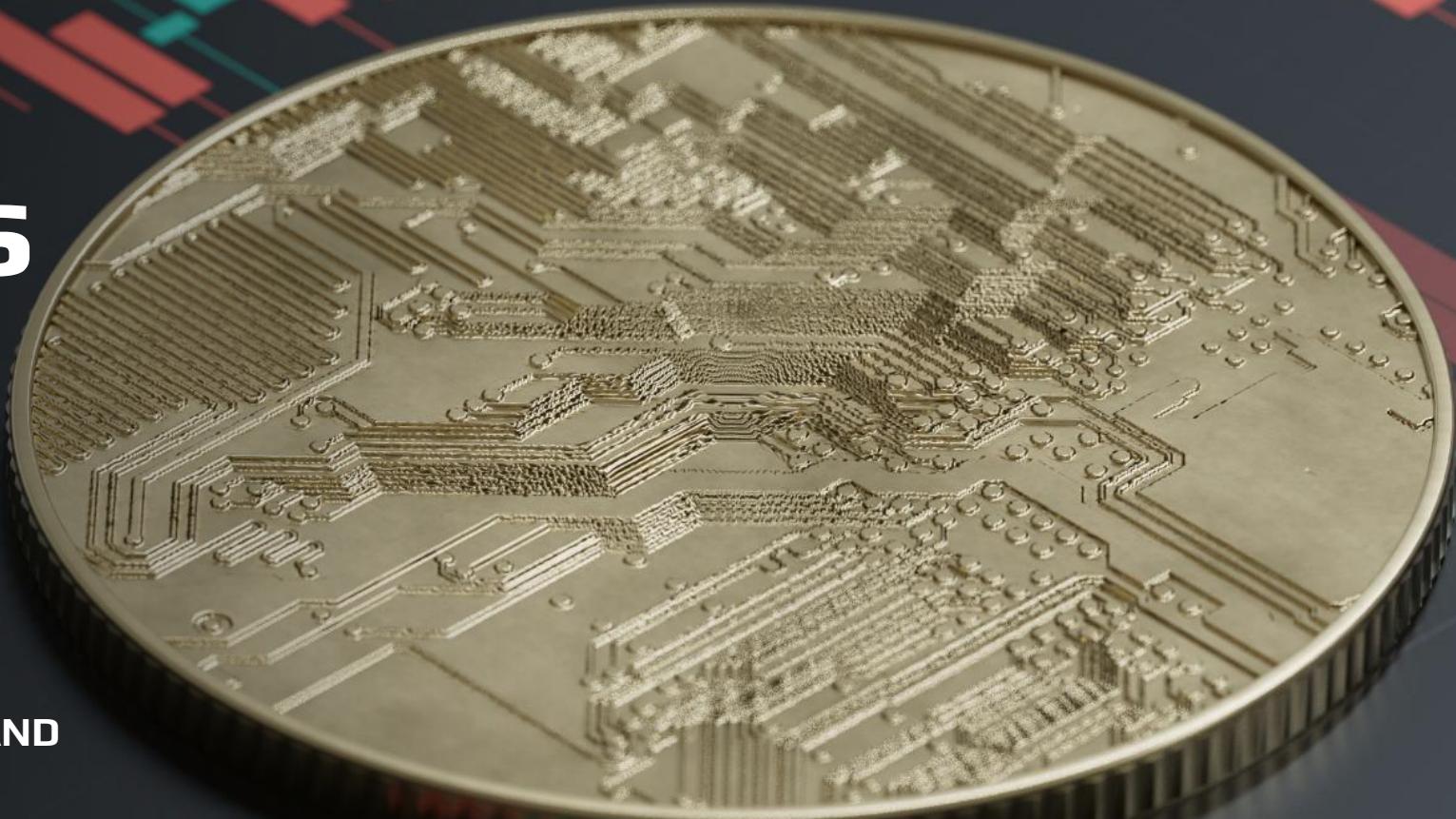


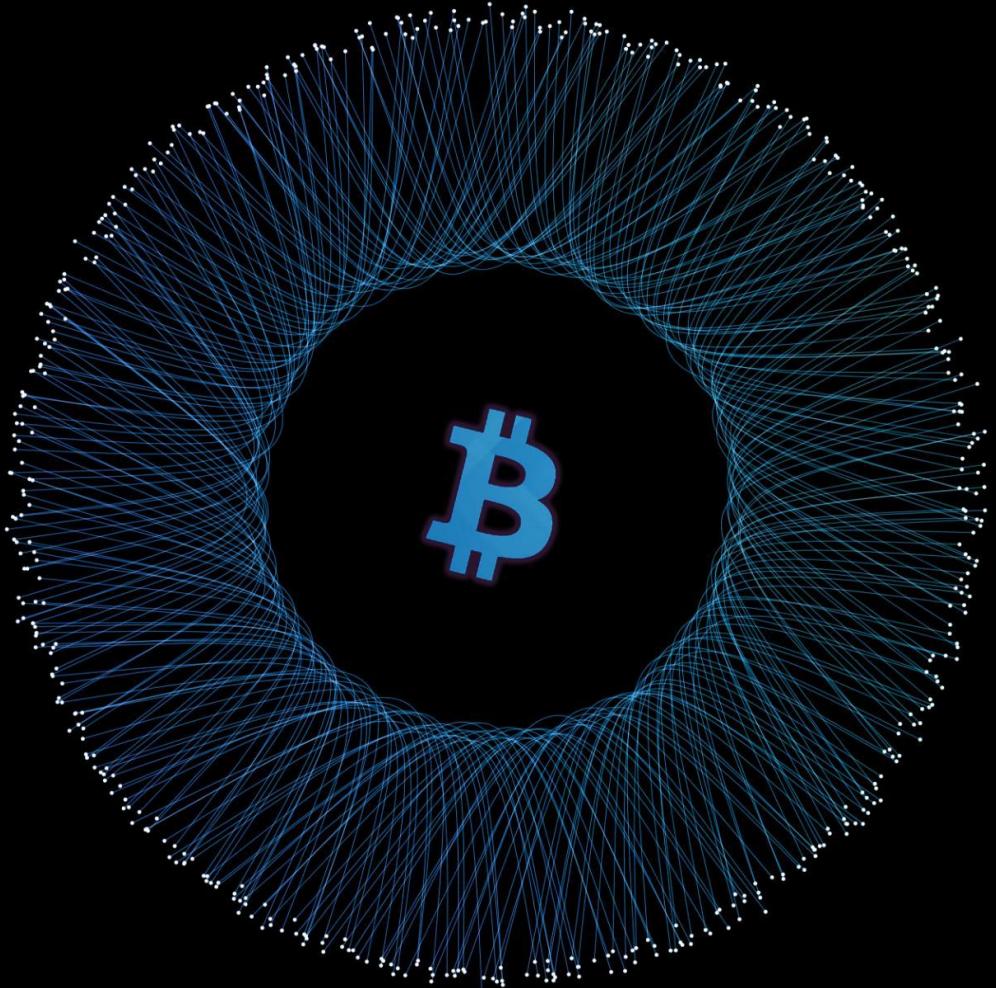
# Does Bitcoin News Affect It's Price?

---

EXPLORING NEWS SENTIMENT AND  
BITCOIN PRICE DYNAMICS

SHAKED SHTAUBER & ADI SEGEV





# Agenda for Today

- Data Sources
- Data Engineering Process
- Results and Findings
- Components and Alternatives
- Conclusion and Future Work

# Data Sources

---



# Bitcoin Price Data

## Google Finance API

Google Finance API connects to data from Google's financial page, providing access to cryptocurrency values and stock information.

## Stock Categories

Data from Google Finance is categorized into US stocks, European stocks, Asian stocks, and cryptocurrency stocks.

## Cryptocurrency Information

The API provides valuable information on the values and trends within the cryptocurrency market, including Bitcoin prices.

# Bitcoin News Articles- Crypto News API



The fields we chose to extract from the API are:

- Title
- Date
- Source\_name (where the article was published)
- sentiment

## Sentiment Classification

Bitcoin news articles are classified into three sentiment categories: positive, negative, and neutral, providing valuable insights.

### Positive Sentiment

Articles with positive sentiment express optimism and favorable views about Bitcoin, indicating confidence in its future.

### Negative Sentiment

Negative sentiment articles highlight caution, fear, or skepticism regarding Bitcoin, reflecting concerns about its stability.

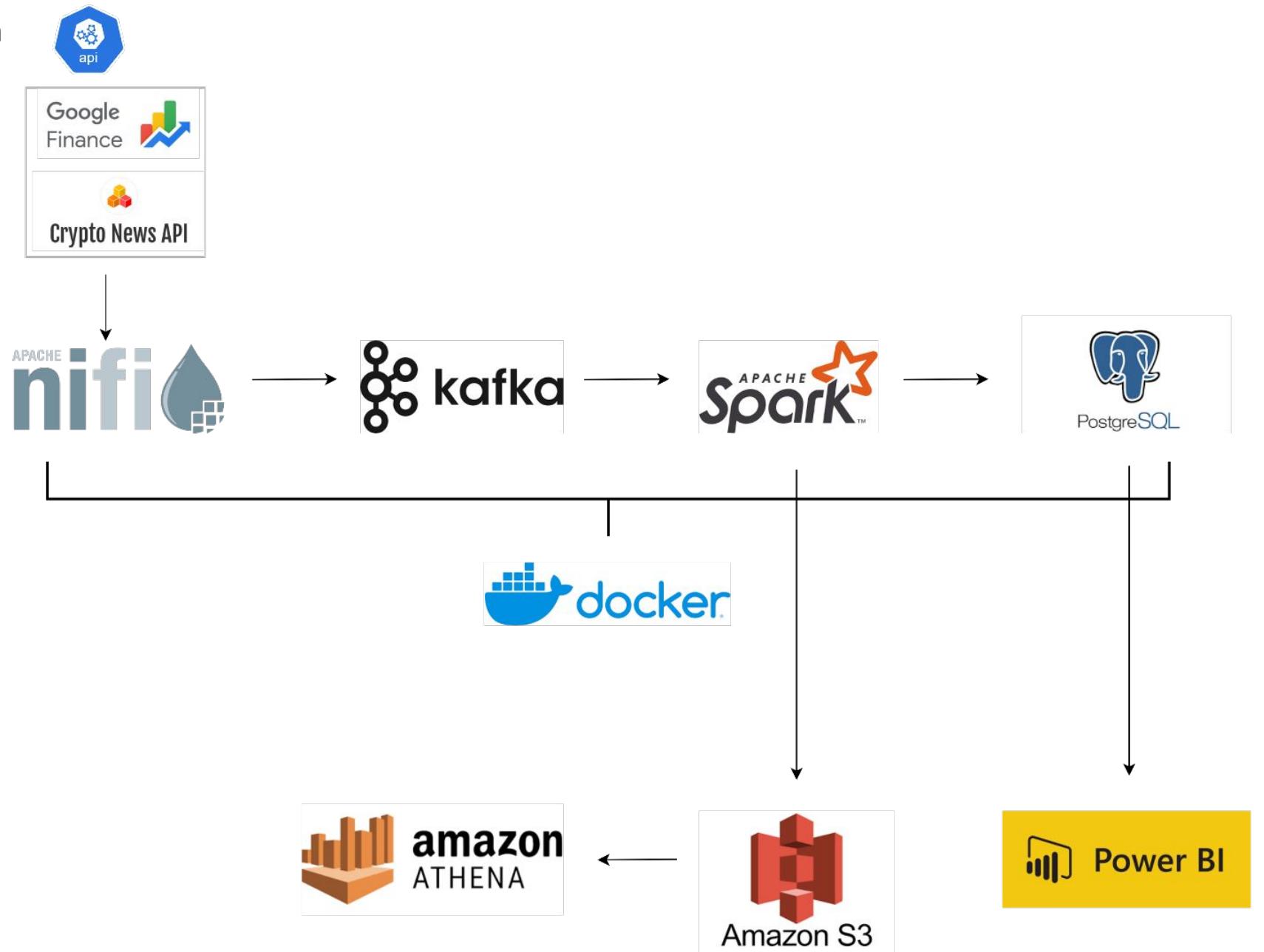
### Neutral Sentiment

Neutral sentiment articles present a balanced view of Bitcoin, offering insights without strong emotional bias.

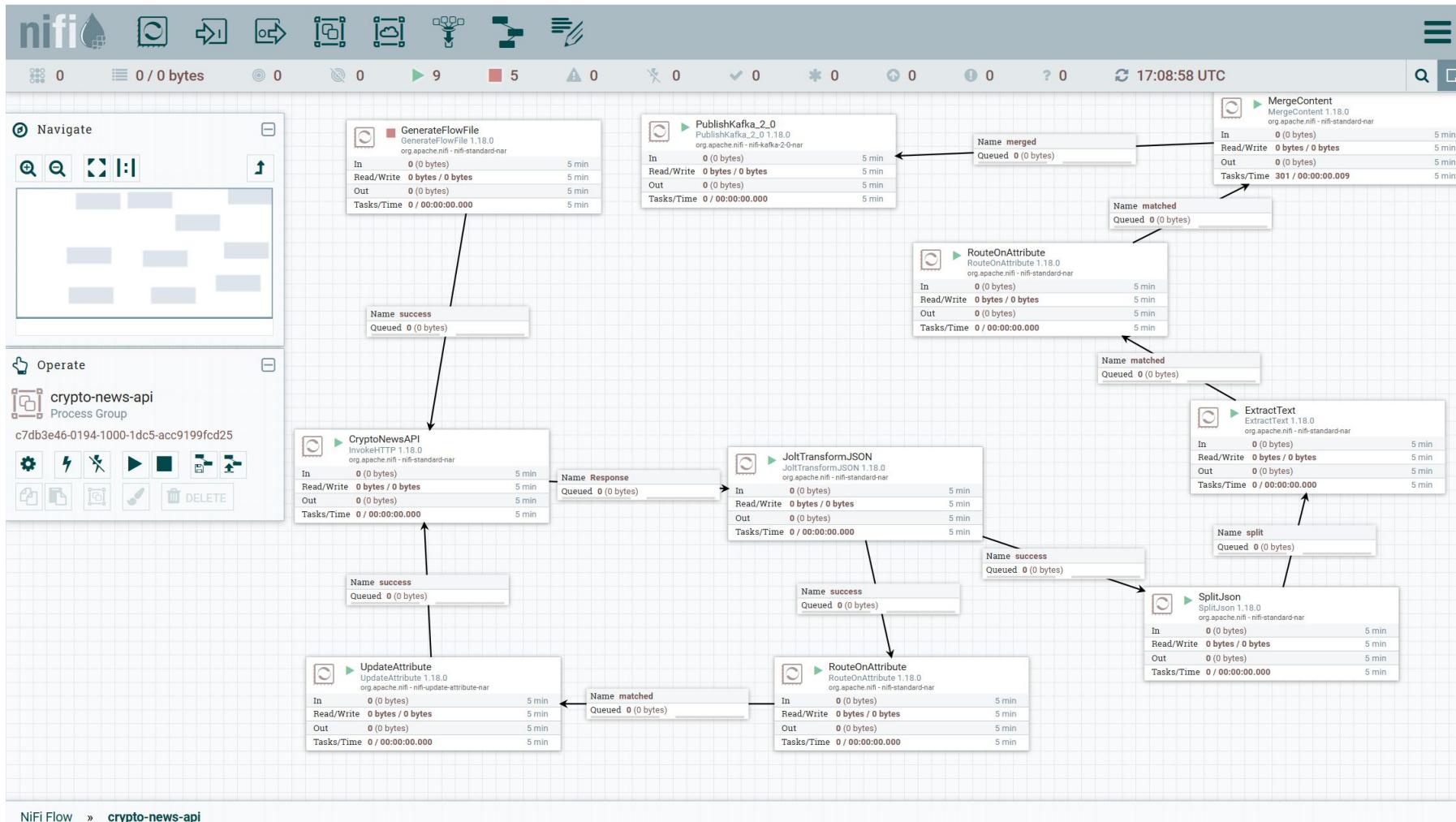
# **Data Engineering Process**

---

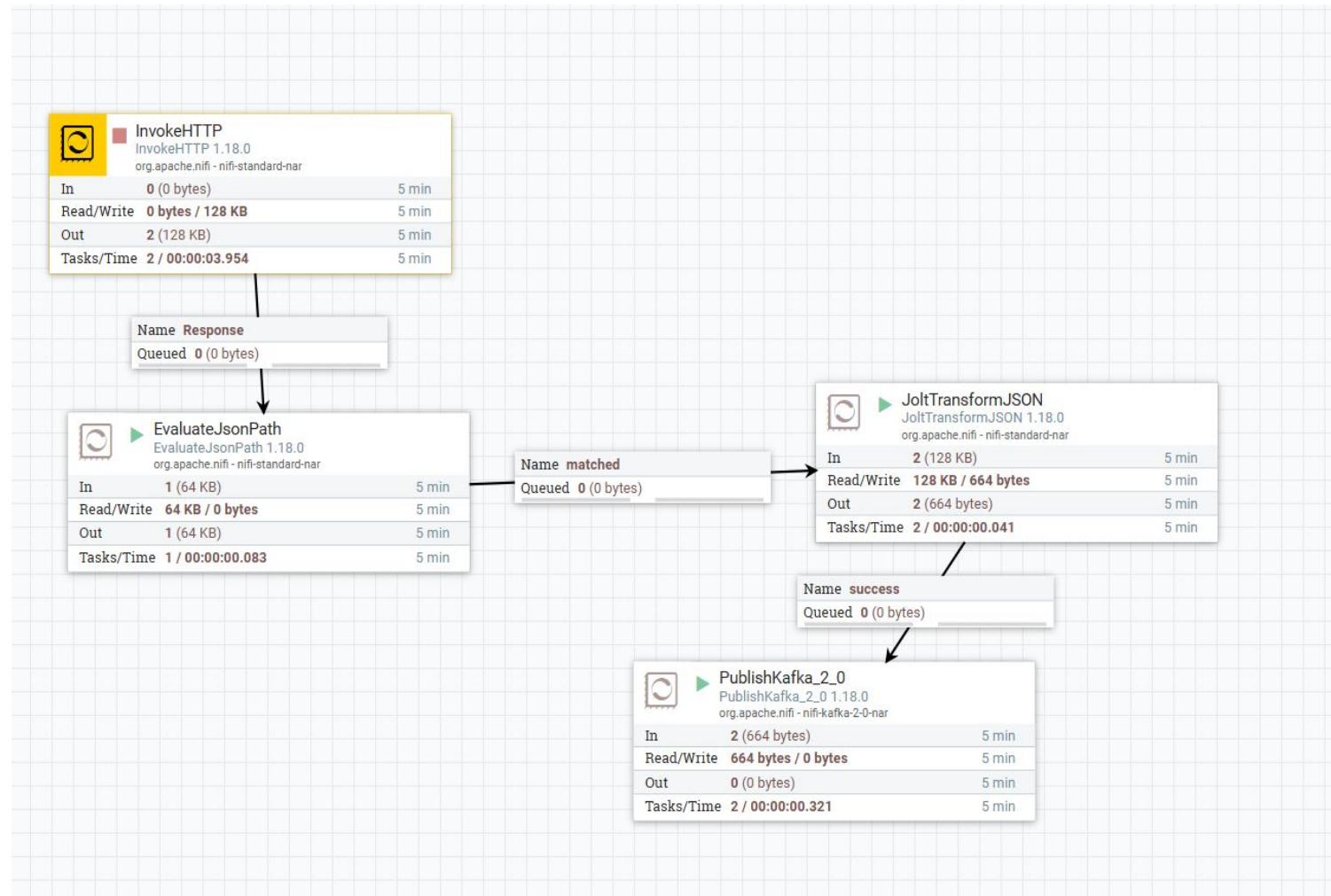
# Architecture



# Crypto News API Flow - Nifi



# Bitcoin Prices - Nifi



# Bitcoin News - write to Postgres

```
project_DE_postgres > ⚡ project_postgres.py > ...
 1  from pyspark.sql import SparkSession
 2  from pyspark.sql.functions import expr
 3  from pyspark.sql import functions as F
 4  from pyspark.sql.functions import col, explode, from_json,to_timestamp, date_format, regexp_replace
 5  from pyspark.sql.types import StructType, StructField, StringType, ArrayType
 6
 7
 8  spark = SparkSession \
 9    .builder \
10    .master("local[*]") \
11    .appName("KafkaToSparkToPostgresandS3") \
12    .config("spark.jars", "/opt/driver/postgresql-42.5.6.jar") \
13    .config('spark.jars.packages', 'org.apache.spark:spark-sql-kafka-0-10_2.12:3.1.2') \
14    .config('spark.jars.packages', 'org.apache.spark:spark-sql-kafka-0-10_2.12:3.1.2') \
15    .getOrCreate()
16
17  data_schema = StructType([
18      StructField("title", StringType()),
19      StructField("source_name", StringType()),
20      StructField("date", StringType()),
21      StructField("sentiment", StringType())
22  ])
23
24  # Read data from Kafka topic
25  kafka_stream = spark.readStream \
26    .format("kafka") \
27    .option("kafka.bootstrap.servers", "course-kafka:9092") \
28    .option("subscribe", "news-api-data") \
29    .option("startingOffsets", 'earliest') \
30    .load()
31
32  # Deserialize the 'value' column (which is in binary format) into a string, and then parse the JSON
33  json_df = kafka_stream.selectExpr("CAST(value AS STRING) AS json_data")
34
35
36  # Extract the "data" array from the JSON
37  json_parsed_df = json_df.select(from_json(col("json_data"), StructType([StructField("data", ArrayType(data_schema))])).\
38    alias("parsed_data"))
39
40  # Explode the "data" array into individual rows
41  flattened_df = json_parsed_df.select(explode(col("parsed_data.data")).alias("news_item"))
42
43  # Select the relevant fields from the exploded rows
44  final_df = flattened_df.select(
45      col("news_item.title").alias("title"),
46      col("news_item.source_name").alias("source_name"),
47      col("news_item.date").alias("date"),
48      col("news_item.sentiment").alias("sentiment")
49  )
50
51
52  final_df = final_df.withColumn("date", regexp_replace("date", "[A-Za-z]+", ""))
53  final_df = final_df.withColumn("date", date_format(to_timestamp("date", "dd MMM yyyy HH:mm:ss Z"), "dd_MM_yyyy"))
54
55  #Write the processed data to a sink (e.g., console for testing or another storage)
56  '''final_df.writeStream \
57    .outputMode("append") \
58    .format("console") \
59    .start() \
60    .awaitTermination()'''
61
62
63  # Function to write data to PostgreSQL
64  def write_to_postgresql(df, epoch_id):
65    if df.count() == 0:
66      print(f"Batch {epoch_id} is empty, nothing to write.")
67    else:
68      print(f"Batch {epoch_id} has {df.count()} records, writing to PostgreSQL.")
69      df.show() # Display the data in the current batch
70      df.write \
71        .format("jdbc") \
72        .option("url", "jdbc:postgresql://postgres:5432/postgres") \
73        .option("driver", "org.postgresql.Driver") \
74        .option("dbtable", "bitcoin_news") \
75        .option("user", "postgres") \
76        .option("password", "postgres") \
77        .mode("overwrite") \
78        .save()
79
80  # Write data to PostgreSQL
81  postgres_query = final_df.writeStream \
82    .foreachBatch(write_to_postgresql) \
83    .start()
84
85  postgres_query.awaitTermination()
```

# Bitcoin News - write to S3

```
project_DE_s3 > ➜ project_s3.py > ...
 1  from pyspark.sql import SparkSession
 2  from pyspark.sql.functions import expr
 3  from pyspark.sql import functions as F
 4  from pyspark.sql.functions import col, explode, from_json,to_timestamp, date_format, regexp_replace
 5  from pyspark.sql.types import StructType, StructField, StringType, ArrayType
 6
 7
 8  spark = SparkSession \
 9 .builder \
10 .master("local[*"]) \
11 .appName("KafkaToSparkToPostgresandS3") \
12 .config("spark.jars", "/opt/driver/postgresql-42.5.6.jar") \
13 .config("spark.jars.packages", 'org.apache.spark:spark-sql-kafka-0-10_2.12:3.1.2') \
14 .config("spark.jars.packages", 'org.apache.spark:spark-sql-kafka-0-10_2.12:3.1.2') \
15 .config("spark.hadoop.fs.s3a.access.key", "-----") \
16 .config("spark.hadoop.fs.s3a.secret.key", "-----") \
17 .config("spark.hadoop.fs.s3a.endpoint", "s3.amazonaws.com") \
18 .getOrCreate()
19
20 data_schema = StructType([
21     StructField("title", StringType()),
22     StructField("source_name", StringType()),
23     StructField("date", StringType()),
24     StructField("sentiment", StringType())
25 ])
26
27 # Read data from Kafka topic
28 kafka_stream = spark.readStream \
29 .format("kafka") \
30 .option("kafka.bootstrap.servers", "course-kafka:9092") \
31 .option("subscribe", "news-api-data") \
32 .option('startingOffsets', 'latest') \
33 .load()
34
35 # Deserialize the 'value' column (which is in binary format) into a string, and then parse the JSON
36 json_df = kafka_stream.selectExpr("CAST(value AS STRING) AS json_data")
37
38
39 # Extract the "data" array from the JSON
40 json_parsed_df = json_df.select(from_json(col("json_data"), StructType([StructField("data", ArrayType(data_schema))])).\
41                             alias("parsed_data"))
42
43 # Explode the "data" array into individual rows
44 flattened_df = json_parsed_df.select(explode(col("parsed_data.data")).alias("news_item"))
45
46 # Select the relevant fields from the exploded rows
47 final_df = flattened_df.select(
48     col("news_item.title").alias("title"),
49     col("news_item.source_name").alias("source_name"),
50     col("news_item.date").alias("date"),
51     col("news_item.sentiment").alias("sentiment")
52 )
53
54
55 final_df = final_df.withColumn("date", regexp_replace("date", "[A-Za-z]+", ""))
56 final_df = final_df.withColumn("date", date_format(to_timestamp("date", "dd MMM yyyy HH:mm:ss Z"), "dd_MM_yyyy"))
57 final_df = final_df.withColumn("partition_date", F.col("date"))
58
59
60 # Write data to S3
61 s3_query = final_df.writeStream \
62 .partitionBy("partition_date") \
63 .format("parquet") \
64 .option("path", "s3a://crypto-news-project/bitcoin_news/") \
65 .option("checkpointLocation", "s3a://crypto-news-project/checkpoints/") \
66 .outputMode("append") \
67 .start()
68
69 s3_query.awaitTermination()
```

# Bitcoin News - Postgres & S3

Grid Text Record

bitcoin\_news 1 ×

select \* from bitcoin\_news | Enter a SQL expression to filter results (use Ctrl+Space)

	A-Z title	A-Z source_name	A-Z date	A-Z sentiment
1	Analysts Predict Bitcoin Surge as Trump's Tariffs Weaken Dollar	Coinpaper	03_02_2025	Positive
2	Monochrome Brings First Aussie Bitcoin, Ethereum ETFs to Singa	Decrypt	03_02_2025	Positive
3	Chance of Bitcoin Tanking to \$75K Doubles as Trump's Tariffs Ig	Coindesk	03_02_2025	Negative
4	Bitcoin Crashes to \$91K as Trump's Trade Tariffs Shake Markets	Tokenpost	03_02_2025	Negative
5	Crypto Market Today Feb 3: BTC Slumps To \$91K, ETH & Meme	Coingape	03_02_2025	Negative
6	South Korea's 'bitcoin kimchi' premium index rises 9.7%	Cryptopolitan	03_02_2025	Positive
7	Bitcoin Bull Market: When Could BTC Peak Based on Past Trends	The Currency Analytics	03_02_2025	Positive
8	Bitcoin transaction count at lowest despite rising bullish signs -	\ AMBCrypto	03_02_2025	Neutral
9	Bitcoin and altcoins crash after Trump's tariff announcements	spc Crypto news	03_02_2025	Negative
10	XRP News Today: XRP Slumps as Tariff Shock Hits Risk Markets;	FXEmpire	03_02_2025	Negative
11	Bitcoin falls to \$92k, US stock market crashes as Trump vows to	Cryptopolitan	03_02_2025	Negative
12	Key Levels to Watch as Bitcoin Price Crashes Below \$100K Today	Coingape	03_02_2025	Negative
13	Crypto prices fall off the cliff on Feb 3: BTC at \$91k, XRP below \$	Invezz	03_02_2025	Negative
14	Bitcoin Price Nosedives Nearly 10%: Panic or Buying Opportunity	NewsBTC	03_02_2025	Negative
15	South Korea's bitcoin 'kimchi premium' soars to 10-month high	The Block	03_02_2025	Neutral
16	Robert Kiyosaki Sees Bitcoin Crash as Opportunity for Big Gains	The Currency Analytics	03_02_2025	Positive
17	Bitcoin Nears Crucial Support at \$90K: Market Faces Uncertainty	Tokenpost	03_02_2025	Negative
18	XRP Plunges 10.4% Amid Market Sell-Off, Bitcoin and Ethereum	Tokenpost	03_02_2025	Negative
19	New Trump tariffs stoke inflation fears, trigger \$2 billion in crypto	Crypto Briefing	03_02_2025	Negative
20	Bitcoin Drops 8% to \$93K as Asia Awakens to Trump's Trade Wa	Coindesk	03_02_2025	Negative
21	Bitcoin Crashes Below \$93K, Triggering \$1.23B in Liquidations as	Bitcoin	03_02_2025	Negative
22	Analyst Predicts Bitcoin Surge as Long-Term Holders Reduce Sal	The Currency Analytics	03_02_2025	Positive
23	US Bitcoin ETFs could pull in over \$50B in 2025, Bitwise says	Cointelegraph	03_02_2025	Positive
24	US Bitcoin ETFs could pull in over \$50B in 2025: Bitwise CIO	Cointelegraph	03_02_2025	Positive
25	HIVE Digital Buys Bitfarms' Paraguay BTC Mine for \$56M	CryptoPotato	03_02_2025	Positive
26	Utah could be first US state to pass Bitcoin reserve bill: Satoshi A	Cointelegraph	03_02_2025	Positive
27	Bitcoin Price Prediction for February 1: Will the Accumulation Peri	The Currency Analytics	03_02_2025	Neutral
28	Bitcoin dips below \$97,000 after Trump orders tariffs, smaller cry	CNBC	03_02_2025	Negative
29	Robert Kiyosaki Warns of Bitcoin Crash as Trump's Tariffs Take El	Bitcoin	03_02_2025	Negative
30	Bitcoin Set for Explosive Surge as Long-Term Holders Stop Sellin	The Currency Analytics	03_02_2025	Positive
31	1 Top Cryptocurrency to Buy Before It Soars 600%, According to The Motley Fool	02_02_2025	Positive	
32	Bitcoin Reserve Debate Divides Eurozone as Some Nations Push	The Currency Analytics	02_02_2025	Neutral
33	Despite a close at \$102K, Bitcoin analysts warn of a potential \$95	TheNewsCrypto	02_02_2025	Negative
34	El Salvador Buys 2 Additional Bitcoin in Single Day Post IMF Dea	TheNewsCrypto	02_02_2025	Neutral
35	Bitcoin Price Drops Amid Trump's 25% Tariffs on Mexico and Cai	The Currency Analytics	02_02_2025	Negative
36	Bitcoin slides below \$100,000 as tariffs rattle markets	Reuters	02_02_2025	Negative

Refresh Save Cancel Export data 200 400+ 400 row(s) for

IST en Writable Smart Insert 3 : 1 : 31 Sel: 0 | 0

WS Search [Alt+S] United States (N. Virginia) Shakedjri

Amazon S3 Buckets crypto-news-project bitcoin\_news/

bitcoin\_news/ Objects Properties

Objects (14) Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

Find objects by prefix Show versions

Name	Type	Last modified	Size	Storage class
_spark_metadata/_	Folder	-	-	-
partition_date=01_02_2025/_	Folder	-	-	-
partition_date=02_02_2025/_	Folder	-	-	-
partition_date=03_02_2025/_	Folder	-	-	-
partition_date=04_02_2025/_	Folder	-	-	-
partition_date=05_02_2025/_	Folder	-	-	-
partition_date=06_02_2025/_	Folder	-	-	-
partition_date=07_02_2025/_	Folder	-	-	-
partition_date=08_02_2025/_	Folder	-	-	-
partition_date=09_02_2025/_	Folder	-	-	-
partition_date=10_02_2025/_	Folder	-	-	-
partition_date=11_02_2025/_	Folder	-	-	-
partition_date=12_02_2025/_	Folder	-	-	-
partition_date=31_01_2025/_	Folder	-	-	-

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

# Bitcoin Prices - write to Postgres & S3

```
1 from pyspark.sql import SparkSession
2 from pyspark.sql.functions import from_json, col, to_date
3 from pyspark.sql.types import StructType, StructField, StringType, DoubleType
4 import shutil
5 import os
6
7 # Clear checkpoint directories
8 if os.path.exists("/tmp/checkpoints/s3"):
9     shutil.rmtree("/tmp/checkpoints/s3")
10 if os.path.exists("/tmp/checkpoints/postgres"):
11     shutil.rmtree("/tmp/checkpoints/postgres")
12
13 # Initialize Spark session with Kafka, Hadoop AWS, and PostgreSQL packages
14 spark = SparkSession.builder \
15     .appName("KafkaToS3AndPostgres") \
16     .config("spark.jars.packages", "org.apache.spark:spark-sql-kafka-0-10_2.12:3.5.4,org.apache.hadoop:hadoop-aws:3.3")
17     .config("spark.hadoop.fs.s3a.access.key", "XXXXXXXXXX") \
18     .config("spark.hadoop.fs.s3a.secret.key", "XXXXXXXXXX") \
19     .config("spark.hadoop.fs.s3a.endpoint", "s3.amazonaws.com") \
20     .getOrCreate()
21
22 # Define Kafka parameters
23 kafka_bootstrap_servers = "course-kafka:9092"
24 kafka_topic = "FinancialBTC"
25
26 # Define S3 parameters
27 s3_bucket = "financialbtc1"
28 s3_path = f"s3a://{s3_bucket}/kafka-data/"
29
```

```
30 # Define PostgreSQL parameters
31 postgres_url = "jdbc:postgresql://postgres:5432/postgres"
32 postgres_properties = {
33     "user": "postgres",
34     "password": "postgres",
35     "driver": "org.postgresql.Driver"
36 }
37
38 # Define schema for the Kafka message
39 schema = StructType([
40     StructField("created_at", StringType(), True),
41     StructField("stock", StringType(), True),
42     StructField("link", StringType(), True),
43     StructField("serpapi_link", StringType(), True),
44     StructField("name", StringType(), True),
45     StructField("price", DoubleType(), True),
46     StructField("price_movement_percentage", DoubleType(), True),
47     StructField("price_movement_value", DoubleType(), True),
48     StructField("price_movement", StringType(), True)
49 ])
50
51 # Read data from Kafka
52 kafka_df = spark.readStream \
53     .format("kafka") \
54     .option("kafka.bootstrap.servers", kafka_bootstrap_servers) \
55     .option("subscribe", kafka_topic) \
56     .load()
57
58 # Parse the JSON data
59 parsed_df = kafka_df.selectExpr("CAST(value AS STRING)") \
60     .select(from_json(col("value"), schema).alias("data")) \
61     .select("data.*")
62
63 # Filter for Bitcoin data and include created_at
64 filtered_df = parsed_df.filter(col("stock") == "BTC-USD")
65
```

# Bitcoin Prices - write to Postgres & S3

```
66 # Write data to S3 with partitioning by date
67 s3_query = filtered_df.withColumn("date", to_date(col("created_at"))) \
68     .writeStream \
69     .format("parquet") \
70     .option("path", s3_path) \
71     .option("checkpointLocation", "/tmp/checkpoints/s3") \
72     .partitionBy("date") \
73     .start()

75 # Write data to PostgreSQL
76 def write_to_postgres(batch_df, batch_id):
77     try:
78         batch_df.show() # Show the DataFrame to verify the schema and data
79         null_count = batch_df.filter(batch_df.stock.isNull()).count()
80         print(f"Batch {batch_id} has {null_count} null rows.")
81         if null_count == 0:
82             batch_df.write.jdbc(url=postgres_url, table="bitcoin_data", mode="append", properties=postgres_properties)
83             print(f"Batch {batch_id} written to PostgreSQL successfully.")
84         else:
85             print(f"Batch {batch_id} contains null values and will not be written to PostgreSQL.")
86     except Exception as e:
87         print(f"Error writing batch {batch_id} to PostgreSQL: {e}")
88
89 postgres_query = filtered_df.writeStream \
90     .foreachBatch(write_to_postgres) \
91     .option("checkpointLocation", "/tmp/checkpoints/postgres") \
92     .start()
93
94 s3_query.awaitTermination()
95 postgres_query.awaitTermination()
```

# Bitcoin Prices - Postgres & S3

	AZ created_at	AZ stock	AZ link	AZ serpapi_link	AZ name	AZ price	AZ price_move	AZ price_m	AZ
69	2025-02-12 11:25:59 UTC	BTC-USD	<a href="https://www.google.com">https://www.google.com</a>	<a href="https://serpapi.com">https://serpapi.com</a>	Bitcoin	96285.76	0.38927590000C	373.17	U
70	2025-02-12 11:25:59 UTC	BTC-USD	<a href="https://www.google.com">https://www.google.com</a>	<a href="https://serpapi.com">https://serpapi.com</a>	Bitcoin	96285.76	0.38927590000C	373.17	U
71	2025-02-12 11:25:59 UTC	BTC-USD	<a href="https://www.google.com">https://www.google.com</a>	<a href="https://serpapi.com">https://serpapi.com</a>	Bitcoin	96285.76	0.38927590000C	373.17	U
72	2025-02-12 11:25:59 UTC	BTC-USD	<a href="https://www.google.com">https://www.google.com</a>	<a href="https://serpapi.com">https://serpapi.com</a>	Bitcoin	96285.76	0.38927590000C	373.17	U
73	2025-02-12 11:25:59 UTC	BTC-USD	<a href="https://www.google.com">https://www.google.com</a>	<a href="https://serpapi.com">https://serpapi.com</a>	Bitcoin	96285.76	0.38927590000C	373.17	U
74	2025-02-12 11:25:59 UTC	BTC-USD	<a href="https://www.google.com">https://www.google.com</a>	<a href="https://serpapi.com">https://serpapi.com</a>	Bitcoin	96285.76	0.38927590000C	373.17	U
75	2025-02-12 11:25:59 UTC	BTC-USD	<a href="https://www.google.com">https://www.google.com</a>	<a href="https://serpapi.com">https://serpapi.com</a>	Bitcoin	96285.76	0.38927590000C	373.17	U
76	2025-02-15 10:16:37 UTC	BTC-USD	<a href="https://www.google.com">https://www.google.com</a>	<a href="https://serpapi.com">https://serpapi.com</a>	Bitcoin	97547.83	0.03443599	33.58	U
77	2025-02-15 10:16:37 UTC	BTC-USD	<a href="https://www.google.com">https://www.google.com</a>	<a href="https://serpapi.com">https://serpapi.com</a>	Bitcoin	97547.83	0.03443599	33.58	U
78	2025-02-15 10:16:37 UTC	BTC-USD	<a href="https://www.google.com">https://www.google.com</a>	<a href="https://serpapi.com">https://serpapi.com</a>	Bitcoin	97547.83	0.03443599	33.58	U

Objects (5)

[Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#) [Actions ▾](#)

[Create folder](#) [Upload](#)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	<a href="#">spark_metadata/</a>	Folder	-	-	-
<input type="checkbox"/>	<a href="#">date=2025-02-12/</a>	Folder	-	-	-
<input type="checkbox"/>	<a href="#">date=2025-02-13/</a>	Folder	-	-	-
<input type="checkbox"/>	<a href="#">date=2025-02-14/</a>	Folder	-	-	-
<input type="checkbox"/>	<a href="#">date=2025-02-15/</a>	Folder	-	-	-

# Bitcoin News - Atena

The screenshot shows the Amazon Athena Query editor interface. The left sidebar displays the 'Tables' section with one entry: 'bitcoin\_news'. The main area is titled 'SQL' and shows a completed query execution. The 'Query results' tab is selected, displaying 10 rows of news articles. The columns are labeled '#', 'title', 'source\_name', 'date', and 'sentiment'. The results are as follows:

#	title	source_name	date	sentiment
1	Crypto Prices Today Feb 8: BTC Touches \$95K Low, Altcoins Show Mixed Actions	Coingape	08_02_2025	Negative
2	Bitcoin Shows Strong Bullish Pattern at \$97K – Could \$145K Be Next	The Currency Analytics	08_02_2025	Positive
3	A Missouri Bill Proposes Establishing a Bitcoin Reserve Fund for State Investments	TheNewsCrypto	08_02_2025	Positive
4	Utah Moves Closer to Establishing a Bitcoin Reserve	Tokenpost	08_02_2025	Positive
5	CME Hits Record \$285M in Crypto Trading as Bitcoin Futures Surge	Tokenpost	08_02_2025	Positive
6	Michael Saylor's Strategy Pushes Bitcoin Investment as BTC Briefly Hits \$100K	Tokenpost	08_02_2025	Positive
7	Missouri Eyes Bitcoin Reserve to Future-Proof State Finances	Tokenpost	08_02_2025	Positive
8	Bitcoin Drops as Trump's Tariff Warning Fuels Market Uncertainty	Tokenpost	08_02_2025	Negative
9	Bitcoin Price Briefly Hits \$100K as Retail Investors Boost Accumulation	Tokenpost	08_02_2025	Positive

At the bottom of the interface, there are links for 'CloudShell', 'Feedback', '© 2025, Amazon Web Services, Inc. or its affiliates.', 'Privacy', 'Terms', and 'Cookie preferences'.

# **Results and Findings**

---



# Discussion on Predictive Power

## Sentiment Analysis

In our recent analysis, we examined the correlation between the sentiment of news articles about Bitcoin (positive or negative) and its price fluctuations over the past 14 days. The data indicates that there is no significant correlation, either positive or negative, between the sentiment of these articles and Bitcoin's price movements. This suggests that the news articles do not have a substantial impact on Bitcoin's market volatility.



## Market Reactions: Bitcoin Prices vs. News Sentiment

Date

1/31/2025

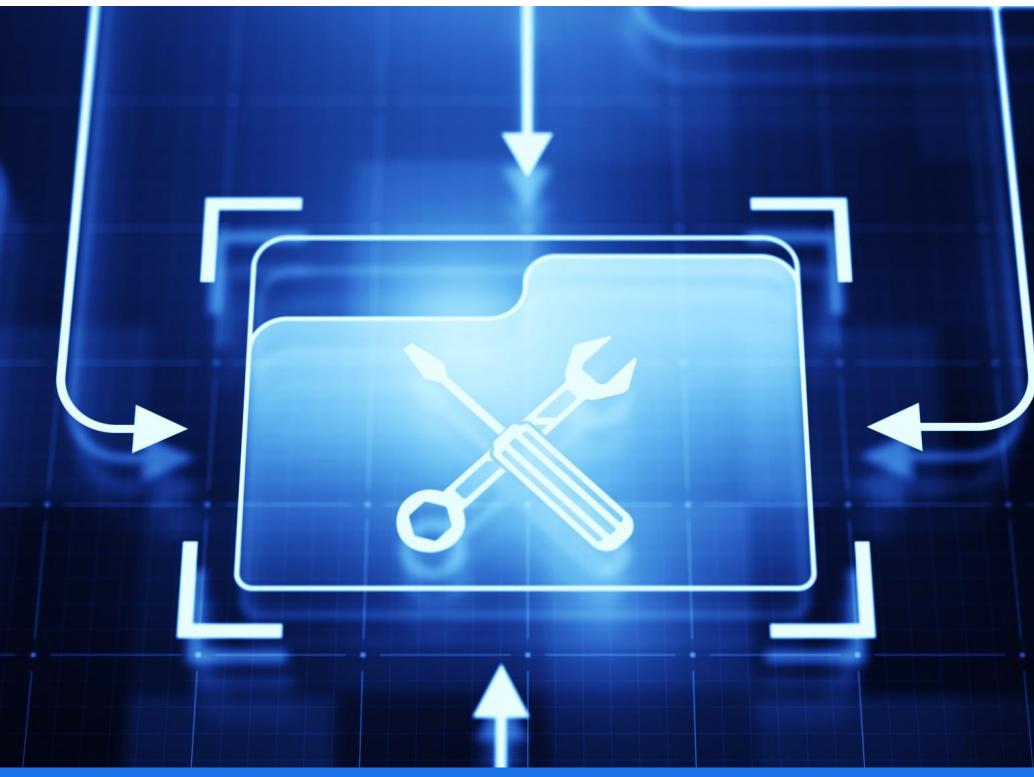
2/28/2025

Date	Today Price	Previous Price	Up/Down Prices	Number of News	Positive News	Negative News	Positive/Negative News	Correlation
1/31/2025	\$104,009	\$105,403	⬇️ (\$1,394)	75	47	9	Positive	✗ False
2/1/2025	\$102,346	\$104,009	⬇️ (\$1,663)	85	39	25	Positive	✗ False
2/2/2025	\$99,656	\$102,346	⬇️ (\$2,690)	97	36	46	Negative	✓ True
2/3/2025	\$92,883	\$99,656	⬇️ (\$6,772)	197	69	95	Negative	✓ True
2/4/2025	\$100,156	\$92,883	⬆️ \$7,273	46	21	13	Positive	✓ True
2/5/2025	\$98,136	\$100,156	⬇️ (\$2,020)	192	131	38	Positive	✗ False
2/6/2025	\$97,649	\$98,136	⬇️ (\$487)	172	109	31	Positive	✗ False
2/7/2025	\$97,520	\$97,649	⬇️ (\$130)	165	98	35	Positive	✗ False
2/8/2025	\$96,362	\$97,520	⬇️ (\$1,158)	112	54	41	Positive	✗ False
2/9/2025	\$97,282	\$96,362	⬆️ \$919	122	57	44	Positive	✓ True
2/10/2025	\$96,933	\$97,282	⬇️ (\$349)	470	276	102	Positive	✗ False
2/11/2025	\$98,077	\$96,933	⬆️ \$1,144	350	238	56	Positive	✓ True
2/12/2025	\$95,573	\$98,077	⬇️ (\$2,504)	634	368	186	Positive	✗ False
2/13/2025	\$96,901	\$95,573	⬆️ \$1,328	84	57	9	Positive	✓ True
2/14/2025	\$97,514	\$96,901	⬆️ \$613	155	108	26	Positive	✓ True
2/15/2025	\$97,655	\$97,514	⬆️ \$141	85	46	18	Positive	✓ True
2/16/2025	\$97,372	\$97,655	⬇️ (\$283)	67	32	13	Positive	✗ False
2/17/2025		\$97,372	⬇️ (\$97,372)	15	4	7	Negative	✓ True
Total	\$1,666,025	\$105,403	\$1,560,622	3,123	1,790	794	Positive	True

# **Components and Alternatives**

---

# Apache NiFi



## **Data Integration Tool**

Apache NiFi automates data flow between systems, simplifying data integration for various applications.

## **Visual Interface**

NiFi features a user-friendly drag-and-drop interface that makes data collection and transformation intuitive.

## **Real-Time Streaming**

The tool supports real-time streaming, allowing users to process data as it is generated effectively.

## **Advanced Data Routing**

NiFi provides fine-grained data routing capabilities, enabling users to prioritize and control data flows seamlessly.



# Apache Kafka

## Event Streaming Platform

Apache Kafka is designed for high-throughput, low-latency data streaming, enabling real-time data processing.

## Scalability and Fault Tolerance

Kafka handles large volumes of data with fault tolerance and scalability, ideal for modern applications.

## Kafka vs Kinesis

Kafka is often preferred over Kinesis due to better scalability, performance, and cost-effectiveness in self-managed environments.



# Apache Spark

# Big Data Processing

Apache Spark is designed for fast and efficient big data processing and analytics, making it a preferred solution for data professionals.

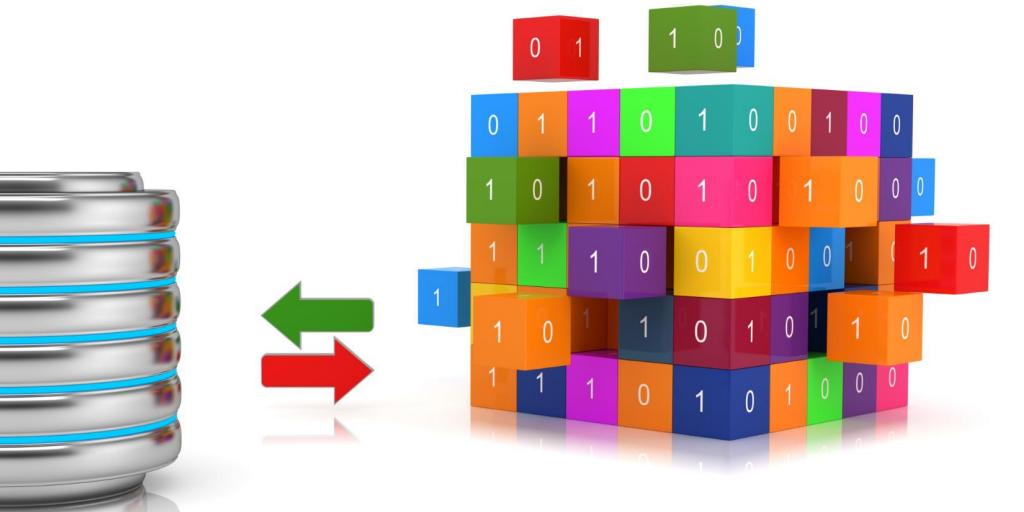
## Comparison with Hadoop

Spark outperforms Hadoop by processing data in memory and offering a simpler API interface across multiple programming languages.

## Machine Learning Capabilities

Spark includes MLlib, a built-in machine learning library optimized for large-scale data, which Hadoop lacks natively.

# PostgreSQL



## Database Management System

PostgreSQL is a powerful, open-source, object-relational database management system known for its robustness and flexibility.

## Postgres vs MariaDB

PostgreSQL offers a richer set of features compared to MariaDB, focusing on extensibility and compliance with SQL standards.

## Handling Complex Queries

PostgreSQL is optimized for complex, analytical queries and large datasets, supporting both SQL and NoSQL data models.

## Community and Ecosystem

With a larger, mature community, PostgreSQL offers more tools, libraries, and extensions for solving complex problems.

# Amazon S3



## Scalable Cloud Storage

Amazon S3 offers scalable, cloud-based storage for any amount of data, accessible from anywhere.

## Integration with AWS

S3 integrates seamlessly with various AWS services, enhancing its functionality for businesses.

## Cost-Effective Storage Solutions

S3 offers various storage options, including S3 Glacier for cost-efficient archiving and data management.

## Security and Compliance

S3 provides advanced security features and compliance options, making it suitable for enterprises.

It is a leading tool in the industry and we wanted to gain experience using it.

# Power BI



## Business Analytics Tool

Power BI is a business analytics tool that helps users visualize and analyze their data effectively.

## Interactive Dashboards

The tool enables users to create interactive dashboards to transform data into insightful visualizations.

## User-Friendly Design

Power BI is designed for both technical and non-technical users, making data analysis accessible to everyone.

## Data-Driven Decisions

With Power BI, organizations can make informed decisions based on data-driven insights and reports.

# **Conclusion and Future Work**

---

# Limitations

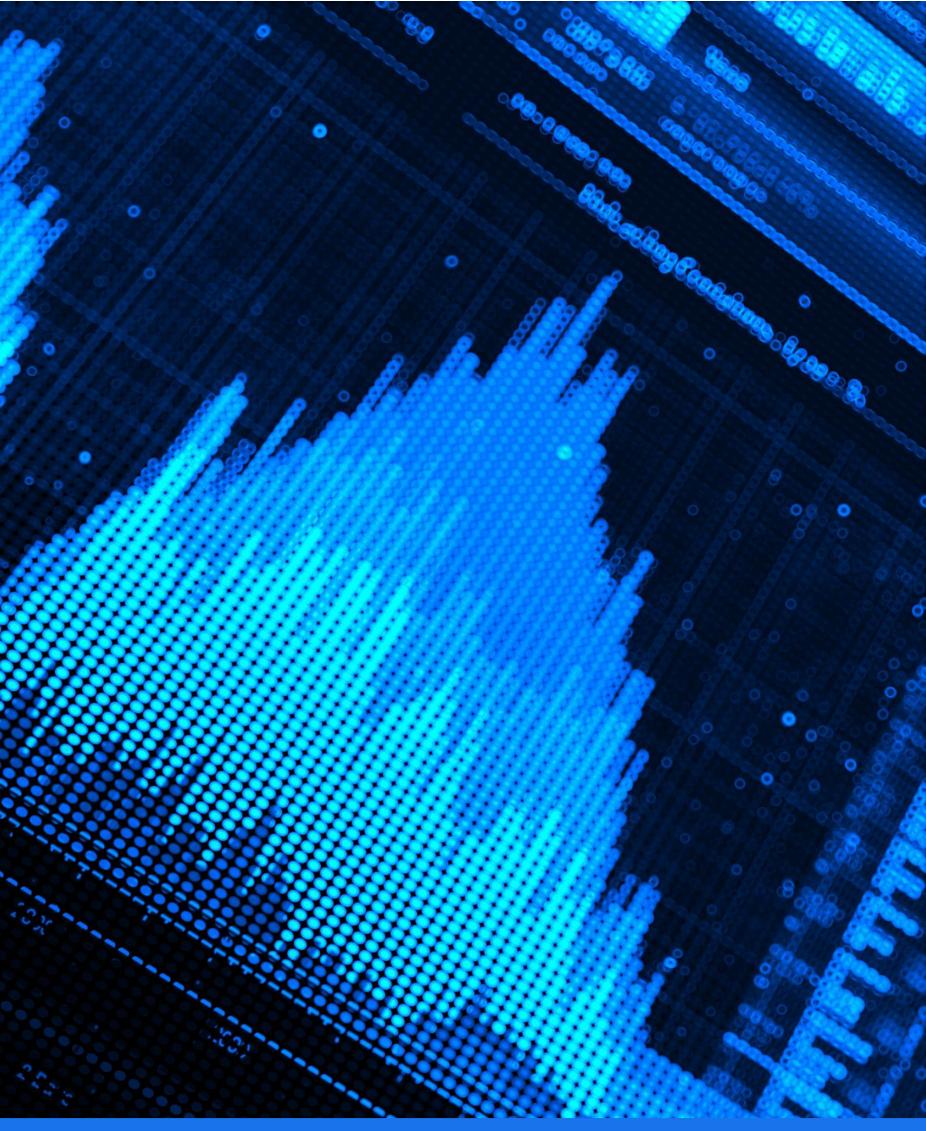
## Nuance in Sentiment Analysis

Sentiment analysis models might fail to capture the full complexity and nuance of news articles, leading to incomplete insights.

## External Influences on Bitcoin

The price of Bitcoin is impacted by various external factors, which can overshadow the effects of news sentiment.





# Future Enhancements

## **Additional Data Sources**

Incorporating social media posts and discussions from platforms like Reddit and Twitter can enrich the data landscape for analysis.

## **Advanced Machine Learning Models**

Utilizing advanced machine learning algorithms can significantly improve the accuracy of price predictions in various markets.

## **Long-term Price Trends**

Expanding models to include macroeconomic indicators will help in understanding and predicting long-term price trends more effectively.