Project Name: CryptoGambit

Team:

1. Prasad Hanwadikar [hanwadip@oregonstate.edu]
2. Tanya Khemani [khemanit@oregonstate.edu]
3. Manideepa Saginatham [saginatm@oregonstate.edu]

Project Goal:

Primary goal of this project is to capture cryptocurrency price data for multiple currencies from various exchanges and analyze it to identify if there is any scope for arbitrage strategy. It will also be used to build a crypto-portfolio valuation dashboard (web UI).

Data API:

We are planning to use CoinAPI.io to gather required cryptocurrency data to perform the stated analysis in the project goal. This API provides data from multiple exchanges and for multiple crypto assets. We will consume the RESTful API in JSON format. Various data sets which will be consumed include information about assets, exchanges, pricing and trade volumes.

Further Questions:

1. Provide a brief description (about a paragraph) about cryptocurrency

   A cryptocurrency (or crypto currency) is a digital asset designed to work as a medium of exchange that uses cryptography to secure its transactions, to control the creation of additional units, and to verify the transfer of assets. Cryptocurrencies are classified as a subset of digital currencies and are also classified as a subset of alternative currencies and virtual currencies. Cryptocurrencies use decentralized control as opposed to centralized electronic money and central banking systems. The decentralized control of each cryptocurrency works through a blockchain, which is a public transaction database, functioning as a distributed ledger.

2. What is the format and scale of data you will be analyzing in this API? How are these data generated? Is it a real-time and distributed information? Please provide some examples/samples of your data.

   CoinAPI.io is a platform which provides fast, reliable and unified cryptocurrency data from multiple exchanges. It exposes RESTful API which will be used for capturing cryptocurrency quote data from various exchanges. Data format will be JSON.

   Metadata or Master data:

i. Exchanges: This is the list of cryptocurrency exchanges for which quote data will be fetched. Approximately these will be 90 distinct records.


```
[
  {
    "exchange_id": "BITSTAMP",
    "website": "https://www.bitstamp.net/",
    "name": "Bitstamp Ltd."
  },
  {
    "exchange_id": "BITFINEX",
    "website": "https://www.bitfinex.com/",
    "name": "Bitfinex"
  }
]
```


ii. Assets: This is the list of cryptocurrency assets for which quote data will be fetched. Approximately these will be 1400 distinct records.


```
[
  {
    "asset_id": "BTC",
    "name": "Bitcoin",
    "type_is_crypto": true
  },
  {
    "asset_id": "USD",
    "name": "US Dollar",
    "type_is_crypto": false
  }
]
```


iii. Symbols: A symbol is a tuple of format {exchange_id}_**SPOT**_{asset_id_base}_{asset_id_quote} which basically helps to uniquely identify any asset pair per exchange. Quote price (ask and bid) is always tagged against the symbol_id. Approximately these will be 10k records.

```
[
    {
        "symbol_id": "OKCOIN_CNY_SPOT_BTC_CNY",
        "exchange_id": "OKCOIN_CNY",
        "symbol_type": "SPOT",
        "asset_id_base": "BTC",
        "asset_id_quote": "CNY",
        "data_start": "2013-06-12",
        "data_end": "2017-11-01",
        "data_quote_start": "2015-02-15T12:53:50.3430000Z",
        "data_quote_end": "2017-10-31T16:05:27.0259310Z",
        "data_orderbook_start": "2015-02-15T13:05:36.1970000Z",
        "data_orderbook_end": "2017-10-31T16:06:17.7759609Z",
        "data_trade_start": "2013-06-12T14:24:24.0000000Z",
        "data_trade_end": "2017-11-01T00:00:04.3916158Z"
    },
    {
        "symbol_id": "HUOBI_SPOT_BTC_CNY",
        "exchange_id": "HUOBI",
        "symbol_type": "SPOT",
        "asset_id_base": "BTC",
        "asset_id_quote": "CNY",
        "data_start": "2015-03-29",
        "data_end": "2017-11-01",
        "data_quote_start": "2015-03-29T21:46:06.2630000Z",
        "data_quote_end": "2017-11-01T10:26:55.0275881Z"
```


Quote Price data:

Quote data is basically best Ask and Bid price per symbol (currency pair per exchange). This data will be analyzed to find for any arbitrage trade opportunity. Data is fetched real time at the time when API is called but note it is not continuous stream. CoinAPI.io also provides real-time streaming data over WebSocket API however it requires a paid subscription and won't be used for this project. Approximately quote data will be 10k records.

```
[
  {
    "symbol_id": "COINBASE_SPOT_ETH_USD",
    "time_exchange": "2017-07-09T11:15:04.2760000",
    "time_coinapi": "2017-07-09T11:14:30.8303308",
    "ask_price": 245.020000000,
    "ask_size": 110.698159470,
    "bid_price": 245.010000000,
    "bid_size": 8.051452000
  },
  {
    "symbol_id": "BITSTAMP_SPOT_BTC_USD",
    "time_exchange": "2017-07-09T11:14:39.2755858",
    "time_coinapi": "2017-07-09T11:14:39.4025858",
    "ask_price": 2536.000000000,
    "ask_size": 1.186175560,
    "bid_price": 2534.560000000,
    "bid_size": 0.895395050
  }
]
```


Exchange data:

Exchange data is basically weighted average USD price (among all exchanges) for the queried cryptocurrency. This data can be used to discover the aggregate portfolio value of any user having multiple cryptocurrencies. Data is fetched real time at the time when API is called but note it is not continuous stream. Exchange data is fetched per cryptocurrency, so number of API calls depends on the distinct cryptocurrencies owned by the user.

```
{
  "asset_id_base": "BTC",
  "rates": [
    {
      "time": "2017-07-09T11:03:39.3143739Z",
      "asset_id_quote": "ETH",
      "rate": 0.095830155
    },
    {
      "time": "2017-07-09T11:04:50.6070000Z",
      "asset_id_quote": "JPY",
      "rate": 291117.96954209290933276041852
    },
    {
      "time": "2017-07-09T11:04:50.8100000Z",
      "asset_id_quote": "USD",
      "rate": 2559.50629169160446304389846 51
    }
  ]
}
```


3. What challenges do you think you will be faced while integrating this data?

    a. Deducing the logic to identify arbitrage opportunities in the quote data
    b. Need of some other kind of data which we are not aware of for now

4. How do you want to analyze it? How will your final output like? Is it a diagram or …?

    We will load the captured data in some RDBMS system like MySQL. Data will be analyzed with SQL queries and analysis will be shown on web or thick client UI. Arbitrage opportunities and portfolio valuation will be shown as two separate UI grids.

    Arbitrage Grid Columns:

    a. T1_Exchange
    b. T1_Currency
    c. T1_Buy_Sell
    d. T1_Amount
    e. T2_Exchange

f. T2_Currency
g. T2_Buy_Sell
h. T2_Amount

Portfolio Grid Columns:

a. Coin
b. USD price per unit
c. Holding Units
d. Holding USD Value