Rapport projet technique :

CryptoBot avec Binance

Introduction :

Les objectifs et les jalons attendus de ce projet sont détaillés de manière exhaustive dans un document fourni par l'équipe Datascientest (CryptoBot avec Binance). En résumé, ce projet vise à mettre en place une chaîne de traitement des données, englobant les composants clés suivants :

1. **Récupération des données :** Le projet implique l'extraction de données à partir d'une source API externe, avec un accent particulier sur l'obtention de données en temps réel et historiques liées aux valeurs boursières de diverses entreprises.
2. **Stockage des données :** Les données collectées seront stockées localement, garantissant leur disponibilité pour des traitements et analyses ultérieurs.
3. **Accessibilité des données :** Un objectif central de ce projet est de rendre ces données accessibles aux éventuels utilisateurs internes, facilitant leur utilisation dans diverses applications et processus de prise de décision.
4. **Utilisation des données :** Les données extraites et stockées serviront à plusieurs fins, allant de la formation de modèles d'apprentissage automatique à leur affichage de manière significative et informative, permettant une prise de décision et des insights efficaces.
5. **Conditionnement et Conteneurisation du Code :** Pour améliorer la réutilisation et l'extensibilité, l'ensemble du code développé dans ce projet sera systématiquement conditionné et conteneurisé, permettant un déploiement et une réplication efficaces.

Sprint 1 : Collectes de données

Toutes les API fournissent essentiellement le même type de données. Les principales différences résident dans les conditions d'inscription et les limites de nombre d'utilisation de l'API pour les plans gratuits.

Nous vous listerons donc les sources des données ainsi que les méthodes que nous pensons employer pour la première partie de collecte.

Binance : https://www.binance.com/

- Inscription requise : Oui (adresse e-mail + nom + numéro de téléphone + carte d'identité)
- Limite d'API : 6000 requêtes par minute (voir ici)
- Documentation de l'API : Oui

FMP Cloud : https://fmpcloud.io/

- Inscription requise : Oui (Google OAuth2 OU adresse e-mail/mot de passe)
- Limite d'API : 250 requêtes / jour (ici)
- Documentation de l'API : Oui

Alphavantage : https://www.alphavantage.co/

- Inscription requise : Oui (adresse e-mail)
- Limite d'API : 5 requêtes / minute, 500 requêtes / jour (voir ici)
- Documentation de l'API : Oui

Yahoo Finance

- Inscription requise : Non
- Limite d'API : Pratiquement illimitée
- Documentation de l'API : Non disponible, mais peut être déduite rétroactivement à partir de : https://github.com/ranaroussi/yfinance

Notre collecte de données se divise en deux catégories distinctes : les données en streaming et les données historiques.

1. **Données streaming**

Nous avons récupéré nos sources sur le site de Binance en utilisant le Websocket API.
Les données brutes en JSON de streaming peuvent être récupérées grâce au websockets suivants suivis par un endpoint :

- **wss://stream.binance.com:9443**
- **wss://stream.binance.com:443**

Exemple :

```
symbol='btcusdt'
interval='1m'
socket = f'wss://stream.binance.com:443/ws/{symbol}@kline_{interval}'
```

Comme montré ci-dessus, Python est le choix actuel après y avoir installé la librairie Websocket_client. Il faut noter qu'on peut également récupérer ces données sur linux en exécutant le fichier contenant le programme de requête.

Les données qui nous intéressent ici particulièrement sont celles qui correspondent au endpoint `/ws/{symbol}@kline_{interval}`. Cet endpoint remonte en streaming les données relatives à l'évolution des prix, volumes et autres…


```
{
        "e":"kline",
        "E":1699298147850,
        "s":"BTCUSDT",
        "k":{
                "t":1699298100000,
                "T":1699298159999,
                "s":"BTCUSDT",
                "i":"1m",
                "f":3270591725,
                "L":3270592323,
                "o":"34951.09000000",
                "c":"34959.97000000",
                "h":"34963.03000000",
                "l":"34951.08000000",
                "v":"22.74315000",
                "n":599,
                "x":false,
                "q":"795079.22724770",
                "V":"13.80648000",
                "Q":"482636.87185850",
                "B":"0"
        }
}
```



```
{
  "e": "kline",     // Event type
  "E": 123456789,   // Event time
  "s": "BNBBTC",    // Symbol
  "k": {
    "t": 123400000, // Kline start time
    "T": 123460000, // Kline close time
    "s": "BNBBTC",  // Symbol
    "i": "1m",      // Interval
    "f": 100,       // First trade ID
    "L": 200,       // Last trade ID
    "o": "0.0010",  // Open price
    "c": "0.0020",  // Close price
    "h": "0.0025",  // High price
    "l": "0.0015",  // Low price
    "v": "1000",    // Base asset volume
    "n": 100,       // Number of trades
    "x": false,     // Is this kline closed?
    "q": "1.0000",  // Quote asset volume
    "V": "500",     // Taker buy base asset volume
    "Q": "0.500",   // Taker buy quote asset volume
    "B": "123456"   // Ignore
  }
}
```


On peut définir la périodicité souhaitée aussi et ça se présente de la sorte :

Si par exemple, on définit un intervalle de 1mn, le streaming continue à diffuser par seconde mais à la 1ère minute le paramètre X donne la valeur True ("X": True).


```
{"e":"kline","E":1697378084696,"s":"BTCUSDT","k":{"t":1697378040000,"T":1697378099999,"s":"BTCUSDT","i":"1m","f":3239495866,"
L":3239496264,"o":"26855.71000000","c":"26851.62000000","h":"26855.72000000","l":"26851.59000000","v":"7.87392000","n":399,"x":
false,"q":"211456.14475150","V":"3.43784000","Q":"92325.40386270","B":"0"}}
{"e":"kline","E":1697378086922,"s":"BTCUSDT","k":{"t":1697378040000,"T":1697378099999,"s":"BTCUSDT","i":"1m","f":3239495866,"
L":3239496312,"o":"26855.71000000","c":"26850.45000000","h":"26855.72000000","l":"26850.44000000","v":"8.76562000","n":447,"x":
false,"q":"235399.44626800","V":"3.78680000","Q":"101695.28822610","B":"0"}}
{"e":"kline","E":1697378089749,"s":"BTCUSDT","k":{"t":1697378040000,"T":1697378099999,"s":"BTCUSDT","i":"1m","f":3239495866,"
L":3239496321,"o":"26855.71000000","c":"26850.44000000","h":"26855.72000000","l":"26850.44000000","v":"8.79655000","n":456,"x":
false,"q":"236229.93044170","V":"3.79325000","Q":"101868.47362860","B":"0"}}
{"e":"kline","E":1697378092584,"s":"BTCUSDT","k":{"t":1697378040000,"T":1697378099999,"s":"BTCUSDT","i":"1m","f":3239495866,"
L":3239496351,"o":"26855.71000000","c":"26850.26000000","h":"26855.72000000","l":"26850.25000000","v":"10.97979000","n":486,"
x":false,"q":"294850.76357160","V":"3.82051000","Q":"102600.41674360","B":"0"}}
{"e":"kline","E":1697378094641,"s":"BTCUSDT","k":{"t":1697378040000,"T":1697378099999,"s":"BTCUSDT","i":"1m","f":3239495866,"
L":3239496362,"o":"26855.71000000","c":"26850.25000000","h":"26855.72000000","l":"26850.25000000","v":"12.56012000","n":497,"
x":false,"q":"337283.01919480","V":"3.82458000","Q":"102709.69730180","B":"0"}}
{"e":"kline","E":1697378096747,"s":"BTCUSDT","k":{"t":1697378040000,"T":1697378099999,"s":"BTCUSDT","i":"1m","f":3239495866,"
L":3239496373,"o":"26855.71000000","c":"26850.26000000","h":"26855.72000000","l":"26850.25000000","v":"12.61435000","n":508,"
x":false,"q":"338739.10832490","V":"3.83184000","Q":"102904.63018940","B":"0"}}
```


2. Données historiques :

Les données sont collectées sur le site de Binance via les API. Les endpoints de base sont listés ci-dessous :

https://api.binance.com
https://api-gcp.binance.com
https://api1.binance.com
https://api2.binance.com
https://api3.binance.com
https://api4.binance.com

Pour accéder aux données on utilise l'url des API de bases ci-dessus suivi des endpoints suivants :

- `/api/v3/aggTrades`
- `/api/v3/avgPrice`
- `/api/v3/depth`
- `/api/v3/exchangeInfo`
- `/api/v3/klines`
- `/api/v3/ping`
- `/api/v3/ticker`
- `/api/v3/ticker/24hr`
- `/api/v3/ticker/bookTicker`
- `/api/v3/ticker/price`
- `/api/v3/time`
- `/api/v3/trades`
- `/api/v3/uiKlines`

Nous comptons dans le cadre de ce projet utiliser le endpoint **/api/v3/klines** pour collecter les données historiques des monnaies.

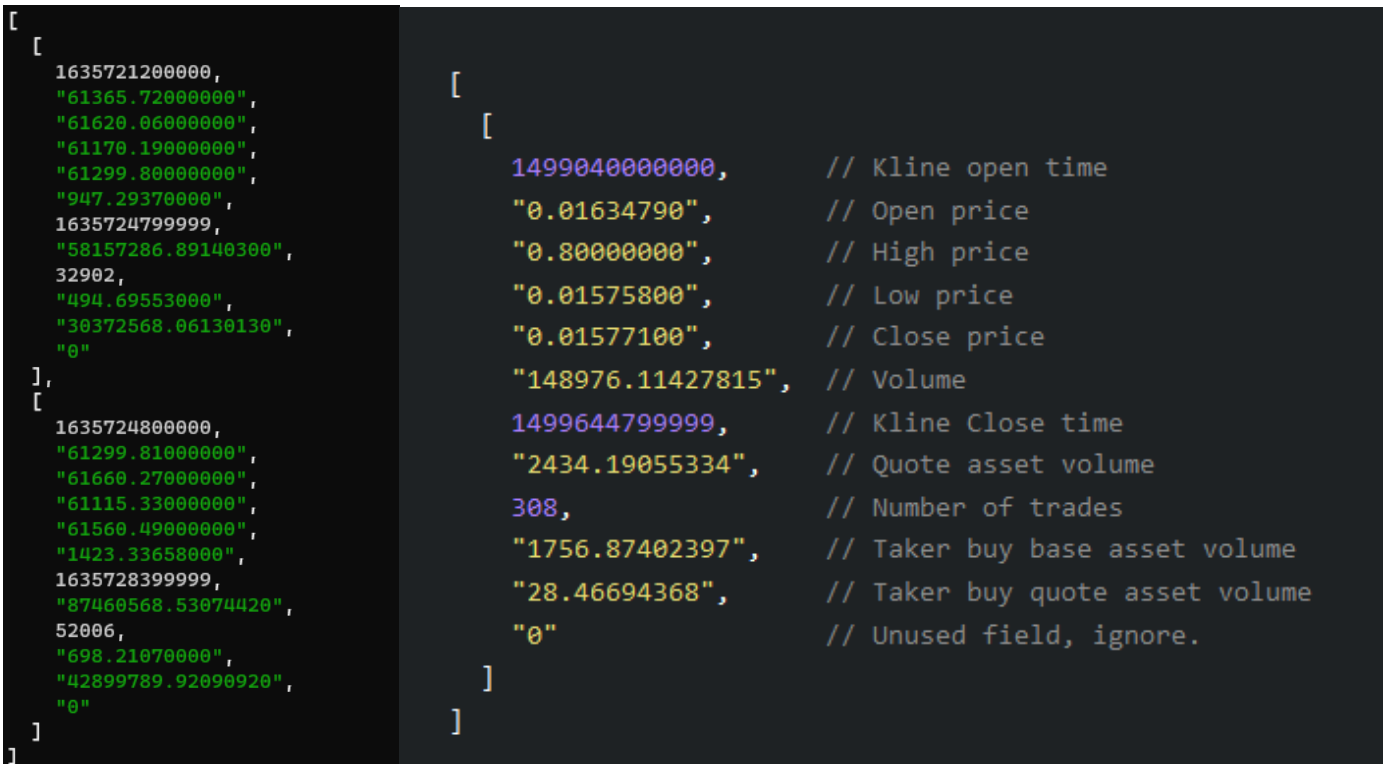
On utilisera également le endpoint **/api/v3/exchangeInfo** pour avoir les informations sur les règles d'échange et sur les monnaies.

Les données historiques sont collectées sur un intervalle allant du 1 Novembre 2021 au 31 octobre 2023 soit une durée totale de deux ans. Nous allons travailler sur deux granularités distinctes, une granularité de un jour et une granularité de 1h. Etant donné que pour une requête sur un API les données sont limités à 1000

valeurs maximum, On a du réaliser plusieurs requêtes via un script bash et les rassembler sur un seul fichier json via la commande `jq -s 'add' *.json`.

Ci-dessous un exemple de requêtes linux utilisé pour collecter les données :

```
Curl -X GET
"https://api.binance.com/api/v3/klines?symbol=BTCUSDT&interval=1h&startTime=163
5721200000&limit=2" | jq .
```



```
[
  [
    1635721200000,
    "61365.72000000",
    "61620.06000000",
    "61170.19000000",
    "61299.80000000",
    "947.29370000",
    1635724799999,
    "58157286.89140300",
    32902,
    "494.69553000",
    "30372568.06130130",
    "0"
  ],
  [
    1635724800000,
    "61299.81000000",
    "61660.27000000",
    "61115.33000000",
    "61560.49000000",
    "1423.33658000",
    1635728399999,
    "87460568.53074420",
    52006,
    "698.21070000",
    "42899789.92090920",
    "0"
  ]
]
```

```
[
  [
    1499040000000,      // Kline open time
    "0.01634790",       // Open price
    "0.80000000",       // High price
    "0.01575800",       // Low price
    "0.01577100",       // Close price
    "148976.11427815",  // Volume
    1499644799999,      // Kline Close time
    "2434.19055334",    // Quote asset volume
    308,                // Number of trades
    "1756.87402397",    // Taker buy base asset volume
    "28.46694368",      // Taker buy quote asset volume
    "0"                 // Unused field, ignore.
  ]
]
```