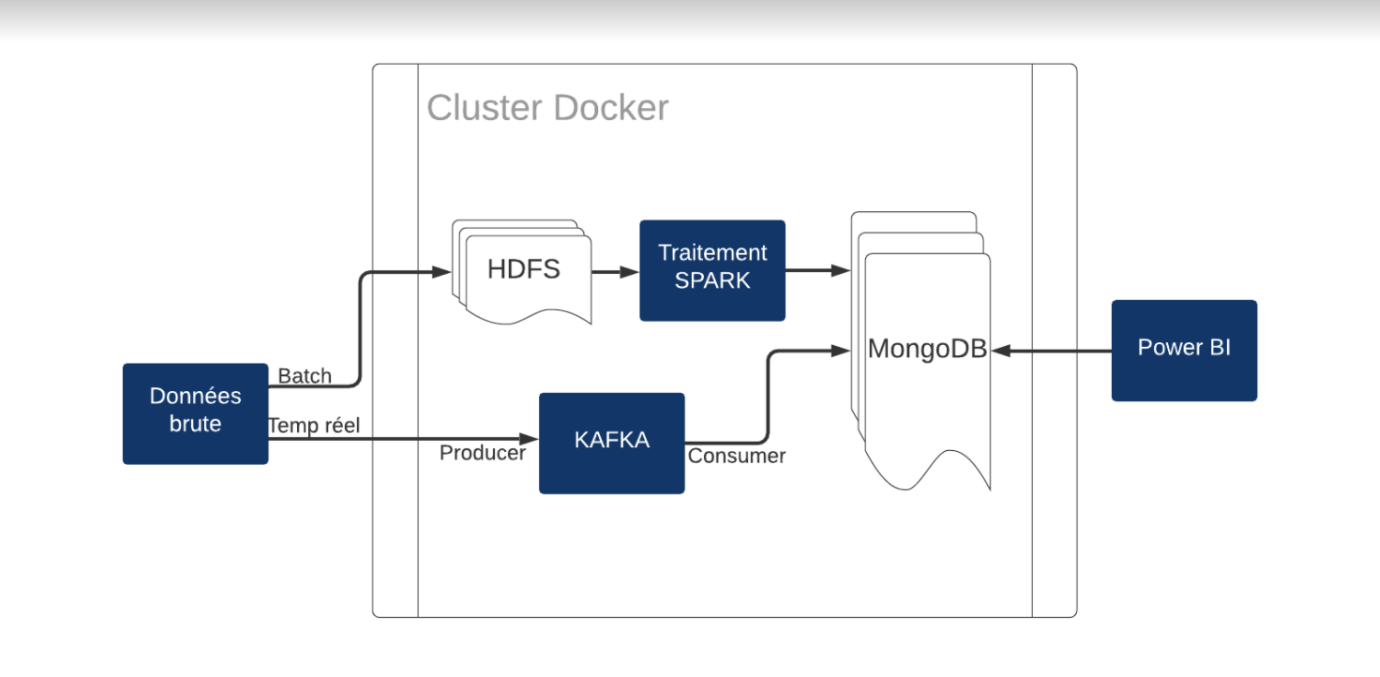
Les deux sources citées précédemment nous permettrons d’effectuer une analyse en temps réel (streaming) de l’évolution des cryptos et de faire un lien entre le nombre de citations dans les news et l’évolution d’une crypto monnaie.

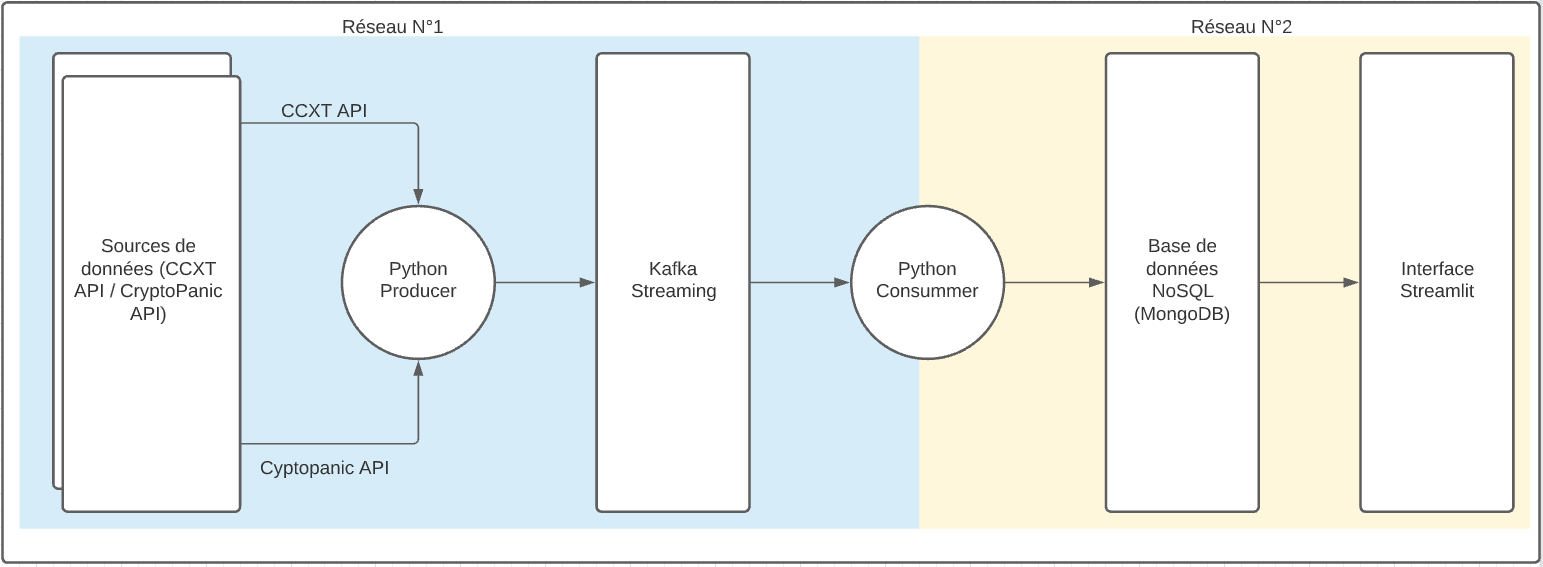
Notre interface permettra de visualiser :

* Le cours d’une crypto monnaie en temps réel
* Une comparaison entre une ou plusieurs cryptomonnaie
* Un wordcloud mettant en avant les cryptos monnaies les plus citées
* Un tableau récapitulatif du nombre de news par crypto ainsi qu’un récapitulatif du nombre de like/dislike.

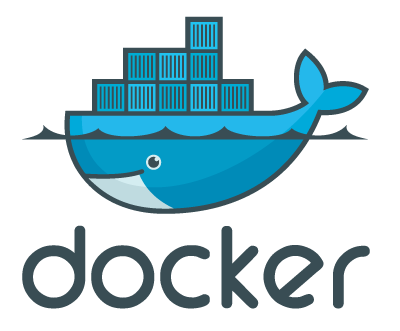
1. Solution technique

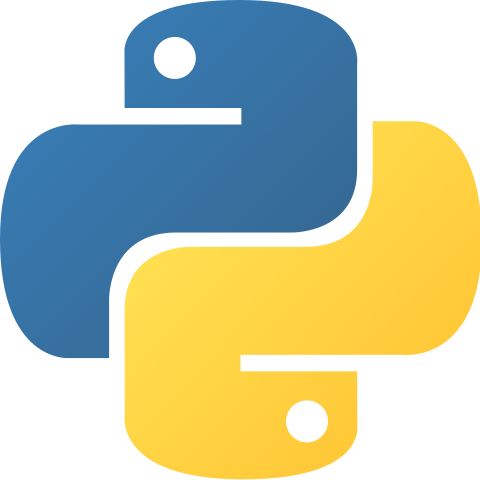
L’intitulé du projet donnait un exemple d’architecture permettant de répondre à la problématique du projet. L’architecture présentée était la suivante:



Nous nous sommes donc basés sur cette architecture afin de mettre en place notre solution. Dans notre cas les données arrivent en temps réel nous nous sommes donc orientés vers l’architecture en temps réel. Dans notre cas la solution mise en place est la suivante:

1. Outils utilisés





**Python :** Dans notre cas python est l’outil principal de notre projet. Nous l’avons utilisé afin de mettre en place nos scripts tel que le producer et le consumer par exemple. L’ensemble du projet repose sur les scripts python que nous avons mis en place.

**Streamlit :** Nous permets la mise en place d’une interface de data vizualisation en temps réelle

**Docker :** Nous permets de faciliter le déploiement de notre application dans plusieurs container logiciel. Il est un atout majeur dans notre projet car l’ensemble de notre solution est basée sur des containers docker.

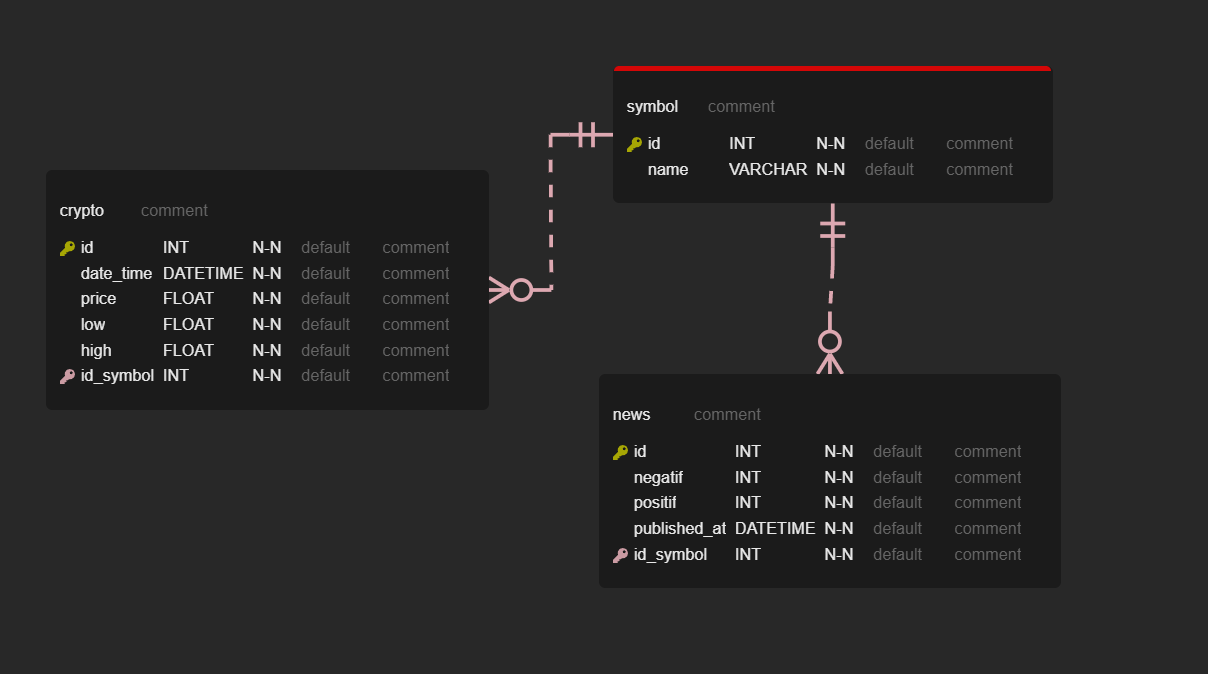
**Kafka :** Nous permets un traitement rapide et efficace des flux de données.

**MongoDB :**

1. Données

**Note base de données MongoDB est composée de 3 collections:**

* **crypto, s’occupe de stocker les données numériques liées à chaque cryptomonnaies, les données stockées proviennent de l’API CCXT**
* **symbol, s’occupe de stocker le nom de chaque crypto monnaie, elle est aussi une table de liaison entre crypto et news. Elle nous permets de réccupérer le nom de la crypto monnaie grace à son ID**
* **news, s’occupe de stocker les données textuelles liées à chaque crypto monnaie, les données stockées proviennent de l’API Cryptopanic.**



Les requêtes reçues et stockées par la base de données arrivent dans le format suivant:

Une image contenant texte

Description générée automatiquement

1. Traitement des données

Le traitement des données est assez simple, nous réccupérons un volume brute de données que nous réduisons en function de nos besoins avant le stockage en BDD.

Le traitement de nos données est effectué sur notre consumer Kafka, une fois le traitement effectué nous stockons le résultat en BDD.

1. Interface