



Université Abdelmalek Essaâdi

École Nationale des Sciences Appliquées de Tétouan

Filière : Big Data & IA

Rapport du Projet Informatique Décisionnel

**Processus ETL pour l'étude du marché des
crypto-monnaies**

Réalisé par :

Hamza Kholti
Anouar Bouzhar

Proposé par :

Mr. Marzouki FAOUZI

Année Universitaire : 2024-2025

Rapport rédigé dans le cadre du projet de fin de semestre.

Table des matières

| | | |
|----------|--|----------|
| 1 | Démarche d'implémentation choisie | 7 |
| 1.1 | Analyse des besoins | 7 |
| 1.2 | Modélisation et conception du projet | 8 |
| 1.3 | Collecte des données | 8 |
| 1.4 | Stockage initial | 8 |
| 1.5 | Transformation des données | 9 |
| 1.6 | Chargement dans le DW | 9 |
| 1.7 | Visualisation | 9 |
| 1.8 | L'architecture des outils utilisés | 10 |

Liste des tableaux

Table des figures

| | | |
|---|--|----|
| 1 | Architecture utilisé | 10 |
| 2 | Schéma en étoile du data warehouse | 11 |
| 3 | Visualization Ergonomie d'ensemble des grandeurs | 16 |

Introduction

Dans un monde dominé par les données, les cryptomonnaies s'imposent comme des actifs financiers à la fois innovants et complexes, représentant des opportunités lucratives mais aussi des risques élevés. En tant qu'ingénieurs d'État spécialisés en Big Data, Intelligence Artificielle (AI), Business Intelligence (BI) et sciences des données, nous avons été mandatés par un groupe d'investisseurs désireux de maximiser leur potentiel d'investissement dans ce domaine en pleine expansion.

Notre mission consiste à concevoir et mettre en place une solution d'intelligence décisionnelle complète, intégrant les concepts fondamentaux du Big Data pour gérer efficacement les grands volumes de données (Volume), leur diversité (Variety) et leur vitesse (Velocity). Grâce à une architecture technologique avancée, nous collectons, traitons et analysons des données en temps réel issues de multiples sources telles que les plateformes d'échange, les flux d'actualités et les réseaux sociaux.

En utilisant des technologies modernes comme Python, des API de données, des bases SQL et des modèles prédictifs basés sur l'IA, notre solution exploite le Big Data pour extraire des insights précis et exploitables. Les fonctionnalités incluent une analyse en temps réel, des visualisations interactives et des prédictions fiables, permettant aux investisseurs de prendre des décisions éclairées tout en minimisant les risques associés à la volatilité des marchés des cryptomonnaies.

Contexte du Projet

Les cryptomonnaies ont révolutionné les marchés financiers en introduisant une nouvelle classe d'actifs décentralisés et transparents via la technologie blockchain. Cependant, leur volatilité reste un défi majeur, influencée par des mouvements de marché rapides, des tendances économiques globales et les sentiments exprimés sur des canaux comme les réseaux sociaux. Cette incertitude rend la prise de décision particulièrement complexe pour les investisseurs.

En tant qu'experts en BI et en Big Data, nous avons été sollicités pour accompagner une entreprise d'investisseurs cherchant à investir efficacement dans les cryptomonnaies tout en minimisant les risques. Notre rôle est de concevoir une solution robuste et évolutive capable d'exploiter des volumes massifs de données issus de diverses sources : plateformes d'échange, indicateurs on-chain, flux d'actualités économiques et analyses de sentiment.

Grâce à l'intelligence artificielle et à des modèles prédictifs avancés comme les LSTM, notre solution proposera une analyse prédictive des fluctuations du marché, permettant aux investisseurs d'anticiper les tendances et d'identifier les meilleures opportunités. Par ailleurs, l'utilisation de tableaux de bord interactifs et personnalisables assurera une prise de décision rapide et intuitive, adaptée aussi bien aux investisseurs novices qu'aux experts financiers.

Avec une méthodologie centrée sur les besoins métiers et une architecture technologique avancée, notre solution ambitionne de devenir un outil stratégique pour garantir des investissements rentables et éclairés dans le domaine des cryptomonnaies.

Objectifs

L'objectif principal de ce projet est de développer une solution BI performante, capable de répondre aux besoins stratégiques et opérationnels de l'entreprise, tout en assurant une prise de décision basée sur des données fiables. Plus précisément, la solution doit permettre :

- **Collecte et traitement des données en temps réel** pour assurer une surveillance continue des marchés et des indicateurs clés.
- **Prédiction des fluctuations du marché à court et moyen termes** grâce à des modèles IA avancés comme les LSTM.
- **Analyse multidimensionnelle des données** pour explorer les corrélations entre différents actifs, données économiques/ boursières, volumes échangés et sentiments de marché.
- **Visualisation interactive et personnalisée** pour faciliter la compréhension des données par les décideurs.
- **Réduction des risques financiers** grâce à des alertes automatisées et des analyses de scénarios.
- **Interopérabilité et scalabilité** pour permettre une intégration fluide avec les systèmes existants et une évolution future de la solution.

En alliant technologie et expertise métier, cette solution vise à transformer les données en un véritable levier stratégique pour l'entreprise.

Analyse Préliminaire

1. Parties Prenantes et Besoins Généraux

Le projet répond aux besoins de plusieurs parties prenantes :

- **Investisseurs individuels** : Accès à des outils intuitifs pour suivre les tendances.
- **Fonds d'investissement** : Tableaux de bord avancés et prédictions fiables.
- **Analystes financiers** : Visualisations précises pour créer des rapports stratégiques.
- **Entreprises Fintech** : Analyses sectorielles pour innover dans leurs services.

Les besoins communs incluent :

- **Données fiables et actualisées en temps réel.**
- **Modèles prédictifs pour anticiper les tendances.**
- **Outils BI interactifs pour une analyse approfondie.**

2. Besoins Fonctionnels et Techniques

La solution doit intégrer :

- **Collecte des données** : À partir de sources variées (plateformes d'échange, réseaux sociaux, API).
- **Analyses descriptives et prédictives** : Indicateurs techniques et modèles IA pour prévoir les mouvements du marché.
- **Visualisation** : Tableaux de bord personnalisables et rapports automatisés.

3. Besoins Non Fonctionnels

- **Performance** : Traitement rapide de gros volumes de données.
- **Évolutivité** : Intégration de nouveaux actifs et indicateurs.
- **Fiabilité** : Disponibilité constante pour éviter les interruptions.

4. Contraintes

- **Coûts élevés** liés à l'infrastructure.
- **Dépendance aux API tierces**, parfois limitées en termes de données.

Technologies Choisies

Le choix des technologies repose sur des critères de performance, d'évolutivité, de flexibilité et de robustesse, afin de répondre aux besoins spécifiques de traitement, d'analyse et de restitution des données dans le domaine des cryptomonnaies :

- **Python** : Outil central pour le développement des scripts ETL, la construction d'API dédiées et le traitement des données en parallèle grâce au multiprocessing. Python joue également un rôle crucial dans l'intégration des différentes composantes du système, comme l'automatisation des workflows et l'implémentation des algorithmes d'analyse.
- **APIs de données** : Des interfaces comme celles des principales plateformes d'échange (Binance, Kraken) fournissent en temps réel des données critiques telles que les prix, les volumes, les tendances et les dominances des cryptomonnaies.
- **Bases de données SQL (MySQL)** : MySQL est utilisé pour le stockage des données brutes provenant des sources externes avant leur traitement. Il constitue la première étape dans la chaîne d'extraction, transformation et chargement (ETL), permettant de structurer les données pour une intégration efficace avec le Data Warehouse.
- **Big Data avec Apache Hive** : Utilisé comme entrepôt de données (Data Warehouse) pour centraliser, stocker et traiter les grands volumes de données structurées issus de l'ETL. Hive permet d'exécuter des requêtes complexes et d'offrir un accès rapide aux données historiques pour des analyses approfondies.
- **Systèmes d'exploitation Linux (Ubuntu Server)** : Ubuntu Server a été choisi pour son efficacité, sa sécurité, et sa compatibilité avec les outils Big Data et les frameworks d'IA. Il garantit une infrastructure stable et scalable, essentielle pour gérer les charges de travail importantes en production.
- **Power BI** : Une plateforme puissante pour la visualisation des données, permettant de créer des tableaux de bord interactifs qui facilitent l'exploration des insights par les investisseurs. Les visualisations dynamiques sont adaptées aux besoins de divers niveaux d'utilisateurs, des novices aux analystes expérimentés.
- **Modèles prédictifs avec LSTM et transformers** :
 - **LSTM** : Implémentés via TensorFlow, ces modèles capturent les dépendances temporelles pour anticiper les fluctuations du marché.
 - **Transformers et NLTK** : Utilisés pour l'analyse des sentiments, ces outils traitent les flux textuels non structurés provenant des réseaux sociaux, des actualités et d'autres sources pour extraire des signaux de marché pertinents.

Ce choix cohérent de technologies assure une gestion fluide de la chaîne de traitement des données, tout en garantissant une solution performante, évolutive et adaptée aux besoins des investisseurs.

Valeur Ajoutée

La solution BI apporte une valeur stratégique considérable :

[label=0., leftmargin=1.5cm, itemsep=5pt]**Réduction des risques financiers** : Prédictions précises et alertes automatisées. **Optimisation des profits** : Identification des opportunités rentables via des analyses approfondies. **Amélioration de la prise de décision** : Des insights clairs et exploitables. **Adaptabilité et évolutivité** : Une architecture capable de s'adapter aux évolutions du marché. **Automatisation des tâches** : Gain de temps et réduction des erreurs humaines.

Perspectives

Pour les futures évolutions, nous prévoyons de :

[label=0., leftmargin=1.5cm, itemsep=5pt]**Améliorer l'intégration en temps réel** pour une gestion optimale des événements critiques. **Automatiser la prise de décision** grâce à des modèles d'apprentissage avancés. **Renforcer la sécurité des données** par des mécanismes de cryptage et d'authentification. **Étendre l'accessibilité** via une application mobile dédiée. **Optimiser les performances** en adoptant des architectures distribuées et évolutives.

Ces évolutions garantiront une solution robuste et adaptée aux besoins changeants des investisseurs.

1 Démarche d'implémentation choisie

La mise en œuvre d'une solution BI (Business Intelligence) robuste et tolérante aux pannes exige une conception rigoureuse ainsi qu'un choix judicieux des outils adaptés au contexte du projet. Notre démarche a été guidée par l'objectif de répondre efficacement aux problématiques de traitement en temps réel, d'hétérogénéité des données et de visualisation performante.

Pour atteindre ces objectifs, nous avons suivi une méthodologie structurée, comprenant les étapes suivantes :

1.1 Analyse des besoins

La première phase de notre projet consistait à clarifier les besoins fonctionnels et techniques afin de garantir une solution adaptée. Cette analyse nous a permis de définir :

- **Les objectifs principaux** :
 - Suivi et analyse des performances des cryptomonnaies.
 - Prédiction des tendances basées sur des analyses de sentiment.
 - Proposition d'un système évolutif et en temps réel pour une prise de décision rapide.

- **Les contraintes spécifiques :**
 - Gestion d'un flux constant de données issues de multiples sources.
 - Maintenance de la cohérence et de l'intégrité des données dans des environnements hétérogènes.
 - Optimisation des délais de traitement pour éviter les retards critiques.

1.2 Modélisation et conception du projet

Une modélisation précise est essentielle pour garantir une architecture performante et évolutive. Notre conception repose sur une architecture ETL (Extract, Transform, Load) et inclut :

- **Sources de données :**
 - *API CoinMarketCap* : Données sur la blockchain (hashrate, transactions par seconde).
 - *yfinance* et *Binance* : Extraction des données liées aux cryptomonnaies, aux actions boursières et aux taux d'intérêt.
 - *API Coingecko* et *Cryptocompare* : Taux de change des cryptomonnaies et données approfondies.
 - *Twitter (X)* et *News API* : Extraction des opinions et actualités pour analyse de sentiment.
- **Outils de stockage et de traitement :**
 - MySQL pour le stockage initial des données brutes.
 - Snowflake comme entrepôt de données pour le traitement analytique.
 - Hive comme alternative locale en cas de contraintes liées à la latence.
- **Visualisation et reporting :** Utilisation de Streamlit et Power BI pour créer des tableaux de bord interactifs et détaillés.

1.3 Collecte des données

La collecte des données représente une étape cruciale de notre processus ETL. Nous avons extrait des données issues de formats et sources hétérogènes :

- Les données blockchain (hashrate, transactions par seconde) via l'API CoinMarketCap.
- Les données financières et économiques, telles que les prix des cryptomonnaies, actions et taux d'intérêt, via les packages *yfinance* et *Binance*.
- Les taux de change des cryptomonnaies et autres métriques via *Coingecko* et *Cryptocompare*.
- Les actualités et opinions des utilisateurs sur Twitter et via *News API* pour les analyser et prédire les sentiments.

Toutes ces données sont collectées en temps réel en utilisant des requêtes Python (**requests**). Ce flux constant est indispensable pour alimenter les processus de traitement et les analyses en temps réel.

1.4 Stockage initial

Pour garantir l'intégrité et la disponibilité des données extraites, celles-ci sont d'abord stockées dans une base de données MySQL. Ce stockage initial sert plusieurs objectifs :

- Prévention de la perte de données pendant les étapes de transformation.

- Réalisation d'opérations préliminaires comme le nettoyage et la validation des données.
- Création d'un point central pour gérer les données brutes avant leur transfert vers l'entrepôt Snowflake.

1.5 Transformation des données

Les données brutes stockées dans MySQL sont ensuite extraites, transformées et chargées dans un format compatible avec le schéma de Snowflake. Les transformations incluent :

- **Nettoyage des données** : Élimination des valeurs aberrantes, des données manquantes et inutilisables.
- **Normalisation et uniformisation** : Alignement des données sur des formats standardisés pour garantir leur homogénéité.
- **Création de colonnes calculées** : Intégration de mesures clés, comme les moyennes mobiles ou les indicateurs financiers, pour enrichir l'analyse.
- **Analyse de sentiment** : Application de modèles de deep learning pour analyser les sentiments des tweets et des articles d'actualité.

Ces transformations, réalisées avec des bibliothèques Python performantes telles que Pandas et `pandas_ta`, permettent d'assurer la cohérence et la qualité des données avant leur chargement dans Snowflake.

1.6 Chargement dans le DW

Une fois les données transformées, elles sont chargées en temps réel dans Snowflake, un entrepôt de données cloud robuste et convivial. Snowflake a été choisi pour :

- Sa capacité à gérer de grands volumes de données de manière efficace.
- Son architecture en cluster, permettant une montée en charge facile.
- Sa compatibilité avec les outils BI modernes comme Power BI.

Cependant, en raison de la latence liée au cloud et de la nécessité d'une exécution rapide, nous avons envisagé Hive comme une alternative locale. Hive, étant local et tolérant aux pannes, offre des avantages en termes de rapidité et de fiabilité dans des environnements à contraintes élevées.

1.7 Visualisation

La visualisation des données est une étape clé du projet. Elle permet de :

- Présenter les résultats de manière claire et intuitive.
- Identifier des tendances et des insights pour une prise de décision éclairée.
- Mettre en valeur l'impact du projet à travers des tableaux de bord interactifs.

Nous avons également testé Streamlit, mais en raison de son manque de convivialité pour les applications en temps réel, nous avons décidé de changer cet outil au profit de Power BI.

- **Power BI** : Choisi pour sa performance, sa diversité de visualisations et ses capacités de connexion à diverses sources de données.

Les visualisations incluent des graphiques dynamiques, des tableaux comparatifs et des analyses détaillées des tendances, permettant ainsi une exploration approfondie des données.

Avec cette démarche structurée, nous avons réussi à créer une solution performante, évolutive et adaptée aux exigences du projet. L'architecture conçue et les outils sélectionnés permettent de garantir une prise de décision rapide et basée sur des données fiables.

1.8 L'architecture des outils utilisés

Pour mettre en œuvre notre solution BI tout en répondant aux besoins analytiques et techniques, nous avons conçu une architecture robuste et adaptée. Les besoins analytiques incluent l'assurance d'une qualité optimale des données, leur uniformité et homogénéité, ainsi que leur visualisation claire et précise. Les besoins techniques, quant à eux, concernent la tolérance aux pannes, le traitement en temps réel des flux de données et l'optimisation des performances.

Pour atteindre ces objectifs, nous avons choisi des outils reconnus comme étant parmi les meilleurs dans leurs catégories respectives. L'architecture ETL utilisée dans ce projet est représentée dans la figure suivante.

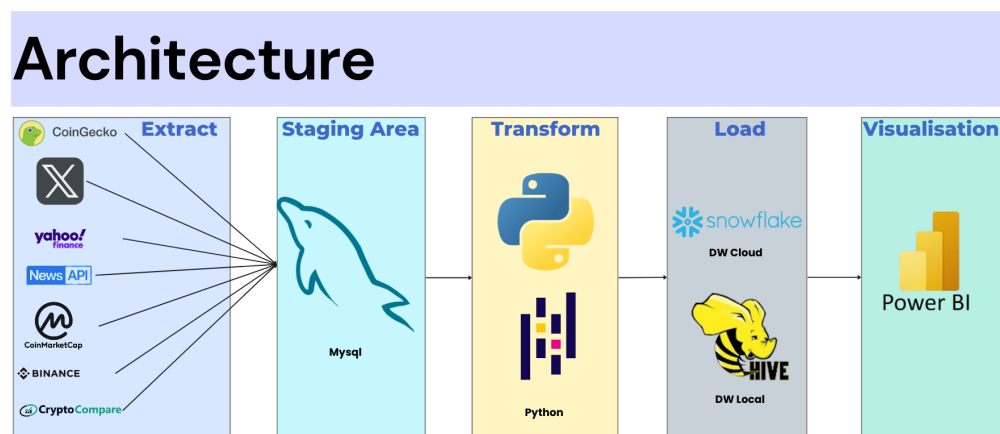


FIGURE 1 – Architecture utilisé

Cette architecture représente un processus ETL (*Extract*, *Transform*, *Load*) conçu pour collecter, transformer, et visualiser des données à partir de différentes sources :

- **Extraction (*Extract*)** : Les données sont collectées en temps réel depuis des sources hétérogènes comme *CoinGecko*, *Yahoo Finance*, et *Twitter (X)* via des APIs.
- **Zone de Transit (*Staging Area*)** : Les données brutes sont temporairement stockées dans une base MySQL pour assurer leur intégrité et préparer leur transformation.
- **Transformation (*Transform*)** : Les données sont nettoyées, normalisées, et enrichies à l'aide de Python (par exemple, pour l'analyse de sentiment ou l'uniformisation des formats).
- **Chargement (*Load*)** : Les données transformées sont stockées dans un entrepôt de données *Hive*, adapté au traitement en temps réel.

- **Visualisation** : Les données sont présentées via des tableaux de bord interactifs créés avec *Power BI*, permettant des analyses stratégiques.

Schéma du Data Warehouse

Après avoir décrit l'architecture générale de la solution, cette section explique le schéma en étoile utilisé pour structurer le Data Warehouse, comme illustré dans la figure ci-dessous (*schéma_DW.png*).

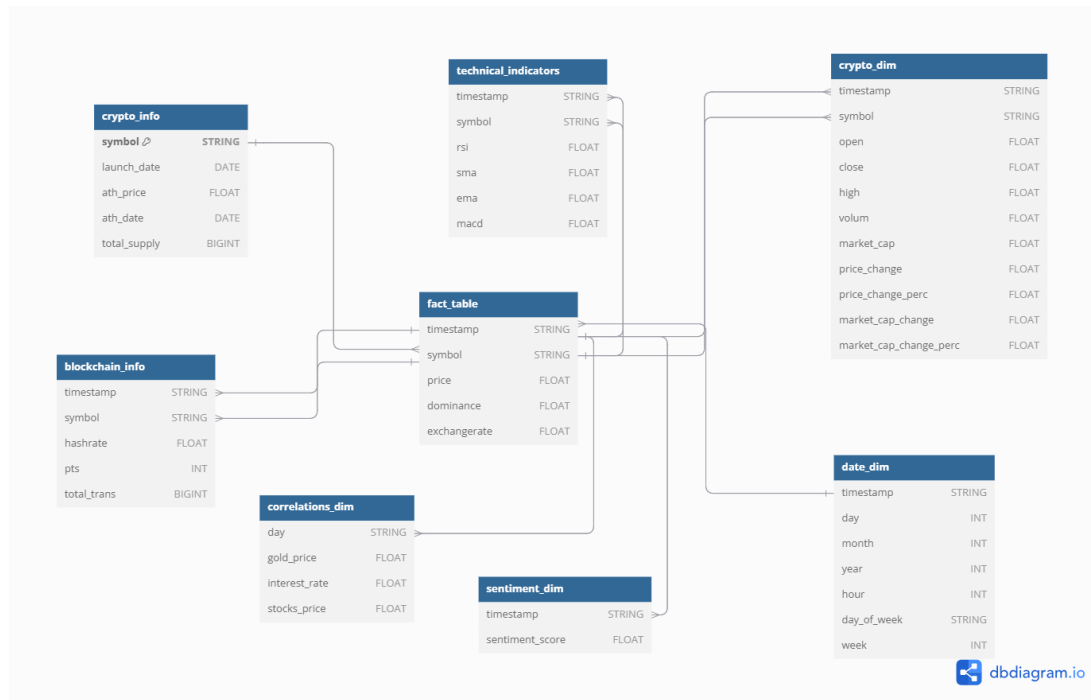


FIGURE 2 – Schéma en étoile du data warehouse

Description du Schéma en Étoile

Le schéma en étoile est une structure simple et efficace pour organiser les données dans un entrepôt de données. Il est constitué d'une **table de faits** centrale connectée à plusieurs **tables de dimensions**. Voici une description des éléments de ce schéma :

1. Table de Faits : fact_table

La table de faits contient les données principales servant d'entrée pour les analyses :

- **timestamp** : L'horodatage des enregistrements.
- **symbol** : Le symbole de la cryptomonnaie analysée.
- **price** : Le prix actuel de la cryptomonnaie.
- **dominance** : La dominance de marché de la cryptomonnaie.
- **exchangerate** : Le taux de change de la cryptomonnaie.

2. Tables de Dimensions

Les tables de dimensions fournissent un contexte et des informations supplémentaires pour enrichir les analyses.

- **date_dim** : Fournit des informations temporelles détaillées, telles que :
 - Jour, mois, année, et heure.
 - Jour de la semaine et numéro de la semaine.
- **crypto_info** : Contient des détails relatifs aux cryptomonnaies, notamment :
 - Date de lancement (**launch_date**).
 - Prix maximum historique (**ath_price**) et sa date (**ath_date**).
 - Offre totale (**total_supply**).
- **blockchain_info** : Fournit des données sur la blockchain associée, telles que :
 - Hashrate (**hashrate**).
 - Nombre total de transactions (**total_trans**).
- **technical_indicators** : Stocke des indicateurs techniques tels que :
 - RSI, SMA, EMA, et MACD, utiles pour l'analyse technique des marchés.
- **correlations_dim** : Permet d'étudier les corrélations entre les cryptomonnaies et d'autres marchés financiers en fournissant des données comme :
 - Prix de l'or (**gold_price**).
 - Taux d'intérêt (**interest_rate**).
 - Prix des indices boursiers (**stocks_price**).
- **sentiment_dim** : Mesure les opinions des utilisateurs et les tendances du marché via un score de sentiment (**sentiment_score**).
- **crypto_dim** : Décrit les caractéristiques détaillées des cryptomonnaies, telles que :
 - Prix d'ouverture (**open**), de clôture (**close**), et variations (**price_change_perc**).
 - Volume d'échange et capitalisation boursière (**market_cap**).

Ce schéma en étoile permet une structuration optimale pour l'analyse des données cryptographiques. La table de faits centralise les mesures clés, tandis que les dimensions offrent un contexte riche, facilitant l'exploration et l'extraction d'insights pertinents.

Les Données Utilisées

Dans ce projet, les données collectées sont essentielles pour assurer une analyse approfondie et des prédictions précises sur les marchés financiers et des cryptomonnaies. Les informations proviennent de plusieurs sources fiables, accessibles via des APIs, couvrant divers aspects des marchés. En plus de l'utilisation de données réelles, une fonction a été développée pour générer des données synthétiques, en simulant des valeurs manquantes selon des distributions prédéfinies. Voici un aperçu détaillé des données utilisées et de la méthode employée pour compléter celles-ci.

1. Données de Cryptomonnaies en Temps Réel (CryptoCompare API)

Une des sources principales pour ce projet est l'API CryptoCompare, qui permet de récupérer les données en temps réel des cryptomonnaies sélectionnées. Un script a été conçu pour récupérer ces informations et les insérer directement dans la base de données, garantissant des données fraîches et actualisées pour les analyses et prédictions.

Source : CryptoCompare API

URL : <https://min-api.cryptocompare.com/data/v2/histoday>

Description : Cette API fournit des informations actualisées sur les prix, volumes, et autres paramètres clés des cryptomonnaies.

Détails des Données Collectées :

- **Timestamp** : Date et heure exacte de l'enregistrement des données.
- **Open, High, Low, Close** : Prix d'ouverture, le plus haut, le plus bas, et de clôture pour chaque coin (par exemple, BTC, ETH).
- **Volume** : Volume échangé pour la période donnée.
- **Market Cap** : Capitalisation du marché estimée en fonction du volume échangé.

Ces données sont collectées pour plusieurs cryptomonnaies clés telles que BTC, ETH, BNB, ADA, SOL et XRP, et sont utilisées pour enrichir la base de données et effectuer des analyses plus précises et actualisées. Cela permet de suivre les fluctuations du marché en temps réel et d'intégrer des informations récentes dans les modèles prédictifs.

2 Données Blockchain (CoinMarketCap API)

Une autre source importante pour ce projet est l'API CoinMarketCap, qui fournit des statistiques détaillées et actualisées sur les blockchains des principales cryptomonnaies.

Source : CoinMarketCap API

URL : <https://sandbox-api.coinmarketcap.com/v1/blockchain/statistics/latest>

Description : Fournit des statistiques détaillées et actualisées sur les blockchains des principales cryptomonnaies.

Détails des Données Collectées :

- **Block Reward Static** : Récompense fixe attribuée pour chaque bloc miné.
- **Consensus Mechanism** : Technologie utilisée pour valider les transactions (Proof of Work, Proof of Stake, etc.).
- **Difficulty** : Mesure de la difficulté actuelle pour miner un bloc.
- **Hashrate 24h** : Puissance de calcul totale utilisée pour sécuriser la blockchain sur 24 heures.
- **Pending Transactions** : Nombre de transactions en attente de validation.
- **Total Blocks** : Nombre total de blocs minés.
- **Total Transactions** : Nombre total de transactions validées.
- **TPS (Transactions Par Seconde)** : Transactions effectuées par seconde sur 24 heures.

3 Données sur la Dominance des Marchés (Binance API)

L'API Binance fournit des informations détaillées sur les prix et volumes des cryptomonnaies, permettant de calculer les dominances de marché.

Source : Binance API

URL : <https://api.binance.com/api/v1/ticker/24hr>

Description : Offre des données détaillées sur les prix et volumes des cryptomonnaies, permettant de calculer les dominances de marché.

Détails des Données Collectées :

- **Prix des Cryptomonnaies** : Dernier prix enregistré pour chaque paire de trading (e.g., BTC/USDT).

- **Volume des Transactions** : Quantité totale échangée sur une paire donnée au cours des dernières 24 heures.
- **BTC Dominance** : Pourcentage de la capitalisation de Bitcoin par rapport à la capitalisation totale du marché.
- **ETH Dominance** : Pourcentage de la capitalisation d'Ethereum par rapport à la capitalisation totale.
- **Altcoin Dominance** : Calculée comme $100\% - (\text{BTC Dominance} + \text{ETH Dominance})$.

4 Données des Taux de Change (CoinGecko API)

L'API CoinGecko permet de récupérer des taux de change actualisés pour les cryptomonnaies et les devises fiduciaires.

Source : CoinGecko API

URL : <https://api.coingecko.com/api/v3/exchangerates>

Description : Fournit des taux de change actualisés pour les cryptomonnaies et les devises fiduciaires.

Détails des Données Collectées :

- **Nom et Unité de la Monnaie** : Exemples : Bitcoin (BTC), Ethereum (ETH).
- **Valeur Actuelle** : Taux de change entre une cryptomonnaie et une devise standard (e.g., USD).
- **Type de Taux** : Crypto à fiat ou crypto à crypto.

5 Données Financières Globales (Yahoo Finance API)

L'API Yahoo Finance permet de récupérer des données financières globales pour l'or, les taux d'intérêt et les indices boursiers.

Source : Yahoo Finance API

URL : <https://finance.yahoo.com> via la bibliothèque yfinance

Description : Fournit des données financières globales pour l'or, les taux d'intérêt et les indices boursiers.

Détails des Données Collectées :

- **Or (Gold Futures)** : Prix d'ouverture, haut, bas, clôture.
- **Taux d'Intérêt** : Rendements obligataires sur différentes maturités.
- **Indices Boursiers (S&P 500)** : Données détaillées pour les indices majeurs (e.g., ouverture, clôture, volume échangé).

6 Données d'Actualités Financières (NewsAPI)

L'API NewsAPI permet de collecter des articles d'actualité liés aux cryptomonnaies et aux marchés financiers.

Source : NewsAPI

URL : <https://newsapi.org/v2/everything>

Description : Collecte d'articles d'actualité liés aux cryptomonnaies et aux marchés financiers.

Détails des Données Collectées :

- **Source de l'Article :** Origine de l'article (e.g., Bloomberg, Reuters).
- **Titre et Contenu :** Résumé ou description principale de l'article.
- **Date de Publication :** Timestamp précis de l'article publié.

7 Données des Réseaux Sociaux (Twitter API)

L'API Twitter permet de collecter les tweets publics contenant des mots-clés liés aux cryptomonnaies et aux blockchains.

Source : Twitter API

URL : <https://developer.twitter.com/en/docs> via la bibliothèque Tweepy

Description : Permet de collecter les tweets publics contenant des mots-clés liés aux cryptomonnaies et aux blockchains.

Détails des Données Collectées :

- **Utilisateur (Source) :** Nom d'utilisateur ou handle Twitter de l'auteur du tweet.
- **Contenu du Tweet :** Texte intégral du tweet publié.
- **Date et Heure :** Timestamp exact de la publication.

8 Génération de Données Synthétiques

En raison des limitations de certaines API et de la nécessité d'obtenir un volume de données plus important, une fonction a été mise en place pour générer des données synthétiques.

Description du Code de Génération de Données Synthétiques : La fonction de génération de données synthétiques est conçue pour simuler les données manquantes dans des domaines comme les blockchains. Elle fonctionne comme suit :

- Elle prend en entrée les derniers enregistrements de données et un horodatage.
- Selon le symbole de la cryptomonnaie (par exemple BTC, ETH), elle génère des valeurs aléatoires pour des paramètres comme le hashrate, les transactions par seconde (TPS), et le nombre total de transactions.
- Les valeurs générées sont ensuite utilisées pour enrichir la base de données, avec un format compatible pour une base de données comme Snowflake.

9 Formats et Validation des Données

Format des Données Collectées : Les données provenant des API et des simulations sont toutes renvoyées sous le format JSON, qui est ensuite traité et validé avant insertion dans la base de données.

Validation des Données :

- **Vérification de la cohérence et de la validité des données :** Chaque entrée est vérifiée pour garantir qu'elle respecte les critères de qualité et de fiabilité des données.

- **Suppression des valeurs incohérentes** : Les valeurs invalides (par exemple, valeurs nulles, négatives, ou incorrectes) sont supprimées.
- **Conversion des timestamps** : Les timestamps sont convertis en format ISO 8601 pour garantir la compatibilité et l'uniformité des données, facilitant ainsi leur traitement et leur analyse.

Mise en Valeur des Données

Les données collectées et synthétiques sont essentielles à l'efficacité de la solution développée. Elles permettent :

- **Analyse Prédictive** : Alimentation de modèles IA comme les LSTM pour anticiper les tendances des cryptomonnaies.
- **Analyse de Sentiment** : Évaluation des tendances du marché grâce aux données des actualités et des réseaux sociaux.
- **Tableaux de Bord Dynamiques** : Surveillance des fluctuations des cryptomonnaies et des actifs numériques.

Grâce à ces données, le projet propose une solution robuste, évolutive, et adaptée aux besoins des investisseurs modernes, combinant précision et rapidité dans la prise de décision. La capacité de traiter et de générer des données en temps réel, associée à l'utilisation de modèles prédictifs et d'analyses de sentiment, permet aux investisseurs de prendre des décisions éclairées et adaptées aux conditions du marché.

Analyse et Description des résultats

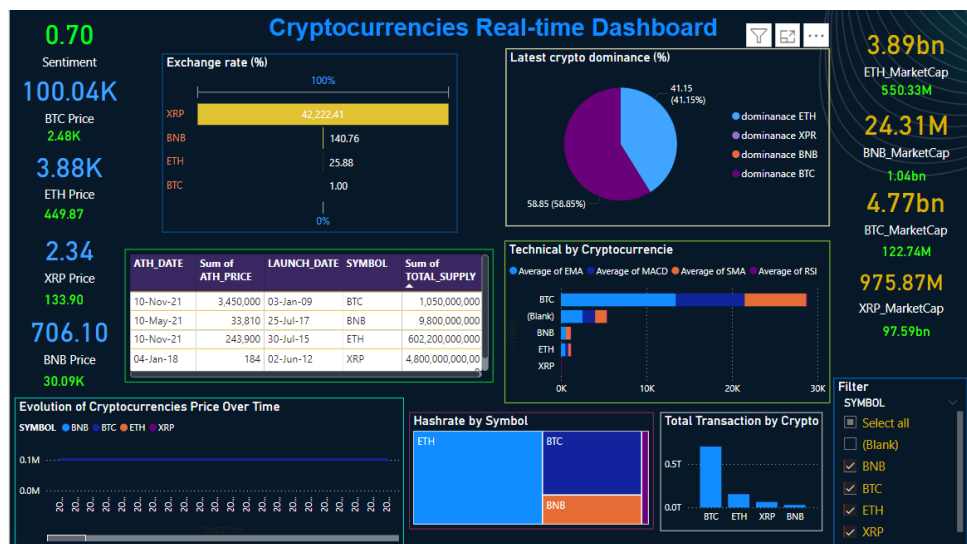


FIGURE 3 – Visualization Ergonomie d'ensemble des grandeurs

1. Prix des Cryptomonnaies (en bleu)

Les prix affichés, tels que :

- Bitcoin (BTC) : 100.04K USD,
- Ethereum (ETH) : 3.88K USD,
- Binance Coin (BNB) : 706.10 USD,

- Ripple (XRP) : 2.34 USD,

indiquent des écarts significatifs entre les cryptos.

Analyse supplémentaire :

- Bitcoin (BTC) : Réserve de valeur numérique, sensible aux événements macroéconomiques.
- Ethereum (ETH) : Pilier des smart contracts et des DApps, reflétant la valeur utilitaire de son réseau.
- Binance Coin (BNB) : Dépendant de l'écosystème Binance, avec un potentiel de croissance.
- Ripple (XRP) : Axé sur les paiements transfrontaliers, mais vulnérable aux régulations juridiques.

2. Capitalisation Boursière (MarketCap)

La capitalisation boursière donne une vision claire de l'adoption et de la taille du marché :

- BTC : 4.77 milliards USD (leader absolu, valeur refuge numérique).
- ETH : 3.89 milliards USD (cas d'usage diversifié).
- BNB : 24.31 millions USD (potentiel de croissance modéré).
- XRP : 975.87 millions USD (impacté par des incertitudes juridiques).

Perspective : Ces données permettent aux investisseurs de choisir leurs investissements selon leurs objectifs.

3. Sentiment Global (0.70)

Un score de sentiment global de **0.70** reflète une perception globalement positive :

- Impact potentiel : Attirer davantage d'investisseurs, augmentant la demande et les prix.
- Une diminution du sentiment pourrait entraîner une baisse rapide des prix.

4. Dominance des Cryptos (Pie Chart)

- BTC (58.85%) : La majorité des capitaux investis en cryptomonnaies.
- ETH (41.15%) : Principal concurrent du BTC.

Cette dominance reflète la confiance des investisseurs dans ces deux actifs.

5. Taux d'Échange (Exchange Rate)

Le taux d'échange aide les investisseurs à évaluer le pouvoir d'achat relatif entre les cryptos et identifier des opportunités d'arbitrage :

- XRP (42,222.41%) par rapport à BTC montre un taux élevé de volatilité.
- BNB (140.76%) : Stable et proche du BTC.
- ETH (25.88%) : Une crypto majeure avec une évaluation bien plus basse que le Bitcoin.

6. Données Historiques et Offre Totale (ATH et Supply)

- Dates ATH : Les pics de prix sont souvent liés à des cycles haussiers spécifiques.
- Offre totale (Total Supply) : BTC, limité à 21 millions, confère une rareté intrinsèque. XRP, avec 4,800 milliards, peut limiter sa hausse potentielle.

7. Performances Techniques (EMA, MACD, SMA, RSI)

- BTC : Indicateurs stables, confirmant son statut de valeur refuge.
- XRP et BNB : Volatilité plus élevée, intéressante pour les day traders.

8. Évolution des Prix

Le graphique temporel des prix montre des tendances historiques qui permettent d'identifier des cycles haussiers et baissiers.

9. Conclusion Stratégique

Ce tableau de bord Power BI fournit une vue d'ensemble holistique du marché des cryptomonnaies. Les investisseurs peuvent :

- . Évaluer les leaders du marché (BTC et ETH) pour une stabilité relative.
- . Rechercher des opportunités de croissance avec des cryptos comme BNB et XRP.
- . Analyser les cycles et sentiments pour ajuster leurs stratégies.

En résumé, cette plateforme favorise la prise de décision stratégique grâce à une intégration fluide des données. L'ajout d'un module de prédiction (par exemple, via des modèles de machine learning) renforcerait encore sa pertinence.

conclusion

Le projet de création d'une solution décisionnelle pour l'analyse du marché des cryptomonnaies met en évidence l'intégration efficace de technologies avancées telles que le Big Data, l'IA et les outils BI, répondant ainsi aux besoins stratégiques des investisseurs. En s'appuyant sur une architecture ETL robuste, une collecte de données en temps réel, des analyses prédictives, et des visualisations interactives, cette solution offre des outils essentiels pour décrypter les dynamiques complexes de ce marché volatile.

Les résultats obtenus démontrent que la combinaison des indicateurs techniques, des analyses de sentiment et des visualisations personnalisées optimise la prise de décision tout en minimisant les risques financiers. La solution se distingue par sa scalabilité, son interopérabilité avec d'autres systèmes, et sa capacité d'adaptation aux évolutions futures du marché.

Cependant, plusieurs axes d'amélioration méritent d'être explorés. L'intégration d'un outil d'orchestration permettrait de créer davantage de visualisations afin de mieux comprendre le marché des cryptomonnaies. De plus, l'utilisation de données réelles issues de la blockchain, telles que le hashrate, les transactions par seconde et le nombre total de transactions, remplacerait avantageusement les données synthétiques pour des décisions encore plus précises et fiables. Enfin, entraîner des modèles prédictifs basés sur l'IA afin d'anticiper les tendances futures renforcerait encore la pertinence et l'efficacité de cette solution.