



IID 2022-2023

Langage RAPPORT PROJET BUSINESS INTELLIGENCE

Implémentation d'une solution ETL pour l'intégration de données de l'API Binance sous Talend

Présenter par :

Mohamed Reda Agbalou

Encadrer par :

Pr.Boukhalfa

Email : a.boukhalfa@usms.ma

Département mathématique et informatique

Informatique et ingénierie de données

Sommaire

Remerciments	5
Introduction	6
Jobs	9
Visualisation Power Bi	12
Prédiction	15

Remerciements :

“

Nous souhaitons exprimer notre profonde gratitude envers le Professeur Alaeddine Boukhalfa, notre instructeur dévoué en BI.

Son expertise, sa passion et son engagement ont considérablement enrichi notre parcours d'apprentissage. Grâce à ses enseignements perspicaces, à ses conseils éclairés et à sa patience infinie, nous avons développé une compréhension approfondie des principes fondamentaux de BI. Les leçons dynamiques et les exemples concrets partagés par M. Boukhalfa ont non seulement facilité la compréhension des concepts complexes de BI, mais ont également suscité un intérêt continu pour ce langage de modélisation graphique. Ses retours constructifs ont été inestimables pour notre développement professionnel, et nous sommes reconnaissants d'avoir eu l'opportunité d'apprendre sous sa direction éclairée.

Merci.

”

- IID2

1. Introduction :

Ce projet consiste à implémenter une solution ETL (Extract-transform-load) sous **Talend Open Studio** pour une intégration de différentes sources de données depuis l'API de binance et notamment avec des scripts en **Python**, pour répondre à un besoin d'une agence de trading, les tâches à réaliser sont :

- Implémenter des jobs
- Réaliser des mapping avec les flux de données
- Orchestrer plusieurs jobs et créations des routines pour transformation de données
- Intégration de la base de données : **MYSQL**
- Définition des schémas de métadonnées
- Automatisation avec des scripts exécutable
- Gestion des grosses sources de données (**Big Data**)

1.1 Les questions que répond notre sujet :

- Quel est l'investissement
- Création d'un model de prédiction et assistant de l'investissement.
- Les statistiques des symboles sur le dernier jour , mois , année.
- Synchronisation de données à temps réelle
- Intégration décisionnelle et opérationnelle

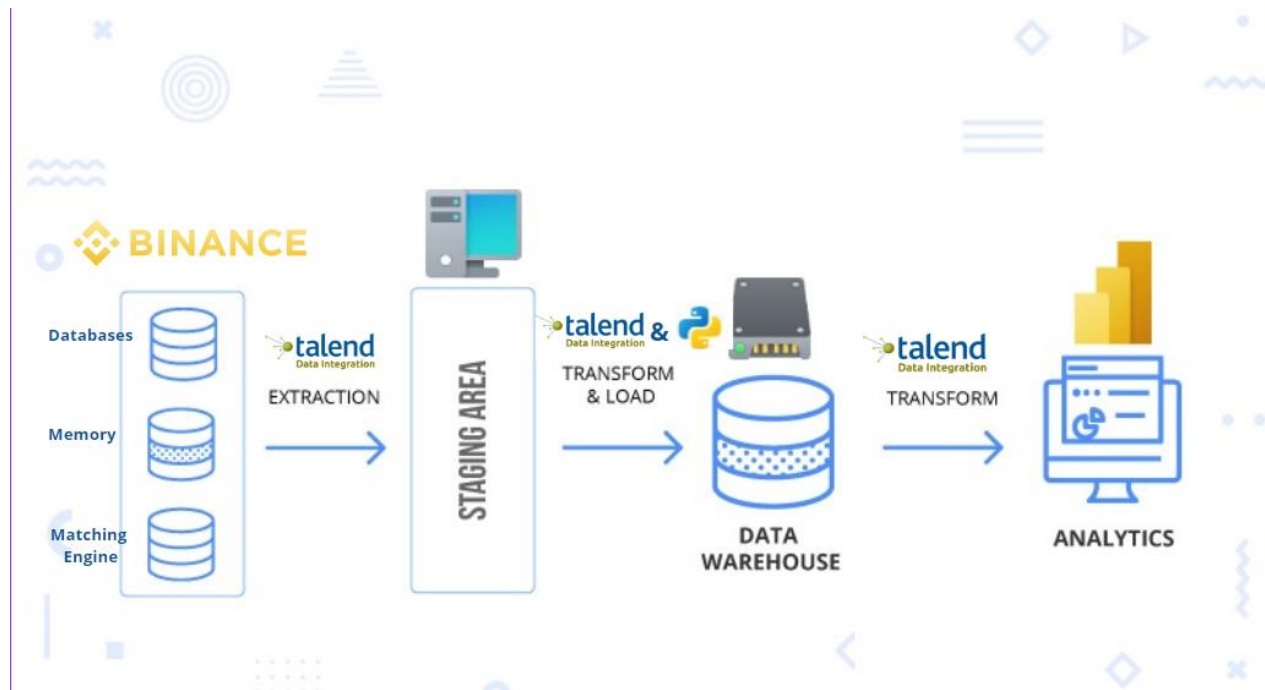
1.2 Context

- Migration de données à un nouveau système
- Stockage de données dans des entrepôts de données
- Consolidation de données (fusion d'entreprises)
- Synchronisation de données à temps réelle
- Lectures de différents DBs ou fichiers (csv, json, xlsx)
- Intégration décisionnelle et opérationnelle

1.3 Keywords

Talend - Python - Jobs - Talend Open Studio - ETL - Crypto - Cryptocurrency -

1.4 Schematic



1.5 Data Set

klines :

- Open/High/Low/Close : Prix à différents moments.
- Volume : Montant de l'actif échangé.
- Number of Trades : Nombre total de trades dans la bougie.

Dates :

- Date : toutes les dates de close price de 8ans dernière .
- Year : l'année de la date.
- Month : le mois de la date.
- Day : le jour de la date.
- Hour : l'heure de la date.

24hrs : • symbol : Symbole de la paire de trading (par exemple, ETHBTC).

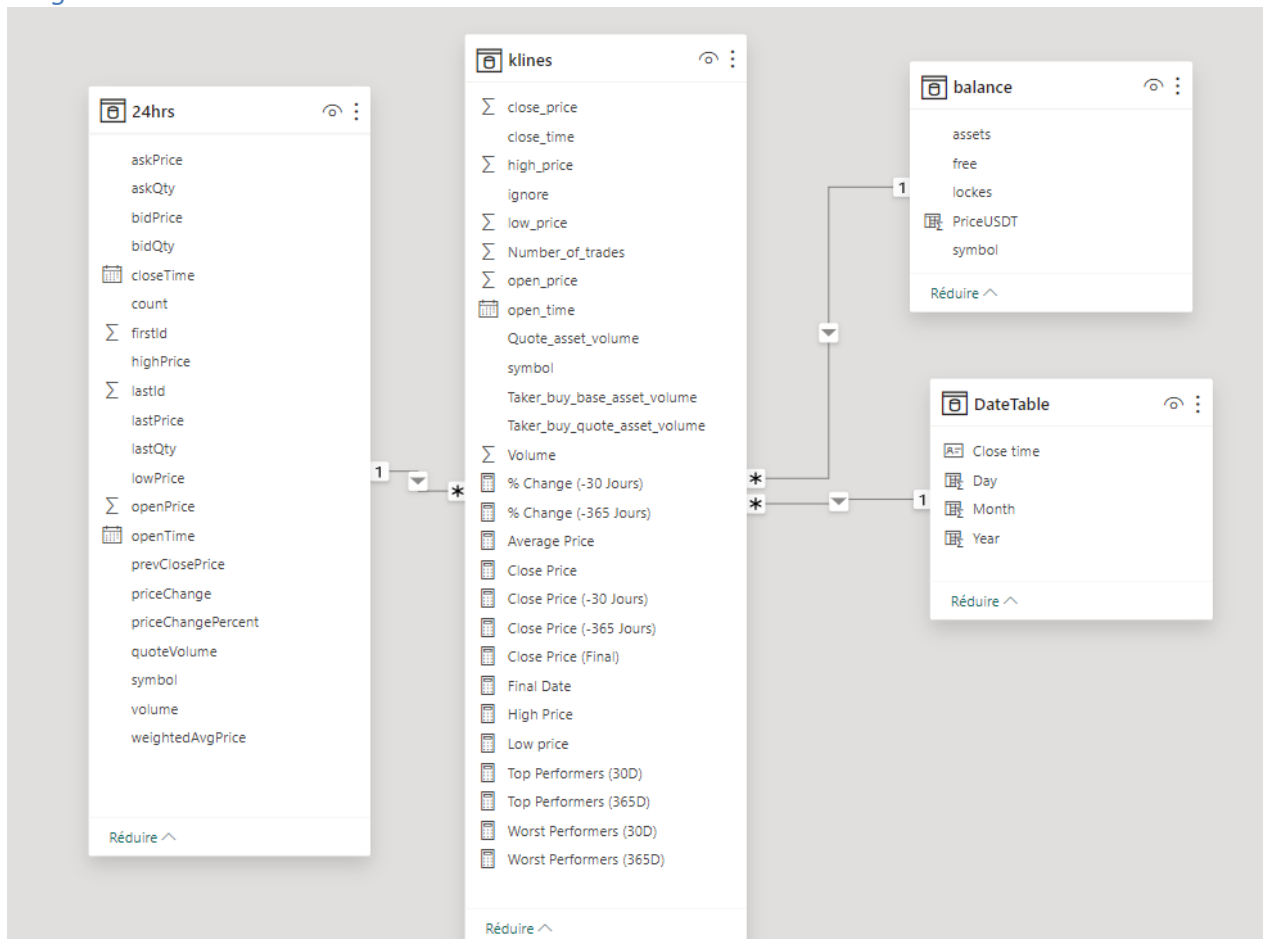
- priceChange : Changement de prix au cours des dernières 24 heures.
- priceChangePercent : Pourcentage de changement au cours des dernières 24 heures.
- lastPrice : Dernier prix négocié.

- volume : Volume total échangé au cours des dernières 24 heures

balance :

- Asset : le nom de la symbol de notre wallet.
- Free : Total de quantité de l'asset libre pour l'échange.
- Lock : Total de quantité de l'assets bloquées de l'échange.

1.6 Diagramme en étoile

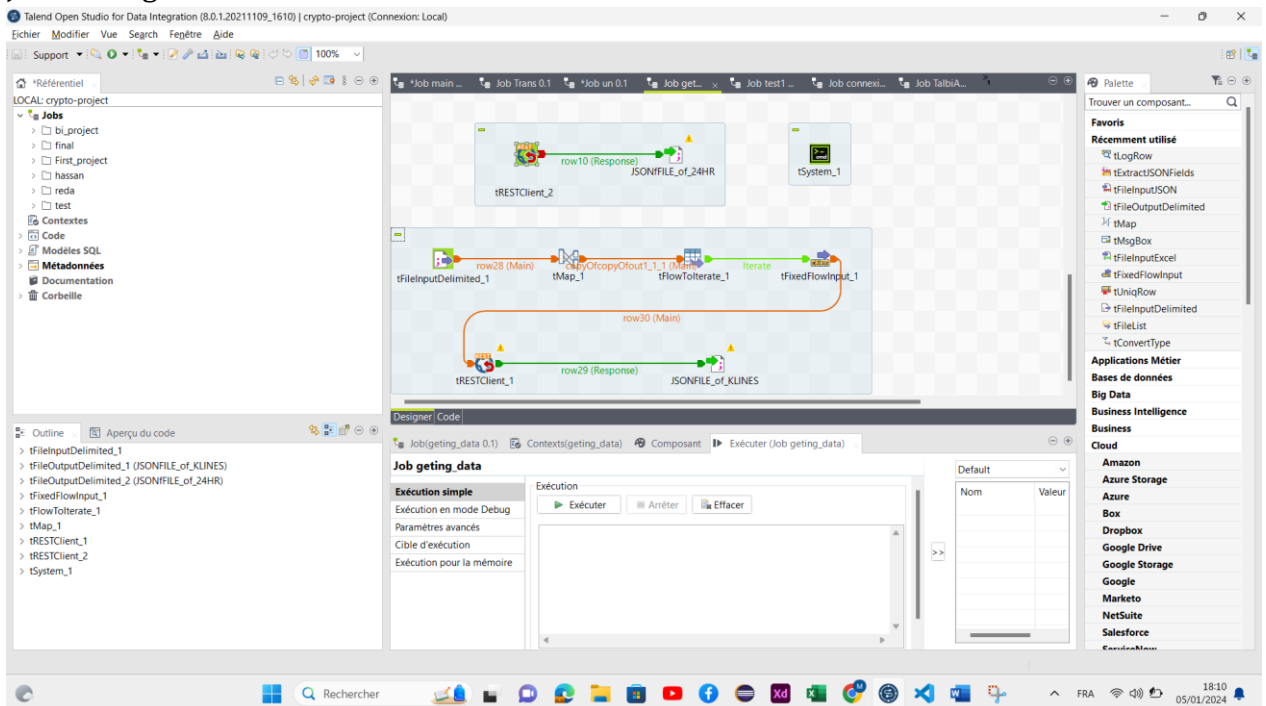


2. Jobs

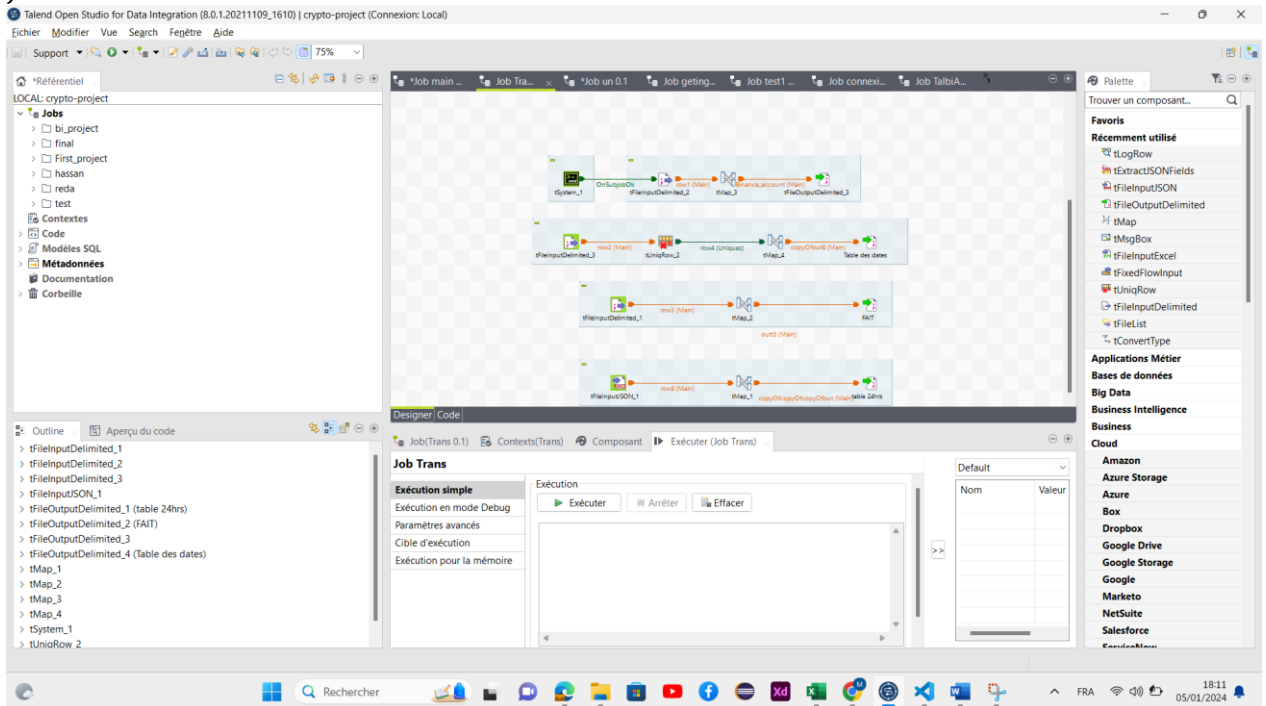
Jobs	Composant_Inputs	Composant_Outputs	Description du job
Getting Data	TrestClient or the endpoints : /api/v3/klines /api/v3/ticker/24hr /api/v3/account	24hr.json binance_account_info.json path ./jsonfiles_klines/	Getting the Data from binance API f
Transformation des données	24hr.json binance_account_info.json Allklines.json file_for_dates.csv	balance.csv klines.csv dates.csv 24hr.csv	The transformation phase involves transforming, cleaning, and normalizing our data.
Saving in the staging area	balance.csv klines.csv dates.csv 24hr.csv	Our BD	Alimentation de la table BD MYSQL klines and dates and 24hr balance à partir des fichiers csv
Main	All Jobs	TmessageBox	Execution des de processus

1.7 Visualisation des processus:

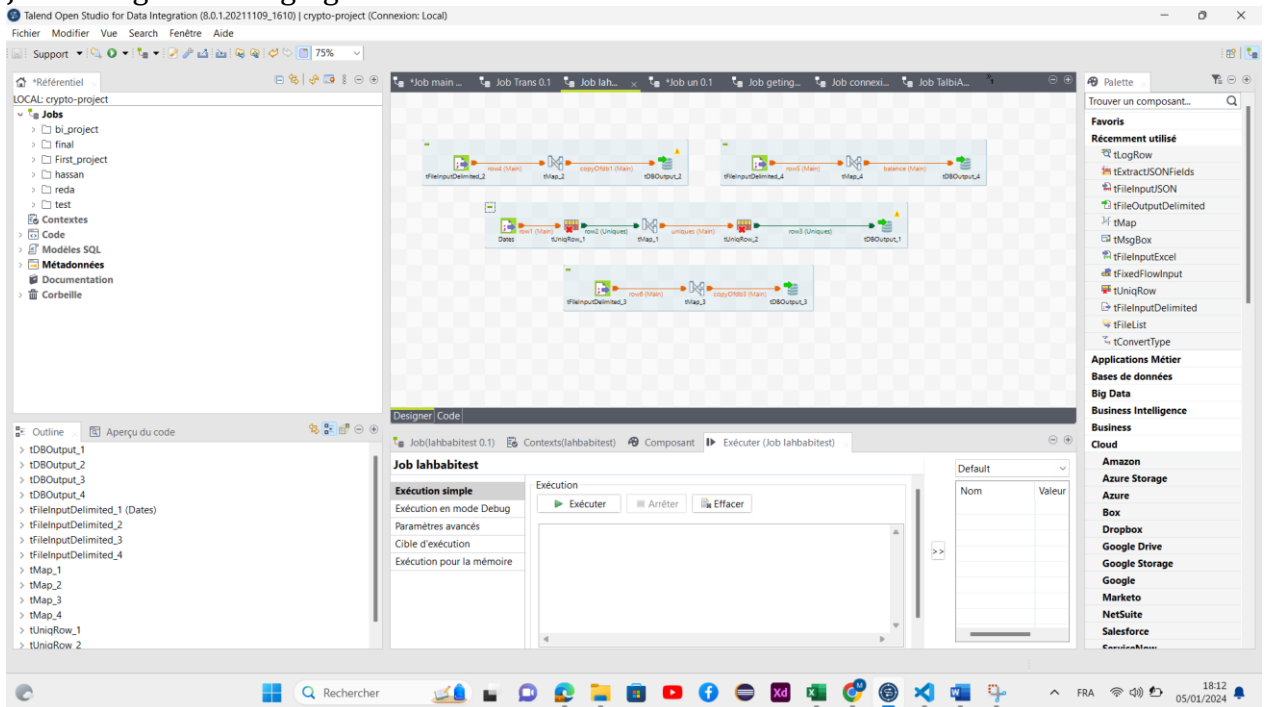
- Job Getting Data :



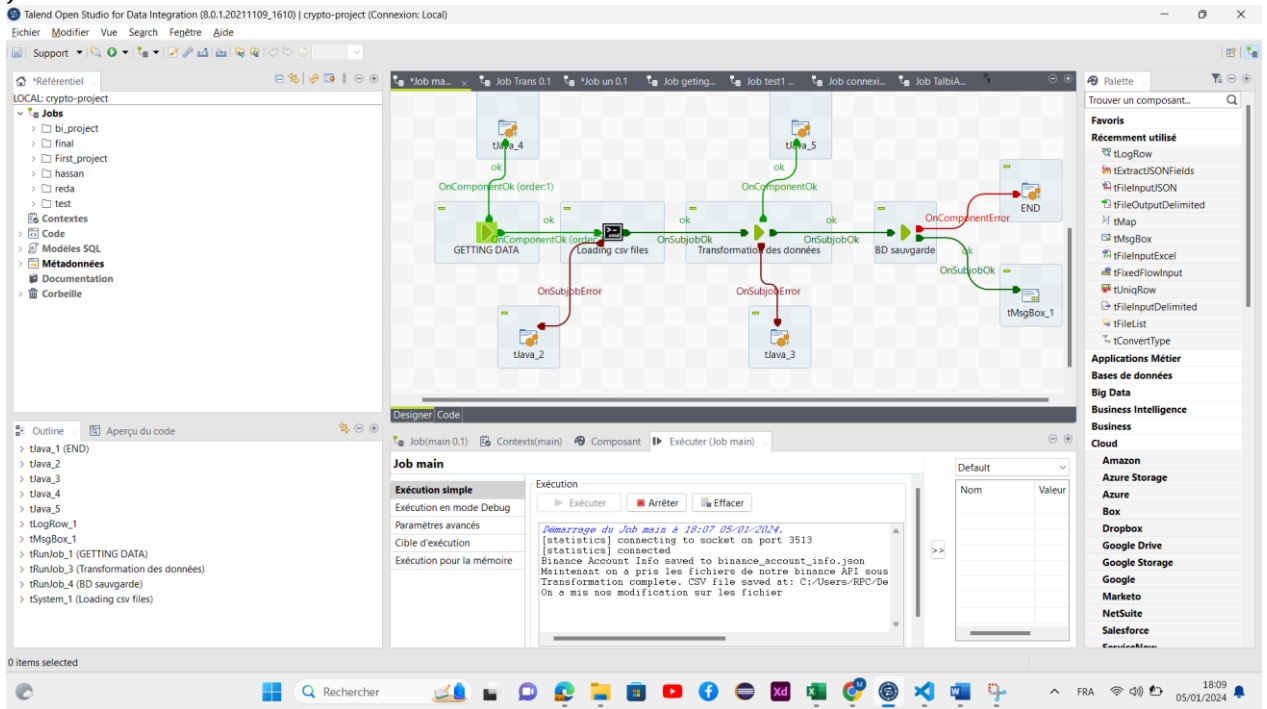
- **Job Transformation des données :**



- **Job Saving in the staging area:**



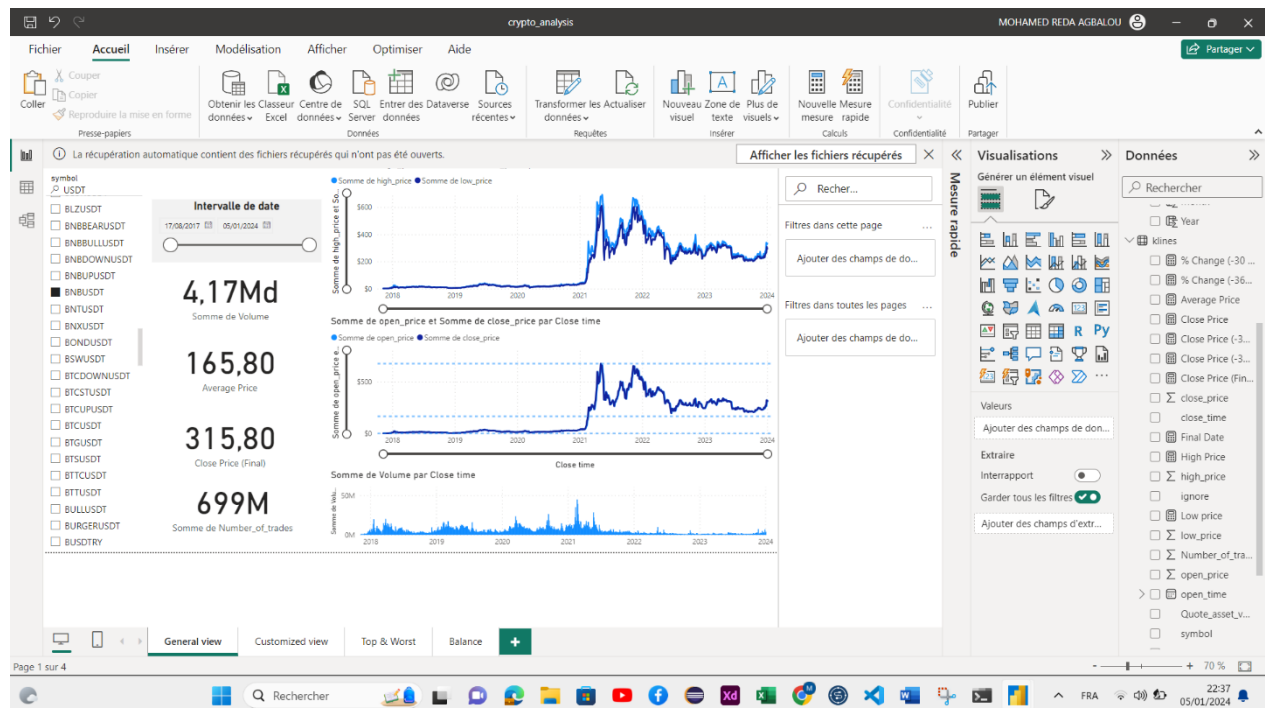
●



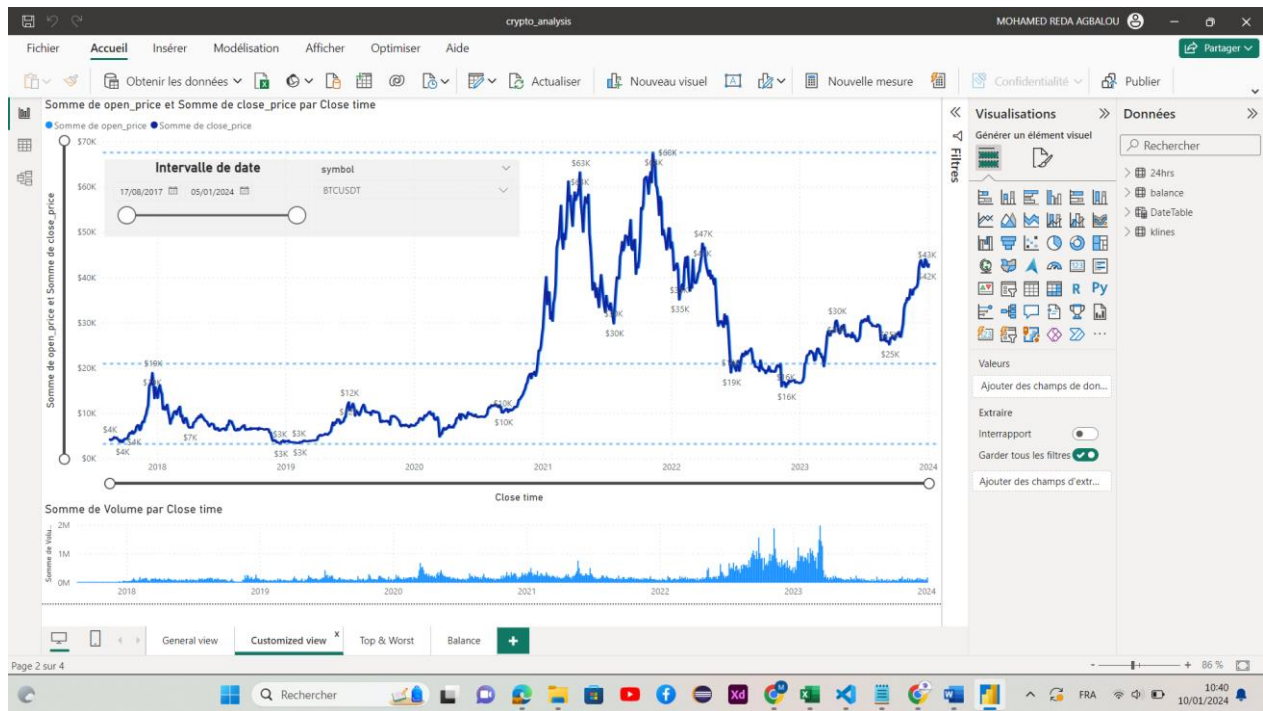
3. Visualisation Power Bi

Nos données sont prises directement de notre staging area, en utilisant un connecteur Mysql dans Power Bi.

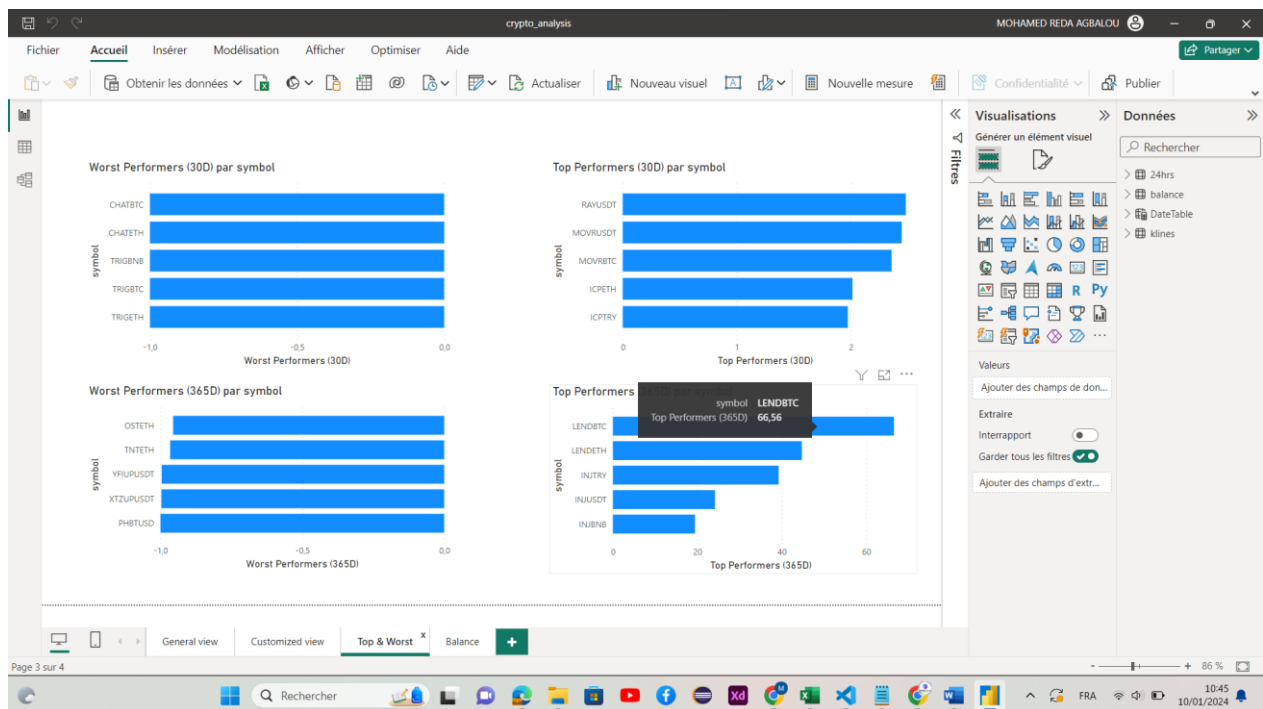
General view Dashboard : une vue général de sur toutes les symbols sur un interval de temps dont : leur prix d'ouverture et cloture , le minimum et le maximum , le volume durant cet intervalle , sans oublié d'autre indicateur comme la moyen de prix , prix final , nombre total des échanges et la somme de volume.



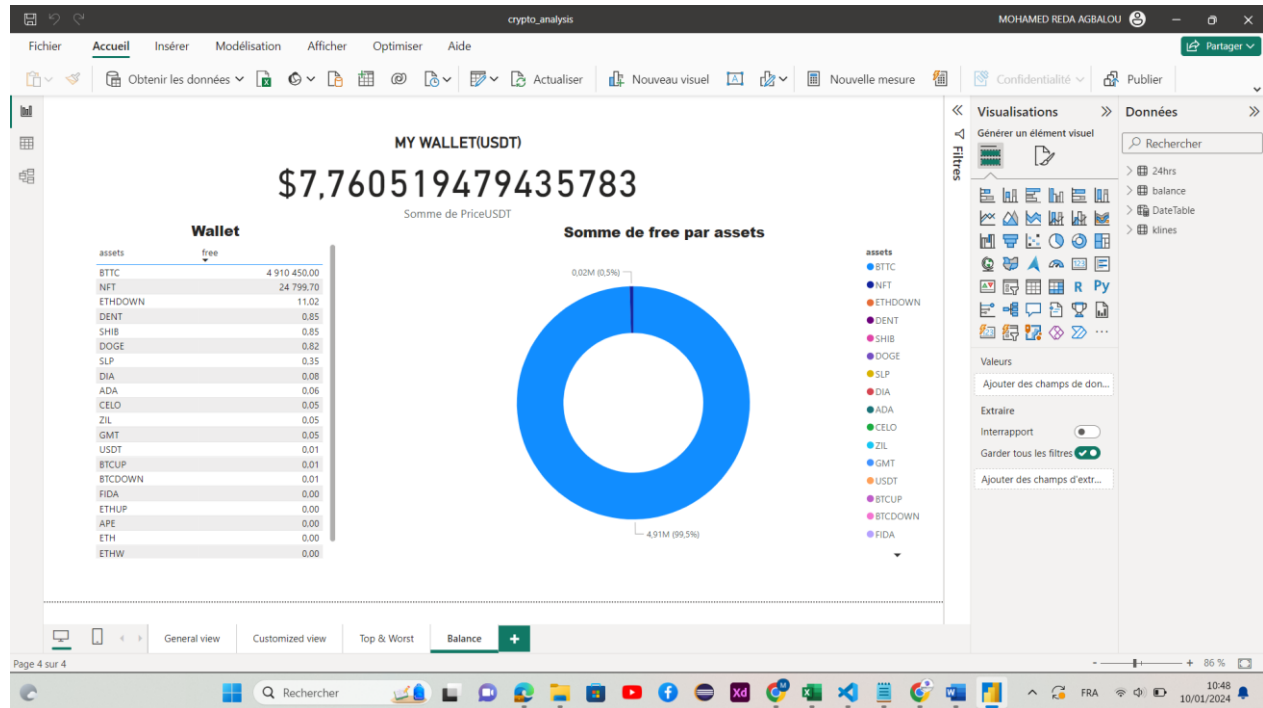
puis un dashboard personnalisé (Customized View) ou le client peut analyser un symbol de son choix d'une manière plus visible :



Le DashBoard Top&Worst représente le classement des 5 meilleurs et mauvais investissements sur le mois dernier et l'année dernière :



Et finalement une représentation de notre portefeuille c  d la quantit   de chaque symbole qu'on a :



4. *Prédiction* : Mon projet étale aussi la création d'un model Machine Learning basé sur nos données afin d'accompagner le client dans ses investissements.

