

Thalia Project Report

Team Charlie

**Martti Aukia, Albert Boehm, Arthur-Louis Heath,
George Stoian, Daniel Joffe, Weronika Kakavou,
Marcell Veiner**



University of Aberdeen
Wednesday 26th February, 2020

Acknowledgement

TO DO: Thank stackoverflow and everyone else here.

Contents

1	Introduction	3
1.1	Project Overview	3
1.2	Motivation/Rationale	3
1.3	Project management strategy	3
1.4	Budget	4
2	Background and Competitors	5
2.1	Portfolio Visualizer	5
2.2	Other Competing Software	6
2.3	Summary	8
3	Requirements	9
3.1	Functional Requirements	9
3.2	Non-functional Requirements	11
3.3	Use Cases	12
3.4	Feasibility Analysis	12
4	Design	13
5	Coding and Integration	14
5.1	Overview	14
5.2	Planning	14
5.3	Schedule	16
5.4	Key Implementation Decisions	17
5.4.1	Web Framework	17
5.4.2	GUI	17
5.4.3	Business Logic	17
5.4.4	Database & Finda	18
5.4.5	Harvester	18
5.5	Integration	18
5.5.1	Continuous Integration	19
5.6	Hosting and Deployment	20
6	Testing and Evaluation	22
7	Conclusions and further work	23
8	Appendices	24

1 Introduction

1.1 Project Overview

The aim of this project was to create a portfolio backtesting software, which enables creating custom portfolios and measuring their performance with different backtesting functions. The user can pick from a variety of assets, some of which are Equities, Fixed Income, Currencies, Commodities, and Cryptocurrencies. Risk metrics and performance are then visualized for the given asset allocation.

1.2 Motivation/Rationale

Since retail investing is a growing market, our target audience consists of individual investors, who instead of seeking the full package that comes with financial advising, would like to take over the wheel and assess the viability of their investment strategies themselves. As retail investors are non-professionals and invest comparatively small amounts, with financial advice services being non affordable for those individual clients, our goal was to create a product that would not only be more affordable for small retail investors, but would also include a variety of international assets, which most existing backtesting software fails to provide [?].

1.3 Project management strategy

The creators of Thalia are:

- Martti Aukia, *Team leader*
- Arthur-Louis Heath, *Deputy team leader* Leader
- Albert Boehm, *Chief editor*
- Marcell Veiner, *Product design lead*
- George Stoian
- Daniel Joffe
- Weronika Kakavou

Our goal was the creation of a high quality industrial prototype of the Thalia backtesting software based on the identified project requirements [?] and optional features [?]. The team held regular meetings, both with the project guide, Dr Nigel Beacham, and the Dr Ernesto Compatangelo. During the analysis stage meetings took place weekly and were aimed to discuss ideas and requirements. Whereas later, during the implementation stage the team held meeting at least once per week, some of which were aimed to discuss the design for the tool with the inclusion of some coding sessions.

As in the past[?], our workflow was centered around the GitHub platform and the tools it provides. Furthermore, we continued to follow the Egalitarian Team structure, as this worked very well during the first term of the course. In our Technical Report [?] we also discussed the use of effort-oriented metrics, i.e. story points, which were assigned to tickets based on the time needed and the functionality of the task. However, the following term we concluded that in many occasions these metrics failed to successfully measure the size of a task, as they were artificially assigned. Based on this observation, we decided not to make use of them for the rest of the development.

As we were fortunate enough to gain an additional member this term, some of our initial effort was focused on introducing her to the project and team dynamics. Given our access to a large team, it was common to see coding tickets assigned to pairs and groups rather than individuals. We believe this not only resulted in writing better-quality code, but in faster team communication when making design decisions.

1.4 Budget

As previously discussed, we did not have any budget restrictions other than time. All of us agreed to allocating a minimum of 10 hours per week to development efforts and discussed personal expectations well in advance. We estimate the value of our products goodwill to be approximately 13000£. We base this estimate on the developer time allocated to this project, the total duration of 20 weeks, and average developer salaries in the UK[?] (taking into account that we have no industry experience).

2 Background and Competitors

As a result of our initial competitor analysis, we may group the competing software into two categories. For a comprehensive list of available backtesters please see [?]. The first consists of various third-party trading software, such as Fidelity, MetaTrader and NinjaTrader that offer backtesting features as well. Services in this category include features such as buying and selling of financial assets, prediction of future prices and storing and managing of users' money. Although it may be beneficial to consider some trading features provided by the members of this group, these would only be part of a future development process, and not in the scope of this course.

The second category consists of pieces of software that do not offer trading as a service, and solely focus on backtesting. Thus, for now, we shall only consider the feature set provided by the second category. In the following paragraphs we will consider some of the design decisions made by competing software, together with the brief analysis and conclusion on each feature.

2.1 Portfolio Visualizer

Our main competitor of the second category is an online backtesting tool named Portfolio Visualizer[?]. During the inception phase of development we heavily relied on this website for writing the requirement analysis. Let us now briefly dissect what it has to offer.

Upon opening the website we are greeted with a brief description of the domain, together with the following input form:

Time Period ⓘ	Year-to-Year ▼
Start Year ⓘ	1985 ▼
End Year ⓘ	2019 ▼
Initial Amount ⓘ	\$ 10000 .00
Cashflows ⓘ	None ▼
Rebalancing ⓘ	Rebalance annually ▼
Display Income ⓘ	No ▼
Benchmark ⓘ	Vanguard 500 Index Investor ▼

Figure 1: Portfolio Visualizer - Input (source: <https://www.portfoliovisualizer.com/>)

As we can already see, the key elements of portfolio analysis are provided. Considering the functionality of our prototype as a baseline (that is testing an allocation of assets with a fixed initial investment on a fixed time period), we see that in addition users are able to do the following set of features:

- Set the endpoints of the investment period, up to months.
- Set the initial investment.
- Specify a regular cashflow and its frequency.
- Select a rebalancing strategy.
- Select a benchmark strategy for comparison.
- Compare multiple strategies at the same time.
- Adjust to inflation.

- Select from a set of lazy portfolios.
- Calculate additional metrics.
- Export the results to PDF, Excel, or save link.

At first it may seem as if Portfolio Visualizer meets all the requirements needed for a financial backtester, and indeed our main criticism is regarding the UI, responsiveness and user experience, and is a result of the initial user testing.

The UI design is simplistic and has a non-commercial, bare-bones look, and although this was appreciated while testing the system, it certainly does not improve the user experience. The input, mostly using dropdown menus is straightforward to use, except for the selection of Assets, which we will discuss briefly at the end of this section.

Moreover, users will want to fine-tune their investment strategies, by changing their allocations frequently. The only means to do this using Portfolio Visualizer is to scroll to the top, change the input and rerun the entire simulation. Our goal is to design a more responsive and dynamic system, to ease this procedure.

The website also has user accounts, however little to no data is associated with a users profile, which further decreases the overall user experience. Finally, as a last remark, which holds for most of the back-testing tools we have tried, the website is heavily US biased, and so the selection of asset classes is limited. Furthermore, we have no means of changing the currency, which would also be an important feature for a product accessible through the web, and therefore available to users from all over the globe.

2.2 Other Competing Software

The rest of the alternatives are of a significantly lower quality. In the remaining parts of this section we will briefly consider a few design decisions made by these websites.

Tree-like Asset Selection

One challenging aspect of designing such a system, is the following:

What is the most intuitive way for selecting a portfolio item?

Where a portfolio item could be: an equity, ETF index, a commodity, bond or stock. Within these classes we have more options to choose from. Many backtesters opted for a search bar, which is a sensible approach, but poor in practice. Many portfolio items are named similarly, and for a new user it is a barrier, as they might not know what is available.

One of the better option is what ETFReplay [?] has implemented, which is a tree-like selection form, shown below.

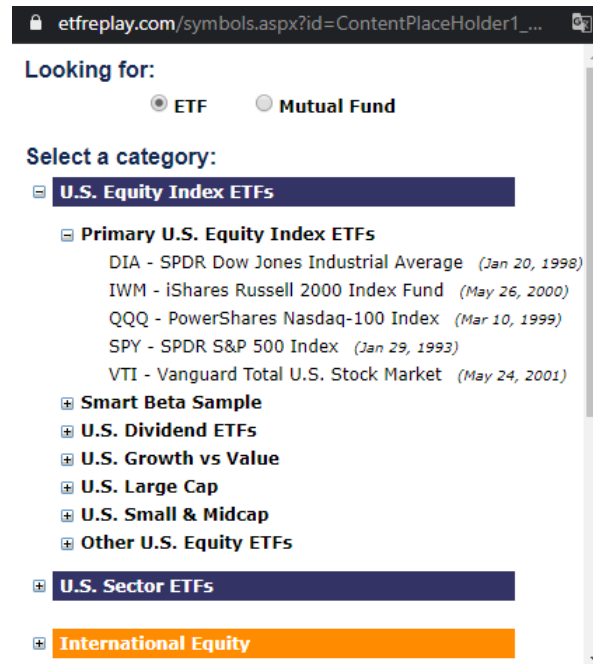


Figure 2: ETFReplay - Tree-like Structure (source: <https://www.etfreplay.com/>)

We feel this was the most intuitive to use, and will thus pursue a similar approach. The only potential issue identified with this approach was that it opens in a new window, which we would prefer to avoid.

Tiles

It may be worth briefly discussing an example of a backtester with a good layout design. We found the simplistic and tiled design of Backtest Curvo [?] a good choice as it gives the website an overall modern and fresh look, whereas most backtesters looked old and out of date. Our goal is to achieve a layout similar to this. For a further analysis on design decisions, please see section 4 of this report.

Simple Logic

In our whitepaper, we briefly discussed implementing a simple scripting language allowing users to simulate dynamic trading strategies [?]. As mentioned, this would not be in the scope of the course, but for inspiration we can have a look at the approach of Stockbacktest [?].

Signals (Need at least one buy or short signal)
 Buy When:

Signal 1:	Upper Bollinger Band (days devs) ▼	Crosses Below ▼	Percentage % ▼	AND ▼
	20 2	none	70	
Signal 2:	Close Price ▼	Is At Least % Above (%) ▼	Percentage % ▼	
	none	15	10	

Figure 3: Stockbacktest - Simple Logic (source: <http://stockbacktest.com/>)

Although not intuitive at all, given its presentation, it shows this feature is also possible. We believe that after nailing down how exactly the user would create the rules, this would be a straightforward task, as the calculations involved are not more complicated than in the static case. As mentioned above, this is something we would like to implement in the future.

2.3 Summary

As we initially tested the competing software during the inception phase of development, we were left with three key observations. These were the following:

- Transparency: Many of the testers' design did not give the impression of being operated by a reputable businesses. In the worst cases they looked transparent or trustworthy (misspelt words, advertisements with questionable subject matter), which is something we should avoid.
- Learning Curve: After getting to know our target audience we concluded that most seem willing to learn how to use a complex system if it is worth it.
- Design of the UI: It was easy to tell which backtesters are still getting updated, by looking at their design. We should aim at looking fresh. In addition, some backtesters offered so many features that every corner of their UI was packed with information, this is something we should also avoid doing.

The analysis was conducted informally with the help of colleagues studying or involved in finance. This allowed us to develop a more neutral and accurate impression, despite the subjective nature of the observations. Having analysed the competing software, we are now better suited to determine the requirements of our system.

3 Requirements

Earlier in the inception phase of the development, we classified our requirements using the established FURPS+ model [?]. Since identifying requirements for our initial project report [?] we have also identified some additional requirements that Thalia should fulfil based on feedback from our project guide. Let us briefly revisit our main functional and non-functional requirements. The items outlined below focus on some of the key requirements we have so far identified as features necessary for providing a compelling product for paying customers.

3.1 Functional Requirements

Portfolio Configuration	
Allocate fixed amount/proportions of the portfolio to given assets	Choose how much each asset contributes to the portfolio's total value using either percentages or raw monetary amounts
Find assets quickly by category or name	When adding an asset the user can search a category for assets or search for a specific asset by its name
Share portfolio	Portfolios can be shared between people using a URL
Edit portfolio	Change asset allocation and their distributions in a portfolio

Portfolio analysis	
Compare portfolios	Use multiple portfolios in a single analysis to see differences in their performance
Use a selection of lazy portfolios	Select an existing common portfolio to compare against, such as common index funds (e.g. Vanguard 500 Index Investor or SPY)
Plot portfolio as a time-series	View portfolio performance as a line graph for a quick overview
Specify a time frame for the analysis	Select start and end dates for portfolio analysis
Choose rebalancing strategy	Optionally choose a strategy for buying and selling assets to meet your strategy e.g. buying and selling stocks each year to ensure the value of portfolio stays at 60% stocks and 40% bonds (i.e. maintain the initial allocation)
Change the distribution of assets in a portfolio using a slider	A slider for each asset to quickly increase or decrease its proportion of the total value
Edit portfolio analysis	Change parameters for portfolio's analysis after running it (e.g. date range or rebalancing strategy)

View results	
See key numerical figures	Show important numerical metrics for a portfolio's performance such as Initial Balance, Standard Deviation, Worst Year, Sharpe Ratio, and Sortino Ratio
See both real and nominal values	See portfolio's value as both adjusted and not adjusted for inflation
A breakdown of portfolio value at specific points of time	See what the value of the portfolio is at some point in time (e.g. January 3rd 1997)
Export result of analysis	Exports results to PDF for sharing and offline reading

User accounts	
Combine portfolios	Combine two portfolios' assets into one single portfolio
Save portfolio analysis for later	Save portfolio analysis parameters to the account so you can rerun it with a single click
Delete saved portfolio analysis	Remove a stored portfolio analysis from your account
Manage portfolio analyses	Edit saved portfolio analysis with different assets, distributions or other parameters
Sign-up, log in and log out	Basic authentication

Assets	
Choose assets from European market	Data for European assets were found to be lacking in competing products
Choose assets from Equities, Fixed Income, Currencies, Commodities, and Cryptocurrencies	Coverage of some of the largest asset classes

3.2 Non-functional Requirements

1. Usability:

- The product must be easily usable for users who already have some financial investment experience.
- The basic backtesting interface needs to look familiar to people already experienced with it.
- The product must have detailed instructions on how to use its advertised functions.
- All major functions must be visible from the initial landing page.
- Must work in both desktop and mobile browsers.
- The results page should scale with mobile.

2. Reliability:

- The product must have a greater than 99% uptime.
- All our assets need to have up to date daily data where the asset is still publicly tradeable.
- All assets supported by the system must provide all publicly available historical data.

3. Performance:

- The website should load within 3 seconds on mobile [2].
- Large portfolios must be supported - up to 300 different assets.

4. Implementation:

- The system needs to work on a cloud hosting provider.

5. Interfacing:

- The Data Gathering Module must never use APIs stated to-be-deprecated within a month.
- The Data Gathering Module must not exceed its contractual usage limits.

6. Operations:

- An administrator on-call will be necessary for unexpected issues.

7. Packaging:

- The product needs to work inside a Linux container (e.g. Docker).
- All dependencies need to be installable with a single command.

8. Legal:

- All user testing must be done with ethical approval from the University.
- UI must display a clear legal disclaimer about the service not providing financial advice.
- All third-party code should allow for commercial use without requiring source disclosure (e.g. no GPL-3).
- User data handling should comply with GDPR.
- Provided services should not constitute financial advice under UK law to avoid being subject to financial advice legislation and potential liability

9. Accessibility

- Display items should be clearly labelled
- UI should scale to accommodate different screen sizes and aspect ratios
- UI elements and text superimposed over one another should have high contrast in their colors
- UI should allow for the use of assistive technologies to accommodate individuals with accessibility issues

In the literature, the FURPS+ model has been criticised for disregarding developer consideration [?], such as not taking into account portability and maintainability. Furthermore it has been pointed out, that our requirements fail to capture ...

TODO

3.3 Use Cases

TODO

3.4 Feasibility Analysis

TODO

4 Design

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

5 Coding and Integration

After a brief overview of the project and the project plan, this section will focus on the main technologies used in the project and the rationale behind choosing them. Moving on, we will discuss how these components were integrated and eventually deployed.

5.1 Overview

One of the first decision we have made this term was to completely recreate our prototype of Thalia. Firstly, this radical move was the consequence of a new implementation decision (for more detail see 5.4.1). Secondly, as prototypes are meant to be disposable and are designed only to answer key questions about the system [?], the proof of concept served its purpose, giving us a chance to refine the structure and the quality of the code.

Despite the risks posed by using an Software Version Control (SVC) Host such as GitHub, we have decided to continue using it as our software development platform. The reasoning behind this builds on the argument developed in our Technical Report [?], which highlights that our, that our Data Processing Module is a completely separate component in our system which none of the other components is able to access. Additionally, we are storing API keys as environment variables in a file that is not tracked by our SVC system, which minimizes the probability of us exposing sensitive data.

TODO talk about API keys and security measures

We have also decided to develop the application with python as our main choice of programming language. Even though this choice seemed obvious from the beginning, we did consider its main benefits, these would be the following:

- Python is a high level programming language allowing us to better focus on the application.
- A standard choice for prototyping.
- Provides superb third party libraries and frameworks for free.
- Easy to integrate if we were to choose other languages at some point in our development.
- The whole team was already familiar with the language, saving us the precious time needed to learn another programming language.
- Our application does not require an unreasonable amount of computation, so there is no need for a more efficient programming language such as C. 5.4.3 [?].

We will discuss other technologies used in more detail after the discussion on project planning.

5.2 Planning

Early in the inception phase of development we have decided that our goal was not to have fixed responsibilities in our team, allowing everybody to work on each component of the system. This decision has also eased the code review process, as we had no status differences to distort the error-correcting mechanism [?]. Furthermore, since no team member was the sole developer of a system component, this allowed us to direct comments at the code and not the author [?]. For these reasons we have decided to follow the Egalitarian Team structure, and make use of the flexibility offered by it.

Our workflow was centered around the tools provided by GitHub. We used a ticketing system to divide and distribute tasks among team members. These tickets were sometimes given by the team on the weekly meetings, but occasionally they were chosen by the member proposing the feature or change. Our goal with this approach was to divide larger jobs into smaller tasks.

A typical ticket in our project was an encapsulation of a user story, as it consisted of a title, a one liner, value, acceptance criteria, and sub tasks. In the first half of development we also used effort-oriented metrics, called story points to measure the amount of work but we decided to abandon this aspect. An example of a ticket can be seen on Figure4.

mara42 commented on Nov 17 2019 • edited
+
😊
...

As a user, I want to see the results of a backtest, so I can use the tool.

Value

- Working product

Acceptance criteria

- Does pressing submit on the portfolio page return results of the backtest?
- Do the results contain a table of key metrics?
- Do the results contain a graph?

SP:?

Subtasks:

- ☒ Send user input to backend
- ☒ Send backtest data back to frontend
- ☒ Create graph
- ☒ Create table

Figure 4: Ticket Example (source: <https://github.com/>)

TODO update git stats

Throughout the whole project we had a total of 110 tickets, and 71 pull requests. The following graph also shows the number of contributions to master, excluding merge commits.

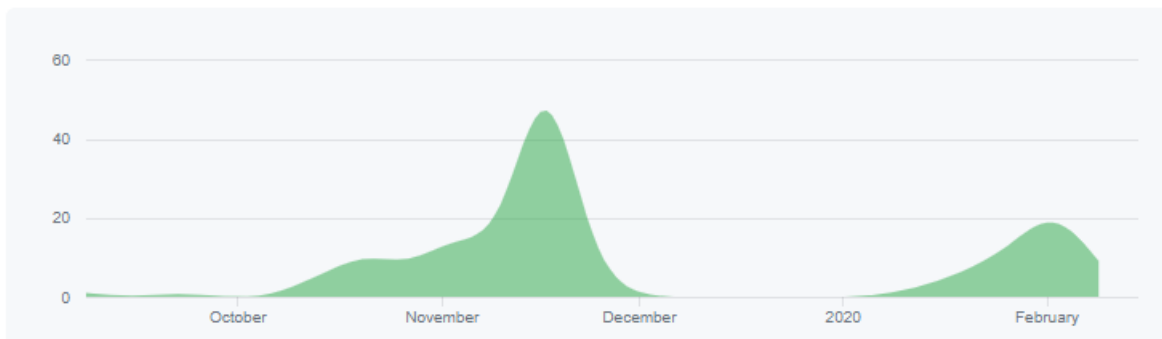


Figure 5: Contributions to the Master branch (source: <https://github.com/>)

Nevertheless, this was not always achievable especially in the beginning of the development when more crucial components of the systems were developed. In these cases, we assigned the ticket to a pair, or group of people. This approach achieved the following:

- Improved the overall code quality and fastened production [?].
- Minimised review time on the long run.
- Distributed the knowledge of large system components amongst a few people instead of one.
- Eased introducing the new team member to the project.

In addition, we also had a scrum board as an overview for the ongoing tickets. Although in the Scrum community there are ongoing discussions about the benefits of a physical scrum board over an online one

[?], given no actual workplace this was not possible to achieve. This allowed us to see which tickets needed to be reviewed and which were ready to be merged. The tickets/cards were distributed into columns, such as To do, In progress, Review in progress, Review complete and Finished. A truncated picture of this scrum board can be seen on figure Figure6.

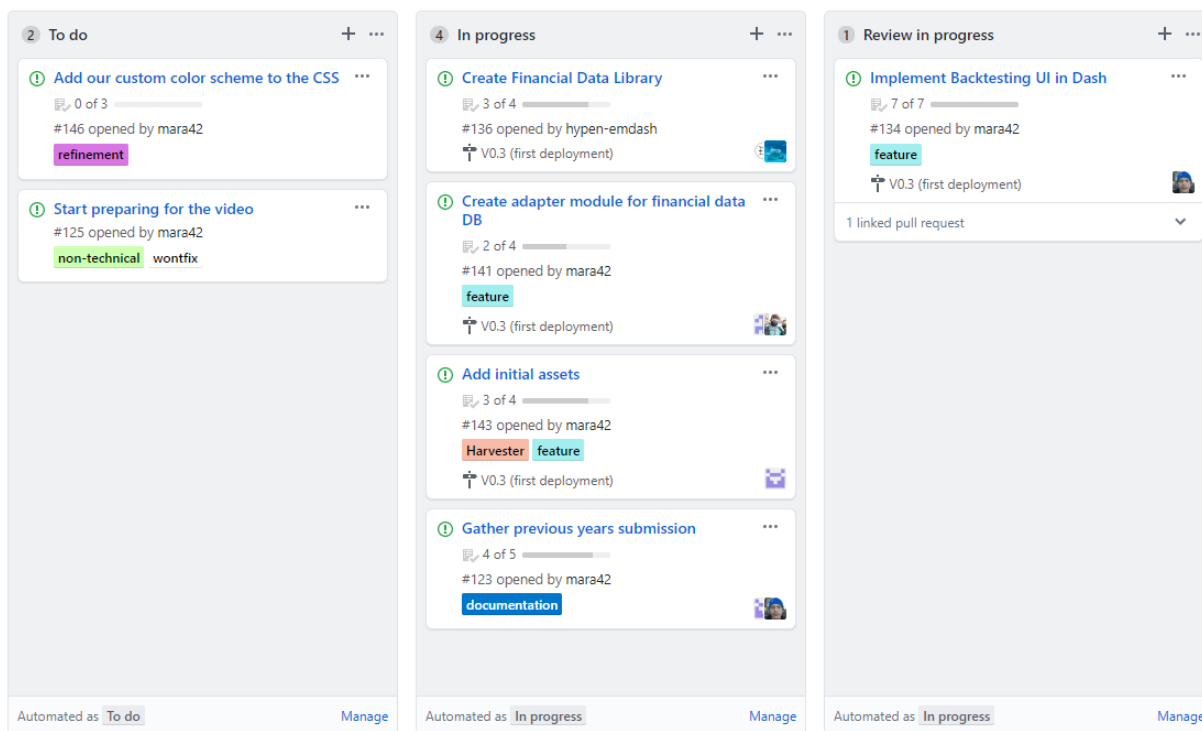


Figure 6: Scrum board - truncated

The workflow then largely depended on the task at hand and the person working on it. It was up to them if they met for pair-programming, or chose to work individually. In the beginning we all agreed on the developing environment and coding standards (for a detailed analysis see the later section 5.5). Since this report was a relevant portion of the workload, we decided to treat it as code. The report was written in L^AT_EX, first creating the overall structure of the document with a main.tex compiling the sections together. This let us to work on different sections separately just like features in our software.

Regardless of the nature of the ticket, the week, or in case of some larger tasks, two weeks, ended by the creator indicating that the changes were ready to be integrated. This was done by publishing the changes and creating a new pull request, pushing the changes onto the master branch. To ensure code quality, we set up a continuous integration (CI) environment (more on that in 5.5). After all checks have passed, the changes were successfully integrated into the code base.

5.3 Schedule

As mentioned in our Technical Report, we introduced some milestones or releases for our product based on incremental sets of features. This was done to help us stay on track with development. The roadmap based on this schedule can be seen on Figure7.

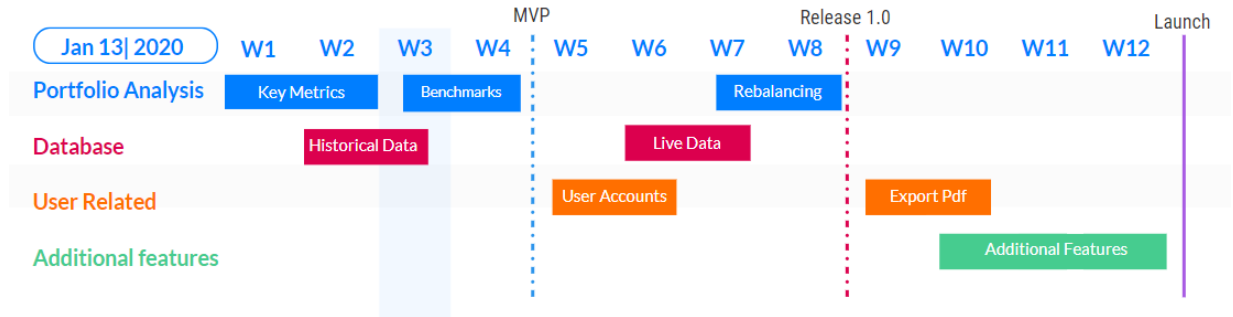


Figure 7: Technical Report: Roadmap - revisited

TODO retrospective on schedule

5.4 Key Implementation Decisions

Let us now discuss the main technologies used in our project, these can be categorised as follows:

5.4.1 Web Framework

Choosing the right web framework was more controversial than originally planned. The two major options for Python are Django and Flask. Although we decided to use Django for our MVP in the first semester, we had to spend a significant portion of the time available learning the framework, so we had to decide whether we were going to stick with it or learn Flask. In the end, we opted to go with Flask for the following reasons:

- Django has one architecture that all projects must share, and we have designed the architecture for our project ourselves. While neither architecture is wrong, the two are not compatible. Flask, on the other hand, is structure-agnostic, so we can lay out the code as we see fit.
- Flask comes with the bare minimum for web-development, which means that we don't need to manage the complexity of any feature we're not using. Django has a more complete feature-set from the beginning. This would be desirable in a large web application, but introduces significant overhead in our case, where the website has only a handful of pages.
- Django all but insists on using its ORM for all database interaction, while we plan to have a more manual approach.
- Our concerns were also confirmed by more experienced web-developers, suggesting simpler alternatives.

As stated this decision has contributed to the opportunity to recreate the basis of our application, and to apply what we have learnt from our prototype.

5.4.2 GUI

5.4.3 Business Logic

The business logic module sits at the very core of our application. We require it for producing a time series of a portfolio's performance as well as selected key risk metrics of a given asset allocation. This output is consumed by our web application and presented in plots and tables as described in 5.4.2.

Research into developing this module was initiated by listing our requirements for its desired behaviour. We identified that it should support:

- Specification of a portfolio as a set of pandas dataframes, the common data exchange format in our application
- Calculation of a portfolio's return over time

- Regular contributions to mimick saving
- Rebalancing strategies to reestablish the desired weighting of a portfolio
- Calculation of key metrics, including the Sharpe and Sortino Ratio, Max Drawdown, Best and Worst Year
- Collecting and reinvesting dividends for equities

Using these requirements, we struggled to find an open-source library that would handle these tasks for us. While backtesting libraries written in Python are available in abundance (e.g. PyAlgoTrade [?] and bt [?]), each of these was lacking in at least one critical aspect. None of them support specifying a portfolio using absolute or relative weights and instead seem to focus on backtesting trading strategies involving just a single asset while using technical indicators. Thus, we made the decision to develop our own library for handling the aforementioned tasks.

The result of this effort is "Anda" (short for analyse data). For each of its functions, Anda takes as input a Strategy object that specifies the entire list of parameters for a backtest, including contribution dates, a list of assets with associated price data, dividends for equities, etc. Calculations are performed on a per-metric basis by separating them into individual functions. This approach has allowed us to tailor the entire business logic module exactly to our needs without having to produce complicated wrapper code for existing backtesting libraries.

5.4.4 Database & Finda

5.4.5 Harvester

5.5 Integration

Writing well documented and good quality code is one thing, but making sure it all works together as a whole is a completely different story. In the first term, many hours have been wasted on trying to integrate different components of the system which did not want to fit together. Even then we had some DevOps tools in place [?], but considering that we had to produce significantly less code and more documentation last term, this was not a priority.

Starting afresh the coming term, we have decided to set up the development environment again. Upon opening the setup.py file, we see a list of required libraries, and the following two lines of code:

```
1 """A setuptools based setup module."""
2 from os import path
3
4 from setuptools import find_packages, setup
5
6 here = path.abspath(path.dirname(__file__))
7
8 install_requires = [
9     "flask",
10    "flask-login >= 0.5",
11    "flask-migrate",
12    "flask-wtf",
13    "pandas",
14    "dash",
15 ]
16
17
18 tests_require = ["pytest", "coverage"]
19
20 extras_require = {"dev": ["black", "flake8", "pre-commit"], "test": tests_require}
21
22 setup(
23     name="Thalia",
24     version="0.2.0",
25     packages=find_packages(),
26     install_requires=install_requires,
27     extras_require=extras_require,
```

Listing 1: setup.py - Development environment

One of the first decisions we had to make, is to decide on a standards coding style. This is exactly what flake8 is for, which we can see amongst the extras in the code snippet above. The original documentation of flake8 defines it as ” [...] is a command-line utility for enforcing style consistency across Python projects. By default it includes lint checks provided by the PyFlakes project, PEP-0008 inspired style checks provided by the PyCodeStyle project, and McCabe complexity checking provided by the McCabe project” [?]. However, as many other developers we also decided to redefine the maximum line-length from 79 to 88 as we found this convention a hindrance.

Another tool used for enforcing uniform style was the black auto-formatter for Python [?], which formatted the code for us upon every save if enabled, and also when committing code. This has been achieved by the use of githooks [?], which are programs that are triggered upon certain git actions. For this we needed the pre-commit package for setting up these actions and writing the configuration file. This ensured that both flake8 and black have been run before publishing changes.

5.5.1 Continuous Integration

Many studies have investigated the positive effects of developing in a continuous integration (CI) environment [?], [?]. Regardless of the exact implementation, its obvious benefit is that it provides security and uniformity for projects. We already made some steps to achieve a uniform style, but had no means to know whether the code published has been also passed its tests. Note in this section we will only discuss testing as a part of CI and not the testing strategy, for that see the section 6.

Another important part of the DevOps toolchain is the use of containers, which is what Docker helps to achieve [?]. Docker helps developers focus on writing code rather than worrying about the system the application will be running on, and also helps to reveal dependency and library issues. As Docker is open source, there are many free to use docker images available online [?]. When choosing the CI environment, Docker support was one of the main requirements.

The most promising candidate for this was CircleCI [?], which is a cloud-based system with first-class Docker support and a free trial. After connecting our GitHub repository to CircleCI, and setting up a configuration, CircleCI now does the following on every pull-request:

1. Sets up a clean container or virtual machine for the code.
2. Checks previously cached requirements.
3. Installs the requirements from requirements.txt
4. Caches the requirements for faster performance.
5. Clones the branch needed to be merged.
6. Runs flake8 one last time and saves results.
7. Runs tests and saves results.
8. Deploys the master branch to Heroku, see 5.6

The outcome of these steps is visible on CircleCI, but more importantly also on GitHub, and it refuses to merge if failing test (or no tests) have been found.

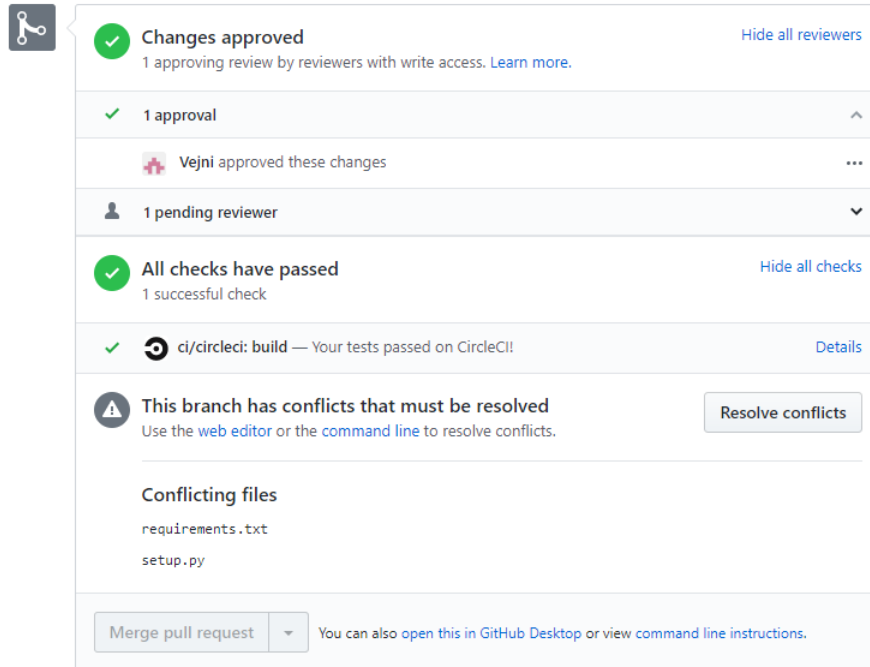


Figure 8: CircleCI on GitHub (source: <https://github.com/>)

With the help of these steps and CircleCI we managed to ensure that the code written is uniform and of good quality. It significantly reduced the time needed for integrating and code reviewing. The last step was now to deploy the system.

5.6 Hosting and Deployment

An overview over the literature covering reveals a plethora of different strategies for deploying and hosting web applications [?]. Our decision of how to choose among them involved the following considerations:

- Price - since we have severe budget constraints, we were looking for a cheap hosting solution
- CircleCI support - The target host should be supported by CircleCI natively to ease development of a continuous deployment (CD) pipeline

The upfront cost of buying physical servers ruled it out as an option for us. Thus, we turned our attention to using a solution that involved deployment to a virtual machine in the Cloud. Many providers for such a service exist, including Amazon Web Services [?] and Microsoft Azure [?]. While these would provide us with extensive control over the hosting process, their use involves a lot of complexity that seemed unnecessary for the initial rollout of a simple application such as Thalia.

Due to native CircleCI support and a free tier service, we ended up choosing Heroku [?] as our initial hosting provider. This enables us to host our application for free in the initial stages of development while providing ample opportunity for horizontal and vertical scaling later on, if it is required.

The benefits of using continuous deployment have been well established for multiple years and involve "the ability to get faster feedback, the ability to deploy more often to keep customers satisfied, and improved quality and productivity" [?]. Using Heroku in combination with CircleCI, our CD pipeline involves the following simple steps:

1. Upon commits to the master branch on GitHub, CircleCI triggers a workflow.
2. The workflow first executes the steps listed in 5.5.1 to ensure the validity of the current codebase state.

3. If this step is successful, the master branch is pushed to a remote repository recognized by Heroku via git.
4. Heroku executes the Procfile script stored in the root of our project to start the application using a gunicorn web server [?].

Our deployment process is thus fully automatized and immune to failing tests, as it will only complete successfully if the application is in a correct state.

6 Testing and Evaluation

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

7 Conclusions and further work

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

8 Appendices

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.