

Research Question

- **Situation: Digital currencies have taken off**
 - Bitcoin launched in 2009 with almost no value and today it is worth nearly \$60,000!
- **Complication**: With so much hype around digital currencies, **prices can be highly volatile.**
 - In the first weekend of 2021, bitcoin rose 20%. The next Sunday, it fell 20%
- **Challenge**: Savvy investors can exploit market volatility to generate significant returns; however, **they need real-time financial data at their fingertips to monitor performance and identify potential trading signals**



To address this challenge, our team set out to build a **streaming digital currency data pipeline** so that an investor can **have full access to real-time data for decision-making**

Bitcoin Rises Above \$50,000. Where Does It Go From Here?



Taylor Tepper
Forbes Advisor Staff

Apr 10, 2021, 04:50am EDT | 188,098 views

'Moon Very Soon' — Cryptic Elon Musk Spurs Bitcoin On As Price Suddenly Blasts Past \$60,000 And Ethereum Hits Fresh High

Billy Bambrough, Contributor

and blockchain can change the world.

Bitcoin Hits Tipping Point After Skyrocketing On Investment Mania



DERIVATIVES, SELL-SIDE April 8, 2021 10:02 AM GMT

Machine learning futures algo trading surges at JP Morgan

Peter Ward, global head of futures and options electronic execution at JP Morgan, tells Hayley McDowell that buy-side adoption of its reinforcement learning FICC futures algorithms has surged in recent years, accelerated by the market volatility in 2020.

By Hayley McDowell

Bitcoin evangelists and recent converts, having shared their excitement on social media, are now being targeted by scammers. (Mareen — stock.adobe.com)

'Volatility is your friend': A former Goldman Sachs hedge-fund boss says crypto investors should embrace bitcoin's wild price swings - and explains why it's one of the asset's key features

Shalini Nagarajan
© Apr 11, 2021, 01:30 PM

SHARE

Access digital currency data in real-time with



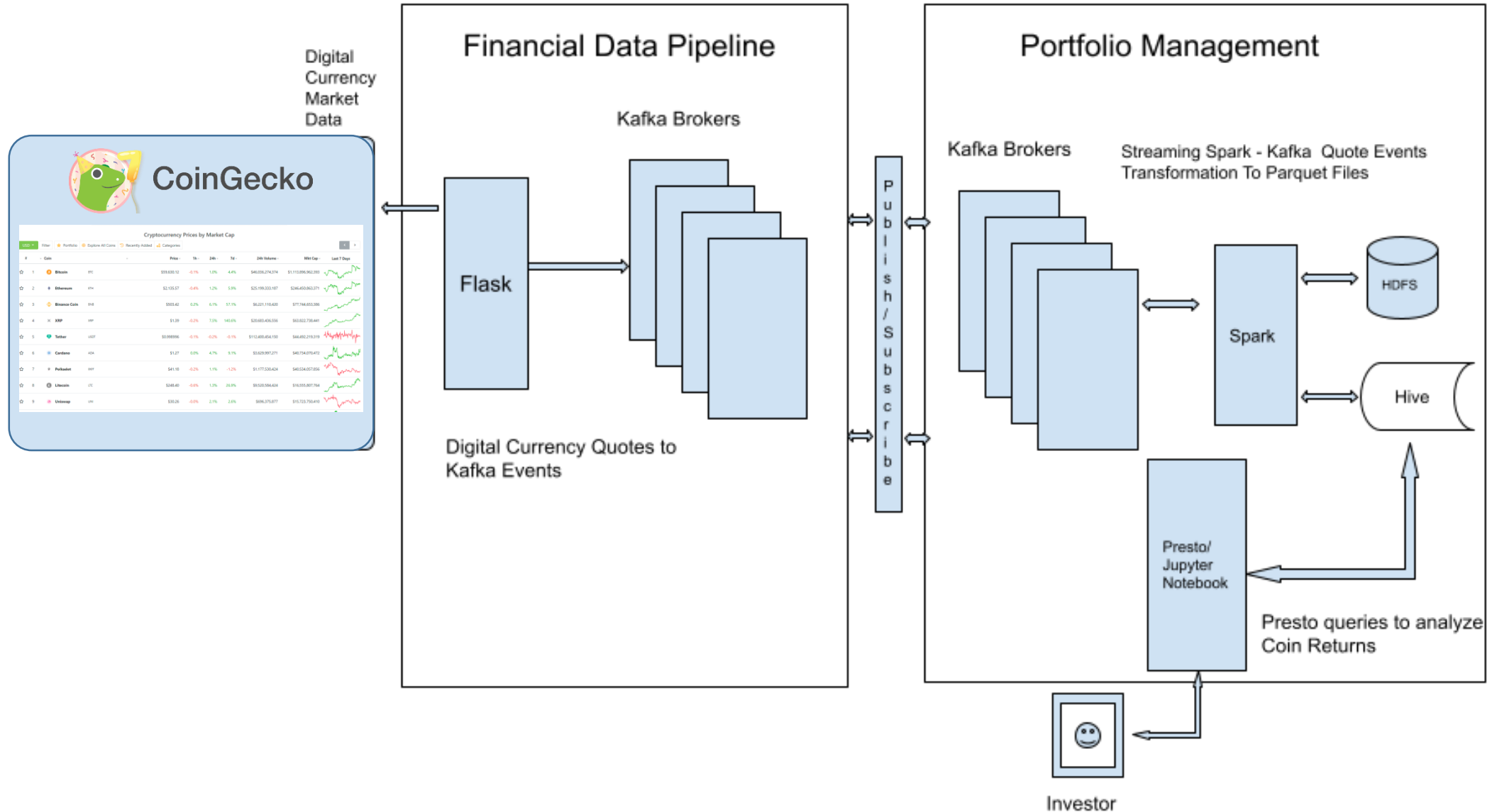
CoinGecko API

- **CoinGecko provides digital currency data** such as live quotes, trading volume, and historical data on more than 7,000 coins, such as bitcoin, litecoin, ethereum
- The **CoinGecko API is free** and allows up to **10 calls per second** which is more than enough to set up a real-time streaming pipeline

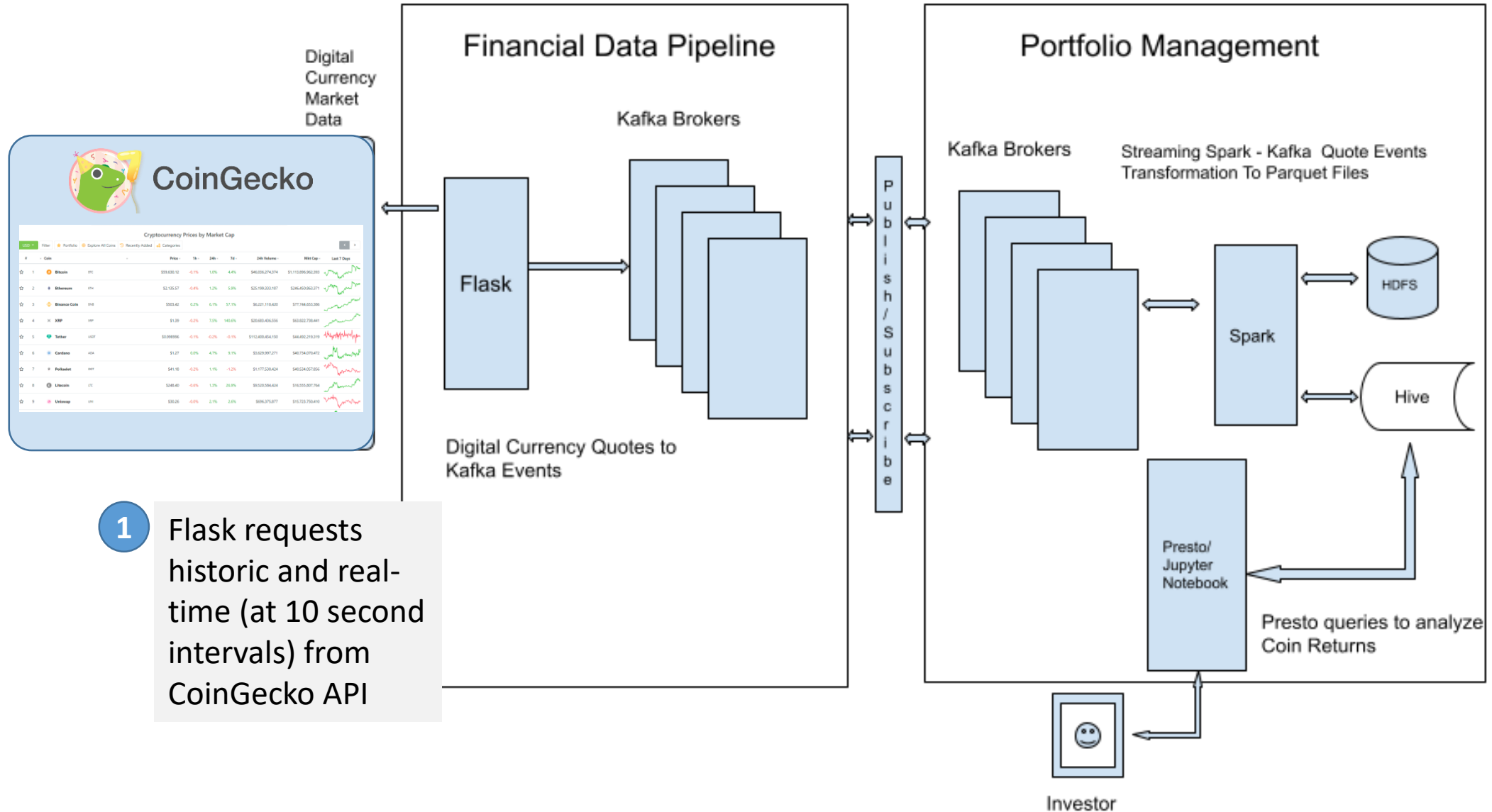
The screenshot shows the CoinGecko website homepage. At the top is a navigation bar with links: Home, Markets, DeFi, NFT, Portfolio, Exchanges, News, Tools, Resources, Learn, and Community. A search bar is on the right. Below the navigation bar is a status bar displaying: Coins: 6680, Exchanges: 447, Market Cap: \$2,111,713,533,543 2.7%↑, 24h Vol: \$215,461,127,082, Dominance: BTC 52.8%, ETH 11.7%, and ETH Gas: 65 gwei. The main hero section has a dark background with the text "The world's most comprehensive cryptocurrency API" and a green "Get Started" button. Below this is a three-column layout highlighting the API's features: Free, Reliable, and Comprehensive.

Free	Reliable	Comprehensive
Because we believe data should be democratized	Solid team of cryptocurrency specialists working hard to bring top notch cryptocurrency data	All the info you'll ever need at your fingertips
<ul style="list-style-type: none">✓ 100% Free crypto API✓ No keys required✓ Publicly available	<ul style="list-style-type: none">✓ In operation since early 2014✓ Generous rate limits with up to 100 requests/minute✓ Extensive market data	<ul style="list-style-type: none">✓ Track over 7,000 coins such as bitcoin, litecoin, ethereum, and more from more than 400 exchanges and growing✓ Comprehensive historical data✓ 21 endpoints and more coming soon

Streaming pipeline set up



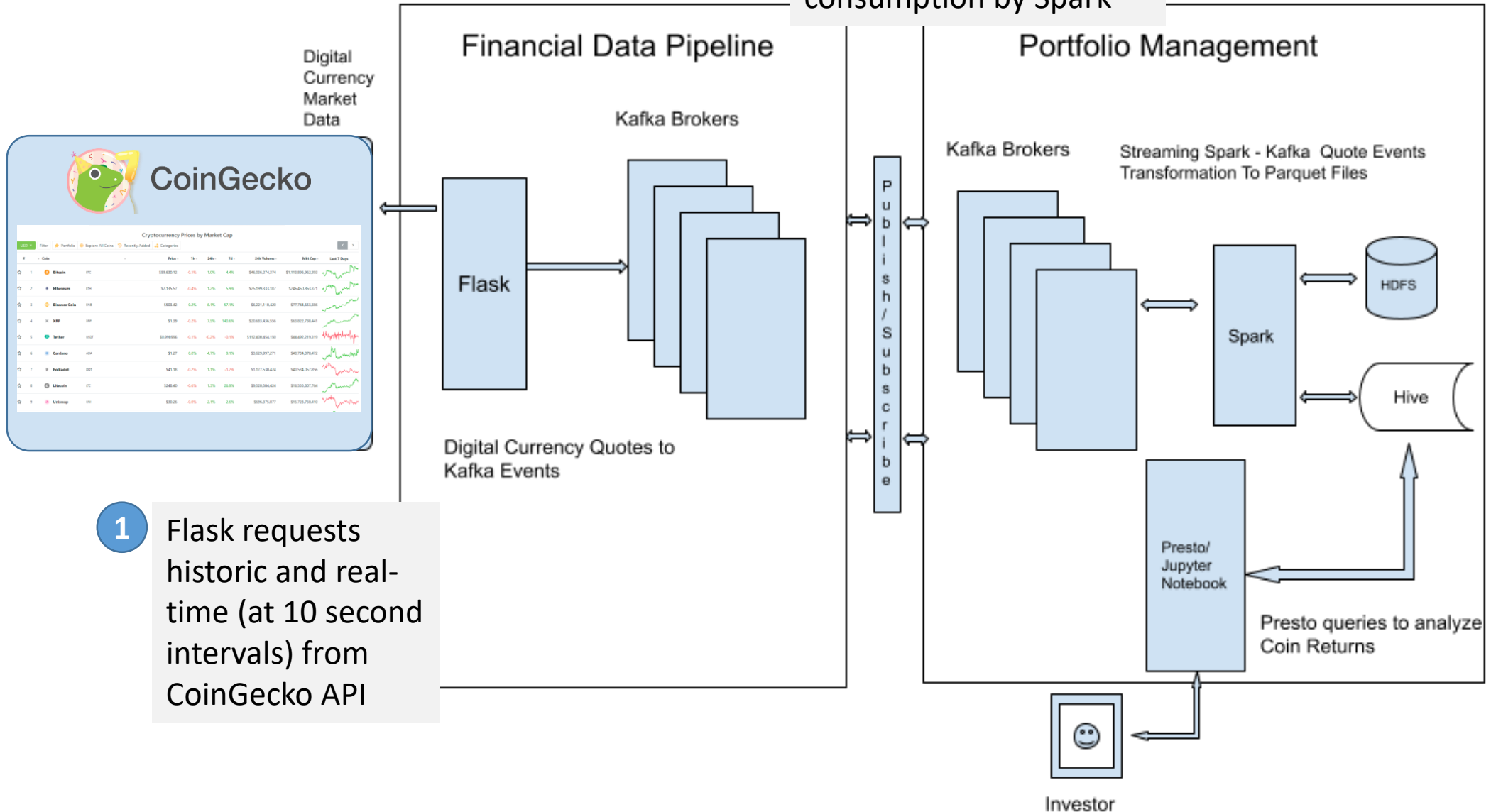
Real-time data stream: Financial Data Pipeline



Real-time data stream: Financial Data Pipeline

2

Events from Flask are logged (aka produced) to Kafka broker, ready for consumption by Spark



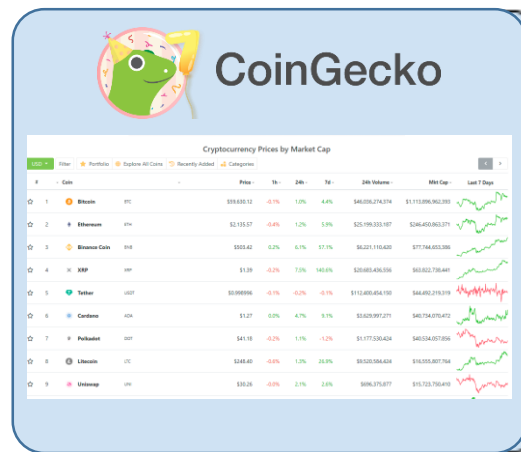
Real-time data stream: Portfolio Management

2

Events from Flask are logged (aka produced) to Kafka broker, ready for consumption by Spark

3

Spark reads JSON files from Kafka broker, defines schema and writes the data as parquet files to HDFS at 10 second intervals



1

Flask requests historic and real-time (at 10 second intervals) from CoinGecko API

Digital Currency Market Data

Financial Data Pipeline

Kafka Brokers

Flask

Digital Currency Quotes to Kafka Events

Publish / Subscribe

Portfolio Management

Kafka Brokers

Streaming Spark - Kafka Quote Events Transformation To Parquet Files

Spark

HDFS

Hive

Presto/ Jupyter Notebook

Presto queries to analyze Coin Returns



Investor

Real-time data stream: Portfolio Management



1 Flask requests historic and real-time (at 10 second intervals) from CoinGecko API

2

Events from Flask are logged (aka produced) to Kafka broker, ready for consumption by Spark

3

Spark reads JSON files from Kafka broker, defines schema and writes the data as parquet files to HDFS at 10 second intervals

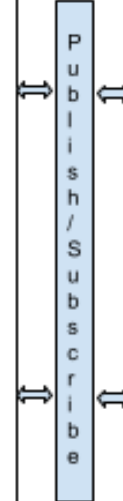
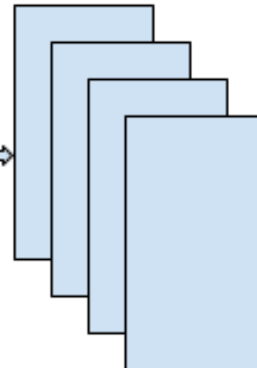
Digital Currency Market Data

Financial Data Pipeline

Kafka Brokers



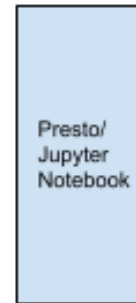
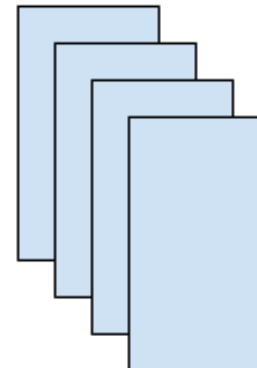
Digital Currency Quotes to Kafka Events



Portfolio Management

Kafka Brokers

Streaming Spark - Kafka Quote Events Transformation To Parquet Files



Presto queries to analyze Coin Returns

4

Hive writes parquet files to table ready for Presto query



Investor

Real-time data stream: Portfolio Management



1 Flask requests historic and real-time (at 10 second intervals) from CoinGecko API

2 Events from Flask are logged (aka produced) to Kafka broker, ready for consumption by Spark

3 Spark reads JSON files from Kafka broker, defines schema and writes the data as parquet files to HDFS at 10 second intervals

Digital Currency Market Data

Financial Data Pipeline

Kafka Brokers

Flask

Digital Currency Quotes to Kafka Events

Kafka Brokers

Portfolio Management

Streaming Spark - Kafka Quote Events Transformation To Parquet Files

Spark

HDFS

Hive

5 Through Jupyter notebook, investor queries data with Presto to analyze real-time data

4 Hive writes parquet files to table ready for Presto query

Presto/ Jupyter Notebook

Presto queries to analyze Coin Returns



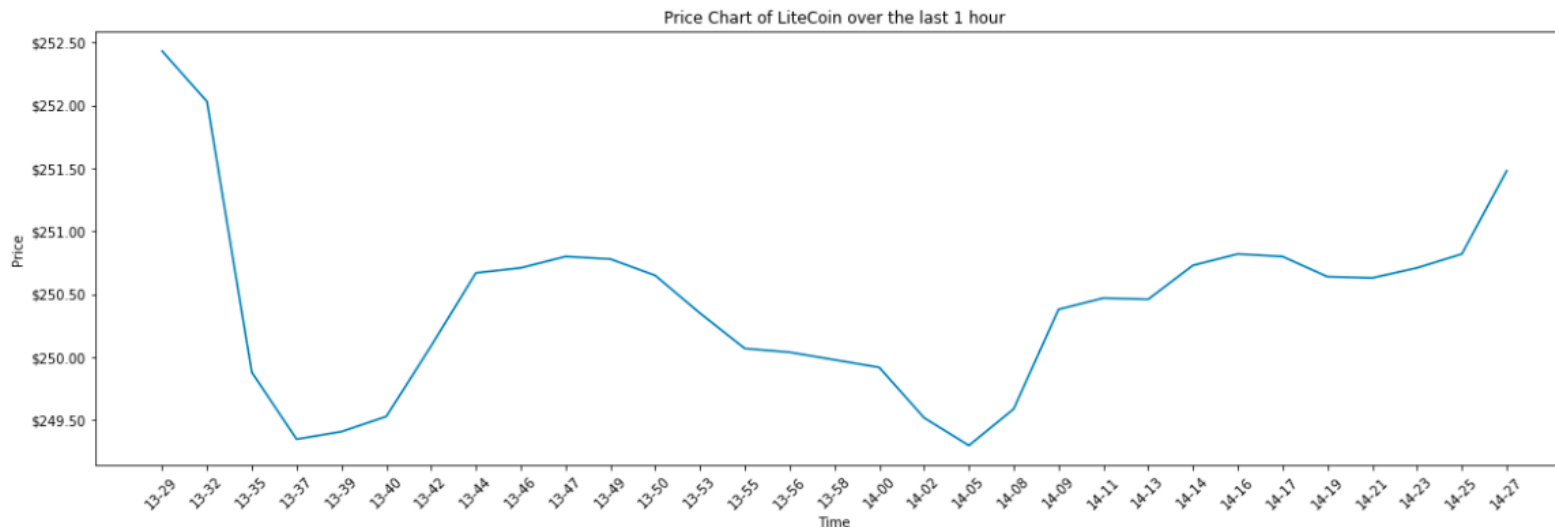
Investor

Pipeline provides real-time data for investors & traders

Example Analytics: Current coin prices and Returns over any time period

	Coin ID	Start Date	End Date	Initial Price	Current Price	Number of Days in Period	Return
0	ethereum	2015-08-07	2021-04-11	USD 2.83	USD 2,139.06	2,074	75,441.92%
1	bitcoin	2013-04-28	2021-04-11	USD 135.30	USD 59,518.00	2,905	43,889.65%
2	litecoin	2018-07-08	2021-04-11	USD 84.50	USD 252.75	1,008	199.11%

Example Analytics: Trend of coin prices over last hour



Pipeline use cases:

- **Investors** can monitor coin prices and returns of investments
- **Algo traders** can use data to identify potential signals to trade on