Algorithmic Trading System

Financial Data Collection, Database and Integration Prototype using Jupyter Notebooks

- Ajitesh Parihar, COSC, Okanagan College

Okanagan COLLEGE

Intro

- Financial data needed for training ML model
- Problems with manual collection of this data
- Demo of Database User Interface and Integration for Prediction Prototype

Okanagan COLLEGE

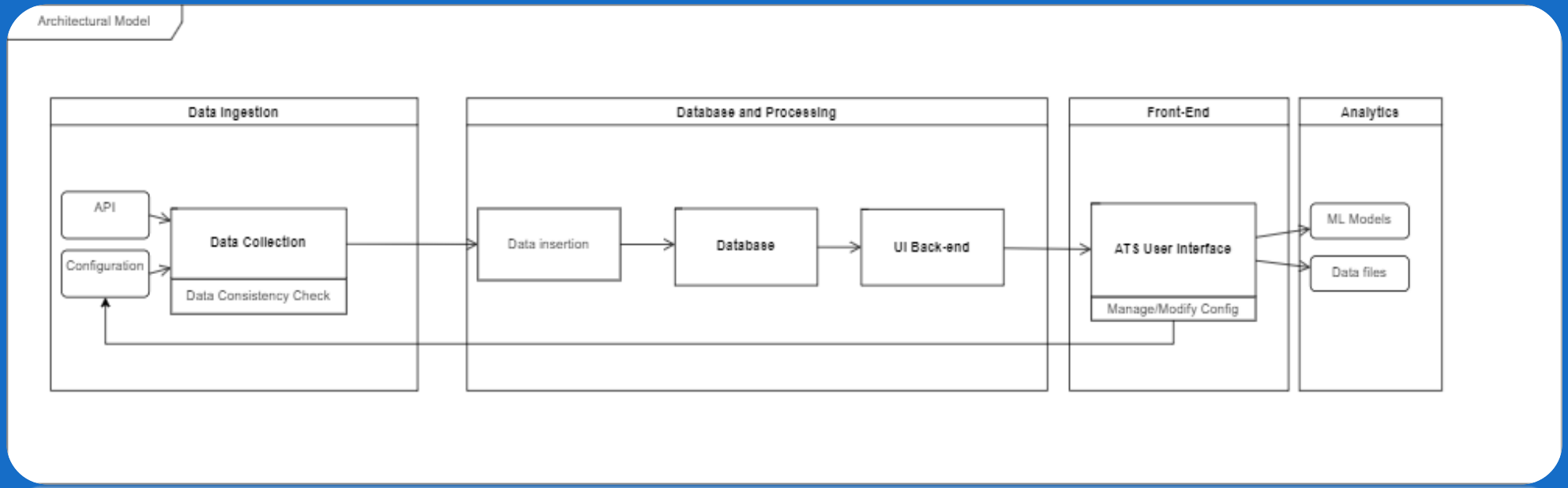
Intro

- Importance of a good dataset

Objective

- Creating a reliable datastore for efficient data collection
- Web User Interface for ease of use

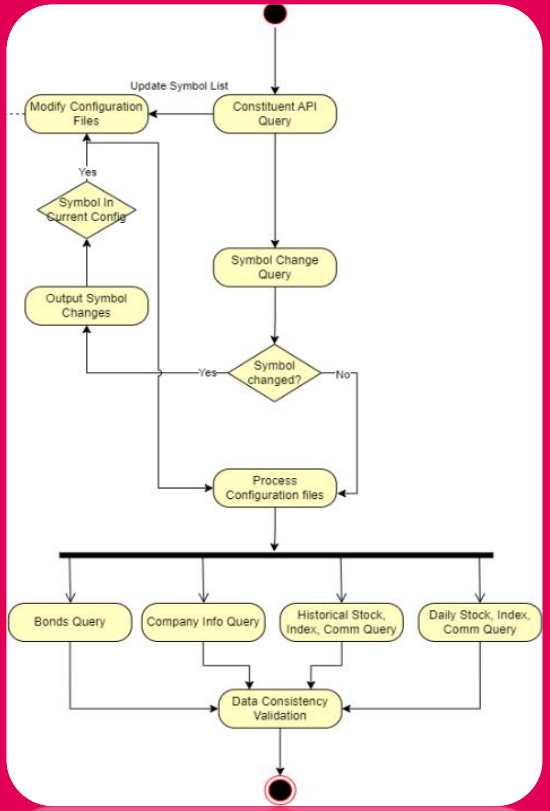
System Architecture



Data Collection

- Sourcing data from financial data APIs
- YAML files for configuring types of data collected
- Modular bash and python scripts

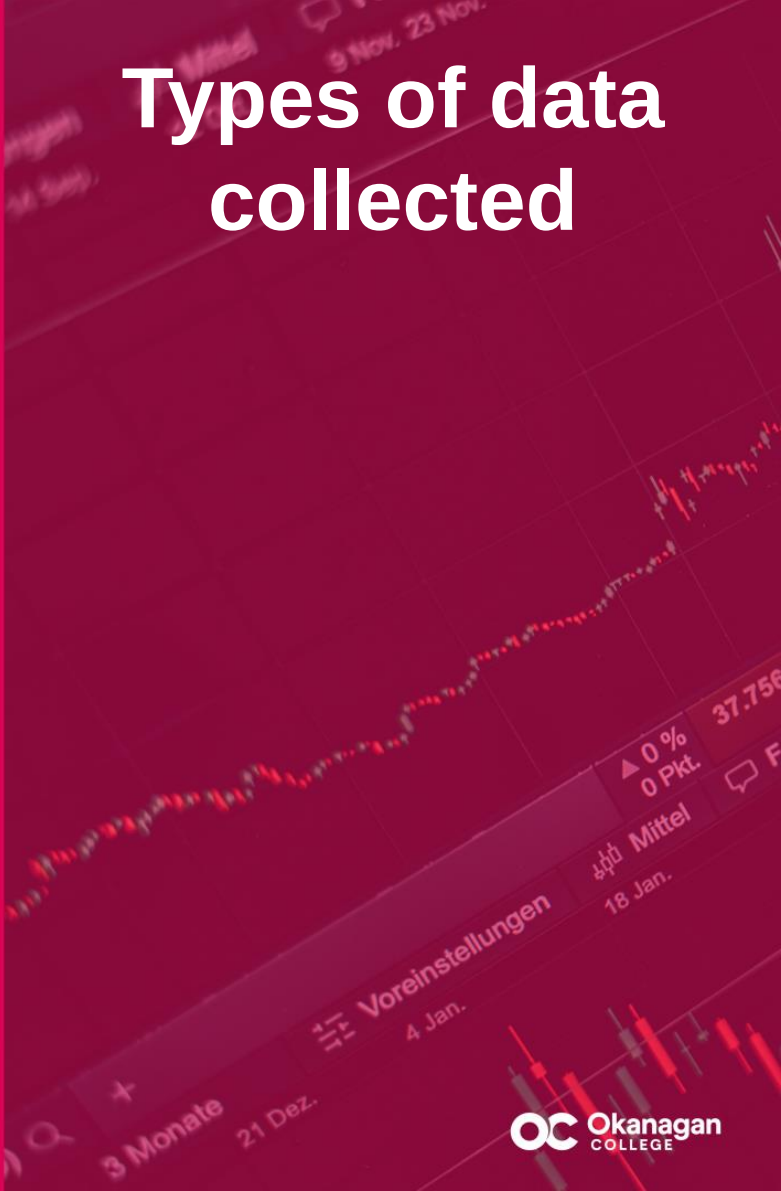
Data Collection Process



Okanagan COLLEGE

Types of data collected

- Stocks
- Indexes
- Commodities
- Bonds

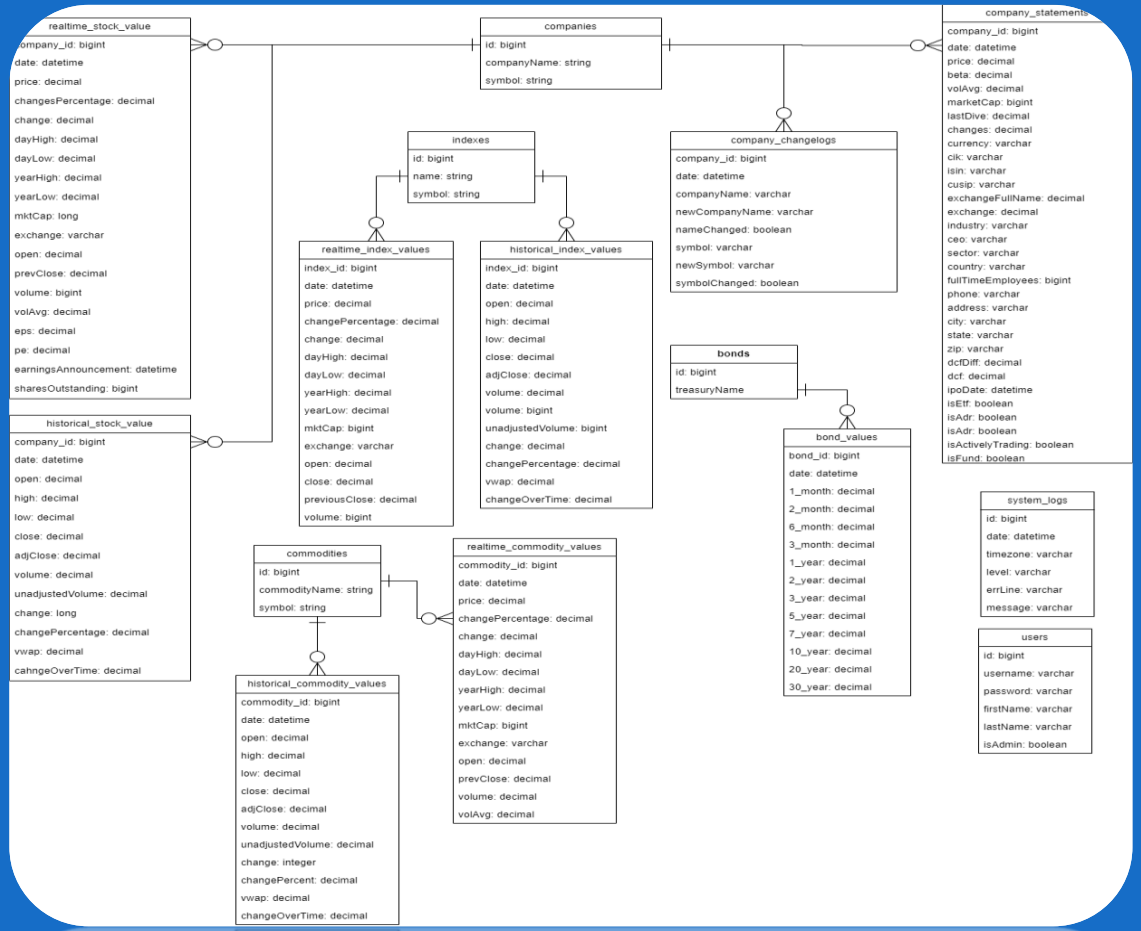


Okanagan COLLEGE

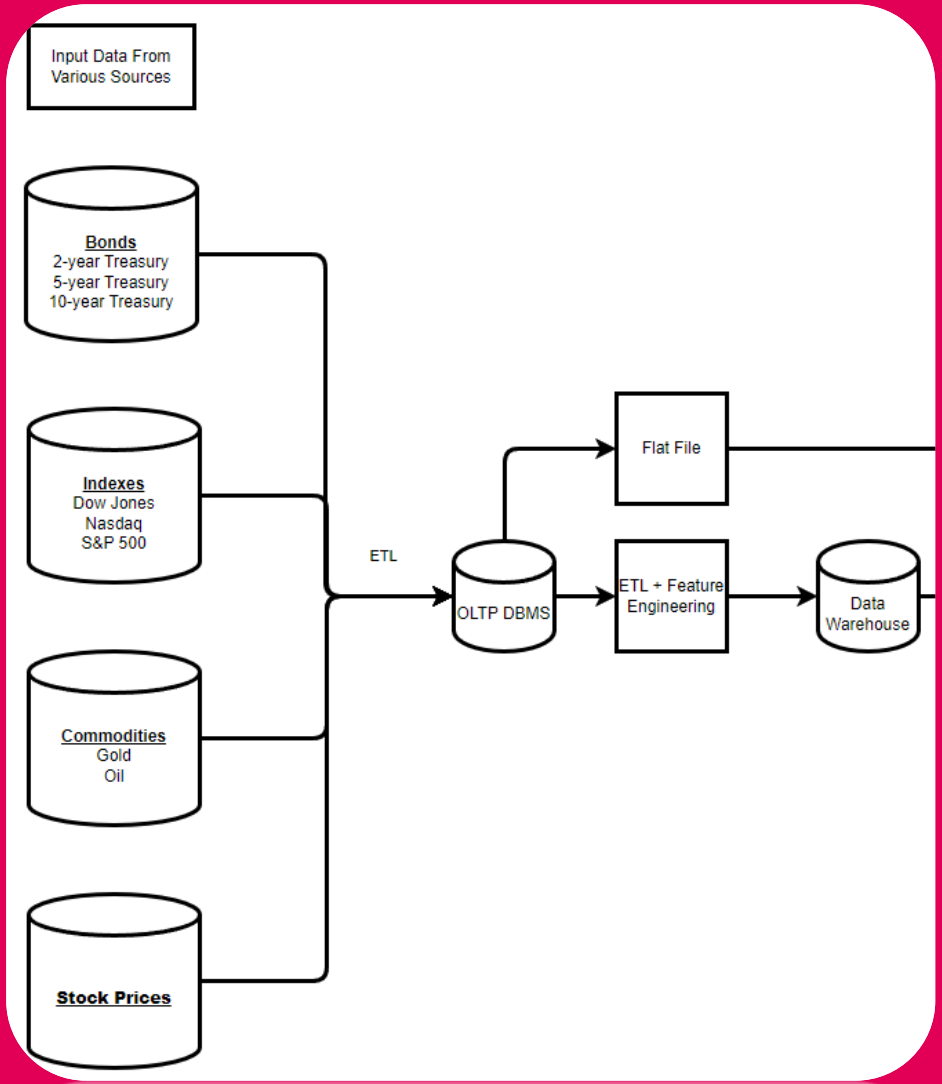
Database Design

- OLTP DB requirements
- MySQL for the current stage
- Plans to develop data warehouse to handle larger scales of data

Database Schema



Okanagan COLLEGE



Data Warehouse and Data Extraction, Transformation, and Loading

Okanagan COLLEGE

Process automation

- Scheduled collection and storage of data using cron jobs
- Customizable configuration

User Interface

- Configuration
- Job scheduling
- Exporting data
- Roles based access

Okanagan COLLEGE

Integration Prototype

- High performance hardware
- Jupyter notebooks
- Database integration
- Historical data analysis
- Preprocessing data for machine learning
- Model training of demo model

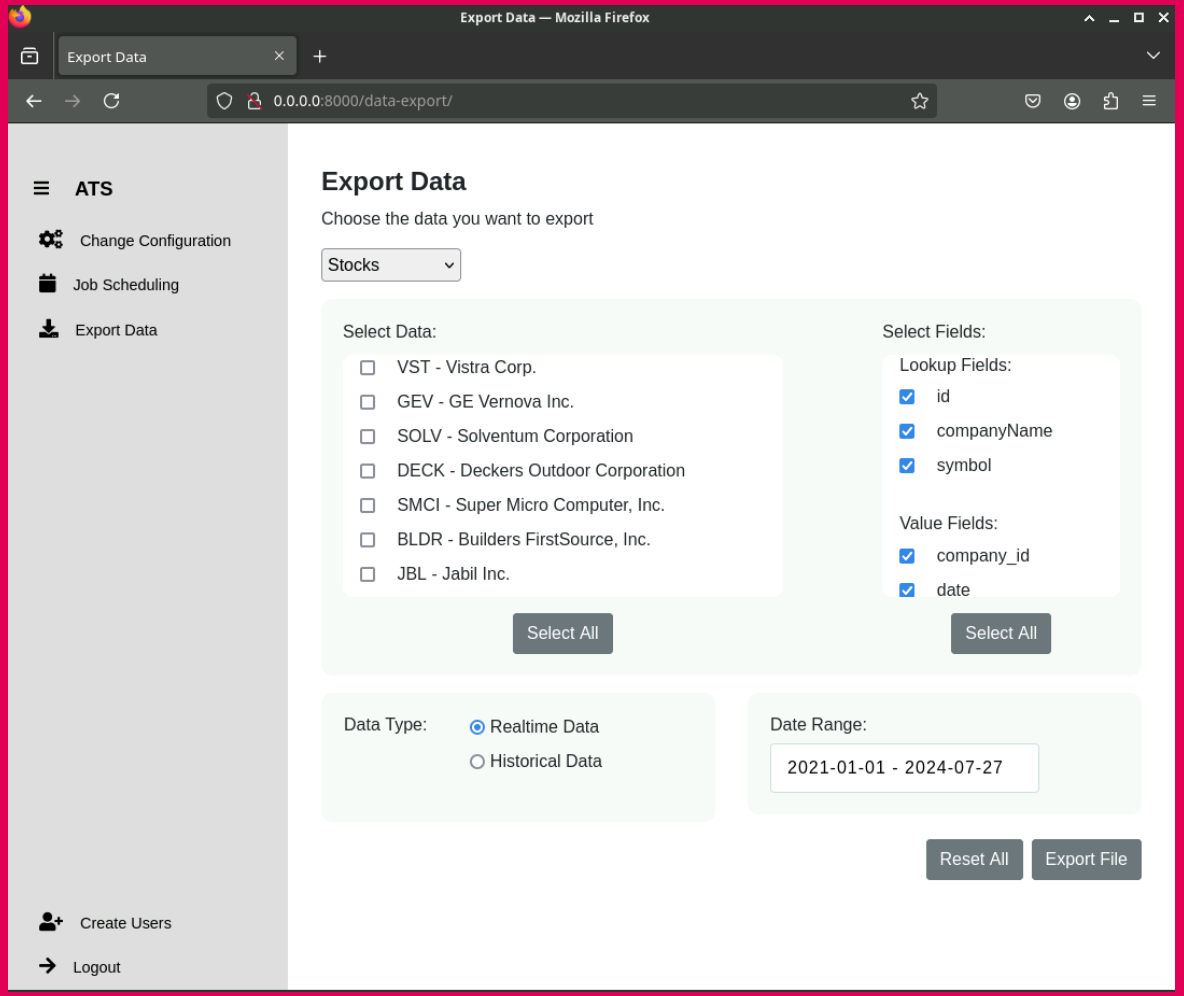
Okanagan COLLEGE

Future Work

- Data Warehouse Development
- UI improvements
- BI tools integrations
- UI integration for stock forecast analysis

Okanagan COLLEGE

UI Design Demo



Okanagan College

Integration Demo

jupyter ATS Last Checkpoint: 8 seconds ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

Algorithmic Trading System Demonstration

This project is being developed for short-term stock price forecasting with collaboration between Langara College, Okanagan College and University of Paris East Creteil. It is inteded to make it easy to analyze historical stock data and train a machine learning model to generate accurate predictions for selected stocks. The project is divided into three categories:

1. Data warehouse with automated data extraction, transformation, and loading of trading data from multiple sources: An OLTP database is developed using processes to collect trading data and transforming to for use in training the machine learning algorithm.
2. A machine leaning model for short-term algorithmic trading forecast: Traditionally, stock price prediction uses technical and fundamental analysis, which has its shortcomings. With adding sentiment analysis to the mix, it is possible to develop a generalized algorithm that can be applied to any stock, future, or traded commodity. Machine learning algorithms such as neural networks, random forecasts, support vector regression, XGBoost, and long short-term memort (LSTM) were evaluated and XGBoost was chosen for further development based on performance analysis.
3. Automation for assessment of algorithmic short-term trading strategies.

In this demonstration, we will analyze the collected data, the transformation process for preparing data to train the ML model, training the ML model, and the prediction accuracy based on several metrics, along with computational performance metrics.

Database Exploratory Analysis

In this section, we will look at the historical data for stocks, indexes, commodities, and bonds. The database provides fast and easy access to the required trading data for analysis.
Begin by loading the following notebook.

[69]:
```
%%capture
%run ./DatabaseExplore.ipynb
```

Load bonds, commodities and index data from database

The database contains historical trading data for the past five years

Fetch the required data from the database:
Choose the date range for which you need to analyze the data for. (Format: YYYY-MM-DD)
Add the stock symbols in the 'symbols' list to choose stocks (case sensitive)
List of stocks

Indexes
The database contains data for the following indexes:

1. S&P 500
2. NASDAQ Composite
3. Dow Jones Industrial Average
4. Russell 3000
5. FTSE 100

Bonds

Okanagan COLLEGE

Conclusion

- Established a reliable datastore and streamlined data workflows
- Enabled users with intuitive tools and customizable configurations for efficient data handling and analysis
- Created integration framework for adaptation into future work

Okanagan COLLEGE

Thank You

Any Questions?

Okanagan COLLEGE