

|  |
| --- |
| Tero Ala-Hulkko |
| Modernization of a legacy codebase |
|  |

|  |
| --- |
| School of Technology |
| 2024 |

VAASAN AMMATTIKORKEAKOULU

Cloud-Based Engineering**TIIVISTELMÄ**

Tekijä Tero Ala-Hulkko

Opinnäytetyön nimi Modernization of a legacy codebase

Vuosi 2024

Kieli English

Sivumäärä 43 + 3 liitettä

Ohjaaja Rayko Toshev

Tämän opinnäytetyön tarkoituksena oli kehittää siirtymisstrategia vanhentuneesta koodikannasta moderniin ratkaisuun. Tutkimuksessa selvitetään syitä, miksi koodikantaa tulee uudistaa, miten se tulisi tehdä ja mitä se tulee maksamaan.

Kohdeyrityksen pääasiallinen ohjelmointikieli on Visual Basic 6 (VB6). Tutkimuksen ensimmäisessä osassa keskitytään VB6:n haasteisiin. Tutkimuksessa selvitetään, millaisia ominaisuuksia VB6:ltä puuttuu ja miten nykyaikaiset työkalut eivät enää tue sitä. Lisäksi tutkitaan riskianalyysin avulla, millaisia haasteita kohdeyritys kohtaisi, jos Microsoft lopettaisi VB6:n tuen tulevissa versioissa.

Toisessa osassa keskitytään siihen, miten markkinointi ja myynti voivat hyödyntää nykyaikaisia ratkaisuja. Kaiken kehittämisen ytimessä tulisi olla liiketoiminta tarpeet ja arvon tuottaminen. Markkina-arvon toteuttaminen on tärkeää suuria projekteja aloitettaessa. Tämä tutkielma sisältää laadullisen haastattelun kohdeyrityksen markkinointi- ja myyntihenkilöstön kanssa.

Tutkimuksen viimeisessä osassa keskitytään siihen, miten kohdeyritys voisi siirtyä nykyaikaiseen koodikantaan. Tutkimuksessa selvitetään ohjelmistoarkkitehtuuria, käytettäviä ohjelmointikieliä ja muita hyödynnettäviä teknologioita. Tutkimuksen lopputuote on kohdeyritykselle siirtymätiekartta.

Avainsanat Vanha koodikanta, Visual Basic 6, Modernisointis

VAASAN AMMATTIKORKEAKOULU

Masters in Cloud-Based Engineering

**ABSTRACT**

Author Tero Ala-Hulkko

Title Modernization of a legacy codebase

Year 2024

Language English

Pages 43 + 3 Appendices

Name of Supervisor Rayko Toshev

The purpose of this thesis was to develop a transitioning strategy from a legacy codebase to a modern solution. The study will investigate reasons why a codebase should be modernized, how it should be done, and what it will cost.

The main programming language in the target company is Visual Basic 6. The first part of this thesis focuses on the challenges VB6 has. This study will investigate what kinds of features VB6 lacking and how modern tools are no longer tailored towards it. This study will also include a risk analysis on what kind of challenges would the target company face if Microsoft dropped VB6 support from its future releases.

The second part focuses on how can marketing and sales leverage modern solutions. At the core of any development should be business need and value generation. It is important to realize market value when beginning large projects. This thesis includes a qualitative interview with the marketing and sales personnel from the target company.

The final part of this thesis focuses on how the target company could achieve transitioning to a modern codebase. This thesis will investigate software architecture, which programming languages should be used, and what other technologies and methods can be employed. The final product of this thesis is a migration roadmap for the target company.

Keywords Legacy codebase, Visual Basic 6, Modernization

**CONTENTS**

Tiivistelmä

ABSTRACT

[1 Introduction 7](#_Toc180319898)

[1.1 Background 7](#_Toc180319899)

[1.2 Research methods 8](#_Toc180319900)

[1.3 Structure of the thesis 9](#_Toc180319901)

[2 LegaCy woes 11](#_Toc180319902)

[2.1 Visual Basic 6 deprecation 11](#_Toc180319903)

[2.1.1 VB6 shortcomings 11](#_Toc180319904)

[2.1.2 32-bit environment 11](#_Toc180319905)

[2.1.3 IDE 12](#_Toc180319906)

[2.1.4 Business impact 12](#_Toc180319907)

[2.2 Forced to upgrade 13](#_Toc180319908)

[3 Upgrading is difficult 15](#_Toc180319909)

[3.1 High risk of failure 15](#_Toc180319910)

[3.2 Wrong approach 15](#_Toc180319911)

[4 modernizing the architecture 17](#_Toc180319912)

[4.1 Starting point 17](#_Toc180319913)

[4.2 Ensuring high quality codebase 18](#_Toc180319914)

[4.2.1 Continuous inspection 18](#_Toc180319915)

[4.2.2 Testing the code 19](#_Toc180319916)

[4.2.3 Single-responsibility principle and code isolation 21](#_Toc180319917)

[4.2.4 Code review 22](#_Toc180319918)

[4.3 Software architectures 22](#_Toc180319919)

[4.3.1 Monolithic architecture 23](#_Toc180319920)

[4.3.2 Separated front-end and back-end 24](#_Toc180319921)

[4.3.3 Service-oriented architecture and micro services 25](#_Toc180319922)

[4.4 Migrating the architecture 26](#_Toc180319923)

[4.4.1 Strangler fig pattern 27](#_Toc180319924)

[4.4.2 Transitional architecture 28](#_Toc180319925)

[4.4.3 Targets for displacement 28](#_Toc180319926)

[4.5 Web service architectures 29](#_Toc180319927)

[4.5.1 Representational state transfer 29](#_Toc180319928)

[4.5.2 GraphQL 30](#_Toc180319929)

[4.6 Programming language 30](#_Toc180319930)

[5 decision time 31](#_Toc180319931)

[5.1 Architecture 31](#_Toc180319932)

[5.1.1 Backend 32](#_Toc180319933)

[5.1.2 Frontend 33](#_Toc180319934)

[5.2 Language and IDE 34](#_Toc180319935)

[5.2.1 Language 34](#_Toc180319936)

[5.2.2 Project type 34](#_Toc180319937)

[5.2.3 Unit test 35](#_Toc180319938)

[5.3 Technologies 35](#_Toc180319939)

[5.3.1 API strategy 35](#_Toc180319940)

[5.3.2 Continuous inspection 36](#_Toc180319941)

[5.4 Methodologies 37](#_Toc180319942)

[6 proof of concept 38](#_Toc180319943)

[References 39](#_Toc180319944)

[Appendices 43](#_Toc180319945)

**LIST OF FIGURES AND Code Samples**

[Figure 1 Structure of the thesis 9](#_Toc180319888)

[**Figure 2** Continuous integration workflow 18](#_Toc180319889)

[**Figure 3** Unit test result window 19](#_Toc180319890)

[**Figure 4** Test-driven development cycle 20](#_Toc180319891)

[**Figure 5** Refractoring for code isolation 21](#_Toc180319892)

[**Figure 6** Monolithic architecture 23](#_Toc180319893)

[**Figure 7** Architecture with separated front-end and back-end 24](#_Toc180319894)

[**Figure 8** Service oriented architecture 25](#_Toc180319895)

[**Figure 9** Modernization over time using strangler fig pattern 27](#_Toc180319896)

[**Figure 10** 3-tier architecture 31](#_Toc180319897)

[**Code Sample 1** Simple test to verify sales basket total is added correctly 16](#_Toc167654574)

# Introduction

The term “Legacy code” is not clearly defined. Some say it is code without tests. Some extreme definitions say code becomes legacy code as soon as it is written. The definition I like the best is “Legacy code is valuable code that you’re afraid to change” (Carlo, 2024). This definition leaves room for improvement.

Legacy code comes with several hindrances that have business impact. Legacy code cause longer development times, longer QA times, and difficulties implementing continuous integration practises. The development team can simply do less in a given amount of time. This in turn can have customers feeling ignored and unhappy. (Magalhães, 2020)

This thesis explores what kinds of issues can occur when sticking with legacy codebases. The focus is on a codebase written in Visual Basic 6. VB6 is a programming language that was developed by Microsoft in 1998. This thesis will cover what kind of support is expected from Microsoft, how modern tools apply for VB6, and some modern programming concepts that VB6 fails to deliver.

The goal of this thesis is to create a plan forward for the target company’s codebase. Literature review will be undertaken to gain information about modern practices, migrating codebase to a different language, and modern software architectures.

## Background

Winpos released its first POS system in the early 1990s. In the beginning the POS machines were large and the software within was concise. Since then, Winpos has grown to be one of the leading POS system providers in the Nordics. Winpos offers solutions to a varied customer base, such as restaurants, ferries, theme parks, and several more sectors, all requiring their own set of features. One of Winpos’s strengths has been the ability to extend the software to serve new clientele.

Over the years Winpos too has needed to make technological leaps in order to provide modern solutions. Today Winpos provides highly scalable systems for large restaurant- and ferry chains, mobile applications, and cloud-based solutions. Even though core codebase has been rewritten multiple times to accommodate changing times, developing modern concepts often results in more work needed than it should.

For this thesis, an interview was conducted with the sales- and marketing teams at Winpos. The aim was to identify business needs, potential pain points, and avenues for further growth. Additionally, the technical team was consulted on several occasions. The information gathered was used to design a suggestion for a new architecture for the Winpos software.

## Research methods

The research methods used in this thesis are generally qualitative. Software architecture is very contextual, and decisions made always reflect on the environment. The project for making changes in architecture and codebase will likely take even years to complete. In an agile and iterative process, changes can be made during the project. It is best to apply the best practices and latest methods as a starting point.

This thesis will include a proof-of-concept on how to separate a section of the main program. This separated section will implement many of the methods and strategies discussed in the thesis. To guide the work, following research questions will be answered:

1. How to migrate from legacy software into modern software?
2. What can be done to ensure codebase remains maintainable?
3. What is the business impact of a legacy codebase, and conversely what kinds of benefits could a modern software architecture bring?

## Structure of the thesis

This thesis consists of eight chapters. The introduction-chapter aims to define the focus of the thesis, give an overview of the target company and its needs, and present research methods. Chapter two highlights some of the difficulties of VB6 and what kind of business impact it can entail. In the third chapter, some reasons why businesses get stuck using an outdated system are presented. The fourth chapter evaluates strengths and weaknesses of common architecture patterns, introduces ways to maintain a healthy codebase, and discusses how a software migration could be executed. In the fifth chapter, a practical way forward is established, including choosing the architecture, technologies, and methodologies. The sixth chapter is about making a proof-of-concept using the choices in the previous chapter. Chapter seven displays a road map from beginning of the software migration process to a final modernized codebase. The final chapter is reserved for reflection.

Kuva, joka sisältää kohteen teksti, kuvakaappaus, Fontti

Kuvaus luotu automaattisesti

Figure 1 Structure of the thesis

# LegaCy woes

The second chapter of this thesis digs into some clear issues and risk factors staying with Visual Basic 6, and legacy systems in general, faces. This chapter will cover both technical and financial challenges that need to be addressed.

## Visual Basic 6 deprecation

Microsoft has reduced the amount of support they are providing for VB6. In an article Microsoft published, they say they are committed to “It just works” compatibility for VB6 Windows systems up to Windows 11. The article has been revised several times to include the latest Windows version. (Microsoft, 1.4.2024)

Despite the promising name of the support scheme, the support only extends to the VB6 runtime files. Runtime files will work for a minimum of 5 years after the release of Windows 11. After 5 years, Microsoft promises 5 years of extended support. The support offered by Microsoft is limited to serious regressions and critical security issues. Development platform for Visual Basic 6 has been out of support since 2008. (Microsoft, 1.4.2024)

### VB6 shortcomings

Due to the lack of further development, several modern concepts and architectures are not supported on VB6. These shortcomings can make developing applications more difficult and costly. For example, multi-threading is not supported by VB6. Multi-threading is an important feature to fully make use of a computer’s resources, or to build listeners that launch tasks as they get requested. In software development, nothing is impossible, but poor starting point increases the costs.

### 32-bit environment

Visual Basic 6 and any application written on it, can only be run as a 32-bit application (Microsoft, 1.4.2024). 32-bit applications have several disadvantages compared to the newer 64-bit architecture. One disadvantage a developer may run into is the limited memory capacity of the 32-bit architecture. 32-bit applications can only reserve up to 4GB RAM. This can be a limiting factor and a liability if applications cannot handle the limited environment. (GeeksForGeeks, 5.2.2024)

### IDE

Several modern IDEs support direct integration with the version control system GIT. Since git was developed later than the end of support for VB6 IDE, they cannot be integrated directly. This is the case for several other modern third-party technologies as well. Several third-party extensions exist for VB6. These are often used to enable some very rudimentary features that the VB6 IDE is missing. These features include the functionality of mouse scroll wheel and the inclusion of tabs in the IDE.

### Business impact

Finding skilled professionals becomes more difficult as the popularity of a technology decreases. This can make finding candidates for hire take longer and increase candidate’s expectations of salary. Additionally, the total cost of ownership increases as legacy systems are more difficult to maintain.

Codebase with a lot of technical debt makes it more difficult to implement new features. Repairing technical issues with a codebase itself has risks involved. Common issues caused by repairing technical debt include accidentally removing features still in use and regression. Inability to quickly implement features runs the risk of being left behind competitors. (Birchal, chapter 1)

TIOBE is an organization that tracks and analyses the popularity of programming languages. TIOBE’s data is based on availability of skilled engineers, courses, and third-party vendors. According to their statistics, Classic Visual Basic has fallen from 2% popularity in 2015, to 1% popularity in 2024. They define Classic Visual Basic to include both VB6 and VBA, a programming language used withing applications such as Excel. Additionally, they make it known that due to ambiguity between all the versions of Visual Basic, there is only 50% confidence in assigning which Visual Basic should be credited. (TIOBE, 5/2024)

The number of job postings for VB6 has decreased. Meanwhile, salaries for VB6 developers have increased faster than some its peers. In the UK, the proportion of job postings in IT sector citing VB6 has decreased from roughly 1.8% to less than 0.1%. In the same period, salaries have increased from roughly 30 000£ to 60 000£. In the same period, C# developer salaries have increased from 40 000£ to 60 000£. (ITJobsWatch, 10.5.2024)

Total cost of ownership adds up as inefficiencies increase. Nearly half of UK employees say they waste more than 3 hours a day due to inefficient systems. This amounts to nearly half of those worker’s working hours. Employees also report dissatisfaction about the tools they use. Maintenance of legacy systems alone can account for 10-15% of a company's budget. These costs arise from cross-platform interfaces, ongoing management, and complex integrations among other things. (Audacia)

## Forced to upgrade

There are several reasons why companies may want to start a modernization project. It is important to identify which needs the business has and work accordingly to remedy the issue. It is important to realize that a modernization project cannot be just about choosing new technologies. Leaving everything else the same eventually leads to the same outcomes.

Developing legacy systems may be costly. Developing minor business needs may end up costing a lot of money. This can lead to situations where small changes are not worth doing on their own. Instead, small changes are lumped together with larger projects. This complicates larger projects further, increasing their risk and cost. (Cartwright, Horn, Lewis)

Mounting costs and difficulties in supporting old systems can be a reason for modernization. System built on older technology may not be problematic because it’s older, it’s a problem when it’s expensive or out of support. Conversely, adoption of newer technologies is not beneficial on its own. Offering competitive advantage compared to competitors, keeping up with competitors, lower running costs, or developmental benefits, such as speed and cost, can be worthwhile reasons for modernization. External pressure from competitors’ ability to offer modern systems can be a crisis for a stale company. Another source of external pressure can come from regulators. (Cartwright, Horn, Lewis)

Modernization in these situations is best done early before it reaches a crisis point. When a crisis point is reached, modernization efforts need to be low hanging fruits and need to cause as little disruption as possible. (Cartwright, Horn, Lewis)

# Upgrading is difficult

Previous chapters established the shortcomings of using dated technologies and how they increase costs for the organization. Even with all the good upgrading would do, risks and costs have deterred many organizations from moving forward with upgrades. This chapter will delve into reasons why companies find it difficult to upgrade architectural designs and some pitfalls to avoid when deciding to upgrade.

## High risk of failure

Complex codebases are difficult to migrate to a newer standard. Older codebases often lack documentation. Details about the system, its integrations, and interactions can only be read from the code. Manually migrating code to a newer language is prone to errors, often leads to missing features, and in general takes a significant amount of effort. This significant amount of effort could instead be going towards new marketable features. This can make the whole project seem like a high-risk low-reward situation. (Mathijs T. W.)

Automating the migration process can help reduce change of failure. Automating the migration can be done iteratively until the correct result is achieved. This can be much faster than learning each part of the code and the manually programming it again. For large projects automation can be the only feasible way to accomplish a migration project. (Mathijs T. W.)

## Wrong approach

Code migration projects can fail for many reasons. Sometimes the failure is a project that get cancelled due to problems. Even a completed migration project may not always provide the expected benefits.

It is difficult to commit to a full rewrite of the current system. During the rewrite, developing of new features may need to be halted, especially for monolithic code architectures. The time that development needs to be halted can be long for large applications. During this time critical bugs may be discovered and some features that have been sold to customers may have to be implemented. This creates the need to fix the same bugs and develop the same features on 2 systems: the production system and the new one, still under construction. Extra development and other delays can stretch the timetable to the point that development of the new system is simply stopped. At worst the company can be left in a situation where they will need to maintain two separate systems and consider building a third system to replace them both. (Carlo)

Instead of fully rewriting a system in one go, the project can sometimes be split into multiple smaller phases. Smaller rewrites can reduce the amount of risk. Each phase should provide value and if the project is discontinued, some benefits will still have been gained. One way to divide the work is to split the codebase into logical components such as billing, till, and article management. (Birchal, chapter 3)

# modernizing the architecture

This chapter introduces today’s modern concepts in programming. It takes a look at how codebase can be kept clean, compares different software architectures, and discusses how software migration can be achieved. Choosing technologies for a modern requires well-reasoned information. The aim of this chapter is to lay a foundation upon which the migration work can rely on.

Software architecture is designed with a purpose in mind, as such it is a reflection of its time. As time passes technology advances and adjustments need to be made to the codebase. Adding to an older codebase creates complexities that might not exist in a more recent codebase. Additionally, the outcomes may not be ideal for modern times. It can be justified to change the underlying architecture during a rewriting process.

## Starting point

Any modernization effort must begin by getting to know the system that is getting modernised. It is important to gather information about the current state of the system, where to focus efforts first, and set up monitoring to verify progress.

Automated tools can provide information about various aspects of the codebase. Automated tools are particularly useful in detecting code that violates styling rules and certain types of bugs in the code. If tools flag a certain area of the code, it can be a good indication to focus efforts there. Tools should not be fully trusted to provide all the information needed and can flag code that is working well as a bug. (Birchal, chapter 2)

Performance of code should be monitored and measured. Performance tests should exist in both production systems and in systems before and after refactoring. Performance tests should be written into code and batches of data can be processed to measure the time it takes for the program to complete. Results of the experiment can be written into a log file for human verification. (Birchal, chapter 2)

## Ensuring high quality codebase

Managing large codebases can be difficult. Technical debt is a measurement of how much bug fixing and refactoring a codebase needs to be considered optimal. Technical debt naturally increases over time as new features are implemented. This section of the thesis explores some of the ways to reduce the amount of technical debt is added over time.

### Continuous inspection

Tools can be integrated into the IDE that flag code that does not abide by rules set by the organisation or code that may introduce bugs and vulnerabilities. By themselves they rely on programmers frequently checking the output of the tools. Continuous inspection helps ensure that no problem flagged by the tools goes unnoticed. Continuous inspection can be a part of continuous integration. Continuous inspection makes use of an automated workflow that gives feedback to the developer upon completion. (Birchal, chapter 2)

The automated workflow (Figure 1) for continuous integration makes use of several different components. When the developer is ready with a feature, they can commit the code to a version control system. Committing code to version control system triggers an automatic build in the build server. Any problems with the build get noted and feedback is sent to the developer. The committed code is optionally rejected, and developer can fix any issues and recommit the code. (Birchal, chapter 2)

Kuva, joka sisältää kohteen teksti, Tarralappu, Suorakaide, kuvakaappaus

Kuvaus luotu automaattisesti

**Figure 2** Continuous integration workflow

### Testing the code

The most important purpose of testing is to find bugs before they make it into a production environment. In the context of this thesis, tests are snippets of code that test the outcome of a function in code. Several different types of tests are available.

Unit testing is about testing the inputs of a function and verifying their outputs. Unit tests can be set up in a way that tests both regular cases and edge cases. Edge cases are typically more laborious to test in other types of testing. To make effective tests, each part of the code should only address one concern. Unit test’s data can be mock data. Mock data is intended to simulate a real database connection. Mock data is data that uses the classes and functions of a real-life scenario, but the data is engineered for testing purposes. (Schneider) In the case of Winpos, instead of testing whether adding payment and ending the receipt, one should test all individual parts of the process, such as adding payment to the basket, sending a receipt to a receipt control unit, and creating a creating a new task in the kitchen display unit.

Code sample 1 and Figure 2 showcase unit testing in a simple way. The code tests another piece of code that adds items to sales basket. Figure 2 shows if the tests were successfully completed. If a programmer makes changes to the sales basket logic and gets the test passed, they should be confident that no logic was broken.

Kuva, joka sisältää kohteen teksti, kuvakaappaus, Fontti

Kuvaus luotu automaattisesti

**Code Sample 1** Simple test to verify sales basket total is added correctly

Kuva, joka sisältää kohteen teksti, kuvakaappaus, ohjelmisto, Tietokonekuvake

Kuvaus luotu automaattisesti

**Figure 3** Unit test result window

One way to ensure tests are implemented systematically, and are effective, is the method test-driven development. In TDD a list of tests is written before the actual function is written. A test is then chosen, and the function is implemented until the test is passed successfully. Both the code written, and the existing code is then refactored to form a cohesive unit. (Fowler)

Kuva, joka sisältää kohteen teksti, Suorakaide, Tarralappu, diagrammi

Kuvaus luotu automaattisesti

**Figure 4** Test-driven development cycle

### Single-responsibility principle and code isolation

Single-responsibility principle is about reducing the effect changes into software have on the codebase. SRP advocates that a class should only have one responsibility, a function, and only one reason for changing it. The benefits of SRP are ease of implementation and help with preventing unexpected results caused by changes. Single-responsibility principle can make the codebase easier to understand. Easily understandable code can help reduce bugs, makes development faster, and improves the developer experience. (Thorben)

Single-responsibility principle can be used as a guide rather than a strict rule. Too many classes can create a feeling where reading the code becomes more difficult as the developer can only see a small part of the code at once. Defining what a responsibility is difficult and subjective.

Breaking code into smaller pieces improves the testability of software. Some types of functions are easier to test using unit testing. By isolating unit testable code from a larger function, unit tests can be made more easily. This also creates a clear boundary between unite testable code and code that needs to be integration tested. (Wagner)

Kuva, joka sisältää kohteen teksti, kuvakaappaus, diagrammi, Fontti

Kuvaus luotu automaattisesti

**Figure 5** Refractoring for code isolation

### Code review

Code review is the practice of reviewing another programmer’s code before it can be merged into the master branch of version control system. Code review is useful for increasing cohesion in the codebase. If a piece of that is being reviewed does not follow the patterns agreed by the team, the reviewer can reject it. Suggestions can also be given if the reviewed code has mistakes in it, this can be particularly useful when onboarding junior programmers. When putting code up for review, the programmer is also sharing information about changes they are making to the codebase. (Birchal, chapter 3)

## Software architectures

Making changes to architecture is a large project. It is paramount to establish clear goals before beginning the work. The goals should be tangible and actionable properties such as splitting the codebase into modules or improvements to the build process. To have a cohesive end product, it is important to agree within the team how the codebase needs to look like in the end. (Birchal, chapter 5)

The target company of this thesis is more or less built as a monolithic application. Architectures most worth considering are keeping the current architecture, a monolithic architecture. Second option is to change the architecture to an architecture with front-end and back-end separated into their own projects. Third option would be to opt for a service-oriented architecture.

Different software architectures bring with different benefits and challenges. Any benefit obtained by changing the architecture needs to outweigh the workload changing the architecture brings. (Birchal, chapter 5)

### Monolithic architecture

Monolithic architecture (Figure 2) is the simplest architecture to develop. This is due to functions being readily available in the same codebase. Monolithic architecture should result in fast applications as there is reduced need to talk to other services or applications. (Birchal, chapter 5)

Downsides of monolithic architecture include unexpected interactions in the codebase. Because codebase is a single entity, making changes to code has more opportunities to unintentionally affect other code. Increasing isolation in the code is one reason to avoid monolithic architecture. Another byproduct of low isolation is catastrophic failures when one part of the code is failing. If the application is run as a server, monolithic architecture is the least efficient to scale up. (Birchal, chapter 5)

Kuva, joka sisältää kohteen teksti, kuvakaappaus, diagrammi, Fontti

Kuvaus luotu automaattisesti

**Figure 6** Monolithic architecture

### Separated front-end and back-end

Presenting user interface and business logic can be separated into their own application tiers (Figure 3). This type of architecture is typically used in web applications. Communication between tiers happens through APIs. Separating the two tiers makes the code easier to understand. User interface and business logic existing in separate applications promotes separation of concerns. Often there are two developing teams focused on user interface and business logic. The technology stack can vary significantly between the two tiers. (Birchal, chapter 5)

Compared to a monolithic architecture, communication though APIs is more complicated. Communication must be carefully engineered for it to work as expected. Additionally, error handling must be much more thorough as communicating over network APIs can cause issues that are out of hands of the programmer. (Birchal, chapter 5)

This approach could be the best middle ground solution for the target company. POS systems require several different user interfaces for mobile, browser, and local installations. One of the recent challenges the target company has had is that the customers expect to be able to access their POS system anywhere. This style of architecture supports running the back end as a service and accessing it through a browser in a customer’s home, for example.

Back-end running as a service with APIs used for communication supports all types of UIs without the need for modification. Local installations would suffer very little as running the front end as well as the back end on the same computer offline is still an option.

Kuva, joka sisältää kohteen teksti, diagrammi, kuvakaappaus, viiva

Kuvaus luotu automaattisesti

**Figure 7** Architecture with separated front-end and back-end

### Service-oriented architecture and micro services

Service-oriented architecture (Figure 4) is composed of several services that are isolated from the rest of the architecture. Developing services in this way makes it harder to accidentally introduce regression in the rest of the ecosystem. Scaling the application is also much improved as you only need to scale the services that need more resources. (Birchal, chapter 5)

Service-oriented architecture adds operational overhead. Developing and testing all the different services becomes more complicated. Additionally, tracking metrics and bugs is more difficult. Because of the isolation in SOA, code is often duplicated in the different services. This is due to multiple teams working in different services have limited knowledge of each other’s work. (Birchal, chapter 5)

Service-oriented architecture is the best choice for many applications deployed in a cloud platform. Software split into multiple services can be cost optimized better than a back end that is built out of one part. Offering software as a service in a cloud platform, however, is about much more than mere software. SaaS model is a decision that affects the way a business operates, and the revenue streams would be significantly different. Multiple services would also want to make use of multiple databases. Other options listed before make use of a single database. Generally, developers should avoid using the microservice pattern, unless they have a really good reason to. Good reasons include things like having more than a million users, or the system is too large and complex to be a monolith, both in a cloud environment. Neither case applies so the use of microservices is not advised.

Kuva, joka sisältää kohteen teksti, diagrammi, viiva, kuvakaappaus

Kuvaus luotu automaattisesti

**Figure 8** Service oriented architecture

## Migrating the architecture

Large-scale modernization projects are complex and in many cases should not be attempted in a single project. Instead, modernization effort should be considered a process that takes place over a long period of time.

### Strangler fig pattern

Strangler fig pattern is an idea in which a portion of a software is “strangled” and migrated to a new architecture. This act reduces the size of the legacy codebase, while the modernized codebase grows. The pattern is repeated until the entire software is migrated. Both new and legacy software will be in use until the migration is fully completed. This will require a mechanism to allow communication to be directed to where it is needed. (Microsoft)

Kuva, joka sisältää kohteen teksti, Suorakaide, kuvakaappaus, diagrammi

Kuvaus luotu automaattisesti

**Figure 9** Communication between legacy and modernized during transition

Strangler fig pattern is a pattern that helps minimize the risks involved in legacy modernization projects. Timeline of the modernization process can be extended without negative business impact. The legacy software and modernized parts can exist side-by-side, routing more and more traffic to the modern platform as it builds. One goal of the strangler fig pattern is that the migration process is not visible to the end user. (Microsoft)

Kuva, joka sisältää kohteen teksti, kuvakaappaus, Suorakaide, diagrammi

Kuvaus luotu automaattisesti

**Figure 10** Modernization over time using strangler fig pattern

### Transitional architecture

During the migration process, legacy applications will need to work together with the new applications. For this to be possible, transitional architecture will need to be in place to enable the communication between the applications. The key feature of transitional architecture is that it is temporary, and only serves to enable the migration process.

What transitional architecture offers, is speed of delivery. New features can be developed for the new platform, while it is still in an unfinished state. New features can include an interpreter for the legacy system that is designed for removal once it’s no longer needed. Another benefit of the transitional architecture is further derisking of the migration.

### Targets for displacement

Identifying which code can be migrated is often called finding seams or introducing seams into the system. These seams can be used to direct the flow to the new system without altering the old system. Seams were originally intended to make the legacy system more testable, but their most valuable use was found in legacy migration. Seams are a great way to gradually migrate behaviour away from the legacy system, by decoupling functionality and migrating it partially to a new system. Even without modernization of the codebase, introducing seams make the old system more testable. (Fowler)

Function calls are common locations of seams. A function can often be migrated to a new system without modifying the original code that calls for it. Issues arise when seams are not present, and a code block’s logic is too large to be migrated. While migrating the code as is, is possible, benefits of migration will suffer when improvements are not made.

## Web service architectures

Modern applications often rely on communication over the internet. The method in which communication is established defines the structure, and protocols of the application. Several approaches have been developed that each come with their own strengths.

### Representational state transfer

Representational state transfer, also known as REST, is a stateless way of achieving server-client communication. Like many strategies, REST allows client-server communication where the two systems don’t need to know each other’s state. Clients call endpoints of individual resources it needs, and the server will fulfil the client's request. (Codeacademy)

In REST, web requests are sent to a specific URL. The URL includes where the server lives and any information it needs to handle the request. Requests are made using HTTP verbs; special keywords that describe the intention of the call, such as GET, and POST. Additionally, headers are used to make the request more accurate. Headers may include pointers like content type that the client accepts as a response. When the request is properly formed and handled, the server will response with a code indicating whether the request was completed successfully or not, and the contents of the response. (Codeacademy)

### GraphQL

GraphQL was initially develop to remedy some of the shortcomings of REST. It has been gaining popularity since its inception in 2012. Its benefits are best realized as a part of a complex system where data is dispersed. Additionally, its benefits become greater as traffic increases. Overall, the implementation is more complex than the one of REST. (Kong)

The main idea behind GraphQL is that it retrieves the exact information a client requests. In REST there is no way for a client to limit what data an API returns, the client may have to make multiple requests for data to get the dataset it needs, in GraphQL one request is sufficient. This results an efficient system as traffic over the internet is reduced, amount of API calls is reduced, and the server workload is reduced. (Kong)

## Programming language

When choosing a programming language, several technical and non-technical factors need to be considered. Non-technical factors are all about the team working on a project. Team’s existing knowledge plays a big role in which language is the best choice. Every new technology has a learning curve, pre-existing expertise reduces friction at the start of the project. Additionally, not all programming languages are equally difficult to get started with. When considering new hires, lower barrier of entry increases the number of candidates that can be considered. (Fauerbach)

From a technical standpoint, language selection depends on the solution that is being built. Different languages are specialized in different types of applications, such as desktop or web applications. (Fauerbach) Another factor that should be considered is how well the language integrates with existing systems.

# decision time

This chapter introduces all the technologies that will be used to make a proof-of-concept about the modernized system. The technologies are intended to reflect upon what a true production system would look like. The decisions were made based on three main points; theory laid out in the thesis, integrability at Winpos, and feedback received from the business and technical departments at Winpos.

## Architecture

When designing software architecture, business need defines the product. At Winpos the system has moulded to its current form as a result of business needs and opportunities. Still the current state makes implementing some modern concepts more difficult than they should be.

Winpos POS system is a large software product with numerous components that need to be supported. The most basic use case for customers is locally making sales using a POS pc, and being able to manage products, run reports, and perform other general maintenance. Additionally, customers would like to be able to do maintenance from home, make sales using mobile devices, and general functionality available in an offline environment. While all the aforementioned is possible it is not without significant development overhead.

What this thesis proposes is conversion of the existing structure to a 3-tier architecture. 3-tier architecture consists of web-, business logic-, and data-tiers. Each tier is physically separated, albeit there is nothing stopping one from running them on a single machine. 3-tier architecture benefits over a lesser tier architecture include faster development, improved reliability, scalability, and security. (IBM)

In figure 8, parts of the POS system have been divided into tiers and layers. Layers represent a logical component of the application. In the presentation layer, a shop manager can, for example, connect through a web service to the Backoffice application to edit product information. Firewalls are present to deny access from unauthorized traffic to the internal network. Changes made to products are forwarded to the Business logic layer. Business logic layer contains the back-end software in one or more individual applications. Data is formatted and saved in the data layer, which houses databases.

Kuva, joka sisältää kohteen teksti, Tarralappu, diagrammi, Suunnitelma

Kuvaus luotu automaattisesti

**Figure 11** 3-tier architecture

### Backend

Winpos POS system today is primarily written using two languages. The core components are written in Visual Basic 6. Significant portion of integrations and other components are written in C#. These lay a foundation where either language can be used. Two options rise above the rest for modernizing the Winpos POS system: adopting Twinbasic or migrating fully to C#.

Twinbasic is a new basic language that aims to be fully backwards compatible with VB6 projects. Twinbasic aims to implement several features that VB6 is missing, such as a modern IDE, 64-bit support, and multi-threading. Twinbasic is still in development, with version 1 anticipated in 2024. (Twinbasic)

Twinbasic is an option because of the backwards compatibility with VB6. In a perfect scenario, upgrading to a modern language and IDE would be a simple conversion of the existing codebase. Twinbasics small development team and its early release status brings some risks as the main development language. Some features might not be fully functional, and it might lack wide third-party support tools and libraries.

Another option would be to use Microsoft’s Visual Studio IDE and C#. Visual Studio is currently in use at Winpos, as several applications have been written in C#. There is no reliable way to translate VB6 code C# code. To maintain full functionality code must be rewritten to the new language. Any rewrite-project is bound to be resource intensive and time consuming.

Effort that is required to fully migrate to C# will likely be greater. Despite this, rewriting to C# would likely be a better choice due to its potential to yield superior results. Visual Studio has garnered a well-established, robust ecosystem. Most modern tools and libraries are readily available to be integrated into Visual Studio. For Winpos specifically C# is the superior choice as there is little learning curve for new tools and environments. Additionally, for a developer, a backend written in one language makes it easier to switch between different projects.

### Frontend

Backend built using restful APIs supports any number of frontends. The focus of this thesis is on the core shop interface. In a shop environment, a browser-based user interface is suitable, as the same interface can be used remotely over the internet. Crucially, in the interim period, the existing user interface needs to be connected to the new backend. In the proof-of-concept section of this thesis, connectivity of a browser-based GUI and the connectivity of existing GUI is presented.

Web-based interfaces at Winpos today are written in Angular. Angular is a modern and widely adopted framework of TypeScript. It is designed to make building user-interfaces both over the web and natively. Due to strong existing knowledge, angular is a good choice to move forward with. Another option could be to use Blazor. Blazor is developed by Microsoft and utilizes the .NET Core framework. A shared codebase with the back-end application could be seen as a benefit.

## Language and IDE

Changing the architecture to a 3-tier system requires that the existing software needs to be split into front- and back-end applications. Each side needs their own developer tools. Learning curve for developers was one of the most important factors when choosing languages, as long as the language was able to provide the necessary features.

### Language

For proof-of-concept, C# will be used to create the backend. C# was chosen due to its history withing Winpos, long-term support, and proven track record. Another factor why C# was selected is its Visual Studio integrated development environment, or IDE. Visual Studio offers easy access to many of the topics discussed in this thesis, such as support for unit testing, and integrated continuous inspection tools.

For proof-of-concept, Angular will be used due to its suitability for the job and history within Winpos. Angular is more flexible as to which tools a developer would like to use. One good option is to use Visual Studio Code, or a text editor such as VIM or NotePad++.

### Project type

ASP.NET Core is an extension to the .NET Core platform. It is specifically intended to be used for building web applications.

### Unit test

Choosing the right unit testing framework is a decision that will have an impact on the software for years. Lots of code will need to be written into the unit test project so choosing the correct project type is impactful. Today there are three popular unit testing frameworks for C# and .NET; MSTest, xUnit, and nUnit. Although they are all unit testing frameworks, they all have their strengths over one another.

The main benefits of MSTest are its easy integration to Visual Studio environment, support from Microsoft by regularly offering updates, and support for data-driven tests. Main drawbacks are its lack of extensibility options, and slowness in large projects. NUnit is highly customizable for an organization’s specific needs. Test execution can be customized, and custom extensions can be self-written. XUnit is the most performant option and has an active open-source community for updates. XUnit is also the most extensible option using plugins and other extensions. (Ghinaiya)

For proof-of-concept, xUnit will be used to create tests. Winpos POS system is a large project with many functions that need to be tested. Fast tests, combined with ready extensions for many needs, and are compelling reasons to choose xUnit over the other options. It also synergises well with integration testing applications.

## Technologies

When choosing technologies, longevity, easy integrability to existing systems, and external systems, and time to market were of the highest importance.

### API strategy

The back end will be written as REST APIs. Compared to alternatives, REST APIs offer an easier learning curve and a proven track record. REST APIs are endpoints and resources that can be called from over the internet. These endpoints and resources are intended to allow easy information swapping between different systems.

Today, a lot of developer’s time is spent making integrations to other systems. Winpos has integrations to many services in different financial sector operators. Integrating other systems to Winpos requires a significant effort. Easy integrability may bring new business opportunities. This is where REST APIs can shine. REST APIs allow easy integrability for both in-house front ends such as, browser and mobile, as well as external integrations.

### Continuous inspection

Code quality analysis is best achieved by using an integrated extension in an IDE, and automated inspection server in conjunction with other CI/DC tools. A popular continuous inspection tool is SonarQube. SonarQube can be run, as an automated inspection tool, on its own server that analyses code as it is committed to version control. Analytics can also be directly integrated to an IDE so the developer can self-monitor their work.

One of the key principles using SonarQube is the principle of clean as you code. Clean as you go refers to writing new code, and altering existing related code, that is verified to be up to standards. Automated tooling uses quality gates that ensure that code that does not pass certain quality criteria cannot be merged to the larger codebase. Quality gates can be configured by the organization with minimum passing quality.

For proof-of-concept, the integrated components of SonarQube will be used. The generic settings are a good starting point as it will report about poor code, unsafe methods and keep track of consistent naming of variables.

## Methodologies

For the longevity of any codebase, it is vital to adopt practises that promote well written code. Starting a new project is an opportune time to apply methodologies and tools that help ensure the code is written well and stays high quality throughout the development process.

In a rewriting project, the end result of the new code is particularly well known. This makes it a good opportunity to practise test-driven-development.

# proof-of-concept

This chapter presents a proof-of-concept way of splicing a piece of code from the legacy monolith and migrating it to an API-based backend. Connectivity to the backend is demonstrated from both a new Angular.js-based browser frontend and how to use the new API from the existing solution during the transitioning period. Many of the topics discussed in previous chapters are also presented in practise, such as test-driven-development, creating the project files and must have infrastructure, as well as IDE-based code quality validation.

## Founding the project

A solution needs to be created that will house two projects. One project contains the new backend code, and the other contains unit tests for the code.

The backend project type is ASP.NET Core Web API. When creating such a project, several options are given. In this PoC, the latest framework, .NET 8.0, is selected. OpenAPI support is enabled to leverage Swagger in development. Swagger is designed to simplify API development by allowing easy testing of APIs without a ready UI. Another option that is enabled is the support for containers. Containers are virtual operating systems that run the application being developed. The main benefits of using containers are that the program will run on the same operating system regardless of where it is installed, resource efficiency when deployed in a cloud environment, and security by isolation.

Kuva, joka sisältää kohteen teksti, kuvakaappaus, numero, Fontti

Kuvaus luotu automaattisesti

**Figure 12** Backend initial setup

Secondly project with type xUnit Test Project is added to the same solution. This project will contain all tests needed for the backend application. The backend project is added as a dependency to the unit test project, so its code is visible to the unit test project.

Kuva, joka sisältää kohteen teksti, kuvakaappaus, Fontti, numero

Kuvaus luotu automaattisesti

**Figure 13** Initial project structure including boiler plate classes

## Splicing a feature

## Web Frontend

## Integrating to legacy

# References

Carlo, N. What is Legacy Code? Is it code without tests?. Understand Legacy Code. Retrieved 2024-05-05. https://understandlegacycode.com/blog/what-is-legacy-code-is-it-code-without-tests/

Magalhães, G. Legacy Code: a dead end for your project?. Medium. Retrieved 2024-05-05. https://medium.com/what-really-matters/legacy-a-dead-end-for-your-project-dc466d24a3a5

Support statement for Visual Basic 6.0 on Windows. Microsoft. Retrieved 2024-05-09. https://learn.microsoft.com/en-us/previous-versions/visualstudio/visual-basic-6/visual-basic-6-support-policy

Difference Between 32-bit and 64-bit Operating Systems. GeeksForGeeks. Retrieved 2024-05-10. https://www.geeksforgeeks.org/difference-32-bit-64-bit-operating-systems/

TIOBE Index for May 2024. TIOBE. Retrieved 2024-05-10. https://www.tiobe.com/tiobe-index/

The Classic Visual Basic Programming Language. TIOBE. Retrieved 2024-05-10. https://www.tiobe.com/tiobe-index/classic-visual-basic/

Programming Language Statistics. ITJobsWatch. Retrieved 2024-05-10. https://www.itjobswatch.co.uk/IT-Job-Market/UK/Programming-Languages

VB6 IDE AddIns - What's essential?. VBForums. Retrieved 2024-05-12. <https://www.vbforums.com/showthread.php?804911-VB6-IDE-AddIns-What-s-essential>

Audacia. 2019. The cost of legacy IT. Medium. Retrieved 2024-05-12. https://medium.com/@audaciatech/the-cost-of-legacy-it-1c1fa6217e59

Schuts MTW, Aarssen RTA, Tielemans PM, Vinju JJ. Large-scale semi-automatedmigration of legacy C/C++test code.Softw Pract Exper. Retrieved: 2024-05-12. <https://onlinelibrary.wiley.com/doi/epdf/10.1002/spe.3082>

Carlo, N. The Ship of Theseus to NOT rewrite a legacy system from scratch. Understand Legacy Code. Retrieved 2024-05-13. https://understandlegacycode.com/blog/ship-of-theseus-avoid-rewrite-legacy-system/

Birchal, C. 2016. Re-Engineering Legacy Software. Manning Publications.

Fauerbach, C. 2020. How Do I Choose the Right Programming Language for My Team?. Retrieved 2024-05-13. <https://medium.com/capital-one-tech/lessons-from-the-leadership-trenches-how-do-i-choose-the-right-programming-language-for-my-team-36222aa78b38>

Fowler, M. 11.12.2023. Test Driven Development. Retrieved 25.5.2024. <https://martinfowler.com/bliki/TestDrivenDevelopment.html>

Schneider, J. 24.5.2022. Writing usable code. Retrieved 25.5.2024. https://www.algolia.com/blog/engineering/writing-usable-code/

Wagner, L. 13.7.2023. Mock Your Mocks. Retrieved 9.5.2024. https://blog.boot.dev/clean-code/writing-good-unit-tests-dont-mock-database-connections/

Thorben. 28.2.2024. SOLID Design Principles Explained: The Single Responsibility Principle. Retrieved 9.5.2024. https://stackify.com/solid-design-principles/

What is three-tier architecture?. IBM. Retrieved 22.8.2024. https://www.ibm.com/topics/three-tier-architecture

Twinbasic. Retrieved 22.8.2024. https://twinbasic.com/

Cartwright I, Horn R, Lewis J. Patterns of Legacy Displacement. 5.3.2024. Retrieved 30.9.2024. https://martinfowler.com/articles/patterns-legacy-displacement/

Microsoft. Strangler Fig pattern. Retrieved 30.9.2024. <https://learn.microsoft.com/en-us/azure/architecture/patterns/strangler-fig>

Cartwright I, Horn R, Lewis J. Transitional architecture. 28.3.2022. Retrieved 30.9.2024. <https://martinfowler.com/articles/patterns-legacy-displacement/transitional-architecture.html>

Fowler, M. Legacy seam. 4.1.2024. Retrieved 6.10.2024. https://martinfowler.com/bliki/LegacySeam.html

Codeacademy. What is REST?. Retrieved 12.10.2024. https://www.codecademy.com/article/what-is-rest

Kong. What is GraphQL? 1.3.2023. Retrieved 12.10.2024. https://konghq.com/blog/learning-center/graphql

Ghinaiya, S. 19.6.2024. Testgrid. Retrieved 20.10.2024. https://testgrid.io/blog/nunit-vs-xunit-vs-mstest/

# Appendices

APPENDIX 1

**WHAT TO INCLUDE TO APPENDICES**

You can enclose as appendices for example a questionnaire used in the study or other material that is related to the study.

Material that the client wished to classify can be enclosed as an appendix; in this case, the appendix is not included in the published version submitted to Theseus or in the hardbound version.

**THE TITLES AND NUMBERING OF APPENDICES**

The appendices are numbered and given a title. When you refer to an appendix in the text, use the correct referencing practice. Remember to mention the number of the appendix.

APPENDIX 2