

Final Project

Crypto Forecast with Twitter Data

E. Rafael Hernandez



CSCI E-88, Fall 2022, Principles of Big Data Processing
Harvard University Extension School
Prof. Marina Popova





















PROJECT GOAL AND PROBLEM STATEMENT

With the recent collapse of the FTX crypto scam, I wonder if data from Twitter related to the top cryptocurrencies can be collected and used to correlate and forecast volatility.

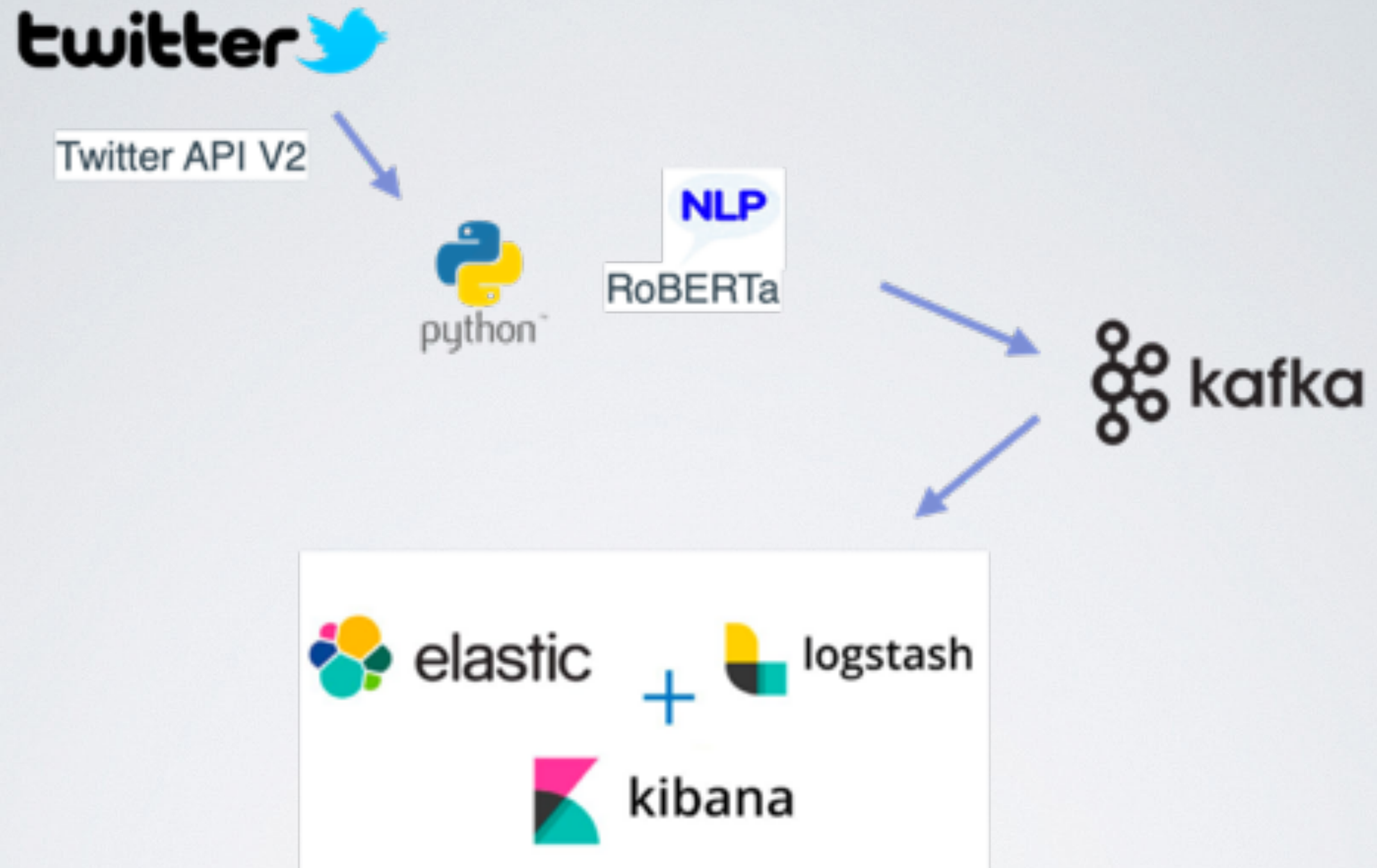
This project aims to demonstrate how to build a system that collects real-time tweets of the existing cryptocurrencies -using keywords. This data can be indexed by its exchange symbols and fed into Elastic Cloud for graphical presentation and analysis.

BIG DATA SOURCE

The data source is Twitter streaming data, filtered to receive tweets related to the top 20 cryptocurrencies.

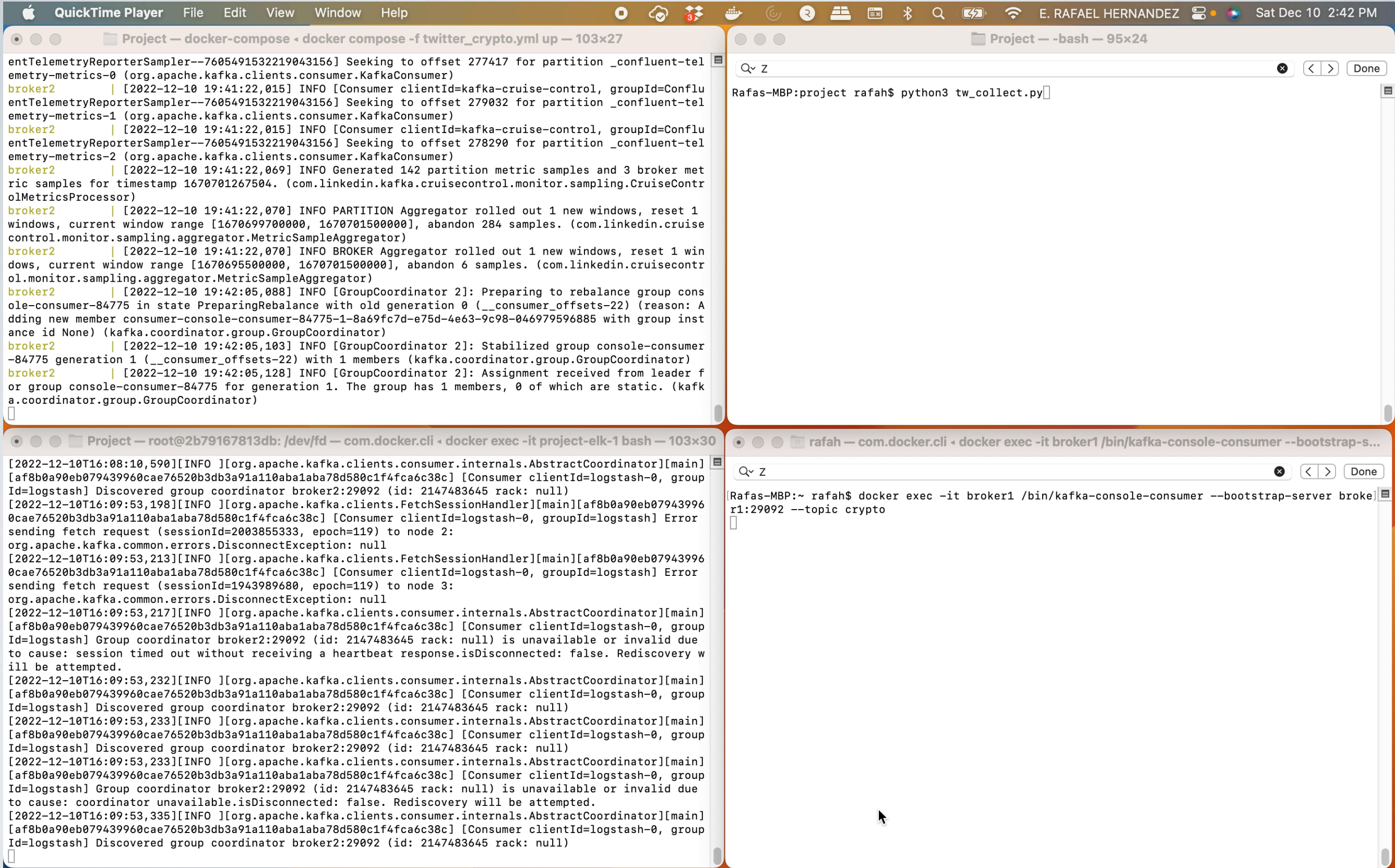
#	Name
☆ 1	 Bitcoin BTC
☆ 2	 Ethereum ETH
☆ 3	 Tether USD
☆ 4	 BNB BNB
☆ 5	 USD Coin USDC
☆ 6	 Binance USD BUSD
☆ 7	 XRP XRP
☆ 8	 Dogecoin DOGE
☆ 9	 Cardano ADA
☆ 10	 Polygon MATIC
☆ 11	 Polkadot DOT
☆ 12	 Dai DAI
☆ 13	 Litecoin LTC
☆ 14	 Shiba Inu SHIB
☆ 15	 TRON TRX
☆ 16	 Solana SOL
☆ 17	 Uniswap UNI
☆ 18	 UNUS SED LEO LEO
☆ 19	 Avalanche AVAX
☆ 20	 Chainlink LINK

From: <https://coinmarketcap.com>

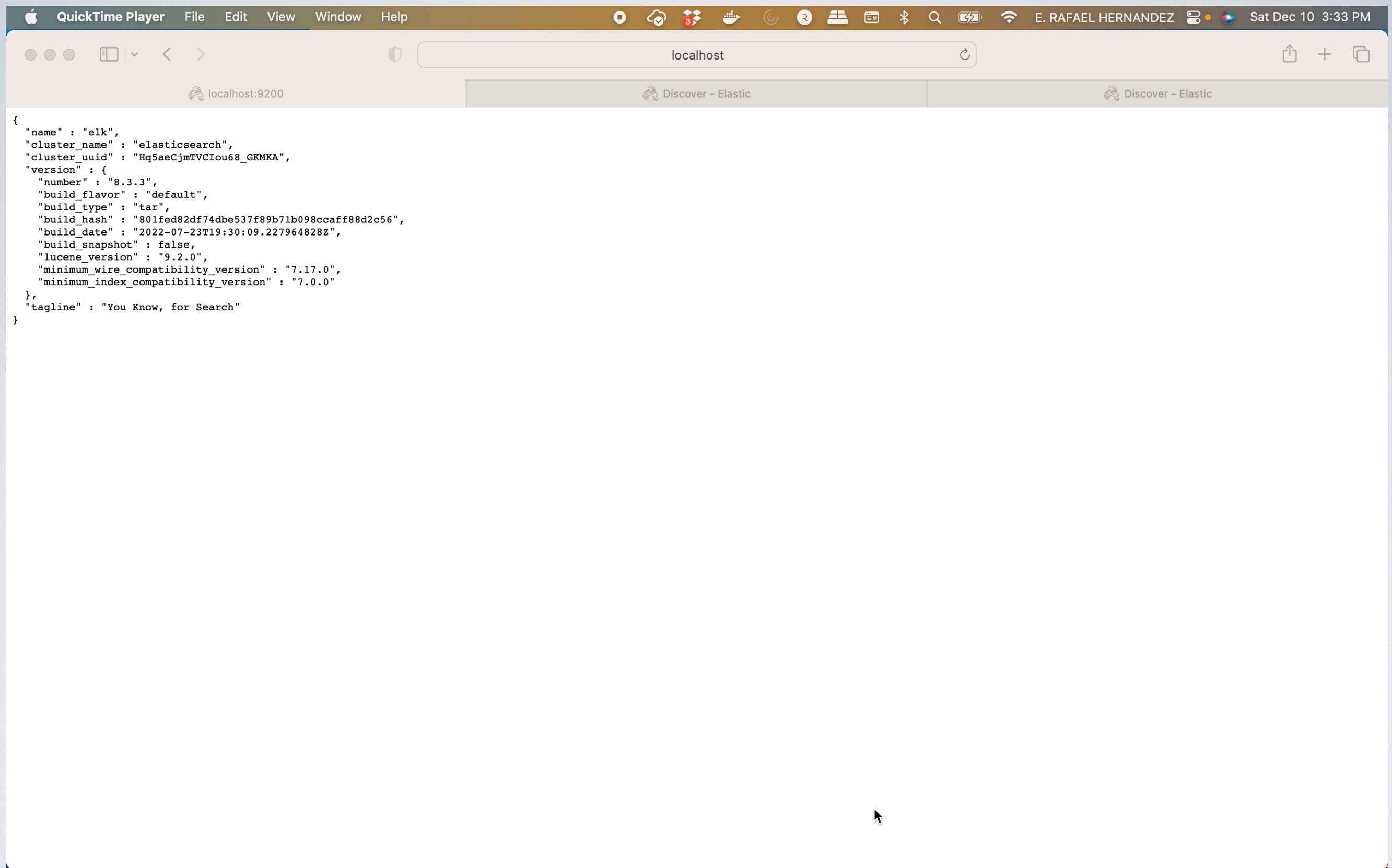


PROCESSING PIPELINE DIAGRAM

QUICK DEMO



DEMO



DEMO

THAN YOU