



REPUBLIQUE TUNISIENNE MINISTÈRE DE  
L'ENSEIGNEMENT SUPÉRIEUR  
ET DE LA RECHERCHE SCIENTIFIQUE  
UNIVERSITÉ TUNIS EL MANAR



FACULTÉ DES SCIENCES DE TUNIS  
DEPARTEMENT DES SCIENCES DE L'INFORMATIQUE

RAPPORT DE PROJET  
ENTREPÔT DE DONNÉES

**De la collecte des données à la  
visualisation :  
analyse décisionnelle de Bitcoin et  
Ethereum**

**Professeur:** Rihab MERSNI

**Réalisé par:**

**Koussay Dhifi  
Yassine Drira**

**Année universitaire:  
2025-2026**

## 1. Introduction

Les crypto-monnaies génèrent aujourd'hui un volume important de données issues de sources variées telles que les marchés, la blockchain et le sentiment des investisseurs.

**L'exploitation efficace de ces données** nécessite une approche décisionnelle structurée.

Le projet **CryptoDW** vise à mettre en place un Data Warehouse dédié à l'analyse de **Bitcoin (BTC)** et **Ethereum (ETH)**, deux crypto-monnaies majeures du marché. Les données collectées sont centralisées, transformées et organisées afin de **faciliter l'analyse et la comparaison des performances**. Le projet s'appuie sur une architecture décisionnelle complète intégrant un **cube OLAP** et **des tableaux de bord Power BI**.

### Questions d'analyse du projet

Pour ce projet, nous avons cherché à répondre aux questions suivantes :

1. Quelle est l'évolution des prix et des volumes de Bitcoin (BTC) et Ethereum (ETH) par année ?
2. Quelle est l'activité on-chain (adresses actives, transactions) pour chaque crypto-monnaie ?
3. Comment se répartit l'activité de trading entre BTC et ETH selon le type d'opération ?
4. Quelle est l'évolution du sentiment du marché au fil du temps ?
5. Existe-t-il une relation entre le sentiment du marché et les variations de prix ?
6. Quelle période a connu la meilleure performance combinée en termes de prix, d'activité on-chain et de sentiment positif ?

## 2. Description et périmètre du projet

Le projet CryptoDW vise à centraliser et structurer les données de Bitcoin (BTC) et Ethereum (ETH) pour faciliter leur analyse. L'objectif est de comparer les crypto-monnaies, suivre l'évolution des prix et volumes, analyser l'activité on-chain et mesurer l'impact du sentiment du marché.

Le périmètre se limite à :

**Crypto-monnaies : BTC et ETH**

**Données : marché, on-chain, sentiment**

**Période : selon disponibilité et fiabilité des données**

### 3. Sources de données

#### 3.1 Types et sources des données

Pour CryptoDW, nous avons utilisé des données fiables et presque prêtes à l'emploi, couvrant la période janvier à novembre 2025, avec trois types principaux :

Données de marché (prix et trades)

The screenshot shows the Binance homepage. On the left, a large yellow banner displays the text "301,440,152 USERS TRUST US" and "The World's Leading Cryptocurrency Exchange". Below this, there are two accolades: "No.1 Customer Assets" and "No.1 Trading Volume", each accompanied by a laurel wreath icon. On the right, there is a "Popular" coin list with the following data:

Coin	Price	Change (%)
BTC Bitcoin	\$87,656.09	+3.19%
ETH Ethereum	\$2,978.55	+6.27%
BNB BNB	\$854.85	+3.64%
XRP XRP	\$1.91	+4.93%
ASTER Aster	\$0.719	+7.80%

Below the coin list, there is a "News" section with several headlines:

- Significant LDO Transfer Between Anonymous Addresses
- Federal Reserve Governor Emphasizes Importance of Easing Policies for Labor Market
- Ethereum Price Movements Could Trigger Significant Liquidations
- Bitcoin(BTC) Drops Below 87,000 USDT with a Narrowed 0.60% Increase in 24 Hours

#### Prices.csv : Binance Klines Monthly



The screenshot shows the Binance Market Data page for the BTCUSDT pair. At the top, it says "Market Data". Below that, the URL is shown: "Home / data / spot / monthly / klines / BTCUSDT / 1mo /". The main content is a table listing various monthly kline files:

Item	Size	Last Modified
..		
<a href="#">BTCUSDT-1mo-2025-11.zip.CHECKSUM</a>	0.1 kB	2025-12-01T08:14:38.000Z
<a href="#">BTCUSDT-1mo-2025-11.zip</a>	0.2 kB	2025-12-01T08:14:38.000Z
<a href="#">BTCUSDT-1mo-2025-10.zip.CHECKSUM</a>	0.1 kB	2025-11-03T08:44:12.000Z
<a href="#">BTCUSDT-1mo-2025-10.zip</a>	0.2 kB	2025-11-03T08:44:12.000Z
<a href="#">BTCUSDT-1mo-2025-09.zip.CHECKSUM</a>	0.1 kB	2025-10-08T11:53:21.000Z

Méthode : nous avons téléchargé les fichiers mensuels pour chaque crypto séparément puis fusionné l'ensemble pour obtenir un dataset complet sur toute la période

# Trades : Binance AggTrades Monthly



Item	Size	Last Modified
<a href="#">BTCUSDT-aggTrades-2025-11.zip.CHECKSUM</a>	0.1 kB	2025-12-01T09:30:23.000Z
<a href="#">BTCUSDT-aggTrades-2025-11.zip</a>	587.6 MB	2025-12-01T09:30:20.000Z
<a href="#">BTCUSDT-aggTrades-2025-10.zip.CHECKSUM</a>	0.1 kB	2025-11-03T10:07:47.000Z
<a href="#">BTCUSDT-aggTrades-2025-10.zip</a>	561.8 MB	2025-11-03T10:07:44.000Z
<a href="#">BTCUSDT-aggTrades-2025-09.zip.CHECKSUM</a>	0.1 kB	2025-10-06T11:34:42.000Z

## API : blockchain metrics

CoinGecko API & Coin metrics API



## Build, Analyze & Scale with Crypto's Most Trusted & Comprehensive Data API

Get complete coverage of the crypto market delivered by CoinGecko API,  
the world's largest independent crypto data authority since 2014.

[Get Your API Key Now](#)

[Explore Docs >](#)

Get started for free, no credit card required



```
def fetch(asset):
    url = "https://community-api.coinmetrics.io/v4/timeseries/asset-metrics"
    params = {
        "assets": asset,
        "metrics": "TxCnt,AdrActCnt,BlkCnt,FeeMean,FeeTotUSD",
        "frequency": "1d",
        "start_time": "2025-01-01" # forcing 2025 for consistency
    }
```

## Données de sentiment:

News : simulation à partir de la base **CryptoSourceDW**



Source:kaggle (crpto news+) & Bitcoin News Data (GitHub) (2025)

The screenshot shows a GitHub repository page for 'mouadja02/bitcoin-news-data'. The repository is public and has one branch ('main') and no tags. It contains three files: '.github/workflows', 'datasets', and 'snowflake-data-sync.yml'. The '.github/workflows' file was updated 5 months ago, and the 'datasets' file was updated 3 days ago. A commit from 'actions-user' was made on December 16, 2025, updating daily data. The URL of the repository is <https://github.com/mouadja02/bitcoin-news-data>.

kaggle

Competitions Datasets Models Code Discussions Blog Courses ...

Search

Sign In

Register

## Level up with the largest AI & ML community

Join over 28M+ machine learners to share, stress test, and stay up-to-date on all the latest ML techniques and technologies. Discover a huge repository of community-published models, data & code for your next project.

Register with Google

Register with Email



## Crypto News +

Crypto news articles containing title, text, and the sentiment analysis.

**CRYPTO NEWS+** Get the latest articles

Data Card Code (367) Discussion (1) Suggestions (0)

### About Dataset

Usability

8.24

...

```
CREATE TABLE News(
    id INT IDENTITY(1,1) PRIMARY KEY,
    publish_time DATE,
    coin_symbol NVARCHAR(10),
    title NVARCHAR(400),
    subject NVARCHAR(150),
    sentiment_class NVARCHAR(20),
    polarity FLOAT,
    subjectivity FLOAT);
```

# Fear & Greed : Coinglass Fear & Greed Index (JSON transformé)



## 3.2 Problèmes rencontrés

Les données choisies sont de bonne qualité et quasi prêtes à l'emploi, mais quelques ajustements ont été nécessaires :

- Harmonisation des formats et unités entre sources
- Fusion des fichiers mensuels pour chaque crypto afin d'obtenir un dataset continu
- Vérification de la cohérence sur toute la période (janvier à novembre 2025)

## 4. Architecture globale

### 4.1 Vue d'ensemble

L'architecture de CryptoDW suit un flux simple et efficace :

Sources → Staging → Data Warehouse → Cube OLAP → Power BI

- Sources : données marché, on-chain et sentiment, provenant d'API, CSV, JSON ou base externe
- Staging : tables temporaires (Stg\*) pour nettoyer, harmoniser et consolider les données
- Data Warehouse : stockage structuré en schéma en étoile, avec dimensions et faits
- Cube OLAP : pour accélérer les analyses et permettre des calculs rapides
- Power BI : tableau de bord interactif et visualisations décisionnelles

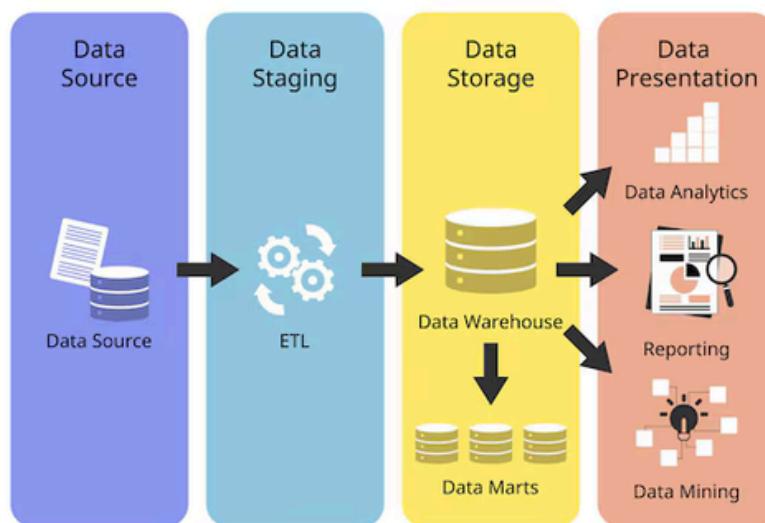
### 4.2 Outils utilisés

- **SQL Server** : stockage et gestion des données
- **SSIS** : processus ETL (extraction, transformation, chargement)
- **SSAS** : création et gestion du cube OLAP
- **Power BI** : tableaux de bord et analyses interactives

#### 4.3 Schéma d'architecture



#### Data Warehouse Architecture



## 5. Processus ETL

**5.1. Étape 1 : Extraction:** Le but de cette étape est la récupération des données à partir de chaque source.

Les données sont extraites de plusieurs sources hétérogènes pour assurer une couverture complète des informations nécessaires au Data Warehouse **CryptoDW** :

Les données extraites sont chargées dans des tables de staging pour préparer les étapes de transformation et de chargement. Chaque source est traitée via un flux de données spécifique dans **SSIS**.

### • Staging

```
-- Staging Price Daily
CREATE TABLE StgPriceDaily (
    full_date DATE,
    open_price DECIMAL(18,8),
    high_price DECIMAL(18,8),
    low_price DECIMAL(18,8),
    close_price DECIMAL(18,8),
    volume DECIMAL(18,8),
    quote_volume DECIMAL(18,8),
    trades_count INT,
    taker_buy_volume DECIMAL(18,8),
    taker_buy_quote DECIMAL(18,8),
    coin_symbol NVARCHAR(20)
);

-- Staging Trades
CREATE TABLE StgTrades (
    aggId BIGINT PRIMARY KEY,
    price DECIMAL(18,8),
    quantity DECIMAL(18,8),
    trade_date DATE,
    buyerMaker BIT,
    coin_symbol NVARCHAR(20)
);

-- Staging News
CREATE TABLE StgNews (
    publish_date DATE,
    coin_symbol NVARCHAR(20),
    title NVARCHAR(400),
    subject NVARCHAR(150),
    sentiment_class NVARCHAR(50),
    polarity FLOAT,
    subjectivity FLOAT
);

-- Staging Blockchain / Metrics
CREATE TABLE StgBlockchain (
    full_date DATE,
    coin_symbol NVARCHAR(20),
    active_addresses INT,
    block_count INT,
    tx_count INT
);

-- Staging Market Live
CREATE TABLE StgMarketLive (
    snapshot_date DATE,
    coin_symbol NVARCHAR(20),
    price_usd DECIMAL(18,8),
    market_cap DECIMAL(38,2),
    volume_24h DECIMAL(18,8),
    change_24h FLOAT
);

-- Staging Mood / Fear & Greed
CREATE TABLE StgMood (
    full_date DATE,
    fear_greed_value INT,
    classification NVARCHAR(80)
);
```

### 5.2 Traitement de la zone de Staging

Toutes les données sources ont été chargées directement dans des tables de staging (**StgPriceDaily**, **StgMarketLive**, **StgTrades**, **StgNews**, **StgBlockchain**, **StgMood**).

Chaque table de staging conserve la structure originale des sources (fichiers CSV,json ou données extraites via API), sans modification du schéma.

La zone de staging joue le rôle d'un espace de stockage temporaire permettant de conserver les données brutes avant leur intégration dans le Data Warehouse.

Étant donné que les données sources étaient déjà propres, complètes et bien structurées (absence de valeurs nulles ou manquantes), aucune transformation majeure n'a été appliquée à ce niveau.

Ce choix de chargement direct permet d'optimiser les performances, de réduire la complexité des flux SSIS et de faciliter la maintenance globale du processus ETL.

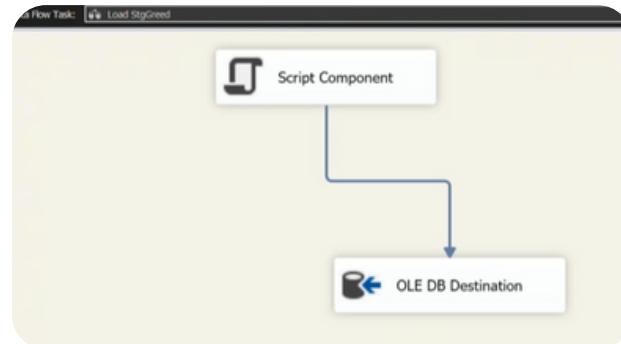


Le chargement de la table **StgMarket** est automatisé toutes les 24 heures via **SQL Server Agent(job)**, qui exécute un **script Python** pour récupérer et insérer les données de marché directement dans la zone de staging.



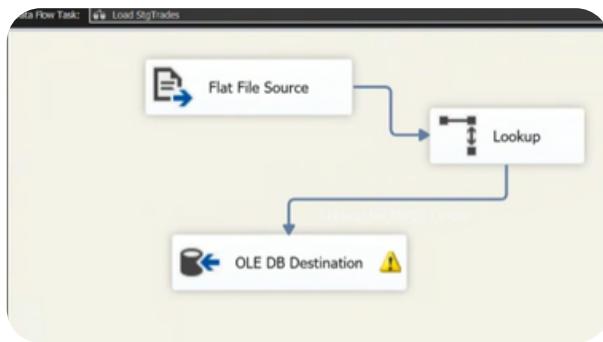
**Stgblockchain**

du fichier plat vers stgblockchain



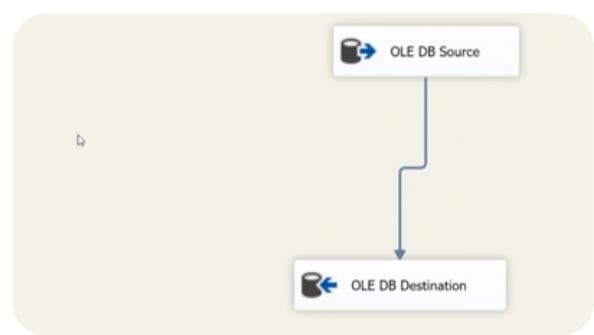
**StgGreed**

Extraction des données à partir du fichier json via un script avec C#



**StgTrades**

Elimination des doublons avant du redémarrer le chargement une autre fois



**StgNews**

Importation des données à partir d'une autre base de donnée externe

## 5.2 Transformation:

- **Dimensions**

```
- CREATE TABLE DimDate (
    date_id INT IDENTITY(1,1) PRIMARY KEY,
    full_date DATE UNIQUE NOT NULL,
    day INT,
    month INT,
    year INT,
    week INT
);

- CREATE TABLE DimCoin (
    coin_id INT IDENTITY(1,1) PRIMARY KEY,
    coin_symbol NVARCHAR(20) UNIQUE NOT NULL,
    coin_name NVARCHAR(100)
);

- CREATE TABLE DimSentiment (
    sentiment_id INT IDENTITY(1,1) PRIMARY KEY,
    sentiment_class NVARCHAR(50),
    min_polarity FLOAT NOT NULL,
    max_polarity FLOAT NOT NULL
);

- CREATE TABLE DimMood (
    mood_id INT IDENTITY(1,1) PRIMARY KEY,
    classification NVARCHAR(50),
    min_value INT,
    max_value INT
);

- CREATE TABLE DimTradeType (
    trade_type_id INT IDENTITY(1,1) PRIMARY KEY,
    label NVARCHAR(20)
);
```

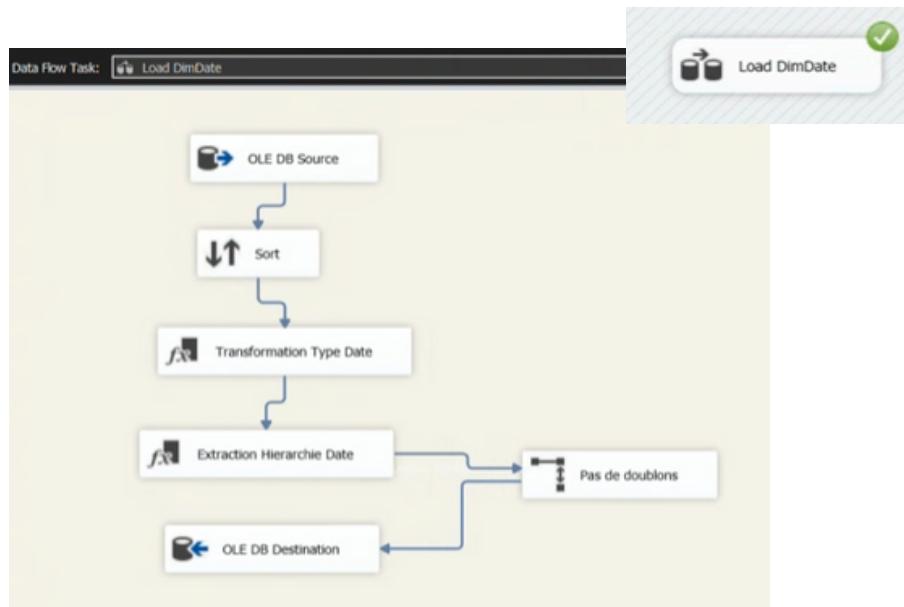
- **DimDate**

La dimension Date (DimDate) est la seule dimension dynamique du Data Warehouse et nécessite un traitement ETL spécifique.

Les dates ont été extraites à partir de plusieurs tables de staging, notamment celles contenant les données journalières de prix, les snapshots du marché, les indicateurs on-chain, les données de sentiment et les transactions.

**Les champs de type timestamp ont été convertis en format DATE standard.**

- **Les étapes ETL appliquées à DimDate sont les suivantes :**
- **Collecte des dates provenant de l'ensemble des sources**
- **Tri des données et suppression des doublons**
- **Conversion du champ date vers un type standard**
- **Extraction des attributs temporels (année, mois, jour, semaine)**
- **Chargement dans la table DimDate**



Certaines dimensions du Data Warehouse CryptoDW contiennent des données de référence stables et ont été chargées directement via SQL, sans utiliser de flux SSIS.

- **DimCoin**

Coin_id	Coin_symbol	Coin_name
1	BTC	BITCOIN
2	ETH	ETHEREUM

- **DimTradeType**

trade_type_id	label
1	BUY
2	SELL

- **DimMood**

mood_id	classification	min_value	max_value
1	ETREME_FEAR	0	25
2	FEAR	26	45
3	NEUTRAL	46	56
4	GREED	55	75
5	EXTREME_GREED	76	100

- **Dimsentiment**

sentiment_id	sentiment_class	min_polarity	max_polarity
1	NEGATIVE	-1	-0.05
2	NEUTRAL	-0.05	0.05
3	POSITIVE	0.05	1

- **Facts**

**Les tables de faits ont été alimentées selon une logique ETL homogène, adaptée à chaque type de données.**

**Pour chaque table :**

**Les clés substituts (date\_id et coin\_id) ont été récupérées via des Lookups vers DimDate et DimCoin.**

**Des transformations simples ont été effectuées : normalisation du texte (UPPER) et conversion des types de données.**

**Aucune opération de nettoyage complexe n'a été nécessaire, les données sources étant déjà propres.**

```

CREATE TABLE FactPriceDaily (
    price_id INT IDENTITY(1,1) PRIMARY KEY,
    date_id INT NOT NULL,
    coin_id INT NOT NULL,
    open_price FLOAT,
    close_price FLOAT,
    high_price FLOAT,
    low_price FLOAT,
    volume FLOAT,
    trades_count INT,
    FOREIGN KEY (date_id) REFERENCES DimDate(date_id),
    FOREIGN KEY (coin_id) REFERENCES DimCoin(coin_id)
);

CREATE TABLE FactTrades (
    trade_id BIGINT IDENTITY(1,1) PRIMARY KEY,
    date_id INT NOT NULL,
    coin_id INT NOT NULL,
    price FLOAT,
    quantity FLOAT,
    trade_type_id INT,
    trade_value AS (price * quantity) PERSISTED,
    FOREIGN KEY (date_id) REFERENCES DimDate(date_id),
    FOREIGN KEY (coin_id) REFERENCES DimCoin(coin_id),
    FOREIGN KEY (trade_type_id) REFERENCES DimTradeType(trade_type_id)
);

CREATE TABLE FactNewsSentiment (
    news_id INT IDENTITY(1,1) PRIMARY KEY,
    date_id INT NOT NULL,
    coin_id INT NOT NULL,
    sentiment_id INT NOT NULL,
    title NVARCHAR(255),
    polarity FLOAT,
    subjectivity FLOAT,
    FOREIGN KEY (date_id) REFERENCES DimDate(date_id),
    FOREIGN KEY (coin_id) REFERENCES DimCoin(coin_id),
    FOREIGN KEY (sentiment_id) REFERENCES DimSentiment(sentiment_id)
);

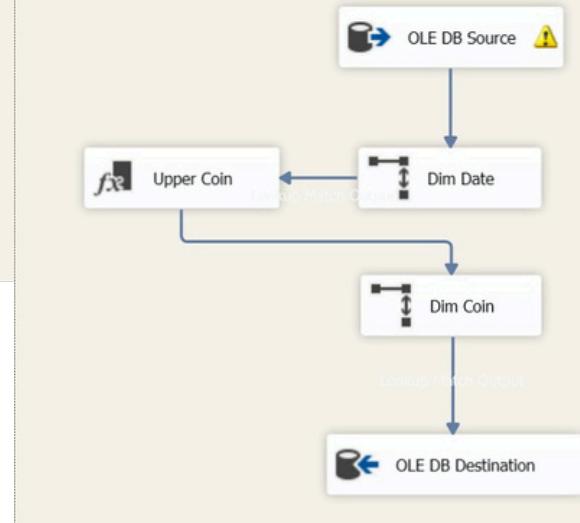
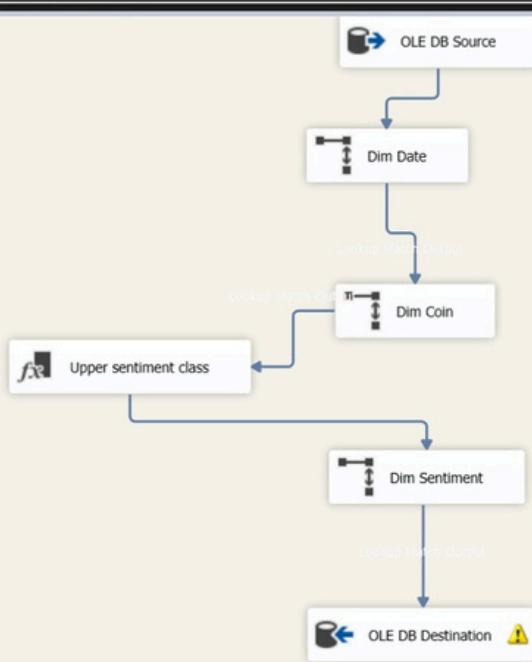
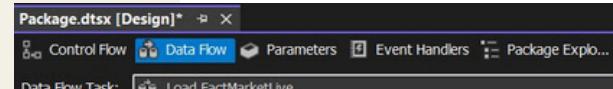
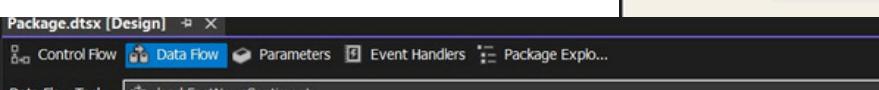
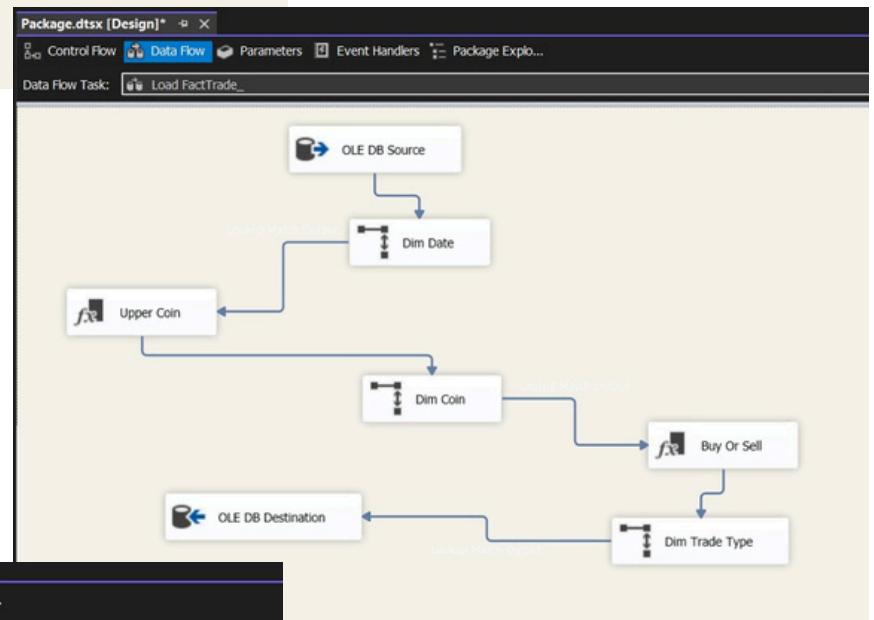
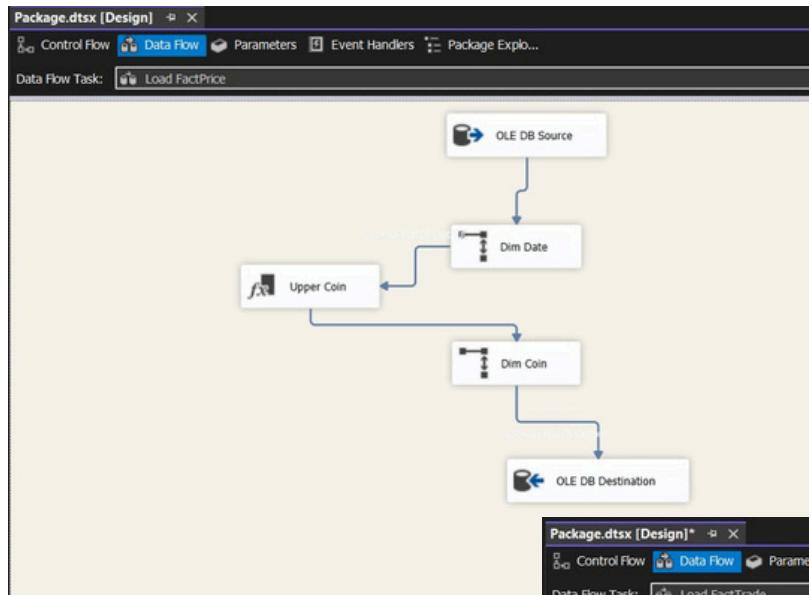
CREATE TABLE FactBlockchainActivity (
    blockchain_id INT IDENTITY(1,1) PRIMARY KEY,
    date_id INT NOT NULL,
    coin_id INT NOT NULL,
    active_addresses INT,
    block_count INT,
    tx_count INT,
    FOREIGN KEY (date_id) REFERENCES DimDate(date_id),
    FOREIGN KEY (coin_id) REFERENCES DimCoin(coin_id)
);

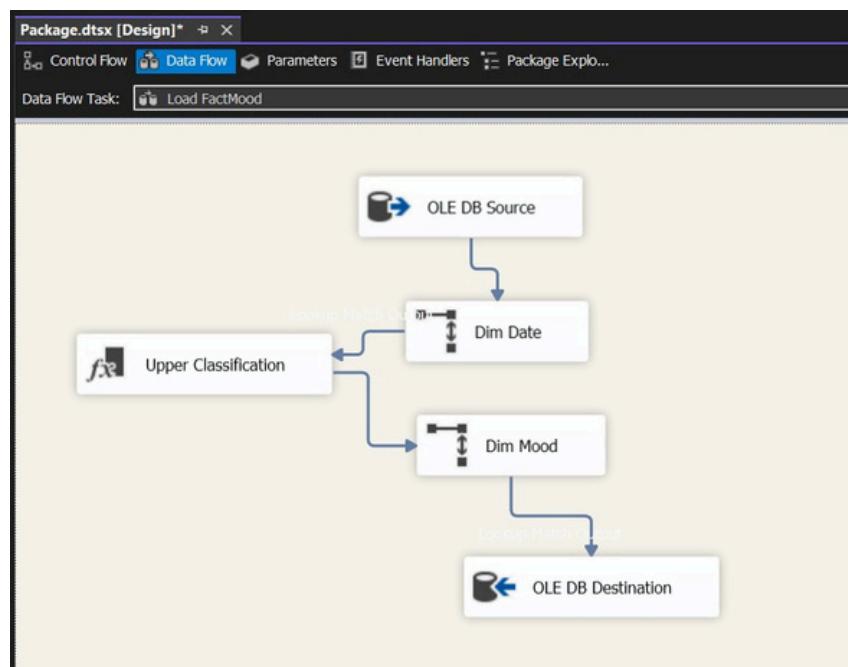
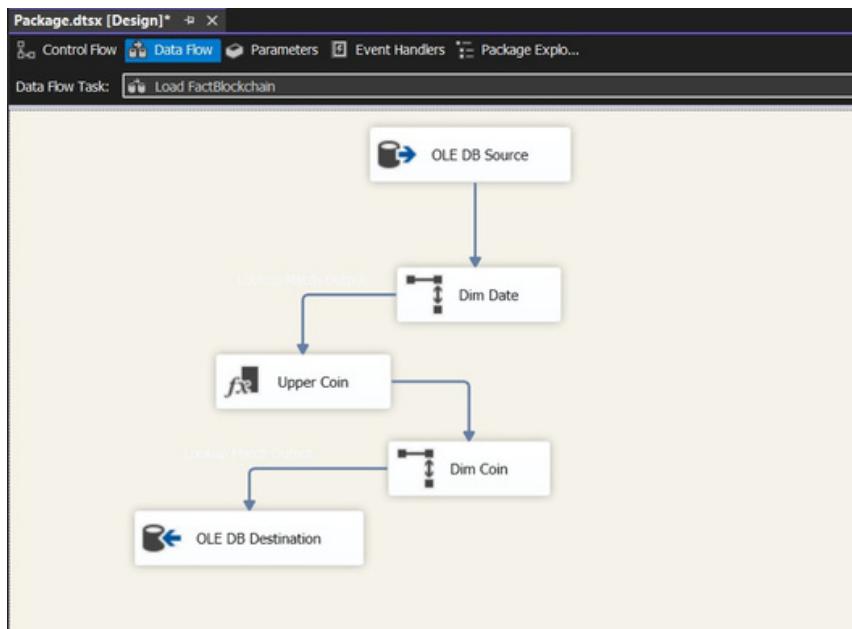
CREATE TABLE FactMarketSnapshot (
    snapshot_id INT IDENTITY(1,1) PRIMARY KEY,
    date_id INT NOT NULL,
    coin_id INT NOT NULL,
    price_usd FLOAT,
    market_cap FLOAT,
    volume_24h FLOAT,
    change_24h FLOAT,
    FOREIGN KEY (date_id) REFERENCES DimDate(date_id),
    FOREIGN KEY (coin_id) REFERENCES DimCoin(coin_id)
);

CREATE TABLE FactMoodIndex (
    mood_fact_id INT IDENTITY(1,1) PRIMARY KEY,
    date_id INT NOT NULL,
    mood_id INT NOT NULL,
    greed_value INT,
    FOREIGN KEY (date_id) REFERENCES DimDate(date_id),
    FOREIGN KEY (mood_id) REFERENCES DimMood(mood_id)
);

CREATE INDEX IX_Price_Date ON FactPriceDaily(date_id);
CREATE INDEX IX_Trades_Date ON FactTrades(date_id);
CREATE INDEX IX_Trades_Coin ON FactTrades(coin_id);
CREATE INDEX IX_Market_Coin ON FactMarketSnapshot(coin_id);
CREATE INDEX IX_Blockchain_Date ON FactBlockchainActivity(date_id);
CREATE INDEX IX_News_Coin ON FactNewsSentiment(coin_id);

```





OLE DB Destination Editor

Configure the properties used to insert data into a relational database using an OLE DB provider.

Connection Manager: Mappings Error Output

Available Input... Available ...

Input Column	Destination Column
date_id	date_id
coin_id	coin_id
<ignore>	trade_id
price	price
quantity	quantity
trade_type_id	trade_type_id
<ignore>	trade_value

Editor

the performance of simple equi-joins between the input and a reference data set.

Available Input... Available Lookup Columns

Lookup Column	Lookup Operation	Output Alias
date_id	<add as new column>	date_id

**Tables de faits et justification :**

### **FactPriceDaily**

**Stocke les prix journaliers (open, close, high, low), le volume et le nombre de transactions.**

**Justification : Analyse de l'évolution historique des prix et des tendances du marché par cryptomonnaie.**

### **FactTrades**

**Enregistre chaque transaction avec type, quantité et valeur totale.**

**Justification : Étude du comportement des investisseurs et de l'activité de trading.**

### **FactNewsSentiment**

**Contient les articles et leur sentiment (polarity, subjectivity).**

**Justification : Relie l'information médiatique aux fluctuations du marché et aux perceptions des investisseurs.**

### **FactBlockchainActivity**

**Regroupe les indicateurs on-chain : adresses actives, nombre de blocs et transactions.**

**Justification : Analyse de l'activité réelle sur la blockchain et corrélation avec le marché.**

### **FactMarketSnapshot**

**Stocke les snapshots en temps réel : prix, capitalisation, volume 24h, variation.**

**Justification : Suivi de la performance du marché et de sa volatilité.**

### **FactMoodIndex**

**Enregistre l'indice Fear & Greed quotidien.**

**Justification : Mesure du sentiment global du marché et anticipation des comportements extrêmes.**

- Intégrité des données et gestion des doublons**

**1. Le tri et l'élimination des doublons lors du chargement de DimDate.**

**2. L'utilisation de Lookups pour associer correctement dimensions et faits.**

**3. Un ordre de chargement strict : dimensions avant faits.**

**4. Cette méthode assure des relations cohérentes entre dimensions et faits dans le Data Warehouse.**

## **Étape 5.3 : Chargement**

Le chargement dans les tables finales est effectué par les requêtes SQL , exécutées dans SSIS via des Execute SQL Tasks après les Data Flow Tasks

## 6. Data warehouse

Suite au processus ETL, les données transformées sont chargées dans un Data Warehouse relationnel structuré selon un schéma en étoile. Ce référentiel central consolide les données provenant de sources hétérogènes et constitue la fondation sur laquelle sera construit le cube OLAP.

### 6.1 Architecture et Rôle du Data Warehouse

Le Data Warehouse occupe une position centrale dans l'architecture décisionnelle du projet:

**Sources de données → ETL (SSIS) → Data Warehouse (SQL Server) → Cube OLAP (SSAS)**  
→ Dashboards (Power BI)

Rôle du Data Warehouse :

- **Centralisation** : Regroupement de toutes les données marché, blockchain, sentiment et comportement
- **Historisation** : Conservation de l'évolution temporelle pour analyses de tendances
- **Préparation** : Structure relationnelle optimale pour la construction du cube OLAP
- **Cohérence** : Source unique de vérité garantissant l'intégrité des données

### 6.2 Modèle Relationnel : Schéma en Étoile

Le Data Warehouse est organisé selon un schéma en étoile (star schema), architecture standard pour l'analyse multidimensionnelle :

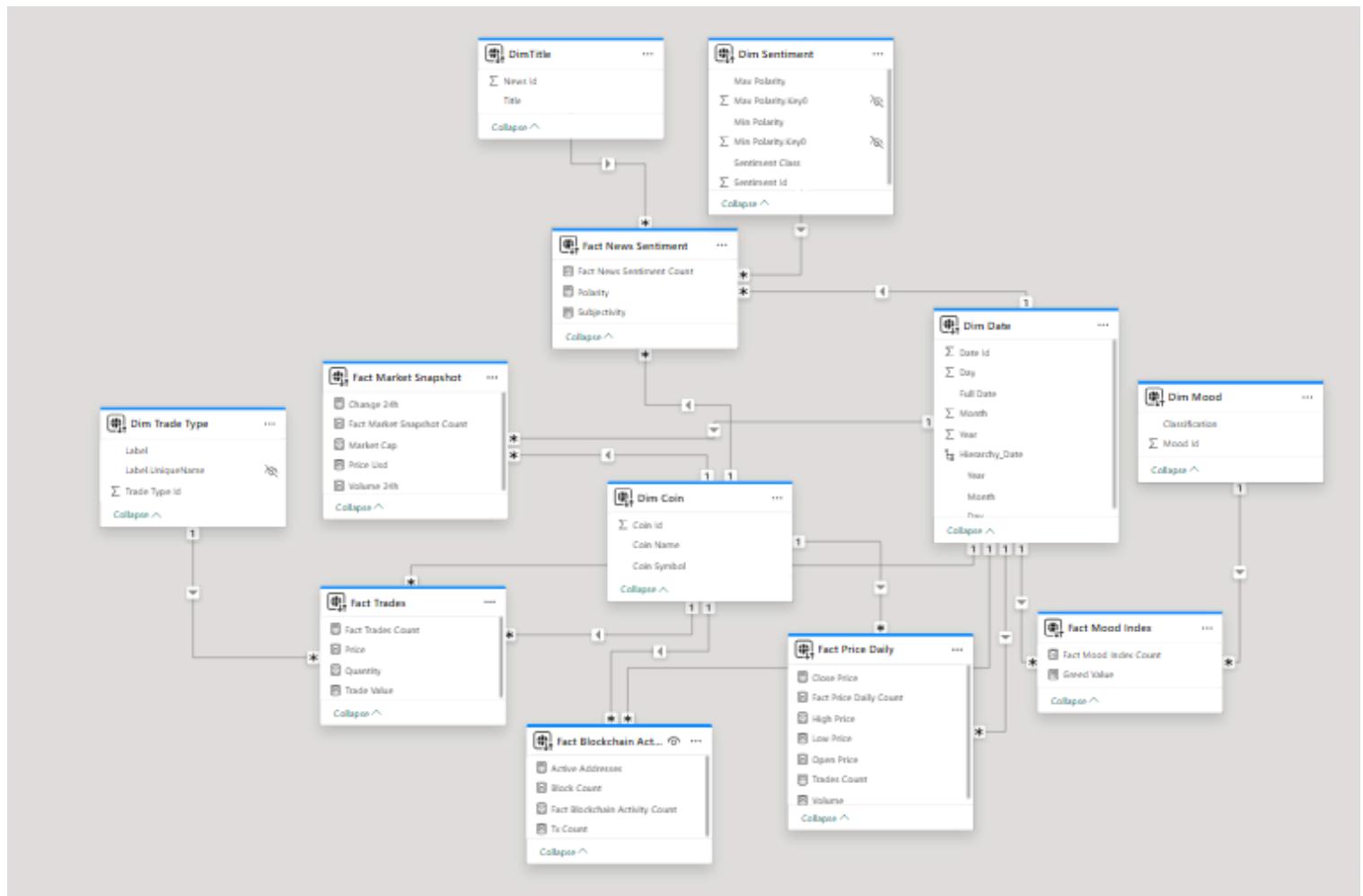
Principes :

- Tables de dimensions : Contiennent les attributs descriptifs (DimDate, DimCoin, DimSentiment, DimMood, DimTradeType)
- Tables de faits : Contiennent les mesures quantitatives et les clés étrangères vers les dimensions
- Relations 1:N : Chaque dimension est reliée à une ou plusieurs tables de faits

Avantages du schéma en étoile :

- Simplicité et intuitivité pour les requêtes analytiques
- Performance optimale avec jointures limitées
- Compatibilité native avec SSAS et les outils BI
- Extensibilité facilitée (ajout de dimensions ou faits)

## Schéma Relationnel Complet



Le schéma ci-dessus illustre la structure complète du Data Warehouse avec :

- **5 dimensions partagées** assurant la cohérence dimensionnelle
- **6 tables de faits interconnectées** représentant différentes perspectives analytiques
- **Relations d'intégrité référentielle** garantissant la qualité des données

Le détail des attributs de chaque table (dimensions et faits) est présenté dans la section 7 "Cube OLAP", où ces structures servent de base à l'implémentation multidimensionnelle.

### 6.3 Processus de Chargement et Qualité des Données

#### Zone de Staging

Avant l'intégration dans le Data Warehouse, les données transitent par une zone de staging comprenant des tables temporaires :

- staging\_prices : Données de prix brutes
- staging\_news : Actualités avec sentiment calculé
- staging\_trades : Transactions individuelles
- staging\_blockchain : Métriques on-chain
- staging\_market : Snapshots de marché
- staging\_mood : Indices Fear & Greed

**Rôle du staging :** Isolation pour validation, application de transformations complexes, détection et gestion des erreurs avant intégration définitive.

### **Changement incrémental:**

Le Data Warehouse est alimenté quotidiennement selon un processus incrémental :

1. Chargement des dimensions (pré-chargées, rarement modifiées)
2. Identification des nouvelles données dans le staging (basé sur date/timestamp)
3. Validation de l'intégrité référentielle (toutes les FK existent)
4. Insertion des nouveaux enregistrements uniquement (évite duplication)
5. Mise à jour des index et statistiques

### **Contrôles de Qualité**

Validations automatiques :

- Cohérence des valeurs ( $\text{high\_price} \geq \text{low\_price}$ ,  $\text{polarity} \in [-1,1]$ )
- Intégrité référentielle (toutes FK pointent vers enregistrements existants)
- Détection d'anomalies (valeurs aberrantes, gaps temporels)
- Vérification de complétude (aucun champ obligatoire NULL)

### **Automatisation:**

Un SQL Server Agent Job orchestre l'actualisation quotidienne complète :

1. Extraction des nouvelles données via APIs et fichiers
2. Transformation et chargement dans le staging
3. Validation et intégration dans le Data Warehouse
4. Traitement du cube OLAP
5. Actualisation automatique des dashboards Power BI

Cette automatisation garantit la disponibilité constante de données à jour pour les analyses décisionnelles.

## 7. Conception et Implémentation du Cube OLAP

Dans le cadre de ce projet, un cube OLAP a été conçu et implémenté à l'aide de SQL Server Analysis Services (SSAS) afin de permettre une analyse multidimensionnelle efficace des données du marché des cryptomonnaies et du sentiment des investisseurs.

### 7.1 Objectif du cube OLAP

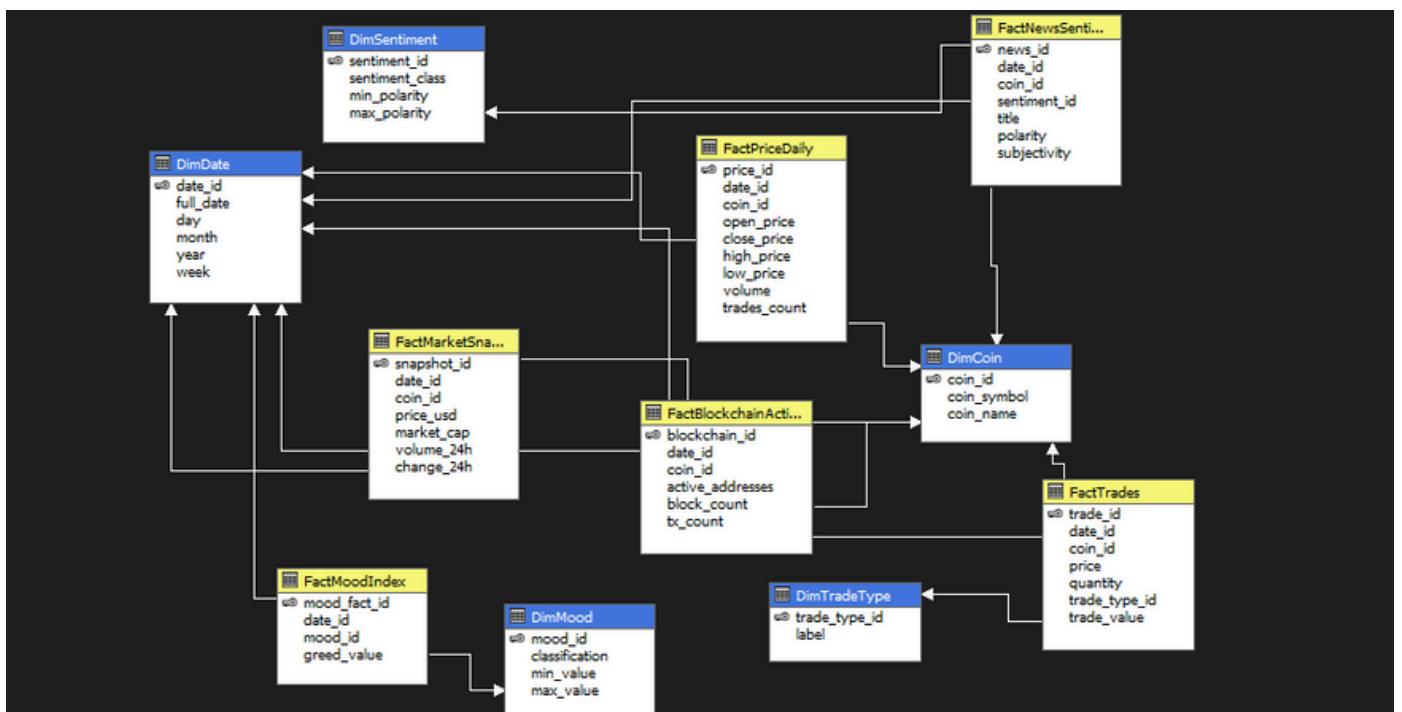
L'objectif principal du cube OLAP est de fournir une structure analytique permettant :

- L'analyse approfondie des prix et volumes de transactions des cryptomonnaies,
- L'étude du sentiment des actualités et leur corrélation avec les mouvements du marché
- L'observation des métriques blockchain et de l'activité on-chain.
- L'évaluation des indicateurs de marché et de la peur/avidité des investisseurs.
- Une exploration interactive des données selon différentes dimensions temporelles et par cryptomonnaie.

Le cube facilite ainsi la prise de décision grâce à des agrégations rapides et des analyses à plusieurs niveaux de granularité, tout en offrant une performance optimale pour les requêtes analytiques complexes.

### 7.2 Architecture et Structure du Cube

Le cube est basé sur un schéma en étoile étendu, construit à partir du Data Warehouse consolidé. L'architecture se compose de plusieurs tables de faits interconnectées et de dimensions partagées, permettant une analyse cohérente et multidimensionnelle.



### 7.3 Tables de faits

Le cube intègre plusieurs tables de faits, chacune représentant un aspect spécifique de l'analyse :

#### FactPriceDaily

- Contient les données quotidiennes des prix OHLC (Open, High, Low, Close)
- Mesures : price\_id, open\_price, close\_price, high\_price, low\_price, volume, trades\_count
- Permet l'analyse des tendances de prix et de la volatilité du marché

FactPriceDaily	
⌚	price_id
	date_id
	coin_id
	open_price
	close_price
	high_price
	low_price
	volume
	trades_count

#### FactNewsSentiment

- Centralise l'analyse du sentiment des actualités
- Mesures : news\_id, polarity, subjectivity
- Relie les actualités aux cryptomonnaies et aux dates pour identifier les corrélations entre sentiment et mouvement de prix

FactNewsSentiment	
⌚	news_id
	date_id
	coin_id
	sentiment_id
	title
	polarity
	subjectivity

### **FactTrades**

- Capture les transactions individuelles
- Mesures : trade\_id, price, quantity, trade\_value
- Permet l'analyse granulaire du comportement des traders via trade\_type\_id

FactTrades	
⌚	trade_id
	date_id
	coin_id
	price
	quantity
	trade_type_id
	trade_value

### **FactTrades**

- Capture les transactions individuelles
- Mesures : trade\_id, price, quantity, trade\_value
- Permet l'analyse granulaire du comportement des traders via trade\_type\_id

FactBlockchainActivity	
⌚	blockchain_id
	date_id
	coin_id
	active_addresses
	block_count
	tx_count

### **FactMarketSnapshot**

- Fournit une vue consolidée du marché
- Mesures : snapshot\_id, price\_usd, market\_cap, volume\_24h, change\_24h
- Permet le suivi des indicateurs macroéconomiques du marché crypto

FactMarketSnapshot	
⌚	snapshot_id
	date_id
	coin_id
	price_usd
	market_cap
	volume_24h
	change_24h

### **FactMoodIndex**

- Capture les indices de sentiment du marché (Fear & Greed)
- Mesures : mood\_fact\_id, mood\_id, greed\_value
- Essentiel pour l'analyse comportementale des investisseurs

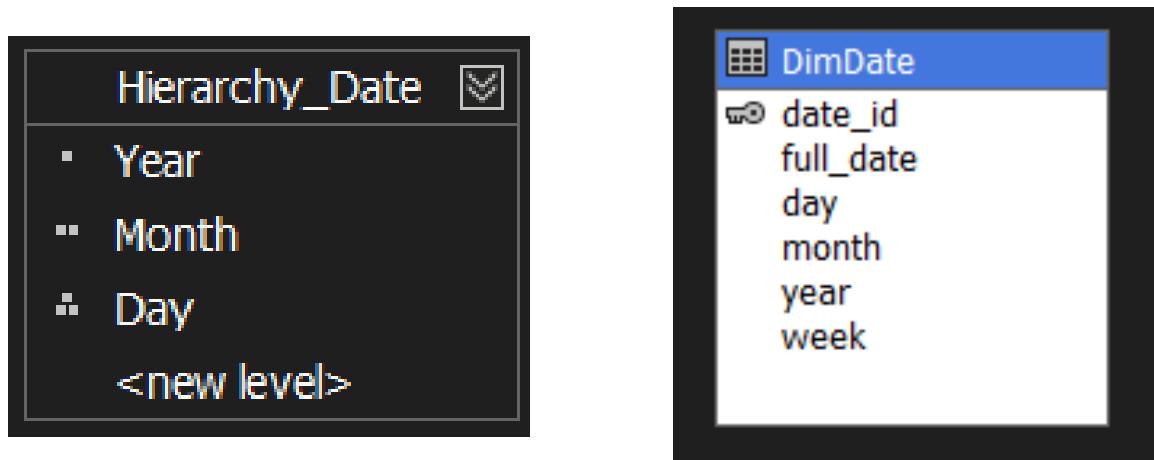
FactMoodIndex	
⌚	mood_fact_id
	date_id
	mood_id
	greed_value

## 7.4 Dimensions principales

Le cube s'appuie sur plusieurs dimensions permettant une analyse multidimensionnelle cohérente :

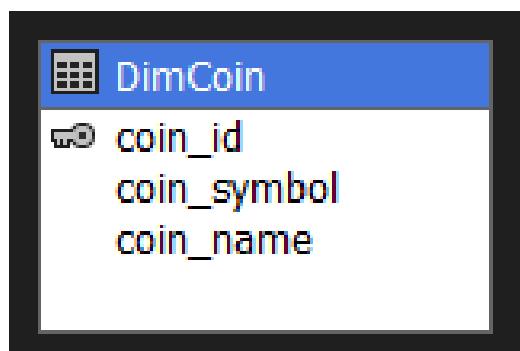
### DimDate (Dimension temporelle)

- Attributs : date\_id, full\_date, day, month, year, week
- Hiérarchie naturelle : Année → Mois → Jour
- Permet l'analyse des tendances temporelles et la comparaison de périodes



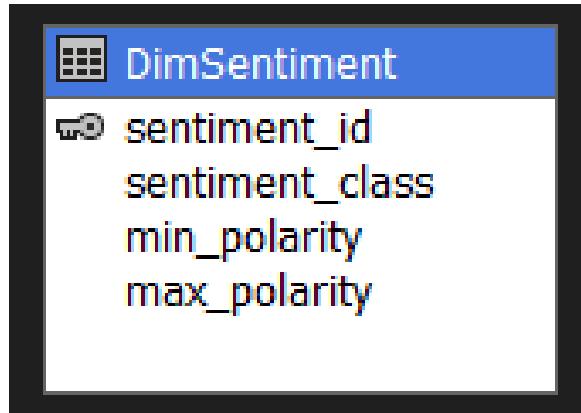
### DimCoin (Dimension Cryptomonnaie)

- Attributs : coin\_id, coin\_symbol, coin\_name
- Permet la segmentation par cryptomonnaie (BTC, ETH, etc.)
- Facilite les analyses comparatives entre différents actifs



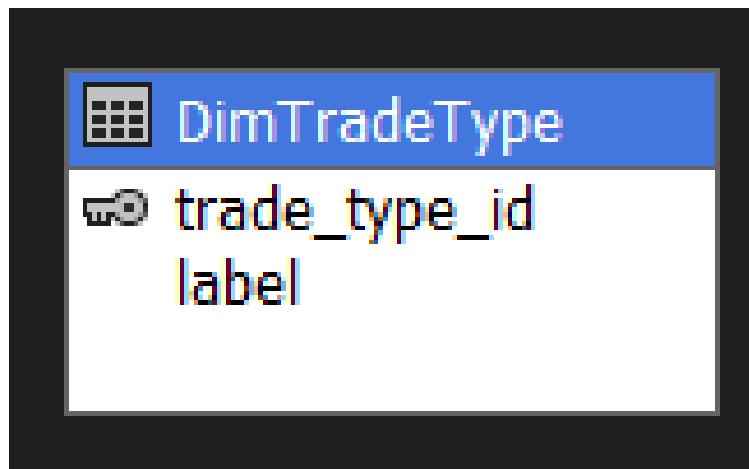
### **DimSentiment (Dimension Sentiment)**

- Attributs : sentiment\_id, sentiment\_class, min\_polarity, max\_polarity
- Classifie le sentiment en catégories (positif, négatif, neutre)
- Permet l'agrégation par niveau de sentiment



### **DimTradeType (Dimension Type de Transaction)**

- Attributs : trade\_type\_id, label
- Distingue les achats des ventes
- Essentiel pour l'analyse du comportement des traders



Ces dimensions sont partagées entre les différentes tables de faits, garantissant la cohérence dimensionnelle et permettant des analyses croisées sophistiquées.

## 7.5 Mesures et Indicateurs Calculés

Des mesures calculées ont été définies dans le cube afin d'obtenir des indicateurs analytiques pertinents. Ces mesures exploitent les agrégations du cube pour fournir des insights actionnables :

### Indicateurs de Sentiment

- Moyenne de la polarité par période/cryptomonnaie
- Moyenne de la subjectivité des actualités
- Distribution du sentiment (pourcentage positif/négatif/neutre)
- Corrélation sentiment-prix

### Indicateurs de Marché

- Volatilité quotidienne (High-Low / Close)
- Variation de prix en pourcentage
- Volume moyen de transactions
- Capitalisation boursière moyenne

### Indicateurs Comportementaux

- Ratio achats/ventes
- Fear & Greed Index moyen
- Nombre de traders actifs

### Indicateurs Blockchain

- Adresses actives moyennes
- Nombre de transactions confirmées
- Taux d'activité du réseau

## 7.6 Relations et Cardinalités

Le cube respecte les relations définies dans le Data Warehouse :

- FactPriceDaily → DimDate (N:1) et DimCoin (N:1)
- FactNewsSentiment → DimDate (N:1), DimCoin (N:1), DimSentiment (N:1)
- FactTrades → DimDate (N:1), DimCoin (N:1), DimTradeType (N:1)
- FactBlockchainActivity → DimDate (N:1), DimCoin (N:1)
- FactMarketSnapshot → DimDate (N:1), DimCoin (N:1)
- FactMoodIndex → DimDate (N:1) via jointure avec la table Mood

Ces relations permettent une navigation fluide entre les différentes perspectives d'analyse tout en maintenant l'intégrité référentielle.

## **8. Dashboard Interactif (Power BI)**

Après la conception et l'implémentation du cube OLAP, un tableau de bord interactif a été développé dans Power BI afin de permettre une exploration visuelle et intuitive des données analytiques. Ce dashboard constitue la couche de présentation finale du projet, transformant les données structurées du cube en insights visuels actionnables.

### **8.1 Objectifs du Tableau de Bord**

Le tableau de bord a été conçu pour répondre à plusieurs objectifs analytiques :

- Suivi des indicateurs clés de performance (KPI) du marché des cryptomonnaies
- Analyse comparative entre Bitcoin et Ethereum
- Corrélation entre le sentiment des actualités et les mouvements de prix
- Visualisation des tendances temporelles et des patterns de trading
- Évaluation des indices comportementaux du marché (Fear & Greed Index)
- Exploration interactive avec filtrage dynamique par période et cryptomonnaie

### **8.2 Architecture Technique**

#### **8.2.1 Structure du Dashboard**

Le tableau de bord est organisé en trois pages thématiques :

- Page 1 : Analyse des Prix et Corrélations
- Page 2 : Analyse des Transactions et du Sentiment
- Page 3 : Indices de Marché et Actualités Détaillées

#### **8.2.2 Connexion au Cube OLAP**

Le dashboard est connecté au cube SSAS via une connexion en direct (Live Connection), garantissant :

- L'accès en temps réel aux données après chaque traitement du cube
- L'utilisation des mesures calculées définies dans SSAS
- Des performances optimales grâce aux agrégations précalculées
- Une cohérence totale avec la logique métier centralisée

#### **8.2.3 Interactivité Globale**

Tous les visuels sont interconnectés avec :

- Filtres globaux : Sélection par cryptomonnaie (BTC, ETH, Select all)
- Slicer temporel : Filtre par mois pour l'analyse de périodes spécifiques
- Cross-filtering : Interactions entre visuels pour une exploration dynamique
- Drill-down : Navigation dans les hiérarchies temporelles (Année → Mois → Jour)

## 8.3 Page 1 : Analyse des Prix et Corrélations

Cette page offre une vue d'ensemble du marché avec un focus sur l'évolution des prix et l'activité blockchain.



### 8.3.1 Indicateurs Clés (KPI)

Today's Market Cap : Affiche la capitalisation boursière totale (\$2.17T)

- Source : FactMarketSnapshot.market\_cap

Average Price By Month : Prix moyen mensuel consolidé (10.97M)

- Calculé à partir de FactPriceDaily.close\_price

### 8.3.2 Visualisations Principales

#### Daily Price (Graphique en Aires)

- Évolution quotidienne des prix BTC et ETH
- Permet d'observer les tendances et la volatilité
- Source : FactPriceDaily avec hiérarchie temporelle

#### Trades Count (Graphique Linéaire)

- Nombre de transactions par jour pour chaque cryptomonnaie
- Révèle les pics d'activité et la liquidité du marché
- Source : FactPriceDaily.trades\_count

#### Volume by Coin Name, Month and Coin Symbol (Barres Groupées)

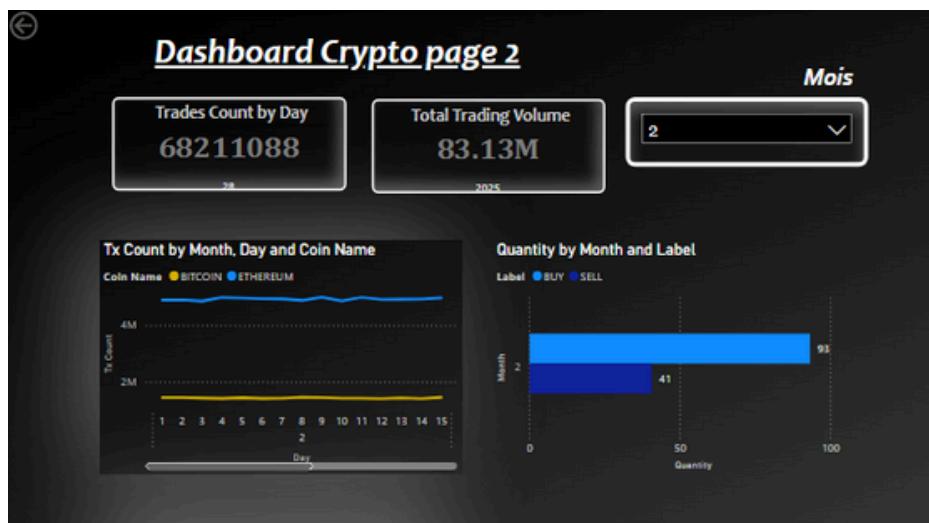
- Comparaison du volume de trading mensuel entre BTC et ETH
- Met en évidence la domination relative en termes de volume
- Source : FactPriceDaily.volume

#### Correlation between Price and Active Addresses (Graphique Combiné)

- Visualisation de la corrélation entre prix (barres) et adresses actives (ligne)
- Permet d'analyser la relation entre adoption on-chain et valorisation
- Source : Jointure FactPriceDaily et FactBlockchainActivity

### 8.4 Page 2 : Analyse des Transactions et du Sentiment

Cette page approfondit l'analyse comportementale des traders et le sentiment du marché.





#### 8.4.1 Indicateurs Clés (KPI)

Trades Count by Day : Nombre total de transactions (68,211,088)

- Source : Agrégation de FactTrades

Total Trading Volume : Volume total en valeur (83.13M)

- Source : FactTrades.trade\_value

#### 8.4.2 Visualisations Principales

##### Tx Count by Month, Day and Coin Name (Graphique en Aires)

- Évolution du nombre de transactions avec hiérarchie temporelle
- Montre la contribution relative de chaque cryptomonnaie
- Source : FactTrades agrégé par dimension temporelle

##### Quantity by Month and Label (Barres Empilées)

- Distribution des transactions par type (BUY vs SELL)
- Ratio 60/40 révélant une pression acheteuse dominante
- Source : FactTrades.quantity par DimTradeType.label

##### Polarity and Subjectivity by Coin Name (Treemap)

- Carte proportionnelle du sentiment par cryptomonnaie
- Répartition entre polarité et subjectivité
- Source : FactNewsSentiment avec moyennes par cryptomonnaie

## Average Polarity (Jauge)

- Indicateur global de sentiment (0.06 - légèrement positif)
- Échelle de -1.00 (négatif) à +1.00 (positif)
- Source : Moyenne de FactNewsSentiment.polarity

## Average Subjectivity (Jauge)

- Mesure d'objectivité des actualités (0.34 - modérément objective)
- Échelle de 0.00 (objectif) à 1.00 (subjectif)
- Source : Moyenne de FactNewsSentiment.subjectivity

## 8.5 Page 3 et 4 : Indices de Marché et Actualités

Cette page se concentre sur les indices comportementaux et l'analyse détaillée des actualités.





### 8.5.1 Visualisations Principales

#### Fact News Sentiment Count by Sentiment Class (Donut Chart)

- Distribution du sentiment : Positive (44.03%), Neutral (32.45%), Negative (23.52%)
- Total de 3,390 actualités analysées
- Révèle un marché plutôt optimiste
- Source : FactNewsSentiment par DimSentiment.sentiment\_class

#### Greed Value by Month and Day (Graphique en Barres)

- Évolution quotidienne de l'indice de cupidité (oscillant entre 28 et 60)
- Illustre les cycles émotionnels des investisseurs
- Source : FactMoodIndex.greed\_value avec hiérarchie temporelle

### **Fact Mood Index Count and Close Price (Graphique Combiné)**

- Corrélation entre classification Fear/Greed (barres empilées) et prix (ligne)
- Permet d'analyser l'impact des émotions collectives sur la valorisation
- Source : Jointure FactMoodIndex et FactPriceDaily

### **News Details Panel (Liste Scrollable)**

- Affichage détaillé des actualités avec leurs métriques de sentiment
- Chaque carte contient : titre complet, polarité et subjectivité
- Permet l'exploration qualitative et la validation des analyses quantitatives
- Source : FactNewsSentiment avec filtrage dynamique

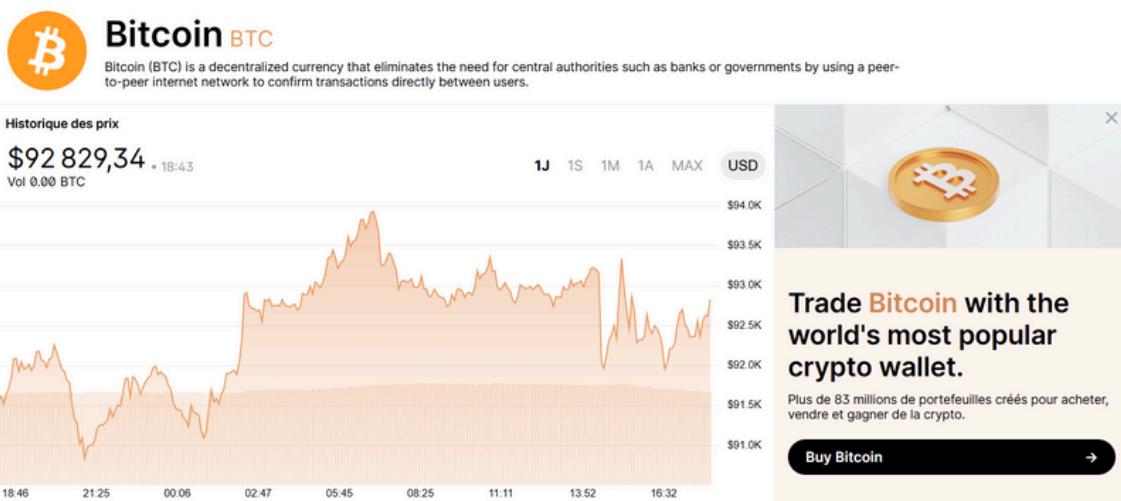
## **8.6 Insights Analytiques Générés**

Le tableau de bord permet de dégager plusieurs insights clés :

- Dominance d'Ethereum en volume : Volume de transactions significativement supérieur à Bitcoin
- Sentiment globalement positif : 44% d'actualités positives avec polarité moyenne de 0.06
- Pression acheteuse dominante : Ratio 60/40 (BUY/SELL) indiquant une phase d'accumulation
- Corrélation prix-activité on-chain : Validation de l'importance des métriques blockchain
- Volatilité émotionnelle : Oscillations importantes du Fear & Greed Index

## **8.7 Conclusion**

Le tableau de bord Power BI constitue l'interface finale du projet, transformant des données brutes en visualisations exploitables. L'architecture en trois pages permet une exploration progressive, de la vue macroéconomique à l'analyse comportementale détaillée. La connexion Live au cube OLAP garantit des performances optimales et une cohérence totale avec la logique métier, tandis que l'interactivité offerte par Power BI permet aux utilisateurs d'explorer les données selon leurs propres hypothèses. Ce dashboard démontre la capacité du système complet à transformer des sources hétérogènes en un outil d'aide à la décision structuré, performant et visuellement engageant.



The image shows a Bitcoin price chart from a trading platform. The chart displays the price of Bitcoin in USD over time, with a current price of \$92,829.34 and a volume of 0.00 BTC. The chart has a light orange background and a dark orange line representing the price movement. A large gold-colored Bitcoin coin icon is positioned in the top right corner. Below the chart, there is promotional text encouraging users to trade Bitcoin with the world's most popular crypto wallet, mentioning 83 million wallets created for buying, selling, and earning crypto. A prominent "Buy Bitcoin" button is located at the bottom right.

**Bitcoin BTC**

Bitcoin (BTC) is a decentralized currency that eliminates the need for central authorities such as banks or governments by using a peer-to-peer internet network to confirm transactions directly between users.

**Historique des prix**

\$92 829,34 • 18:43  
Vol 0.00 BTC

1J 1S 1M 1A MAX USD

\$94.0K  
\$93.5K  
\$93.0K  
\$92.5K  
\$92.0K  
\$91.5K  
\$91.0K

Trade Bitcoin with the world's most popular crypto wallet.

Plus de 83 millions de portefeuilles créés pour acheter, vendre et gagner de la crypto.

Buy Bitcoin →

The figure shows a price chart for Ethereum (ETH) on a digital exchange interface. The chart displays the price of Ethereum in USD over time, with a significant peak around \$3,119.15. A promotional image of a vibrant, futuristic-looking lounge or bar is overlaid on the right side of the chart, featuring palm trees and a large Ethereum logo.

**Ethereum ETH**

Ether (ETH) is the native cryptocurrency that powers Ethereum. It's primarily used to pay transaction fees and the creation of blockchain smart contracts.

**Historique des prix**

**\$3 119,15** • 18:44  
Vol 27 374 404 139 ETH

1J 1S 1M 1A MAX USD

\$3.1K  
\$3.1K  
\$3.1K  
\$3.1K  
\$3.1K  
\$3.1K  
\$3.1K  
\$3.0K  
\$3.0K  
\$3.0K  
\$3.0K

18:46 21:25 00:06 02:47 05:45 08:27 11:10 13:52 16:32

**Trade Ethereum with the world's most popular crypto wallet.**

Plus de 83 millions de portefeuilles créés pour acheter, vendre et gagner de la crypto.

**Buy Ethereum** ➔

The Glassnode Studio dashboard provides a comprehensive overview of the crypto market. Key features include:

- Market Cap**: A line chart showing Market Cap over time, with a current value of \$3.30T and a -11.3% change.
- Futures vs Spot Vol**: A bar chart comparing Futures vs Spot Vol across various assets.
- Altcoin Cycle Index**: A color-coded index ranging from Bitcoin (red) to Altcoin (blue).
- BTC Sharpe Signal**: A color-coded signal for the BTC Sharpe ratio, ranging from Risk Off (red) to Risk On (green).
- Spot Volume**: A heatmap showing spot volume for major coins like BTC, ETH, and SOL.
- Top Gainers - Large**: A table listing the top gainers over 1h, 24h, and 7d, including assets like BCH, LINK, and ETH.
- Sector Performance**: A line chart showing performance trends for sectors like BTC, ETH, L2, DeFi, AI, and Meme.
- Supply In Profit**: A stacked bar chart showing supply in profit for various tokens.

