Warehouse Optimization Model - Digital Twin Simulation

António José da Cruz de Miranda

A dissertation submitted to Faculdade de Engenharia da Universidade do Porto for the Master's degree in Industrial Engineering and Management

Supervisor: Joana Maria Ferreira Alvura da Hora

July 2023

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Abstract

Warehouses are key components of logistics systems and represent a significant part of logistics overall costs. Thus, it is no wonder that organizations heavily relying on warehouses to carry out their operations seek to improve and optimize all aspects of these facilities continuously. Besides continuous improvement, logistics operations may need to swiftly adapt to various disruptions, such as technological breakthroughs, shifts in consumer preferences, and significant events like Brexit.

The present dissertation, developed within a global luxury e-commerce actor, in a broad sense, aims to make an impact on the organization's logistics design capabilities for both continuous improvement and specific initiatives by providing simulation tools and methodologies that enable scenario analysis and solutions development and testing.

By exploring the concept of a Digital Twin (DT) as a virtual replica of a physical warehouse, mirroring its key features and processes in a granular way, this dissertation is dedicated to build a 3D scenario and modeling part of the operations for the biggest warehouse of the company. The simulation model, intended to be reused, improved, and extended, provides a safe environment for testing changes not only in operations but also on the warehouse resources and physical space. The model, developed mainly with the software Anylogic, also relies on a Python environment as an interface for input and output data and also to run Python scripts where specific computations are performed during run time. The integration with Python demonstrates the viability of connecting Anylogic models with other tools thus allowing the integration of other technologies such as Artificial Intelligence (AI), Machine Learning (ML) or data visualization tools.

As an experimental project, the first step was to build a prototype where the necessary features were developed and tested. In a second stage, the definitive model was built and most of the expected features were implemented, in some cases with further refinements.

This project has demonstrated that a comprehensive warehouse simulation model can be developed relying mostly on easy-to-use software such as Anylogic thus justifying the setup of appropriate computational resources for full implementation and further developments. Regarding the business case considered, the preliminary results are promising, justifying thus the full exploration of the underlying concept. It was also evidenced how simulation is the appropriate approach to comprehensively analyze this kind of business case, where proposed operational changes have complex implications that cannot be appropriately considered by relying on analytical methods.

Keywords: Digital Twin, Simulation, Warehouse Optimization, Put-Away, Agile Methodology

Resumo

Entrepostos são uma peça chave dos sistemas logísticos e representam uma porção significativa dos custos logísticos totais. Assim, não surpreende que as organizações muito dependentes de entrepostos para levar a cabo as suas operações procurem otimizar todas as vertentes deste tipo de infraestruturas de forma contínua. Além do melhoramento contínuo, as operações de logística podem necessitar de se adaptar rapidamente a disrupções várias, tal como avanços tecnológicos, alteração das preferências dos consumidores, e eventos extraordinários como o Brexit.

A presente dissertação, desenvolvida no seio de um operador global e-commerce de luxo, pretende gerar impacto nas capacidades de planeamento logistico da organização não só no melhoramento contínuo mas também nas iniciativas pontuais, fornecendo ferramentas e metodologias de simulação que permitam a análise de cenários e o desenvolvimento e teste de soluções.

Explorando o conceito de Digital Twin (DT) aplicado ao desenvolvimento de um modelo do entreposto físico, replicando as suas características e processos chave de modo granular, esta dissertação dedica-se ao desenvolvimento do cenário 3D e da modelação de parte das operações para o maior entreposto da empresa. O modelo de simulação, pensado para ser reutilizável, melhorado, e ampliado, garante um ambiente seguro para testar alterações, não só nas operações, mas também nos recursos e no espaço físico. O modelo, desenvolvido sobretudo por meio do software Anylogic, é suportado por módulos Python que funcionam como interface para input e output de dados e corre código Python para cálculos específicos durante a execução do modelo. A integração com Python demonstra a viabilidade de conectar modelos Anylogic com outra ferramentas permitindo assim a integração de tecnologias adicionais como Artificial Intelligence (AI), Machine Learning (ML) ou ferramentas de visualização de dados.

Tratando-se de um projeto experimental, o primeiro passo foi desenvolver um protótipo no qual funcionalidades foram desenvolvidas e testadas. Numa segunda etapa, o modelo definitivo foi construído e a maior parte das funcionalidades previstas foram implementadas, em alguns casos com melhoramentos adicionais.

Com este projeto demonstrou-se que é possível desenvolver um modelo de simulação abrangente praticamente apenas com recurso a software de simulação fácil de utilizar tal como é o caso do Anylogic, permitindo também justificar a afetação de recursos informáticos adequados a uma implementação completa e a desenvolvimentos adicionais. Quanto ao Caso de Estudo considerado, os resultados preliminares são promissores e justifica-se a continuidade desse estudo. Ficou também evidente que a simulação é a abordagem adequada para analyzar de forma completa, casos de estudo como este em que as alterações operacionais propostas tem implicações complexas que não podem ser adequadamente tidas em conta com base em métodos analíticos.

Acknowledgements

I would like to express my deep gratitude to Filipe Faria, my supervisor in Farfetch, for having given me the opportunity to join his team and embracing a very interesting challenge. I'm also grateful for the knowledge transmitted and guidance provided, for the trust and support and for the freedom regarding my work.

I am sincerely grateful to Farfetch for providing me with this exciting challenge and all the resources to embrace it. And to my fellow Farfetchers, for the cool and enthusiastic environment, for cheering for my success, and also for providing me with support and knowledge every single time I needed, always while having fun!

I am honored to express my gratitude to all professors, whom I have had the privilege to learn from over the past few years, for having prepared me for this challenge and for making this master program truly excellent!

I also wish to convey my profound appreciation to my mentor Joana Hora for the constant availability, approachability, and willingness to provide support and guidance whenever needed. For the knowledge transmitted and interesting discussions. For the belief in my abilities and for continuously motivating me to strive for excellence.

And last but certainly not the least, to my lovely family and close friends for their patience, support, and love.

"In theory, theory and practice are the same.
In practice, they are not."

Albert Einstein

Contents

Acronyms **xv**

1 Introduction **1**
1.1 Company Overview . . . 1
1.1.1 Business Segments . . . 2
1.1.2 Farfetch Digital Platform . . . 2
1.2 Motivation and Objectives . . . 3
1.3 Dissertation Structure . . . 4

2 Literature Review **5**
2.1 Logistics . . . 5
2.1.1 Logistics Strategic Partnerships . . . 6
2.1.2 Logistics Design and Planning . . . 6
2.2 Warehousing . . . 8
2.2.1 Warehouse Classifications and Roles . . . 8
2.2.2 Warehouse Operations . . . 9
2.3 Simulation . . . 11
2.3.1 Types of Simulation . . . 12
2.3.2 Calibration, Verification and Validation . . . 14
2.4 Digital Twin . . . 15

3 Methodology **21**
3.1 Project Main Phases . . . 23
3.2 Model Development . . . 23
3.3 Data Collection . . . 23
3.4 Results Generation and Analysis . . . 24
3.5 Resources Gathering . . . 24

4 Business Case **25**
4.1 Venray Warehouse . . . 25
4.2 Venray Warehouse Operations . . . 29
4.3 Proposed Operational Changes . . . 30

5 Model Presentation **33**

5.1 Prototype Model 33
5.2 Scenario 37
5.3 User Interface 38
5.4 Agents 40
5.5 Model Architecture 40
5.6 Calibration 42
5.7 Final Remarks 45

6 Results and Discussion **47**
6.1 Scope and Assumptions 47
6.2 Results 49
6.3 Discussion 50

7 Conclusion **53**

Bibliography **55**

A Model Artifacts **57**

List of Figures

1.1 Farfetch's revenue breakdown by business segment, data from Farfetch (2023b). . 2

2.1 A framework for logistics design. Retrieved from Rushton et al. (2014, Chapter 6). 7
2.2 The key components of distribution and logistics. Retrieved from Rushton et al. (2014, Chapter 1). 7
2.3 Warehouse Operations Flow. Based on Bartholdi and Hackman (2019) and Rushton et al. (2014). 9
2.4 Decision Tree on how to study a system. Based on Law (2015). 12
2.5 Simulation Modeling Paradigms. Retrieved from Borshchev and Grigoryev (n.d., Chapter 2). 13
2.6 High level view of adjacent subroutines. Retrieved from Chung (2004). 14
2.7 Hypothesis test flowchart. Based on Chung (2004). 15
2.8 Original Digital Twin (DT) Model proposed by Grieves (2015). 16
2.9 DT application papers by application field. Retrieved from X. Liu et al. (2023). . 19
2.10 DT applications by phase. Retrieved from X. Liu et al. (2023). 19

3.1 Agile concept. Retrieved from (Adam, 2022). 22

4.1 Satellite image of the Venray Warehouse and the front view of the main dock. Retrieved from google earth. 25
4.2 Venray Warehouse layout. Based on Image courtesy of Farfetch. 26
4.3 Warehouse location code example. 26
4.4 Schematic layout of the Product Categories locations in floor zero of the Pick Tower. Schematic courtesy of Farfetch. 27
4.5 View of the Warehouse from the mezzanine above the dock. Photography courtesy of Farfetch. 28
4.6 Pictures of different aisles in the Pick Tower. Pictures courtesy of Farfetch. . . . 28
4.7 Simplified representation in a swimlane diagram of the Inbound process. 29
4.8 Simplified representation in a swimlane diagram of the Picking process. 30

5.1 Movements of different elements in the simulation, Inbound and Outbound. . . . 35
5.2 Comparison between two implementation strategies. 35
5.3 First floor of the prototype model 36
5.4 Elevator side of the prototype model. 37

5.5 Screenshot of the 3D Model. 38
5.6 Information displayed on 3D elements. 39
5.7 Simulation Start Screen. 39
5.8 Mosaic of all four views with direct access through the buttons on the menu bar. . 40
5.9 Processes of the three main agents. 41
5.10 Archimate model diagram including Business (yellow), Application (blue) and Technology (green) layers. 41
5.11 Items flow in the Inbound Process. 43
5.12 Valid and invalid elapsed times. 44
5.13 Data split for calibration purposes. 44
5.14 Ideal number of parameters. 45

6.1 Schematic layout of the Product Categories locations in the floor zero of the Pick Tower with the identification of the area under study. Schematic courtesy of Farfetch. 47
6.2 Storage zones covered by the simulations and the pathways where the agents move. 48
6.3 Preliminary results for the categories “footwear” and “accessories” - graph. . . . 50

A.1 List of Agents 58
A.2 Verification Station 59
A.3 Item 59
A.4 Pallets Inbound 60
A.5 Bin 60
A.6 Picker 61
A.7 Stock Manager (agent) 61
A.8 Forklift 62
A.9 Trolley Picking 62
A.10 Trolley Putaway 63
A.11 Elevator 63
A.12 Replacement Runner 64

List of Tables

2.1 Picking time breakdown (Bartholdi & Hackman, 2019). 10
2.2 Digital Twin Levels proposed by Madni et al. (2019). 18

4.1 Warehouse capacities breakdown by category. Based on data provided by Farfetch. 27

6.1 Results of “footwear” and “accessories” categories. 49

Acronyms

1P First Partner.

3PL Third Party Logistics.

4PL Fourth Party Logistics.

ABMS Agent Based Modeling Simulation.

AI Artificial Intelligence.

CGAR Compound Annual Growth Rate.

CPfR Collaborative Planning forecasting and Replenishment.

DEM Discrete Event Modeling.

DT Digital Twin.

ECR Efficient Consumer Response.

FbF Fulfilment by Farfetch.

GMV Gross Merchandise Value.

GPSS General Purpose Simulation System.

IBM International Business Machines Corporation.

IPO Initial Public Offering.

IT Information Technology.

ITS Intelligent Transportation System.

JIT Just In Time.

MAPE Mean Average Percentage Error.

MIT Massachusetts Institute of Technology.

ML Machine Learning.

NASA National Aeronauthics and Space Administration.

NGG New Guards Group.

NYSE New York Stock Exchange.

PLM Product Lifecycle Management.

SKU Stock Keeping Units.

TLC Total Logistics Concept.

UDP Uniform Demand Period.

WMS Warehouse Management System.

1 Introduction

Since its genesis, e-commerce has been experiencing a continuous growth across multiple dimensions, from the number of transactions worldwide and the range of product categories available. Now-a-days, most people of almost any age carry in their pockets the gate to millions of stores where virtually any kind of product, ranging from electronics, fashion or even cars, can be bought from anywhere at any time. This trend not only significantly impacts our daily lives but also shapes the global logistics systems, driving them to adapt to the new needs and challenges. The impact goes beyond logistics network aspects such as warehouse locations and means of transportation, it extends to the warehouse operations as well, where new needs such as e-fulfilment operations and increasingly returns handling arise. Besides tailoring warehouses to these needs, it is imperative to continuously optimize their operations. The rapid advancement of new technologies, from automation and artificial intelligence to data analytics and machine learning, brings with it new methodologies and tools for optimization. This represents a huge opportunity as much as a significant threat given that the players who don't effectively embrace such technologies are doomed to be left behind. This dissertation is focused on warehousing optimization through innovative techniques and was carried out on Farfetch, a global e-commerce platform for luxury fashion.

1.1 Company Overview

Farfetch is an online luxury fashion retail platform launched by Portuguese entrepreneur José Neves in 2008 and headquartered in London, UK (Farfetch, 2023a). As explained by José Neves in Camarotti (2020), most luxury fashion brands don't have scale to operate globally. In addition, fashion specialists usually do not understand technology. On the other hand, platforms such as eBay or Amazon don't understand the nuances of the luxury segment. Thus, in spit of selling products of almost every category, none of the global e-commerce platforms have been successful in selling luxury fashion. In this context, Farfetch was born with the mission of becoming "the global platform for luxury fashion, connecting creators, curators and consumers". Having received the first round of capital in 2010 (Camarotti, 2020), Farfetch was admitted on the New York Stock Exchange (NYSE) in 2018, in an Initial Public Offering (IPO) that valued the company at US $ 4.8 billion and through which US $ 885 million were collected (Negócios, 2018). In 2022, Farfetch revenue was US $ 2,317 billion and Gross Merchandise Value (GMV) of US $ 4,059 billion. The company growth has been impressive, as the mentioned figures were, respectively, US $ 142 million and US $ 382 million in 2015. The product categories sold by Farfetch include

clothing, shoes, bags, accessories, jewelry and beauty, for woman, man and children.

1.1.1 Business Segments

Farfetch organizes and reports its business in three segments: Digital Platform, Brand Platform, and In-Store (Farfetch, 2023b). The Digital Platform, which includes several online sales channels such as Farfetch Marketplace and BrowsFashion.com, generates revenue from the transactions between consumers and sellers on a commission basis. The Brand Platform focuses on the production and wholesale distribution of brands owned and licensed by New Guards Group (NGG). The In-Store segment operates stores such as Browns Stadium Goods and Violet Gray. Figure 1.1 shows Farfetch's revenue in 2022, broken down into its three business segments.

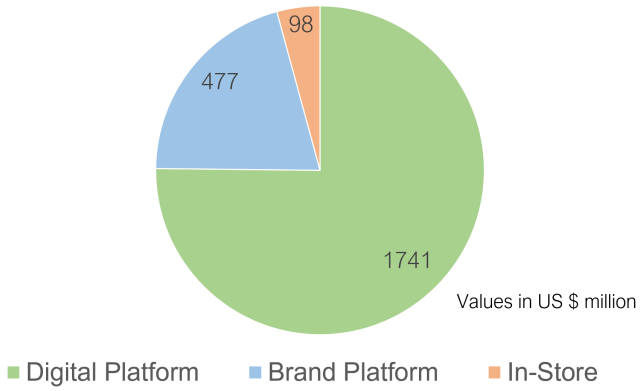


Figure 1.1: Farfetch's revenue breakdown by business segment, data from Farfetch (2023b).

Farfetch core business, Farfetch Digital Platform, generates over 70% of the company's revenue. According to Farfetch (2023b), Farfetch marketplace is the largest global digital luxury fashion marketplace connecting, as of December 31, 2022, consumers in over 190 countries and territories with merchandise in more than 50 countries from over 1400 sellers, including brands, boutiques and department stores.

1.1.2 Farfetch Digital Platform

Farfetch Digital Platform allows consumers, from anywhere in the world, to buy a wide range of luxury items from many different brands. All of that starts when a partner makes a product available online.

On the partners side, the first step is to register the product on the platform providing basic information and requesting a production slot (Farfetch, 2023c). Production is the process of generating all imagery and written information that enables the product to go online. Once approved and scheduled, the product is sent to one of the four Farfetch production studios. Two of these studios are located in Portugal, (Porto and Guimaraes), one in Sao Paulo, Brazil and one in Hong Kong. After production, the items are sent back to the platform partner. Some items such as beauty are not produced by Farfetch, instead manufacturers imagery is used.

Once the product is available online, a customer can place an order and the subsequent steps are the following:

- **Stock Validation**, where the platform partner confirms the availability of the product;
- **Fraud Check**, Farfetch responsibility supported by specialized partners;
- **Shipping Preparation**, where all documentation and labels are print;
- **Parcel Pick**, where the partner schedules the carrier for pick up;
- **Order Delivery**, where the parcel travels from the origin until the customer's doorstep.

For brands owned and licensed by Farfetch's Brand Platform, NGG, stock is held in one of Farfetch's warehouses. Farfetch takes care of all the steps that are otherwise responsibility of the platform partners, in partnership with Third Party Logistics (3PL) operators. These warehouses, located in Italy, Netherlands, United Kingdom, United States, China and Hong Kong not only hold own stock, the so called First Partner (1P) stock, but provide Fulfilment by Farfetch (FbF) services. FbF allows platform partners access to a fully integrated logistics network. Farfetch uses proprietary data and predictive insights to optimize the supply storage location at a regional level and within warehouses. Thus, not only is the seller freed from handling customers' orders but also these customers are served more efficiently and with a higher service level. FbF is a pass-through cost, with no direct economic benefit to Farfetch (Farfetch, 2023b), aiming to facilitate the grow of the Digital Platform segment.

1.2 Motivation and Objectives

Due to the strong growth the hosting company has experienced through the years, there is an increasing need for constantly improving and optimizing operations. Adding to that, amidst the implementation of the new trading rules related to Brexit, the company decided to move UK operations to an EU country to cope with this new reality. The location of choice was Venray, Netherlands where the company settled the biggest warehouse on its network. This move had to happen quickly, and thus there was almost no time for planning and design beforehand. Being the biggest warehouse on the company's network, its inefficiencies have the most impact on the company's operations and on its bottom line. Conversely, operational improvements in this warehouse represent substantial gains for the company. These operational improvements also result in a better service and more sustainability.

The objective of this project is to build a 3D simulation model of the Venray Warehouse on which its As-Is operations and proposed changes (To-Be) can be simulated and analyzed. The model is intended to be maintained over time and continuously improved, updated and extended for future business cases analysis. The model shall also serve as a basis for the development of simulation models for other company's warehouses. As most of the processes involved are usually similar between different warehouses, it will only require changes to the 3D scenario and adaptations of the logic.

The development of the model is accompanied by a business case where changes in the put-away process are proposed. From the beginning of the project, there was already the expectation

that the proposed operational changes would bring new challenges and problems that could more easily be identified and solved through simulation.

Moreover, there are on the pipeline automation projects for the company's warehouses, with studies already ongoing for this one. On the one hand, automation should only be implemented for processes with a high level of maturity as automation is inherently less flexible and thus less able to overcome bad process designs through behavioral changes on-the-fly as humans would do. Automation also means huge investments that shall only be undertaken if there is a good level of certainty that the proposed automation serves its purpose well and does so in such an efficient way that justifies the required investment.

This kind of projects requires multidisciplinary cooperation as stakeholders with different expertise have a say on the improvements to be studied and possibly implemented. In this regard, the simulation model also contributes by allowing a holistic view of the different processes and visually representing the scenarios under study, which functions as a "universal language".

1.3 Dissertation Structure

The present document is organized into seven chapters. **Chapter 2** provides a literature review, delving into different theoretical concepts with relevance to the work developed. This chapter is divided into four sections. The first two, Logistics and Warehousing, are related to the operational realities intended to be studied. The other two, Simulation and Digital Twin, are related to the tools used for the analysis. **Chapter 3** explain the methodology followed providing background in terms of existing frameworks and practical aspects that impacted the definition of the methodology for this work. **Chapter 4** presents a business case that accompanied the development of the model. For better understanding, the chapter is divided into three sections. The first two are dedicated to the presentation of the warehouse and its operations and the last one explains the proposed operational changes. **Chapter 5** presents the model developed within the scope of this dissertation. The chapter dedicates one section to the presentation of a prototype which was a key step towards the final model. The remaining sections are focused on aspects of the definitive model namely, scenario, user interface, agents, model architecture, and calibration. **Chapter 6** focus on the results obtained for the case study and provides a discussion on the main features of it. **Chapter 7** provides a general retrospective of the work developed along with its implications. In addition, it is also identified what were the limitations that impacted the project and a reflection on what might be the next steps moving forward is made.

2 Literature Review

This literature review has the following objectives: introduce theoretical concepts that are the basis for this work, understand the current state of the art in the relevant fields, and lock at previous research works and current trends. The two first sections, Section 2.1 and Section 2.2 are dedicated, respectively to logistics and warehouse. These are the real-world dimensions to be analyzed and studied through Digital Twin (DT) simulation whose the most relevant concepts are tackled in Section 2.3 and Section 2.4.

2.1 Logistics

According to EMR (2023), global logistics market is valued at near US $ 10 trillion (2022) and is projected to experience a 6.3% Compound Annual Growth Rate (CGAR) between 2023 and 2028, with e-commerce being the main driver of this growth.

Logistics is a key economic activity that makes extensive use of human and material resources thus affecting the whole economy. Langevin and Riopel (2005) conveys the idea that as an integral part of everyday life, in modern times logistics influences more than ever a large number of human and economic activities. Rushton et al. (2014) claims that about 30% of the employed workforce is engaged in occupations that are directly or indirectly related to logistics.

The key logistics activities, namely, transport, inventory, and warehousing, have been part of human life throughout history. However, as highlighted by Rushton et al. (2014), it was only in the past few decades that logistics has been recognized as a distinct function. This shift is related to the recognition by academia and industry that a comprehensive view of these activities is necessary to consider how they interrelate and interact with one another and thus achieve overall business success.

According to Ghiani et al. (2004) and Langevin and Riopel (2005), logistics can be defined as the part of the supply chain that deals with the planning and control of material flows and related information. Logistics is concerned with getting the right materials to the right place within the required time frame while optimizing one or more performance indicators.

Ghiani et al. (2004) also remarks that, within the supply chain, products typically move from raw material sources to customers, with the exception of obsolete, damaged, or non-functioning products that need to be returned for repair or disposal. In contrast, information flows in the opposite direction.

2.1.1 Logistics Strategic Partnerships

Logistics encompasses a broad spectrum of activities, including transportation, customs clearance, warehousing, packaging, information systems management, and so forth. Many of these activities require substantial resources and competencies. Most companies are unable to combine all the logistics related resources and capabilities they require. Thus, fully vertical integrated systems are quite rare (Ghiani et al., 2004). As a result, companies need to engage in relationships that can be either merely transaction based and function specific or can be strategic alliances Ghiani et al. (2004), which includes Third Party Logistics (3PL). 3PL is a long-term commitment with an external company to perform all of part of company's product distribution. Ghiani et al. (2004) argues that while suitable whenever the company is not willing to invest on specific infrastructures such as those related to warehousing or transportation, or whenever the company is unable to take advantage of economies of scale, 3PL causes the company to loose control of distribution and may as well generate higher logistics cost. According to Rushton et al. (2014), Fourth Party Logistics (4PL) goes a step further and takes care of both design and management of a client's logistics system providing thus a comprehensive logistics solution.

2.1.2 Logistics Design and Planning

Langevin and Riopel (2005) and Ghiani et al. (2004) differentiate between three levels of logistics decision-making: strategic, tactical, and operational. Strategic decisions encompass the design of logistics system and the acquisition of costly resources. Examples of these decisions, which have a long lasting impact, are facility location, capacity sizing, plant and warehouse layout and fleet sizing. Tactical decisions, on the other hand, involve aspects like production, distribution and resource allocation, and are made on a medium-term basis, typically spanning one to three months. Lastly, operational decisions are made on a daily basis and have a narrow scope such as warehouse order picking or shipment and vehicle dispatching.

Rushton et al. (2014) breaks down logistics design into four elements, namely, process, network, information systems and organizational structures and relates it with logistics strategy as part of corporate strategy, as shown in Figure 2.1. These elements are designed within the scope of the logistics strategy, which, in turn, is aligned with the competitive strategy that results from the corporate strategy. Rushton et al. (2014) also identifies the main activities within logistics and the decisions that need to be made within each, as expressed in Figure 2.2. These decisions are part of the respective functions which need to be planned in a systematic way, considering both their local environment and the wider scope of the distribution system as a whole Rushton et al. (2014).

Within the scope of this work, the objective is to build a model to tackle the "Storage, warehousing and materials handling" component, more specifically with the movement of goods inside the warehouse (material handling).

Rushton et al. (2014) points out that, for designing an effective logistics system, the interrelationships between the different components of the logistics system need to be considered within the context of the broader supply chain. For instance, different packaging which may be cheaper

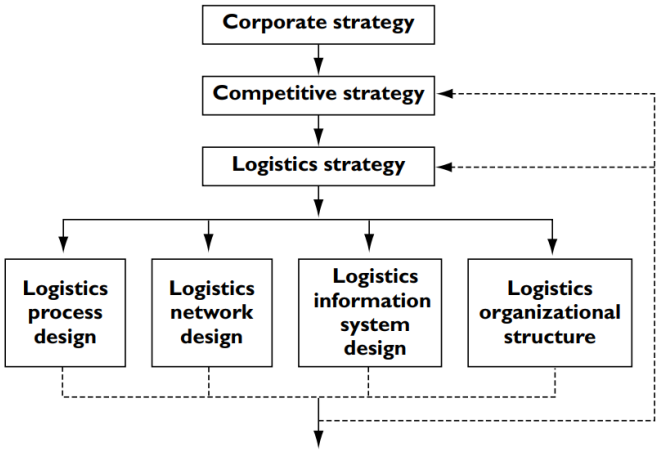


Figure 2.1: A framework for logistics design. Retrieved from Rushton et al. (2014, Chapter 6).

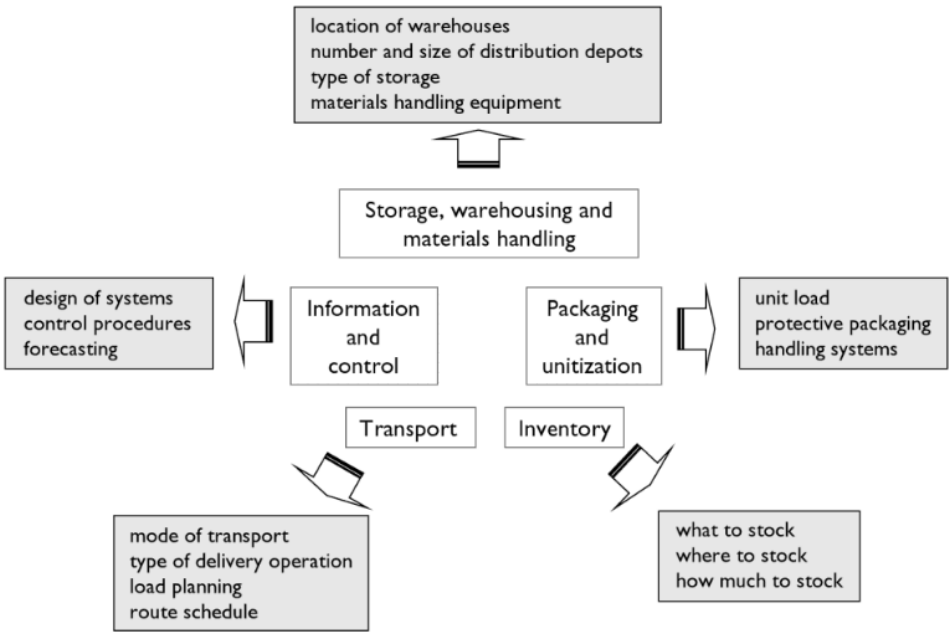


Figure 2.2: The key components of distribution and logistics. Retrieved from Rushton et al. (2014, Chapter 1).

may render transportation less efficient and thus worsens the overall supply chain costs. This concept of considering the whole supply chain as one single integrated system is called Total Logistics Concept (TLC).

2.2 Warehousing

Rushton et al. (2014) recognizes the importance of warehouses in modern supply chains and stresses that these facilities are likely to be found at most of the logistics network stages, from raw materials and production to finished goods. Rushton et al. (2014) acknowledges the existence of many techniques used to reduce the need to hold inventory, such as Just In Time (JIT), Efficient Consumer Response (ECR) and Collaborative Planning forecasting and Replenishment (CPfR) and yet, the reasons for holding inventory still exist. In spite of the typical high expenses associated with labor, capital, and information systems, Bartholdi and Hackman (2019) considers that warehouses bring value mainly in one or both of these ways: to better match supply with customer demand and to consolidate product. The former is related to the function of holding inventory from the time it is produced to the time the customer wants it. In fact, Bartholdi and Hackman (2019) notes that demand can change quickly while supply takes time to react. The latter mostly aims to reduce transportation costs and provide customer service. As an integral component of the supply chain where they belong, warehouses must be designed and operate in alignment with the specific requirements of the entire system (Rushton et al., 2014).

2.2.1 Warehouse Classifications and Roles

Rushton et al. (2014) acknowledges that the nature of warehouses within the supply chain may vary widely and thus there are many different types of classifications that can be adopted. For instance, in terms of ownership, the warehouse object of this project can be classified as a 3PL warehouse as it is owned by a partner. Aside from this, the warehouse can also be classified by "geographic area" as being global, regional, national, or local. In the case of the warehouse under study, it is mostly regional. Some types of classifications may not be as straightforward because the warehouse may lie within different classifications. It is often the case when classifying a warehouse by its function or role. In that regard, Rushton et al. (2014) observes that, more recently, a wide range of names for warehouses have emerged to reflect that mix of roles, e-fulfillment center being one of those.

In fact, the warehouse under study is a fashion e-fulfillment center and encompasses several functions. For instance, this facility can also be regarded as an inventory holding point. As noted by Rushton et al. (2014), in some industries, such as fashion, a particular style may be manufactured on a one-off basis while the product is offered for sale on a continuous basis. Thus, the need to hold inventory in a warehouse. Conversely, the warehouse can also be regarded as an assembly facility. Packaging and labeling, being types of assembly operations, can be tailored for the region/customer the products are sold to (Langevin & Riopel, 2005). Another increasingly unavoidable role for e-fulfillment centers is receiving returned goods. In fact, according to Rushton et al. (2014), e-commerce is associated with a higher rate of returned goods.

2.2.2 Warehouse Operations

As previously discussed, warehouses can have a wide range of different purposes. Yet, Bartholdi and Hackman (2019) and Rushton et al. (2014) recognize a general pattern of material flow, common to most of them. Consequently, the core services and operational activities of a warehouse, which are intrinsically tied to the material flow, are shared by most of them as well.

Figure 2.3 represents the general material flow found in most warehouses. Accordingly, upon arrival, material is subjected to a receiving process and then, through a process called Put-Away, it is stored in a designed zone. Later on, from its storage location, items are picked, packed and then placed orderly in a Marshaling zone. From this point material is ready to be dispatched out of the warehouse.

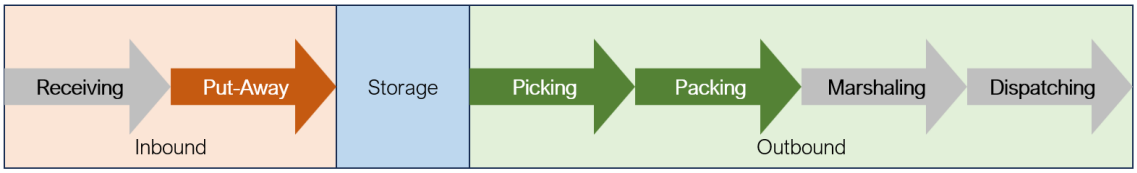


Figure 2.3: Warehouse Operations Flow. Based on Bartholdi and Hackman (2019) and Rushton et al. (2014).

For the special case of Cross-Docking, the steps Put-Away, storage, Picking, and Packing are replaced by Sorting operations which means less time within the warehouse and few number of operations performed. In fact, Bartholdi and Hackman (2019) suggests that, as much as possible, material shall flow continuously through this sequence of processes. Each time it is put down means that it must be picked up again sometime later, which is double-handling, thus resulting in extra-costs.

Receiving

Typically, receiving involves activities such as physical unloading of incoming goods, checking against purchase orders, and recording into the computer system (Rushton et al., 2014), normally by scanning a tag attached to the products (Bartholdi & Hackman, 2019). Besides checking against purchase orders, products may go through inspection for damage, incorrect counts, wrong descriptions, and so forth (Bartholdi & Hackman, 2019; Rushton et al., 2014). Rushton et al. (2014) also remarks that product might be unpacked and repacked in a format suitable for the subsequent warehouse operations (Rushton et al., 2014).

Put-Away

Put-Away is the operation through which products are placed in an appropriate storage location within the warehouse. Bartholdi and Hackman (2019) emphasizes the importance of determining the location to which a product is put away. It directly impacts the cost of Put-Away operations as distance traveled is a direct consequence of the location chosen, but also impacts subsequent

operations such as picking, both in terms of the time it takes to retrieve a product and in terms of costs.

As part of the Put-Away process, it is common practice to scan the locations where the product is being placed. This ensures that the system maintains accurate records of the stock's locations within the required storage area. Furthermore, it enables the system to generate efficient pick-lists, which assist order-pickers in retrieving the product for customer Bartholdi and Hackman (2019).

It is often part of the Put-Away process, to scan the locations where the product is placed. This allows the system to keep track of the locations of all stock inside which is necessary not only to manage available stock locations but also allows the system to construct efficient pick lists to guide the order-pickers in retrieving the product for customers (Bartholdi & Hackman, 2019).

Picking

Picking refers to the process of retrieving items from their respective storage locations to fulfill customer orders. In the warehouse, the outbound processes are initiated upon receiving a customer order (Bartholdi & Hackman, 2019). A typical order may contain a number of order lines, each requesting a certain quantity of a specific item (Rushton et al., 2014). As described by Bartholdi and Hackman (2019), based on the incoming orders, the Warehouse Management System (WMS) produces pick lists to guide the order-picking process.

Bartholdi and Hackman (2019) claims that about 55% of warehouse operating costs come from the picking process. Moreover, Bartholdi and Hackman (2019) provides the breakdown of time spent in the picking process into four categories (traveling, searching, extracting, and paperwork), as shown in Table 2.1. Among these four categories, traveling stands out as being the most time-consuming, representing 55% of the total, while the others vary between 10% and 20%. Since picking operations represent a significant expense in overall warehouse operations, it is no surprise that a considerable effort is usually given to reduce these costs, especially by lessening its traveling time.

Table 2.1: Picking time breakdown (Bartholdi & Hackman, 2019).

Activity	%Order-picking time
Traveling	55%
Searching	15%
Extracting	10%
Paperwork and other activities	20%

The WMS then checks the order against available inventory and identifies any shortages. In addition, the WMS may reorganize the list to match the layout and operations of the warehouse for greater efficiency (Bartholdi & Hackman, 2019).

While explaining how the picking process can be improved, Bartholdi and Hackman (2019) introduces the concepts of Stock Keeping Units (SKU) density or picks per unit of distance along the aisle traveled by an order-picker. Based on this concept we can talk about the pick density of an order. An order with a high pick density doesn't require much travel while one with a low pick density requires visits to widely dispersed locations. As a way of improving pick density, Bartholdi and Hackman (2019) suggests placing the most popular SKUs together and also to batch orders.

Packing

Usually, goods are assembled or packed together after picking (Rushton et al., 2014). This process, normally carried out in a packing station, may also involve final production postponement activities and value added activities such as kitting and labeling. According to Bartholdi and Hackman (2019) packing is a labor intensive process as each item of a customer order must be handled. Notwithstanding, the packaging process represents a convenient opportunity to check that the customer order is complete and accurate.

Marshaling and Dispatching

Marshaling is the last warehouse process before loading materials into the shipping vehicle for dispatching. According to Bartholdi and Hackman (2019), product is likely to be staged if it must be loaded in reverse order of delivery or if shipping long distances.

2.3 Simulation

Banks et al. (2005) defines simulation as the process of replicating the operation of real-world processes or systems over time by running a simulation model. A simulation model can be defined as being a set of rules that dictate how the system transitions from the current state to the next (Borshchev & Grigoryev, n.d.). It often requires a set of assumptions about how it works or behaves which are expressed in mathematical, logical, and symbolic relationships between its entities, or objects of interest, on the system (Law, 2015). The process of setting assumptions involves the use of abstraction, wherein details that are considered irrelevant to the problem are discarded while features that are deemed significant are retained (Borshchev & Grigoryev, n.d.). In cases where the relationships that compose the model are simple enough, obtaining an analytical solution may be viable (Banks et al., 2005; Law, 2015). Nevertheless, the complexity of many real-world systems makes it impractical to evaluate them analytically. As a result, simulation becomes the preferred approach for studying and analyzing such complex systems (Law, 2015). More specifically, Borshchev and Grigoryev (n.d.) identifies a set of features that render dynamic systems unsuitable to be studied by analytical means, namely:

- Non-linear behavior;
- Having memory;
- Non-intuitive influence between variables;
- Time and causal dependencies;
- All the above combined with uncertainty and large number of parameters.

Law (2015) suggests that a model should be used when it is not possible or cost-effective to experiment on the actual system and proposes a decision tree, shown in Figure 2.4, for guiding the choosing of the approach for studying a system. Yet, resorting to simulation requires attention regarding the accuracy and validity of the results thus obtained. When using models in operations research and system analysis, the mathematical ones are the natural choice. Physical models, as

useful as they might be in certain fields, are seldom well-suited in this context (Law, 2015). Within the mathematical models sphere, the choice between analytical methods or simulation shall be made based on the factors mentioned earlier.

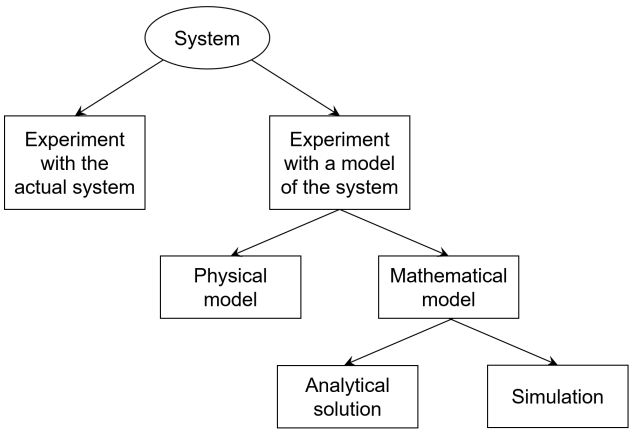


Figure 2.4: Decision Tree on how to study a system. Based on Law (2015).

As a matter of fact, Banks et al. (2005) suggests that in certain cases there are advantages to using both analytical methods and simulation. Indeed, when the two approaches are used, not only they might validate each other's results, but also enable a better understanding of the system. Even when analytical methods are applicable for studying a system, simulation can still provide substantial value due to their numerous advantages. In that regard, Banks et al. (2005) points out that the knowledge gained during the designing of the simulation model can be highly valuable for identifying potential improvements in the system under study. In addition, simulation often provides visualization capabilities that enable the representation of the system's functioning and behavior. When opting for simulation as a modeling method, it is crucial to acknowledge some drawbacks that it may entail. Law (2015) adverts for the amount of computing resources and expertise that simulation modeling may require as it becomes more complex. Besides, simulation analysis can be time-consuming and expensive.

2.3.1 Types of Simulation

Borshchev and Grigoryev (n.d.) identifies three main paradigms for mapping a real-world system into a model: System Dynamics, Discrete Event Modeling and Agent-Based Modeling. Figure 2.5 depicts how the same reality can be represented by each of these paradigms and how they are conceptualized.

System Dynamics was created in the mid-1950s by Massachusetts Institute of Technology (MIT) Professor Jay Forrester (Borshchev & Grigoryev, n.d.; Law, 2015). His innovative approach was to rely on laws of physics, specifically those governing electrical circuits, to model the dynamics of economic and, later on, social systems (Borshchev & Grigoryev, n.d.). This is a top-down approach that focuses on the systems at a more aggregate level (Law, 2015). System dynamics models comprise a combination of feedback loops, stocks, flows, and time delays (Božić,

2023; Law, 2015).

Discrete Event Modeling (DEM) emerged in early 1960s, when IBM engineer Geoffrey Gordon introduced the first version of General Purpose Simulation System (GPSS) which is considered to be the first implementation of this modeling paradigm (Borshchev & Grigoryev, n.d.). DEM considers that the system evolves through instantaneous changes at specific points in time due to the occurrence of events carried out on entities (Borshchev & Grigoryev, n.d.). The representation of processes often takes the form of flowcharts, which include elements like delays, resources utilization, decision branching, splitting, combining, amongst others (Borshchev & Grigoryev, n.d.).

Agent Based Modeling Simulation (ABMS) is a new approach to modeling systems comprised of autonomous, interacting agents (Macal & North, 2009). Until the early 2000s, this modeling and simulation paradigm was predominantly a subject for academic research, with limited practical implementation in real-world settings. Through ABMS, systems are represented more realistically as it eliminates the need for assumptions regarding the high-level behavior of the system which, in turn, results in a more accurate representation.

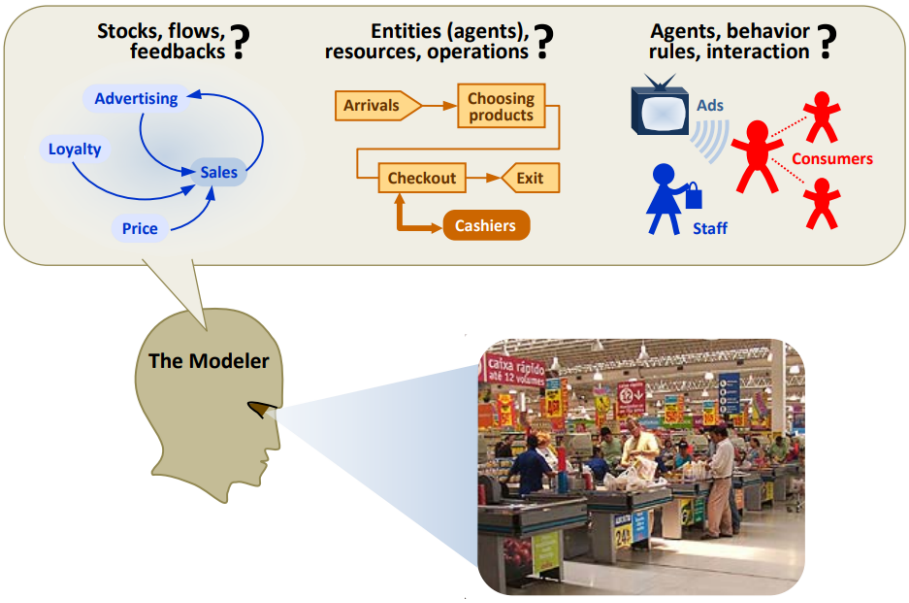


Figure 2.5: Simulation Modeling Paradigms. Retrieved from Borshchev and Grigoryev (n.d., Chapter 2).

Though often the selection of the modeling paradigm is heavily influenced by the modeler's background or the available tools, Borshchev and Grigoryev (n.d.) considers that the specific characteristics of the system being modeled and the goals of the study should be prioritized. Indeed, for each case, there is a paradigm that more naturally represents the problem under study or is better aligned with the objectives. According to Borshchev and Grigoryev (n.d.), there are situations in which different components of the system are best modeled using distinct methods. In such cases, a multi-method model shall be constructed.

2.3.2 Calibration, Verification and Validation

As stated above, the purpose of a simulation model is to replicate how a system or a process evolves over time. The reasons for utilizing simulation can vary widely, ranging from gaining a deeper understanding of a specific system to facilitating informed decision-making. Regardless of the purpose, it is essential that the simulation outputs provide a reliable approximation of the outcomes that would be obtained through real-world experimentation. According to Law (2015), a key challenge for a simulation analyst is determining whether a simulation model is an accurate representation of the actual system being studied, i.e., whether the model is valid. This is the point at which the concepts of verification, calibration and validation come into play.

Verification is the process of ensuring that the "assumptions document" has been correctly translated into a model (Law, 2015). Chung (2004) considers that as soon as the model runs without any errors or warnings, and it incorporates all the components specified under the system definition phase (i.e., runs as intended), then the model verification has been successfully completed. Verification can also be seen as a debugging of the simulation model and, as such, Law (2015) suggests borrowing debugging techniques from the field of software engineering. To facilitate the verification process, Chung (2004) recommends the adoption of divide-and-conquer approaches during development and, in that regard, suggests the organization of code in sub-routines. Figure 2.6 exemplifies this concept by showing sequences of blocks each representing the sub-routine that should contain its own isolated code. Moreover, the use of animation is also key for model verification (Chung, 2004; Law, 2015).

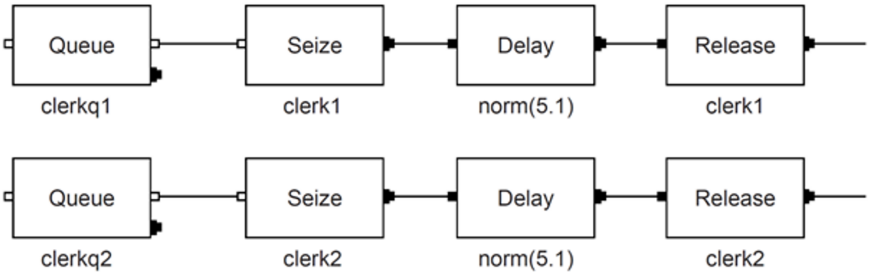


Figure 2.6: High level view of adjacent subroutines. Retrieved from Chung (2004).

On the other hand, validation is the process of determining whether a simulation model is an accurate representation of system in the context of the objectives of the study (Law, 2015). In other words, validation is the process of ensuring that a model reflects reality within a specific level of confidence (Chung, 2004). Law (2015) defines two major types of validation: face validation and statistical validation. Face validity refers to the apparent quality of the model regarding its ability to represent reality whereas statistical validity involves the comparison between system and model output through statistical methods. The two in combination are necessary to assure the accuracy of the model.

Law (2015) presents concepts such as Validation Data Collection and Validation Data Analysis

Process. The former refers to random data generated to feed the simulation model and the latter to the output data. The validation determines whether these datasets resemble the real ones. In both cases, validation data may be based on either individual observations or summary statistics. Law (2015) adverts that the autocorrelation, often present in the datasets, renders some statistical tests invalid as it violates their requirements. For this reason, validation based on individual observations might not be recommended for such cases. Statistical tests seek to determine whether the dataset from the model can be considered "equal" to the real-world dataset. The traditional workflow from statistics used to test whether the two datasets are "draw from the same population" or "are generated by the same phenomenon" are suitable for this purpose. Chung (2004) presents a flowchart (Figure 2.7) for guiding the choice of appropriate statistical test and provides a walkthrough on how to apply each of these tests. The flowchart proposes four different statistical tests that suits different cases depending on whether the real data and model data are both normal, present a natural pairing, and have equal variances.

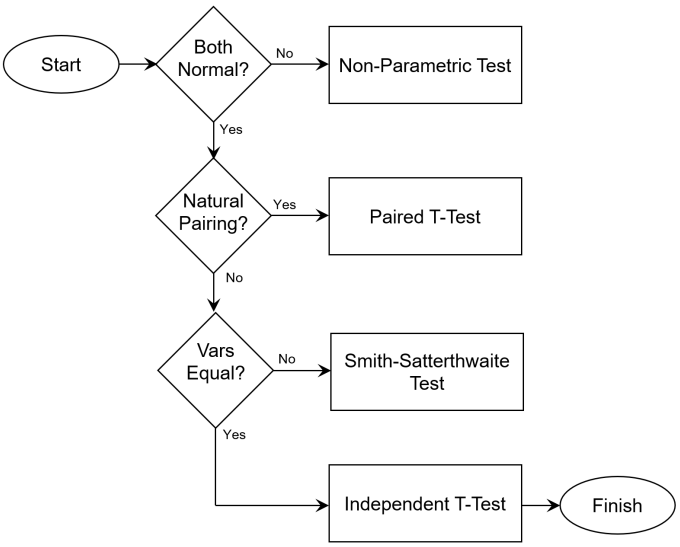


Figure 2.7: Hypothesis test flowchart. Based on Chung (2004).

When the model fails validation, it may need calibration. Calibration is the process of changing the model aiming to reduce discrepancies between the real system output data and the corresponding model output data. Often, model calibration entails changing its parameters' values and shall be performed until the two output datasets agree closely (Law, 2015). However, Law (2015) adverts for the importance of using independent datasets for calibration and validation, otherwise it will be uncertain whether the model is in fact representative of the real system, or instead it only shows a good fit for that particular dataset.

2.4 Digital Twin

It is claimed by Allen (2021) that the concept of DT has its roots in the Apollo 13 mission in 1970. During its journey heading to the moon, one of the spacecraft's oxygen tanks exploded thus

justifying the famous James A. Lovell words: “Houston, we’ve had a problem” (Edwin E. Aldrin et al., n.d.). In an attempt to save the lives of the three man onboard, the flight control team in Houston developed and tested possible solutions on a physical model mirroring the spacecraft conditions. As a result, an improvised air purifier that could be built by the astronauts with materials available to them was developed. Meanwhile, in Houston and Kennedy Space Centers, other crews of astronauts ran and tested procedures to safely bring the Apollo 13 crew back to Earth. Since then, the National Aeronauthics and Space Administration (NASA) has been creating mirrored systems meant to find solutions for problems that happen during missions.

A concept and model of DT was informally introduced by Michael Grieves in 2002 on a presentation about Product Lifecycle Management (PLM) with the title “Conceptual Ideal for PLM” (Barricelli et al., 2019). This concept would then be introduced by Grieves in his University of Michigan Executive Course on PLM. The term “Digital Twin” was publicly introduced by Grieves in his book “Virtually Perfect: Innovative and Lean Products through Lifecycle Management”, though the author attributes it to John Vickers of NASA with whom he has closely worked. Both have used the term in subsequent projects (Grieves, 2015).

The first DT model Figure 2.8 was composed of three primary elements: a real space containing a physical object, a virtual space containing a virtual object and the link for data flow from the real space to the virtual space and vice-versa. This last element allows the convergence and synchronization of the virtual and physical systems (Barricelli et al., 2019).

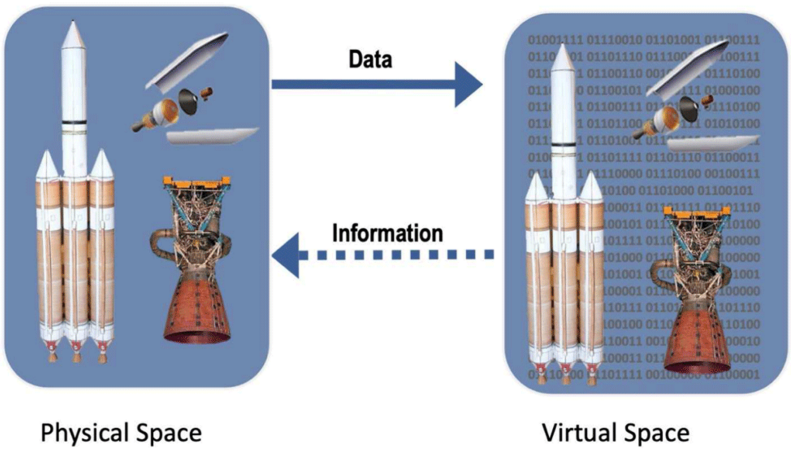


Figure 2.8: Original DT Model proposed by Grieves (2015).

At the time this concept was introduced, digital representations of actual products were relatively new and immature and the information being collected from the physical side was limited and mostly paper-based (Grieves, 2015). In academia, a sudden growth of publications related to DT only started in 2017 and continued through 2018 until eventually, it lost some stamina during the following year (M. Liu et al., 2021). This doesn’t represent a decrease in researchers’ interest in the topic, instead, their attention has been more oriented towards real practice and technologies in industry (M. Liu et al., 2021).

Despite the success of DT technologies and their massive spread in the corporate and business environments, on the literature a detailed description of the characteristics of a generic DT is still

missing as each state-of-the-art paper concentrates on the development of a few components of DT (Barricelli et al., 2019). On the other hand, Grieves (2015) acknowledges that there is no consensus on what the definition of DT should be and criticizes a proposed set of digital constructs as subcategories of Digital Twins. Sharma et al. (2022), identifies a research work that discusses simulation software with Machine Learning (ML) capabilities where it is referred by the authors as a "Digital Twin Approach". This so-called DT doesn't have real-time synchronization connection to the physical product and yet, the use of the DT term can be correct according to some definitions while incorrect according to others. Sharma et al. (2022) considers this as the result of the inconsistency and the lack of a unified architecture and definition of DT concept. Sharma et al. (2022) also notes, through his literature review, the existence of interesting patterns on research papers through the years. In early years, most of the papers would focus the DT concept more on high-fidelity and on multidisciplinary simulation capabilities, with no consideration for real-time connection with the physical object. Later, researchers started to attach importance to dynamic and bidirectional mapping to the physical object but a clear distinction from DT and general computational model simulations was still lacking (M. Liu et al., 2021). In spite of the hype and interest from both industry and academia, no universal definition, implementation framework or protocol exist. On the research side, a comprehensive and in-depth analysis of DT from the perspective of concepts, technologies, and industrial applications is still lacking (M. Liu et al., 2021).

M. Liu et al. (2021) identifies the following four defining characteristics of a DT:

1. Individualized;
2. High-fidelity;
3. Real-time;
4. Controllable.

The first one means that the DT shall resemble, in all aspects, its physical counterpart. The second one states that it should do so as accurately as possible. The third one suggests that DT should represent the physical twin with relatively low latency. This last one means that changes in any of the twins should be reflected in the other.

Madni et al. (2019) purposes a classification in four layers of virtual representation (or maturity), represented in Table 2.2. This classification is based on aspects such as if the physical system has to exist before a DT is created, if the physical system has onboard sensors and processor so it can report performance, health and maintenance data and whether the DT makes use of Artificial Intelligence (AI).

At level 1, called Pre-Digital Twin, is a traditional virtual prototype used during design and is not used through the life of the physical asset. For the second level, the Digital Twin, the physical twin must exist, and the digital representation receives batch updates from the latter regarding its performances, health and maintenance. At level 3, the Adaptive Digital Twin offers interface to both physical and digital twin that is sensitive to users' preferences and priorities. And in the fourth level is the Intelligent Digital Twin which has a high degree of autonomy and can analyze more granular performance, maintenance and health data from the real-world counterpart.

Table 2.2: Digital Twin Levels proposed by Madni et al. (2019).

Level	Model Sophistication	Physical Twin	Data Acquisition from Physical Twin	Machine Learning (Operator Preferences)	Machine Learning (System / Environment)
1 Pre-Digital Twin	virtual system model with emphasis on technology / technical-risk mitigation	does not exist	Not applicable	No	No
2 Digital Twin	virtual system model of the physical twin	exists	performance, health status, maintenance; batch updates	No	No
3 Adaptive Digital Twin	virtual system model of the physical twin with adaptive UI	exists	performance, health status, maintenance; real-time updates	Yes	No
4 Intelligent Digital Twin	virtual system model of the physical twin with adaptive UI and reinforcement learning	exists	performance, health status, maintenance, environment; both batch / real-time updates	Yes	Yes

DT concepts have been applied to a wide range of fields. X. Liu et al. (2023) through literature review identifies the most common fields of application and how often they are referred on research papers Figure 2.9

X. Liu et al. (2023) also identifies the most common types of applications for each application phase, namely Design Phase, Operation Process, Dynamic Interaction Figure 2.10.

As an experimentation tool, a DT is easier to manipulate and study in a controlled environment than its physical counterpart in the operational environment (Madni et al., 2019). This allows for cost-effective exploration of design alternatives and what-if scenarios (Madni et al., 2019), during both the design phase and the exploration of the asset. More systematically, Madni et al. (2019) identifies four advantages of using DT:

1. Allows for validation of the system model with real-world data.

2. Provide decision support and alerts to users over time and adjust/modify physical system parameters to deal with contingencies.

3. Predict changes in the physical system over time and adjust/modify physical system parameters to deal with contingencies.

DT can bring many advantages, most of them often lead to savings in resources such as cost and time (Bazaz et al., 2019). Sharma et al. (2022) points out that real monitoring, simulation, optimization and accurate forecasting are the main advantages of DT but the concept is not mature enough to fully achieve this vision at scale. Notwithstanding the DT global market, valued at US $ 8.60 billion (2022) is projected to reach US $ 137.67 billion by 2030, representing a CGAR of 42.6%.

In spite of these projections, many researchers identify a number of challenges to overcome. For instance, Bazaz et al. (2019) identifies the provision of data as one of these challenges due

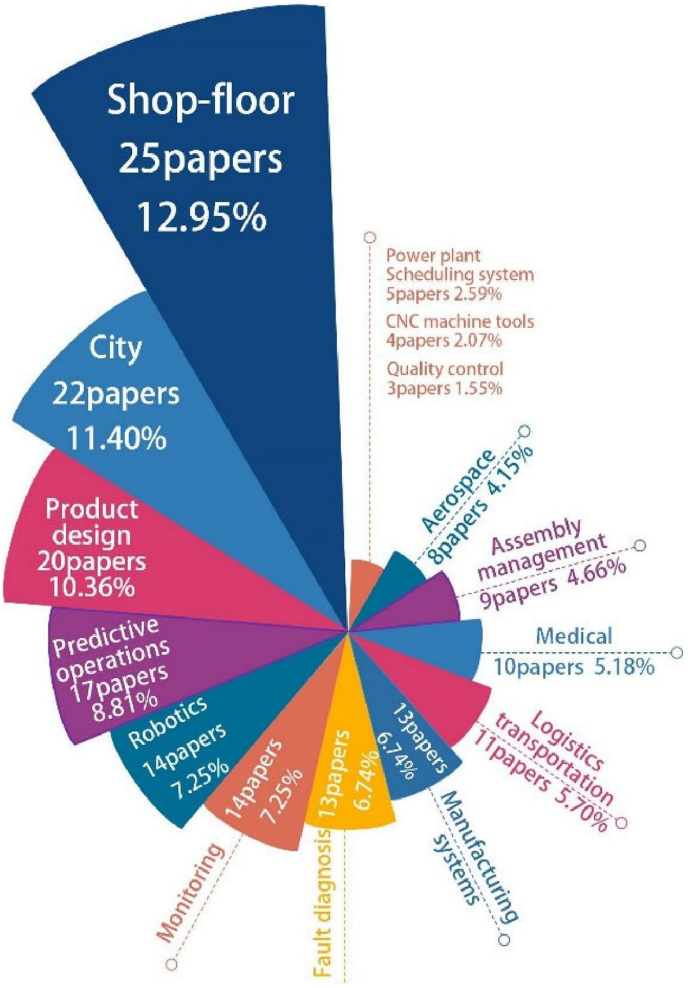


Figure 2.9: DT application papers by application field. Retrieved from X. Liu et al. (2023).

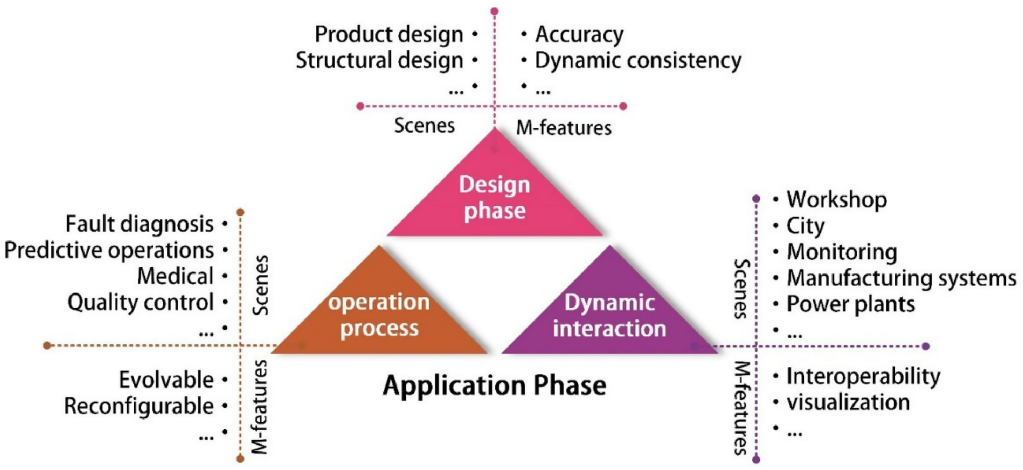


Figure 2.10: DT applications by phase. Retrieved from X. Liu et al. (2023).

to the variety of data in different formats and with different owners. Sharma et al. (2022), acknowledging that DT is a fast evolving concept, points out the lack of a universal DT reference framework, lack of performance metrics, security concerns over shared data and reliance of DT on other fast-evolving technologies are the main obstacles for DT implementation.

3 Methodology

This chapter presents the proposed methodology for the journey toward the established objectives. The methodology encompasses the strategies to develop the Digital Twin (DT) model and supporting frameworks.

To better understand the methodology proposed, the present work can be seen as a software development project, given that the simulation model is, in all aspects, a piece of software. A waterfall model is a classic approach to software development following a strictly defined sequence of phases where the completion of each phase conditions the starting of the next one. This linear approach to project development provides a sense that a systematic and well-defined project progression is assured. That may be the case for projects similar to previous projects or made out of phases/tasks that have been repeatedly performed in the past, and it is expected that, as the project is completed, all of its building blocks will fit together. Notwithstanding, given the complexity and novelty of many software products, a waterfall approach can easily fail. Let's say that the main waterfall phases of a particular software development project are:

1. Requirements gathering;
2. Design;
3. Development;
4. Testing;
5. Deployment.

The software may fail during the testing phase due to bad design. As a result, a redesign would be necessary. In addition, the phases of Development and Testing would need to be repeated with no assurance that the issue would be solved. Or worse, if a technical problem is discovered, it might not be obvious what the origin of it might be, and all the work done in the previous stages may need to be reviewed.

One way of coping with the mentioned problem is to adopt agile methodologies. The main pillar of these methodologies is the delivery of working software in short development cycles called sprints. The first versions of the software, which can be seen as sketches of the final product, have a small number of features, and are simpler. Then through an iterative process, refinements are made and features are added until a final product is obtained. This constitutes a divide and conquer approach that makes complex problems more tractable. In addition, bugs or technical obstacles, that invariably appear along the way, are detected and identified earlier and easily.

Another important issue to tackle is the difficulty of developing the software that the project owner wishes. On the one hand, the requisites established by the project owner may only partially reflect what is expected from the final product. On the other hand, the development team may not accurately understand what is intended to be accomplished. By following a sequential development model such as the waterfall, such discrepancy between what was done versus what should be done is only evident towards the end of the project. Agile methods avoid such an unpleasant end by providing valuable working software that allows mutual feedback between the development team and the project owner. The collaboration between the project owner and the development team allows the latter to understand better what the project shall accomplish. Thus, the development team can help the project owner reformulate the requirements to reflect better what he expects. In addition, as a result of this better understanding, the team can also suggest new features to the project owner that were not yet considered or were believed to be infeasible.

It is also relevant to point out that, throughout the development of the project, as often new technologies and methodologies are adopted, the development team improve their skills and, by following an iterative process it is possible to refine different parts of the projects leveraging this new acquired knowledge.

Figure 3.1 presents schematically a software development process following agile principles. Accordingly, the development is split into sprints. At the end of each sprint, a working version of the product is always delivered. In subsequent sprints, more features and refinements are added to the previous version until reaching the final product.



Figure 3.1: Agile concept. Retrieved from (Adam, 2022).

With that being said, the proposed framework was developed following the Agile Scrum framework. Given that neither the author or other people involved are deeply experienced in this framework and there is also the need, to a certain extent, to keep old ways of working, the framework was not strictly followed. Nonetheless its main principles were adopted with the intention of possibly taking steps in the future to formal adoption.

3.1 Project Main Phases

The current project is organized into the following five phases:

1. Model development;
2. Data collection;
3. Results Generation and Analysis;
4. Resources Gathering;
5. Writing the dissertation.

Each phase has specific objectives to support the project's overall completion. Throughout the development of the project, at each point in time, the assessment of what should be done next was driven by the overall objectives and conditioned by dependencies and by restrictions. For instance, software testing can only be performed after its development. On the other hand, developing huge chunks of code without intermediary testing is not recommended.

Regarding constraints, as an example, at a certain point in time there were technical difficulties with the required software licenses hosted in a virtual machine. As the model could not be developed while the problem persisted, other tasks were carried out.

3.2 Model Development

The Model Development phase comprehends tasks such as building visual elements of the model, defining the model architecture, coding, and testing/debugging. This phase started by developing a prototype model with the following objectives:

1. Review and get familiar with Anylogic functionalities;
2. Assess what might be feasible with Anylogic software and how it could interact or be connected with other software;
3. Test different implementation strategies for the necessary features/agent behaviors.

As soon as the most important and challenging features were tested, the next step was to build the warehouse scenarios under study and implement the previously tested features. During this phase, many features were refined while others considered less risky, were tested and implemented for the first time.

3.3 Data Collection

The Data Collection phase can be divided into two main categories. The first category relates to data gathering, usually from datasets intended to feed the model. The second category refers

to gathering other data, such as pictures, company documentation, or tacit information required to construct model scenarios and to model the agents' behavior.

In relation to the input data for the model, mostly warehouse transactions describing a specific even such as an item has been scanned, an instance of one day was first used. The operations of interest are "Inbound Receiving", "Inbound Put-away", and "Outbound Picking". Later, after being familiar with the data, direct access to the platform used by the warehouse operator was obtained to get similar data for any time span that could be useful.

In relation to the other kinds of data, on a first phase, blueprints of the warehouse were obtained. One of them is an AutoCAD *dwg* file presumably part of the warehouse's building project. The other blueprints, in Excel files, show the disposition of the racks in the warehouse, and identify the locations by their code. Throughout the time the project was developed, it was possible to gather photos taken inside the warehouse by colleagues. That helped both to understand better how the physical space to be modeled looks and how the operations inside are carried out. Conversations with colleagues more familiar with the warehouse were also relevant for that purpose.

3.4 Results Generation and Analysis

The Results Generation and Analysis phase can only be performed for the first time after the model reaches a certain level of development. From that point on, this phase is included in the sprints with the objective of assuring the well functioning of the model and assessing the potential value of the Business Case. Based on the results, decisions are made regarding whether the Business Case shall continue to be explored and the direction of that exploration.

3.5 Resources Gathering

At the beginning of the project, all the necessary resources that can be identified are prepared. However, the need for some resources can only be recognized as the project evolves, given the exploratory and innovative nature of the project. Examples of such resources are software needed for 3D modeling, for reading AutoCAD *dwg* and computational power to run the model. The launch of model versions at the end of each sprint can provide important clues about the expected resource requirements for the following sprint.

As an example, the need for heavy computational capabilities started to be discussed during the first half of the project. Then, the necessary actions for approving and implementing a powerful computer were taken. This process was somewhat new for the people involved, so the sequence of actions required were figured out along the way and with the support of Information Technology (IT) personnel from different departments.

4 Business Case

This chapter introduces the business case that will be addressed using the developed model. The business case is accompanied with the rationale behind it and a brief analysis of its expected impact and implications. Nonetheless, for context and for a better understanding of the business case, a brief introduction to the Warehouse under study is also provided.

4.1 Venray Warehouse

As mentioned previously, the Venray Warehouse is the biggest on the company's network and has been set up in the last couple of years. Figure 4.1 shows a satellite image of the warehouse and a front view of the main dock. The area occupied by the company is delimited by the red line.

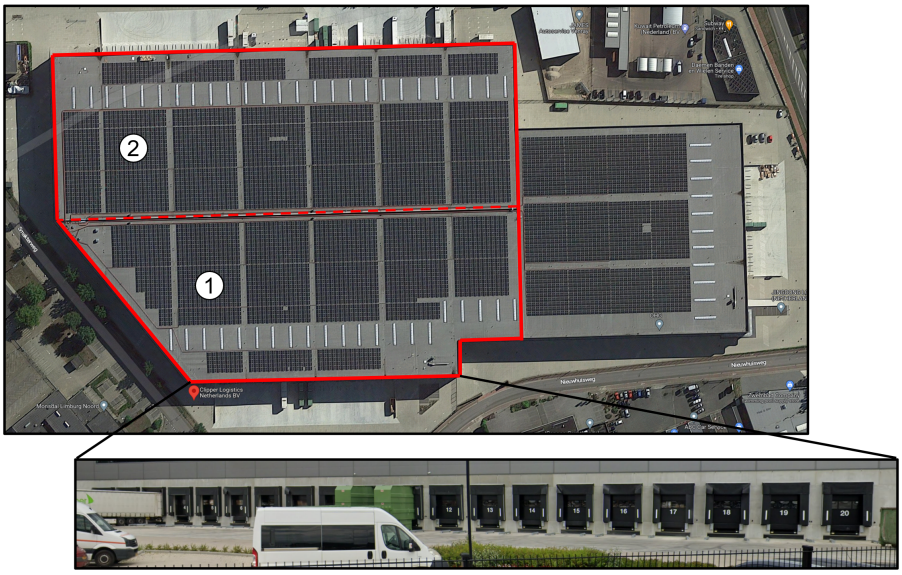


Figure 4.1: Satellite image of the Venray Warehouse and the front view of the main dock. Retrieved from google earth.

In the beginning, only Zone 1 of the Warehouse was rented. Soon after, to cope with the strong growth experienced, it was expanded to include Zone 2. The two parts have a combined blueprint of 36,384m^2 (Zone 1 - 16,560m^2, Zone 2 - 19,824m^2). Figure 4.2 presents the Warehouse layout where its main elements are identified, and those important for the present work are highlighted. Letters from A to F identify storage zones with different characteristics. Letters U, G and H

represent the three first passage points of the incoming material. Conversely, letters I, J and V represent the three last passage points for items that are expedited after being purchased.

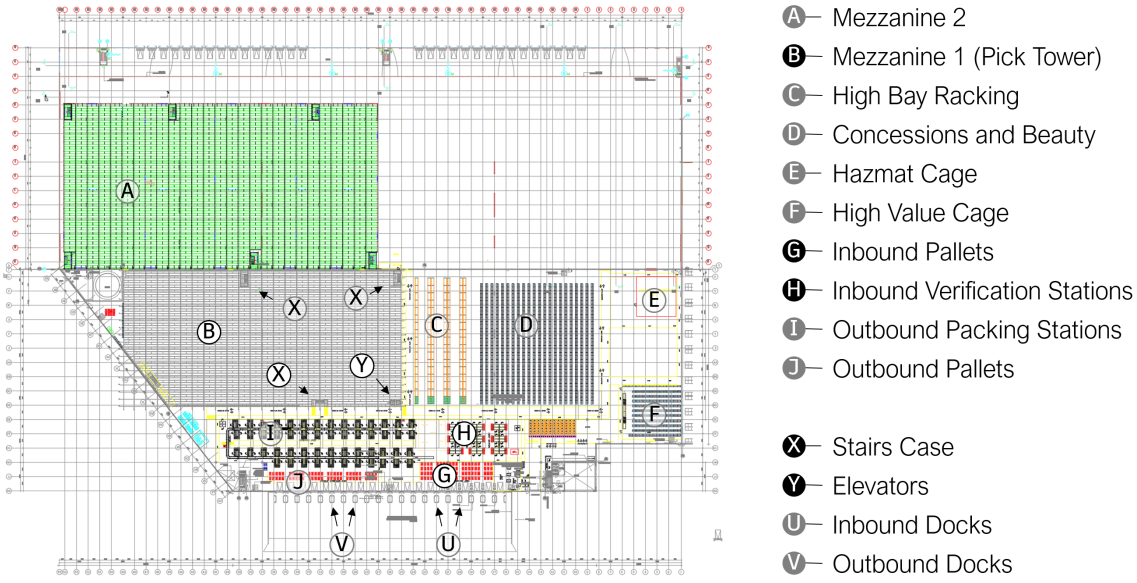


Figure 4.2: Venray Warehouse layout. Based on Image courtesy of Farfetch.

The Warehouse has two mezzanines for storage, Mezzanine 1 and Mezzanine 2, identified, respectively, with letters B and A in Figure 4.2. Each mezzanine adds three more floors to the respective zone. Mezzanine 1 and Mezzanine 2 occupy 5,283m^2 and 7,498m^2, respectively. While part of Zone 2 is still empty and the storage locations on Mezzanine 2 have low occupancy rates, Zone 1 is fully occupied. In Zone 1, besides the Pick Tower (Mezzanine 1), there are specific zones for special items. For instance, for safety reasons, products that have chemical hazards are stored in the Hazmat Cage, while high-value products are stored in the High Value Cage, which can be seen as a big vault.

The location codes, which identify unambiguously every single storage location in the Warehouse, follow a traditional format. Figure 4.3 shows an example of a location code. Accordingly, the first group of characters identify the zone and floor, the second group identifies the aisle, the group 3 specifies the bay within the aisle, and the fourth group identifies the shelf and horizontal position within the bay.

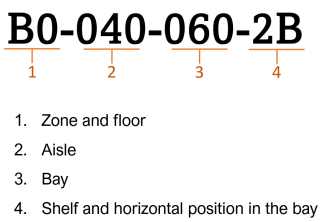


Figure 4.3: Warehouse location code example.

Each zone is divided into categories of items such as footwear or flat apparel. For illustrative

purposes, Figure 4.4 shows the schematic layout of the product categories on the ground floor of the Pick Tower, where each color represents one category. This map was drawn in Excel and each colored cell represents one bay (comprising three horizontal locations), while the column of white cells between columns of colored cells are aisles.

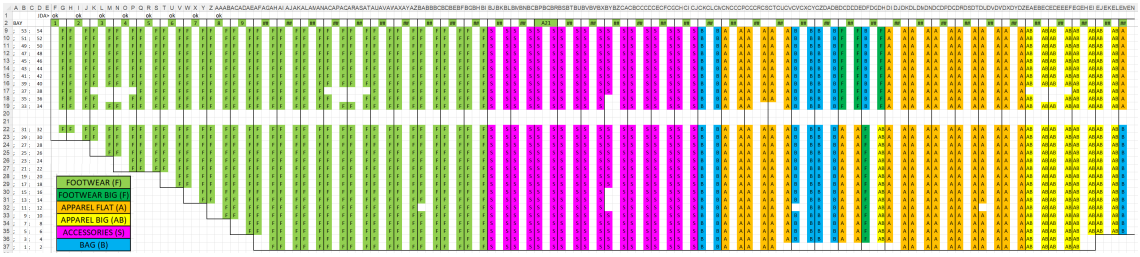


Figure 4.4: Schematic layout of the Product Categories locations in floor zero of the Pick Tower. Schematic courtesy of Farfetch.

On Table 4.1, the Warehouse capacity is broken down by category. Ten categories are identified, plus "Quarantine" which are zones reserved to store items that are offline (not available for purchase). The percentage of allocated spaces is expressed by number of bays, number of locations and also by number of units that can be stored. When applicable, the dimensions of the bins used to store the items are expressed. The last two columns detail the number of locations by bay and also the unit capacity of each location.

Table 4.1: Warehouse capacities breakdown by category. Based on data provided by Farfetch.

							Capacity	
		Bay	Loc	Unit	Shelving	Boxes	Loc/Bay	Unit/Loc
Footwear	F	29.5%	28.9%	17.8%	40 cm	–	15	3
GOH Short	HS	3.8%	1.5%	1.5%	60 cm GOH	–	6	5
GOH Long	H	2.7%	0.5%	0.5%	60 cm GOH	–	3	5
Accessories	S	12.9%	12.6%	13.0%	40 cm	–	15	5
Bags	B	4.7%	4.6%	2.8%	60 cm	30x40x60	15	3
Apparel Flat	A	3.4%	9.9%	12.2%	60 cm	30x40x60	45	6
Apparel Big	A	2.0%	2.0%	1.2%	60 cm	30x40x60	15	3
Beauty	A	16.0%	15.7%	25.8%	40 cm	–	15	8
Hazmat	A	0.0%	0.0%	0.0%	40 cm	30x40x20	32	10
High Value	A	0.0%	0.0%	0.0%	40 cm	30x40x20	32	10
Quarentine	A	25.0%	24.4%	25.1%	Various	–	15	5
TOTAL		100%	100%	100%				

For a better understanding of the Warehouse's inside space, a picture taken from a third mezzanine situated above the docking area is shown in Figure 4.5, providing an overview of the inbound area and two storage zones: Pick Tower (B) and High Bay Racking (C). Most of the operations to be studied happen in zones that appear in this picture. Arriving Material is firstly transported to the Inbound Verification Stations (H), and from there, to the storage locations, which for this work is restricted to the Pick Tower (B). Pictures of aisles inside the Pick Tower, are shown in Figure 4.6, providing a glimpse of different storage systems used for different categories.

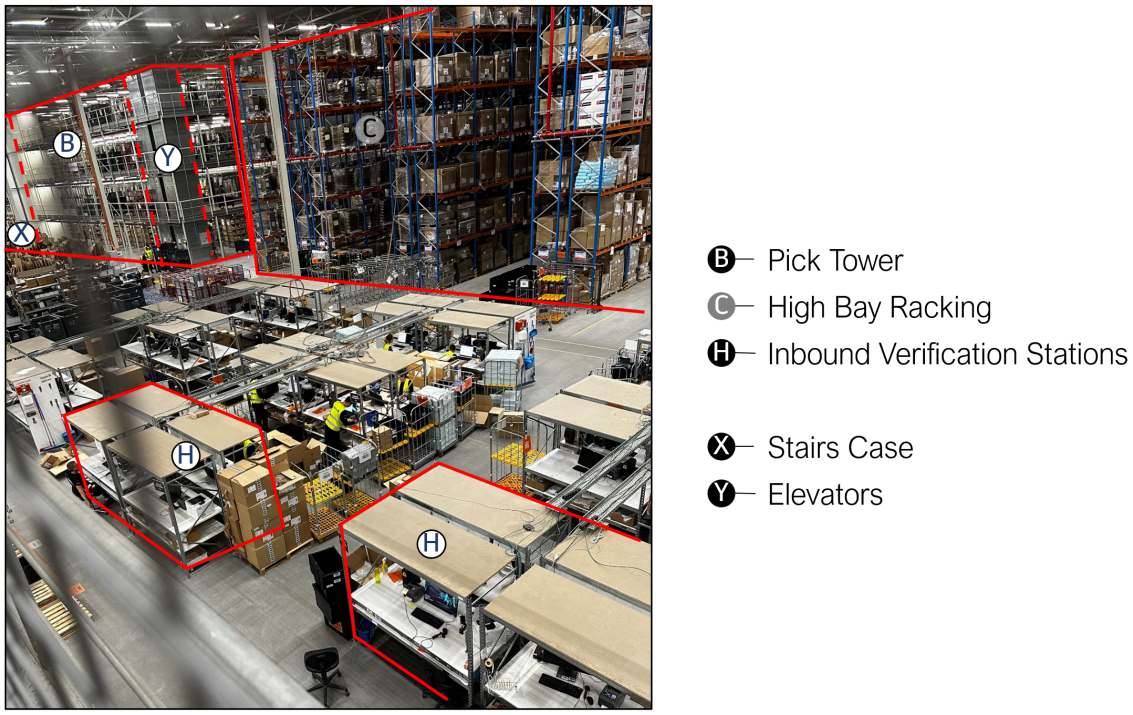


Figure 4.5: View of the Warehouse from the mezzanine above the dock. Photography courtesy of Farfetch.



Figure 4.6: Pictures of different aisles in the Pick Tower. Pictures courtesy of Farfetch.

4.2 Venray Warehouse Operations

Warehouse operations are complex and involve several different processes. For this work, the inbound process is the most relevant. Figure 4.7 shows a swimlane diagram where, in a simplified way, the sequence of tasks of the inbound process, associated with the respective actors, are presented.

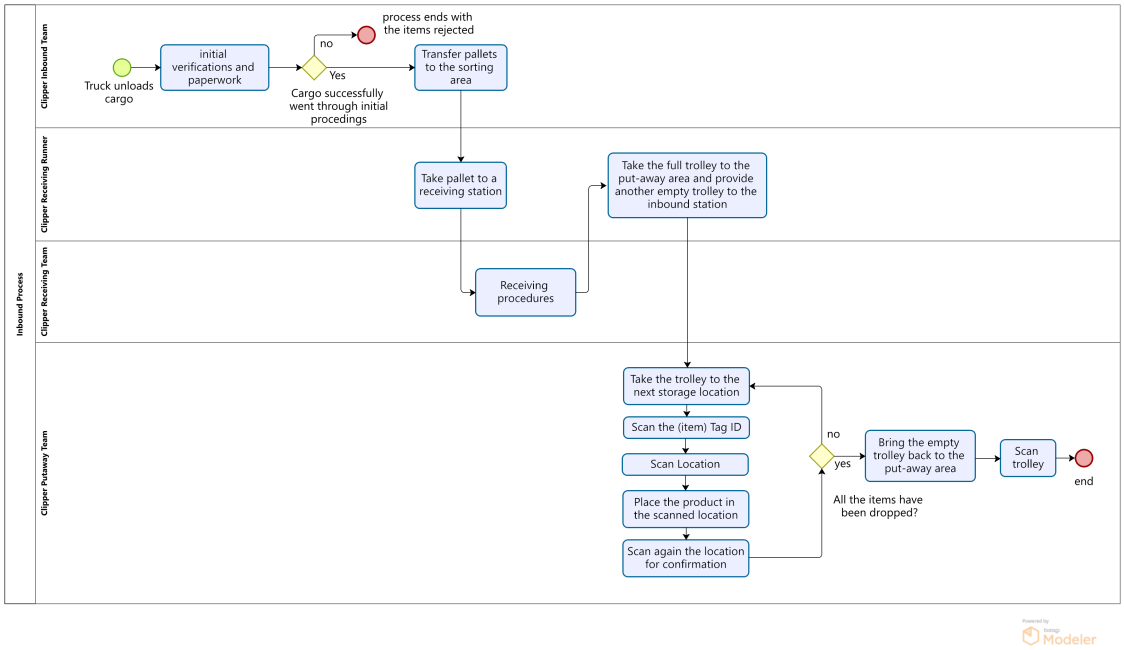


Figure 4.7: Simplified representation in a swimlane diagram of the Inbound process.

The level of detail expressed suits the purpose of this analysis. Thus, from the unloading of the pallets from the vehicle until the pallets are transported, given that the process involves a number of tasks and decision points that are not relevant to this work, few details are expressed. After the pallets are transported from the sorting area to the station, receiving procedures are performed. These procedures aim to confirm that the items are according to description, pass the quality control and insert their information in the JDA[1] by scanning the respective tags and placing them into the trolley. The inbound verification operator repeats this process until the Put-Away trolley is full. At this point, a Runner provides an empty trolley to the station and takes the full trolley to a Put-Away waiting area where a Material Handler[2] takes them to perform a Put-Away wave. This wave consists of placing all the items of a Trolley in their storage locations. For the purpose, the Material Handler takes the Trolley to a suitable location for the next item. Once there, he scans the item, scans the location, puts the item in the location and scans the location again to confirm the item has been placed. This cycle repeats until the trolley is empty. As the Material Handler finishes the wave, he goes back to the Put-Away waiting area, leaves there the empty trolley and takes a full trolley for the next Put-Away wave.

The process from the point where the trolley leaves the inbound verification station until the

[1] James Donald Armstrong Software is used for warehouse management.
[2] Person Responsible for moving stock within the warehouse.

Put-Away wave ends is the most relevant part for the proposed business case. Nonetheless, being an exploratory analysis it is important to consider upstream and downstream tasks.

The picking tasks, though not the main focus of this business case, directly impact the inbound operations because, as a result of picking, locations are free and thus can be used to store incoming items. Another reason to be aware of the picking process, even while analyzing the inbound, is that all these operations happen in the same physical space and some interaction between them may happen. For instance, aisles may become congested by picker operators and thus Put-Away operators are delayed. Figure 4.8 shows, with simplifications, a swinlane diagram of the picking process and the respective actors.

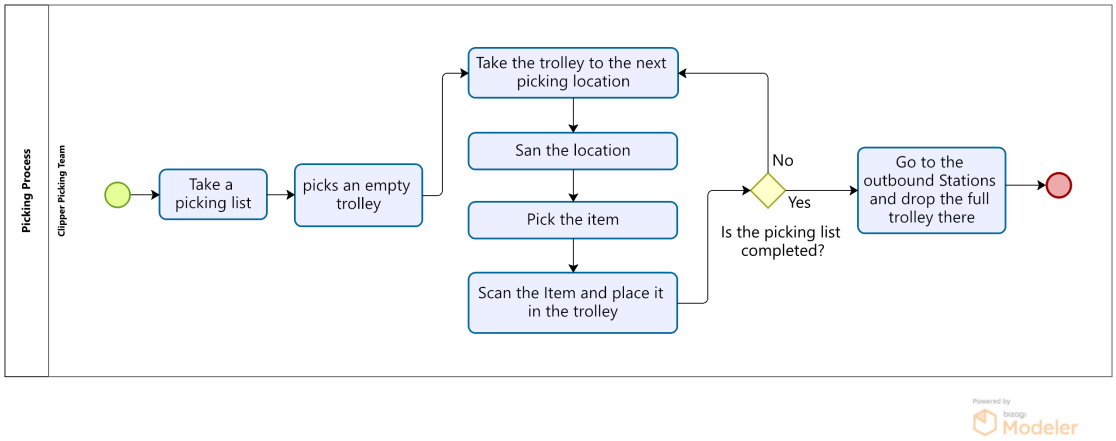


Figure 4.8: Simplified representation in a swimlane diagram of the Picking process.

4.3 Proposed Operational Changes

The first operational change to be analyzed with the model is related to how the Put-Away waves are carried out. Instead of placing items, one by one in bins previously placed on the shelves, the proposed solution is to carry in the trolleys bins full of items and place them on the shelves. The advantage is that, by visiting one location, several items can be stored. As a result, instead of three scans per item, it would be necessary to perform three scans per bin. Taking into account that a bin can contain a certain number of items, depending on their category and volume, the number of scans and also locations visited is reduced considerably. As a result, it is expected that the Put-Away operation will become much more efficient as the same number of items are stored in considerably less time.

A first analysis of the proposed solution brings certain questions. Here are three of these questions:

1. How should the locations for each bin be determined?
2. Should, in a different operation, empty bins be removed and transported back to the inbound stations, or, instead, should they be swapped for full bins?
3. Is it expected a worsening of the effectiveness of the used space and lack of available loca-

tions, as new bins require the place of completely emptied bins? Does it justify a new task of transferring items between bins to clear out the content of some and leave others full?

The first question arises because currently, even though the category locations are respected, Material Handlers search visually for available locations where they can store the items. This procedure is not suitable for a bin-by-bin Put-Away considering that usually is not possible to see whether the bins are empty or not unless they go to the specific location to check it out. For this reason, the system must indicate to the operators a suitable location to store the next bin. Presumably, and with respect to the category locations, this should be done through an ABC analysis.

The second question is related to the need to release locations so a full bin can be placed and is also related to the movement of empty bins back to the inbound verification stations.

Presumably coming back with empty trolleys is always worse than coming back with empty bins in the trolley which is handy given that both bins and trolleys are needed to be filled for the next wave. Notwithstanding, another operational alternative exists. For instance, the combination of picking and Put-Away operations can be analyzed: the same operator in the same wave would perform Put-Away while picking other items in nearby locations. This is not compatible with bringing back the empty bins. In that case, another operator would have to go around picking empty bins. Depending on the type of bins, empty bins may have the particularity of occupying less volume when they are empty either because they can be folded or because they can be piled up, one inside the other. Along with this, the operation of bringing back empty bins could be combined with the operation of transferring items between bins so locations where full bins can be placed are released. This aspect is also related to the question three.

This business case cannot easily be solved by analytical methods for several reasons. Even though the impact of reducing the number of scans could be estimated by a simplified method, the impact of changing the locations, particularly considering that down the line it also impacts the picking process (which in turn impacts back the Put-Away process) would not be accessed. In addition, the increase of the problem complexity due to other issues that become apparent as the analysis proceeds, makes simulation the only option viable to fully analyze the case. Besides, by doing so, new problems and new opportunities for improvement are identified.

5 Model Presentation

The present chapter is dedicated to the presentation of the model developed within the scope of this work. The model, as presented, was considerably restrained by the limited computational resources available throughout the development phase. As a matter of fact, it is expected that a powerful computer will be completely set up for further developments just after completion of the present dissertation and, at that point, all the expected features will be fully implemented. Besides, the model is also intended to be maintained and continuously upgraded with the following objectives:

1. Incorporate increasingly more details, both to refine the modeling of operations already included and to incorporate more operations on the model to allow a comprehensive analysis of the Warehouse functioning.

2. Incorporate changes that might be implemented in the real world. Both, the ones proposed through analysis carried out with this model and any changes that might have been implemented for other reasons.

At this stage, the details included in the model are tailored for the business case proposed.

The first section is dedicated to the presentation of the main requirements and a prototype where the main features were tested. A brief explanation of the modeling approach is provided. The second section presents the definitive model representing the Venray Warehouse and its operations. If the prototype were not developed first, the risk of having a non-working 3D model was high.

5.1 Prototype Model

Aligned with the proposed methodology, a working prototype model was developed. This prototype was based on a scenario as simple as it could be and, through an iterative process, more features were added, and existing features were further enhanced with additional details. The construction of a realistic scenario, that accurately portrays the Warehouse under study requires a substantial investment of time and effort, and should only be done after the main aspects of the model's logic have been figured out.

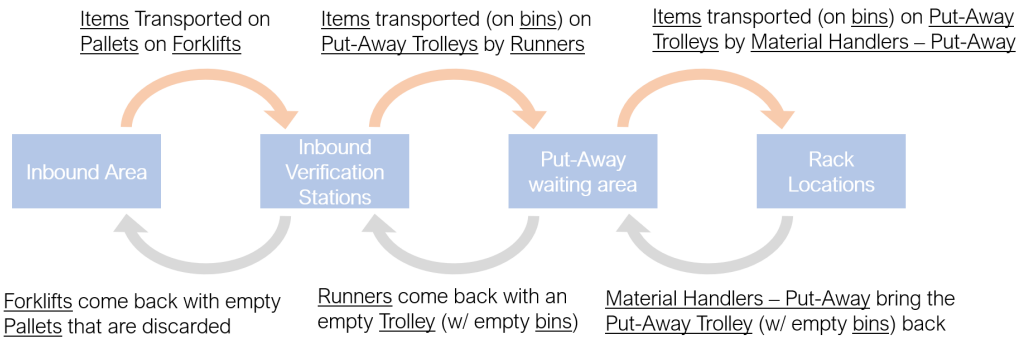
The model is mostly developed with Anylogic simulation software. This software provides well-suited modeling tools for various industries including Supply Chain, Manufacturing, Transportation and Warehouse Operations (Anylogic, 2023). It is important to understand that one

important advantage of using Anylogic is that it is reasonably easy to use as a modeling tool even for business analysts who, though often somehow familiar with programming, are not professional developers. This ease of use is mostly due to the fact that Anylogic features visual programming where logical blocks can be used to model a sequence of tasks that are part of a process. The behavior of agents that might pass through these flowcharts, can be complemented with Java code intended to be executed at certain circumstances. For instance, when an agent enters a flowchart block or when he changes its state. Agents in Anylogic are a Java class specialization and, as such, can have their own methods and variables. For more general purposes, Anylogic offers the possibility of creating Java classes (that are not agents).

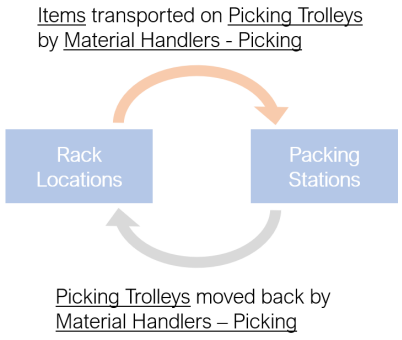
For successfully building an Anylogic model it is fundamental to find an equilibrium between resourcing to few logical blocks that can collectively represent a complete process involving several agents. Or, in contrast, do the same with other blocks or even lines of code that represent smaller steps of the same process. The latter approach offers greater flexibility but comes at the cost of increasing development time. Thus, when creating a model using Anylogic, it is crucial to determine the optimal approach in two dimensions: the granularity of the building blocks used and how they shall be assembled.

For the model development within the scope of this work, the key implementation challenges were related to the material flow in the Warehouse that is intended to be accurately simulated as the proposed business case (and also future ones) are related with the efficiency and effectiveness of processing such flow. Figure 5.1 represents schematically the flow of material that needs to be modeled with adequate level of detail. Rectangles represent different locations and elements that move between these locations are associated with the respective arrows. Starting with the Inbound (Figure 5.1a), while some elements move cyclically between consecutive locations, others move forward to the next one. Starting in the Inbound Area and moving to the Inbound Verification Stations, Items are transported on Pallets by Forklifts. While Items move forward from that point, the Forklifts go back with the Pallets. The former repeats this cycle while the latter is discarded. Between the Inbound Verification Stations and the Put-Away Waiting Area, Items are transported forward (inside bins or not) on Put-Away Trolleys by Runners. Runners then come back with another Put-Away Trolley that is empty (or with empty bins inside). From here to the last destination, Material Handlers - Put-Away transport Items (that can be inside bins) on Put-Away Trolleys to the Rack Locations, and come back with the same Trolley (that can have empty bins inside) with no items. As regard to the outbound (Figure 5.1b), the problem at hand is, in general terms, similar but less complex as the flows of different elements happen all between two abstract locations: Rack Locations and Packing Stations.

For general cases, such as, for instance, a set of pallets that need to be stored in a rack, the traditional way of modeling would be to consider that the pallet is an agent that requests, from a resource pool, a forklift to do the job. In this case, pallets wait for an available resource and, each time that a forklift finishes the transport of one pallet, it goes for the next one. Such implementation would not be appropriate for this model as the agents involved in the material flows have more complex and intricate behaviors. For exemplification, Figure 5.2 shows the Anylogic flowcharts of two implementations: A - the traditional implementation, B - the adopted implementation.



(a) Movement of elements - Inbound.



(b) Movement of elements - Outbound.

Figure 5.1: Movements of different elements in the simulation, Inbound and Outbound.

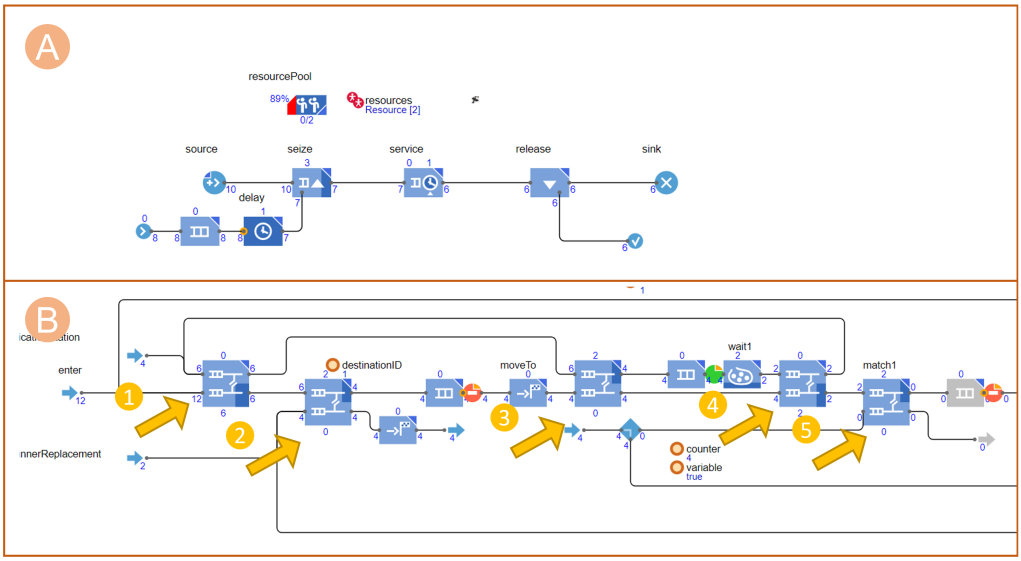


Figure 5.2: Comparison between two implementation strategies.

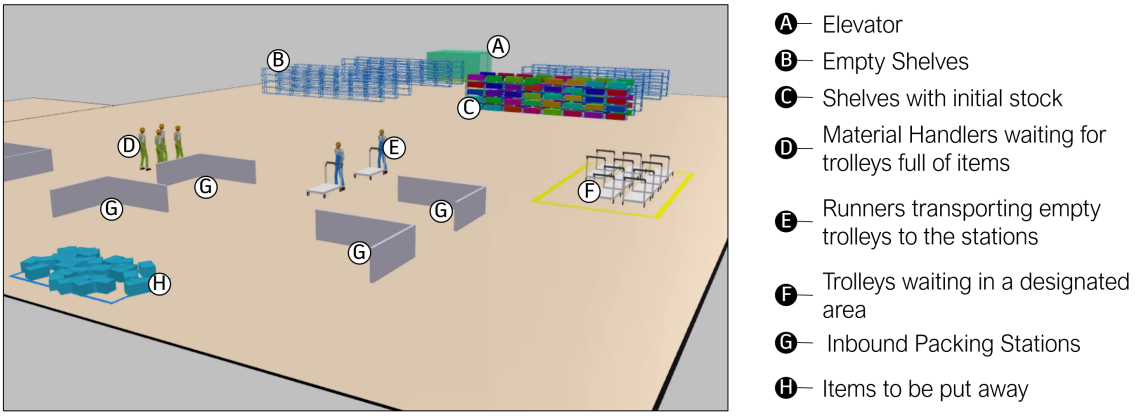


Figure 5.3: First floor of the prototype model

Implementation B corresponds to the material flow from the Inbound Verification Stations to the Put-Away Waiting Area. Instead of using seize/release blocks, more general match blocks are used. The first five of these blocks have the following purpose:

1. The Inbound Verification Station requests an empty Trolley (either the Station waits for an available Trolley or a Trolley waits for an empty Verification Station).
2. A Trolley that has been requested requests an available runner.
3. A Trolley reaches the Inbound Verification Station. The Inbound Verification Station can start transferring items from the pallet to the Trolley and the Runner is released.
4. The Inbound Verification Station releases the Trolley.
5. The released Trolley waits for a Runner to carry it to the Put-Away waiting Area, or Runners wait for Trolleys to be carried out.

The alternative A, showing the traditional implementation technique, is shown for comparison purposes. The inappropriateness of the first option is further exacerbated when considering the subsequent steps that follow those depicted.

All the challenging elements of the model were first developed and tested in a prototype. Figures 5.3 and 5.4 show the last version of the prototype.

Besides the modeling of the general flows of material, attention was also given to details such as the modeling of placement of items on the shelves and the movement between floors. With regard to the placement of items on the shelves, the key points are:

- A Worker with a Trolley stops in front of the designed location.
- A Worker places the item on the designated shelf.

Conversely, with regard to movements between floors, the key points are:

- A Worker knows when to change floors and thus moves toward the elevator.

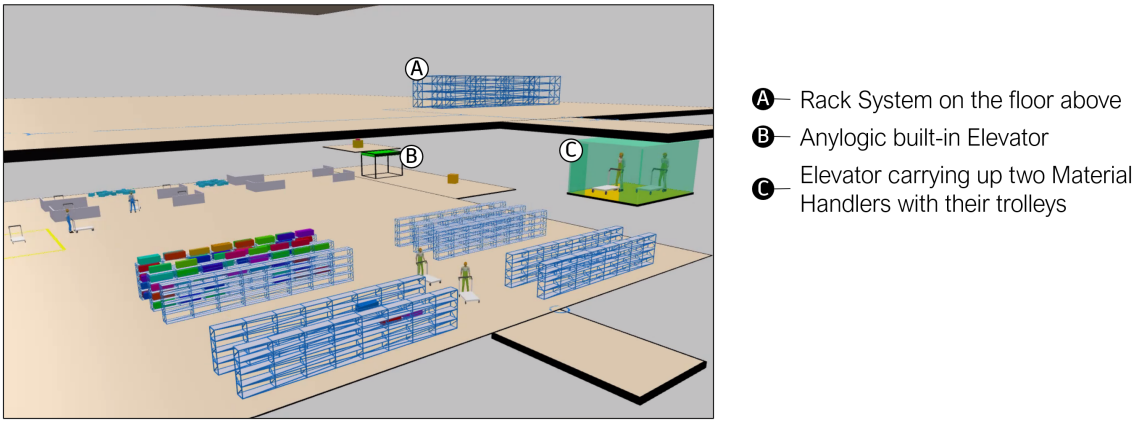


Figure 5.4: Elevator side of the prototype model.

- A Worker waits for elevator to arrive, lets people inside to get out, and then gets in.
- Workers get out after the elevator stops in the destination floor.

Figure 5.4 also shows the Anylogic built-in elevator that was discarded as a possible solution for movements between floors as this elevator requires the traveling agent to be carried out by a transporter. In addition, its behavior is not what is desired: it responds to requests individually and can carry only one item (with its transporter) at time.

Another important feature tested in the prototype was the Python connector that allows the Anylogic model to run a Python environment to which code, defined as string in Java, can be injected allowing for the exchange of information between the Anylogic model and the Python environment. Python modules saved in Python files can also be executed through this environment. There are two main reasons for the interest in this package. Firstly the simulation model deals with a considerable amount of input and output data that requires preprocessing or postprocessing. The basic alternative would be to connect the Anylogic directly to an Excel or text file and, in a different stage, perform the required operations. Another reason is that, during run time, data manipulation is also necessary. For instance, the conversion of Warehouse Location Code to Anylogic locations (identification of the rack, bay in the rack and shelf in the bay, along with the identification of the node in front of it) though, can be done in Java, is more easily coded in Python.

5.2 Scenario

The 3D scenario of the model is a key component of it. It is intended to be a general-purpose scenario to be used for more simulations beyond the ones within the scope of this work. Therefore, it is essential to represent the building comprehensively and at scale. However, the level of detail achieved can be prioritized according to the specific needs of the case studies on the pipeline. To match the needs of the first business case, Pick Tower, the Elevators Inbound Stations and related spaces and walkways have more detail than other features such as the High Value Cage or the Pack Stations. The lower level of detail for elements not relevant for the simulations to be run, besides allowing shorter development times also contribute to economizing computational

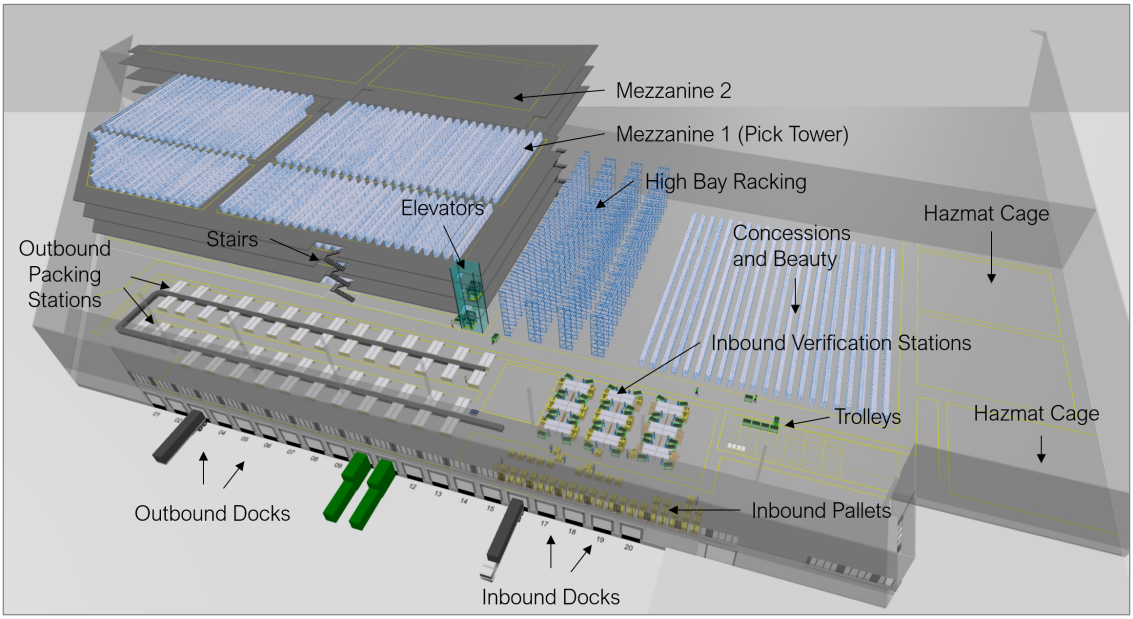


Figure 5.5: Screenshot of the 3D Model.

resources. Figure 5.5 shows an overview of the 3D model where the elements identified previously in Figure 4.2 and in Figure 4.5 are once again identified for comparison.

The accuracy of the measures of the elements' dimensions in the 3D scenario is important for the accuracy of the results. Nonetheless, the importance of the 3D model goes beyond the mere modeling of the distances between elements, it also allows the user to visualize the simulation as it runs. Visualization is important to ensure the model behaves as it should and to analyze operations searching for faults or opportunities for improvements. For these purposes, more than being "photorealistic", the visualization shall be informative. Figure 5.6 shows three screenshots of the modeling demonstrating information displayed on 3D elements.

5.3 User Interface

When initializing a simulation, a start screen where the user can make a number of options is shown. Figure 5.7 shows a screenshot of that start screen. It allows for choosing what solution to simulate (As-Is vs. To-Be) and also to choose parameters such as the number of Inbound Stations, Runners, Material Handlers Put-Away, or Material Handlers Pickers.

During run time it is possible to choose the view, logical view or one of the cameras. To better visualize what is happening on different floors, it is possible to turn floors invisible. Alternatively, it is possible to navigate through the 3D scenario inside each floor. To facilitate this functionality a camera with controls was created. Figure 5.8 presents the four views that can be accessed by clicking in each of the buttons on the menu bar.

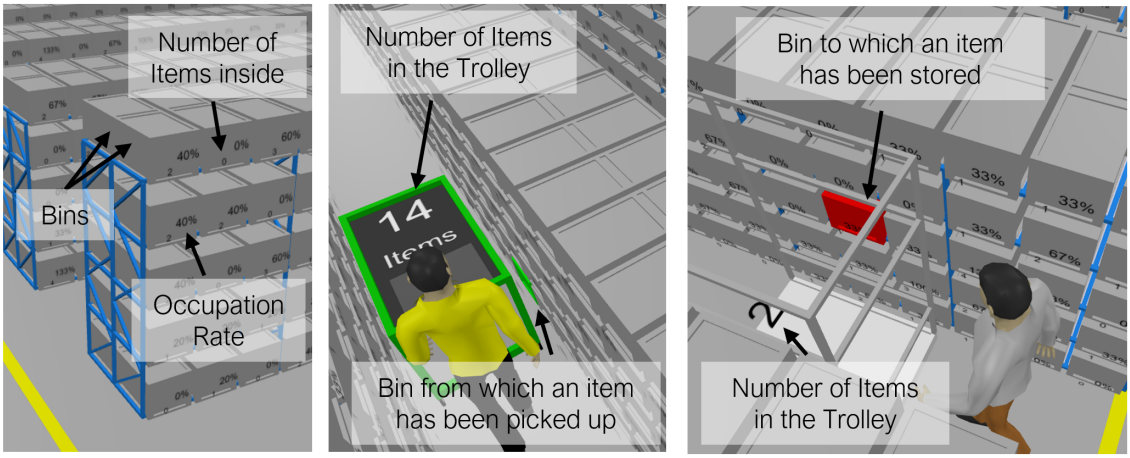


Figure 5.6: Information displayed on 3D elements.



Figure 5.7: Simulation Start Screen.

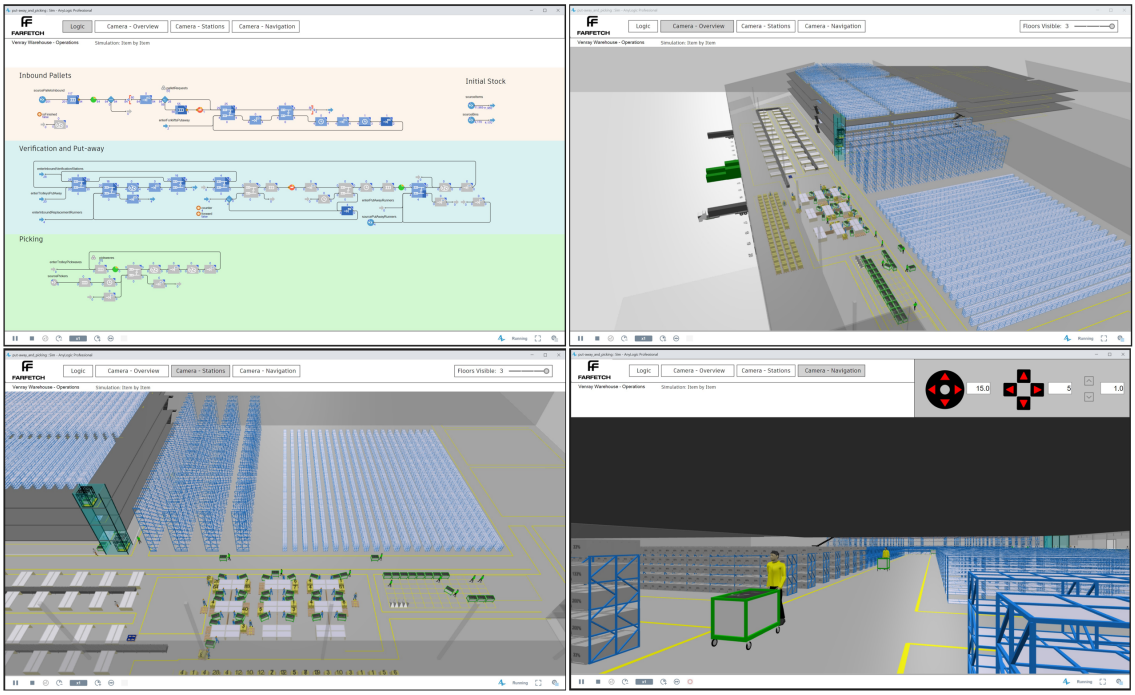


Figure 5.8: Mosaic of all four views with direct access through the buttons on the menu bar.

5.4 Agents

For the case study within the scope of this work there are four main types of workers each with specific roles and thus unique behaviors that need to be modeled, namely Forklifts operators, Runners, Material Handlers - Put-Away, and Pickers. While in the real world, workers may have more or less detailed procedures, in the virtual world, agents do strictly what they are instructed to do. Figure 5.9 shows the flowcharts defining the processes executed by Runners, Material Handlers - Put-Away, and Pickers.

5.5 Model Architecture

The development of the model encompasses creating visual elements, coding component behaviors, and conceptualizing how the different components should be implemented and their interactions. In addition, defining the procedures to be followed when using the model is also part of the model development. The architecture defining the model in these different aspects is represented in Figure 5.10 by an ArchiMate diagram. The model architecture crosses three layers namely, Business, Application, and Technology.

As Figure 5.10 shows, the model relies on two components coded in Python: “Python Module - Data Preparation Simulation” and “Python Module - Run Time”. The purpose of the first one is to help in the data preparation process. In this process data is cleaned (invalid or not understood data entries are corrected or removed), filtered (if only part of the transactions should be included in the model) and the resulting data is stored in a lighter file. This process doesn’t need to be repeated

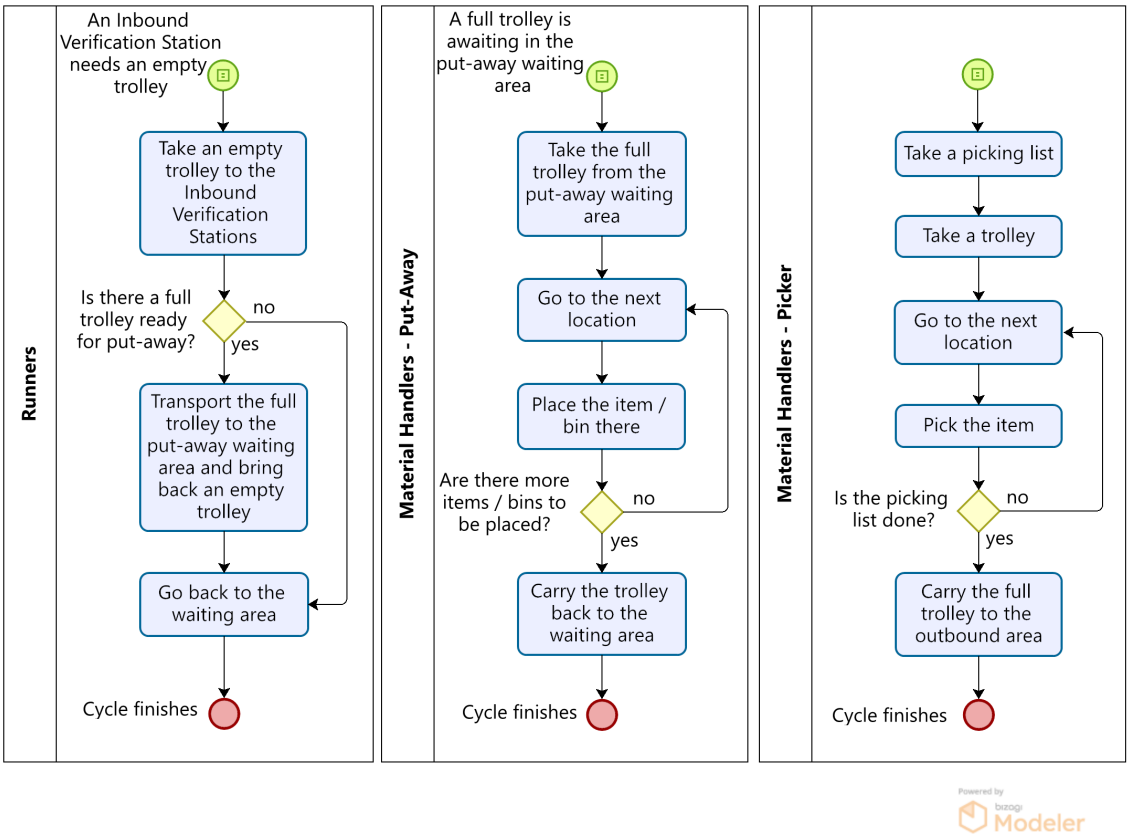


Figure 5.9: Processes of the three main agents.

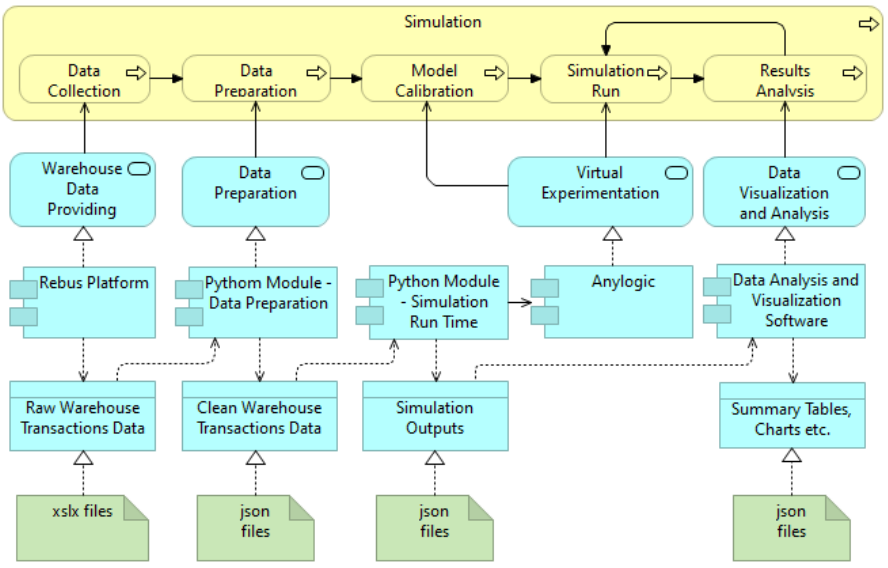


Figure 5.10: Archimate model diagram including Business (yellow), Application (blue) and Technology (green) layers.

for different simulations that have the same input data. The second step is model calibration. The purpose of the calibration is to set the parameters' values, so the model represents reality as accurately as possible. This process certainly doesn't need to be repeated for each simulation that is carried out. Recalibration is only needed when there are reasons to believe that the correct parameters' values might have changed. For instance, for simulations from a scenario that happens in a different season or when changes to the 3D model have been made. After making sure the model is properly calibrated, it can be used for simulation runs and results analyses. These last two steps can be done iteratively when the user is designing new solutions.

For consulting the definition of all agents, excluding Java code, Appendix A can be consulted.

5.6 Calibration

Being the first business case to be tackled focused on the Put-Away wave, the parameters that impact performance are those related to the time required to perform that task. For this model, these parameters are traveling speed and the time it takes to put an item inside the bin (time stopped at each location). The real-world data are time stamps attached to information describing the event registered in that data entry. In the calibration process proposed, the time elapsed between time stamps is the measure that the simulation model aims to approximate by minimizing the Mean Average Percentage Error (MAPE) of these measurements. The choice of measures as granular as the elapsed times between consecutive events, instead of more aggregate ones such as workers daily productivity, is made with the view of providing more credibility to the model and avoid other problems such as undetected redundancy on the parameters. For instance, if a task always takes the same three steps, each with its own parameter representing the required time, and the model only consider the overall time, it will arbitrarily attribute values to these three parameters only assuring that their sum is correct. This granularity is also useful for detecting opportunities to introduce more parameters such as considering different speed values for different types of pathways.

To understand the challenges associated with the calibration process for this model it is crucial to be familiar with the flow of items from pallet to shelf. Figure 5.11, highlights the main stages of the referred flow that is important to be aware of. It shows a screenshot of the model with a view of the Inbound area and its surroundings. The steps in the item's flow that require attention are identified with numbers.

The items start their journey in the pallets according to real data. Thus, items that arrived on the same pallet in the real-world, also are placed in the same pallet in the simulation environment. Then, pallets are transported to a station that has requested one, either because there is not any pallet there or because the pallet that is there is already empty (step 1). Secondly, items are taken, one by one, from the pallets to the Trolley that is in the station (step 2). After being fully loaded, the Trolleys are transported to the locations where the items are dropped.

Because simulation doesn't replicate all details from reality, as the items get ahead in their flow, discrepancies accumulate. Starting with the pallets' composition, normally because of data errors or undesired entries, items are filtered out thus the simulated pallets may not carry exactly



Figure 5.11: Items flow in the Inbound Process.

the same content as the real ones. In addition, in the simulation pallets are queued and transported to the first station that has made a request which may coincide with the station to where the pallet was transported in the real world. As a result of these discrepancies, the Trolley content is often (partially) different from reality. To cope with this issue, before any data filtering (or for that matter, before any processing), waves are identified in the dataset and items are numbered according to their order within the Put-Away wave. Thus, during simulation, each time a time difference between events is calculated it is also determined if these events were also consecutive in real the real world to access whether the simulated time elapsed has a comparable real value. Figure 5.12 aims to help explain this strategy by showing a screenshot of the model where an Agent was performing Put-Away along one aisle. For exemplification, it can be considered that, in the real world, there were five items placed in their respective storage locations. Yet, one of these was filtered out from the dataset. In the simulation, the Trolley stops in P4 and determines the elapsed time P2-P4. However, given the order number associated with these transactions, it acknowledges that they were not consecutive in the real world. As a result, it does not include this elapsed time in the calibration process.

Among the values that can be considered for the calibration, only part of them are used for training while the remaining are used for testing. Thus, for a data split 70-30, each valid entry receives 70% probability of being considered for training. In the end of the regression, it is possible to know the error on the training set and on the testing set. Figure 5.13 shows schematically the data split for calibration.

The accuracy on the test set is what matters to access the accuracy of the model. To understand whether the model is accurate enough or not it is recommended that an aggregate value linked to the measurement under study is used. In fact, it is normal for very granular measurements to have errors as much as 60% and still, for more aggregate values such as daily productivity the error

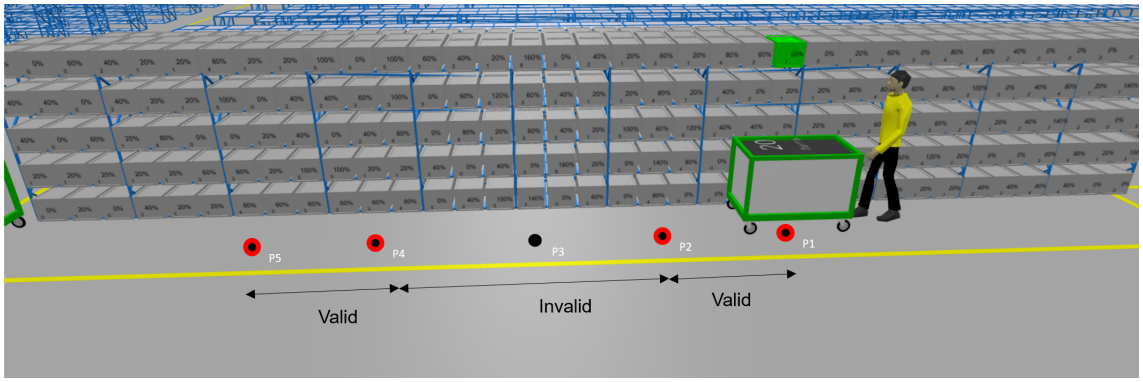


Figure 5.12: Valid and invalid elapsed times.

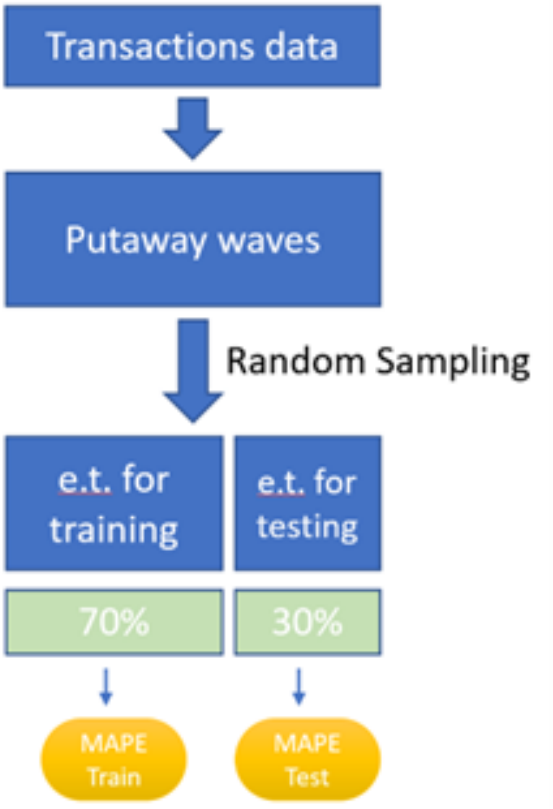


Figure 5.13: Data split for calibration purposes.

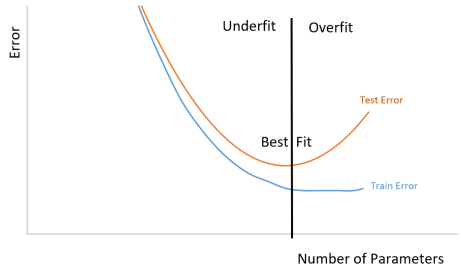


Figure 5.14: Ideal number of parameters.

might be as low as 5%.

Granular parameters that present apparently high errors may also represent an opportunity to increase the number of parameters and thus improve model accuracy. For instance, if the time it takes to travel between positions within an aisle has a high MAPE, it may go lower by considering different parameters for different types of aisles, such as large aisles and thin aisles. If these granular parameters already present low MAPE it is possible to conclude that, even if there are in fact differences between aisles, there is not much to gain by splitting them apart.

When increasing the number of parameters, though the fit may keep improving, the model may start to become overfitted. Figure 5.14 shows what should be expected when the number of parameters is increased. Thus it is important to add progressively the number of parameters to find the suit spot.

5.7 Final Remarks

This chapter presented the Digital Twin (DT) of the Venray Warehouse including its 3D scenario and part of its operations. The model, including all graphical elements, and all Python and Java code, was developed completely from scratch. This model was built at scale based on blueprints and photographs. The Inbound operations were modeled based on internal documents and information provided by colleagues. Throughout the process the model was continuously visually and conceptually validated by experienced company professionals familiar with the Warehouse and its operations and Calibration relied on real data obtained directly through the platform used by the Third Party Logistics (3PL).

6 Results and Discussion

The purpose of this chapter is to present the results obtained for the business case. These results are not definitive as the model was not fully implemented. As expressed in the Chapter 3, it is important to always have a working version of the model throughout its development. At the end of the period of this dissertation the model version is mainly limited by the computational resources available. Nonetheless, the outputs obtained are important to make sure that the model is working as intended and the changes and features on the pipeline are feasible. Furthermore, the results also enable an interim evaluation of the viability and effectiveness of the proposed operational changes. If the preliminary results weren't encouraging enough, it would be better to reformulate the business case or, in an extreme case, to completely reconsider its suitability and perhaps initiate the study of another one, possibly using this same model.

6.1 Scope and Assumptions

The analysis is restricted to a specific area of the Warehouse and encompass only two categories: footwear and accessories. Figure 6.1 shows the schematic layout of the Product Categories locations in the floor zero of the Pick Tower, already presented in Figure 4.4, where the area considered is identified with a red line. Figure 6.2 situates that area of the Warehouse in the 3D scenario and identifies the pathways followed by the agents involved.

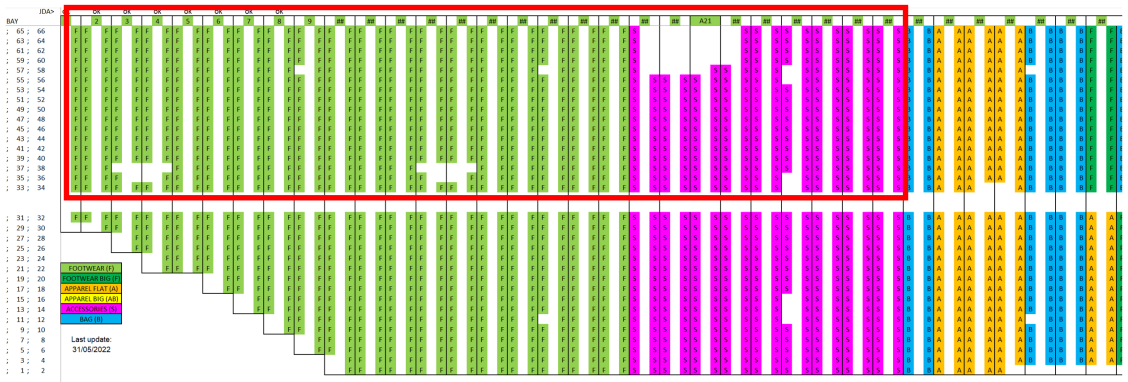


Figure 6.1: Schematic layout of the Product Categories locations in the floor zero of the Pick Tower with the identification of the area under study. Schematic courtesy of Farfetch.

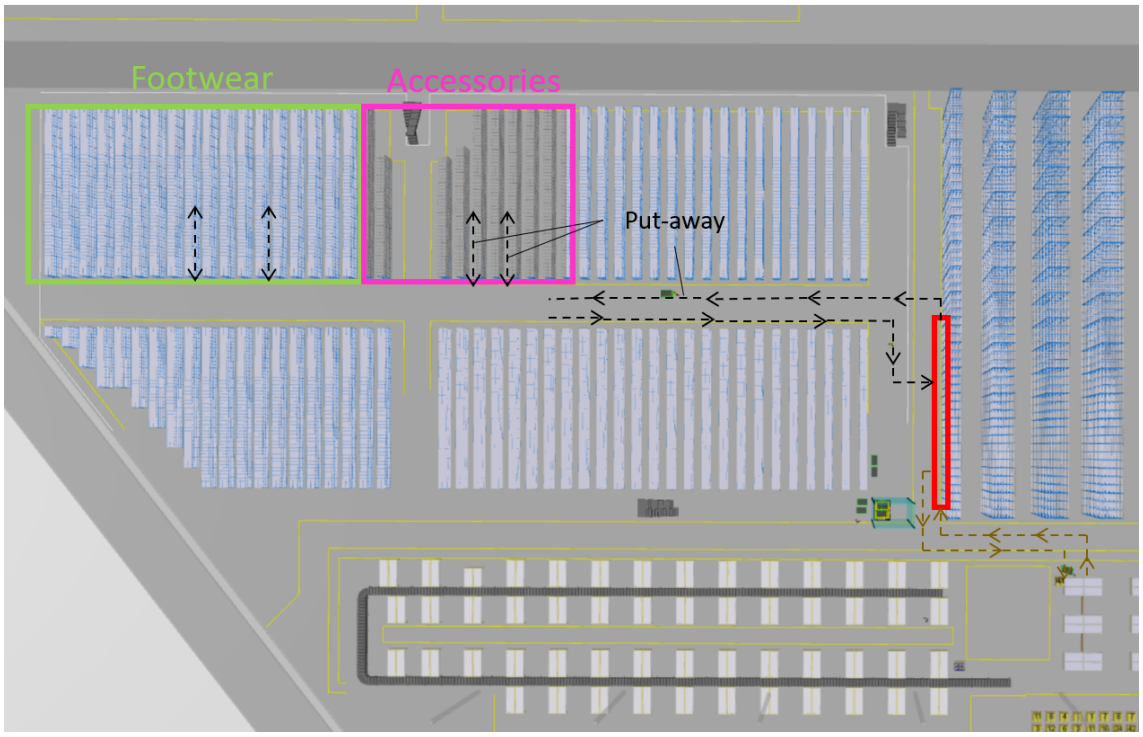


Figure 6.2: Storage zones covered by the simulations and the pathways where the agents move.

For the simulations, the following assumptions were made:

1. The number of Items in the trolleys is the same in both cases (As-Is and To-Be) and depend on the category.
2. Trolleys always carry six bins (To-Be).
3. Stopping time for placing an Item into the bin (As-Is) or for bin swapping (To-Be) is the same and is equal to the value obtained from the calibration process (1.945s).
4. The walking pace of all agents, in all scenarios (As-Is and To-Be) is constant and equal to 1.539m/s, which was also obtained from the calibration process.
5. For locations system-driven determined, the criteria are:
   (a) Nearest aisle;
   (b) Position in the aisle closest to the entrance.
6. Bins have the estimated capacity for each category regardless of the Stock Keeping Units (SKU) inside.

It is also important to refer that the calibration process, so far implemented, performs regression without splitting the data. Introducing changes in the model at this stage of development with the limited resources available is a low-productivity task and thus was, as much as possible, avoided. Nonetheless, given that the data sets used are not too small and only two parameters were estimated, it is reasonable to assume that the values thus estimated are not invalid at all. It is also

important to mention that, once again due to the limited capacity of the computer used, the two categories were simulated separately and the parameters, estimated only considering the footwear category, were the same for both cases.

6.2 Results

Three experiments were carried out for each category. In the first one, the As-Is scenario, representing current Warehouse operations, was considered. For the second experiment, one change was introduced to the As-Is scenario: the locations where Items are stored instead of replicating the real-world, became system-driven, according to the criteria defined in the assumptions. For the third experiment, another change was introduced on the previous experiment: instead of storing Items one-by-one, bins full of Items are placed on the shelves thus storing several Items at once. Real data regarding the Items received and stored were considered for both categories. For footwear, one day of data containing 1791 Items that required 100 simulated waves to be stored, was considered. For Accessories, as the zone considered is smaller and the volume is usually smaller for this category, one month of data was taken. In total the dataset contains 2534 Items of this category that require 85 simulated waves to be stored.

Table 6.1 expresses the results obtained for each experiment in each category. Total time represents the combined duration of all Put-Away waves, in seconds, necessary to store all the Items considered. This time is broken down into "Time Stopped" and "Time Moving" which are expressed in both absolute values and as percentages of the "Total Time". The impact on "Time Stopped", "Time Moving" and, on the "Total Time" are also expressed both with reference to the previous experience (Marginal) and as percentage points, with reference to the As-Is values (Percentage Points). The overall impact on each category is also shown.

Table 6.1: Results of "footwear" and "accessories" categories.

	Footwear Category			**Accessories Category**		
	As-Is		**To-Be**	**As-Is**		**To-Be**
	by item	**sp + by item**	**sp + by bin**	**by item**	**sp + by item**	**sp + by bin**
Total Time	20118	16097	14131	26589	13460	8996
Time Stopped	3489	3489	1163	4936	4936	988
Time Moving	16629	12608	12968	21653	8524	8008
Time Stopped	17.34%	21.67%	8.23%	18.56%	36.67%	10.98%
Time Moving	82.66%	78.33%	91.77%	81.44%	63.33%	89.02%
no. items	1791	1791	1791	2534	2534	2534
Marginal						
Time Stopped		0.00%	-14.45%		0.00%	-29.34%
Time Moving		-19.99%	2.24%		-49.38%	-3.83%
		-19.99%	-12.21%		-49.38%	-33.16%
Percentage Points						
Time Stopped		0.00%	-11.56%		0.00%	-14.85%
Time Moving		-19.99%	1.79%		-49.38%	-1.94%
		-19.99%	-9.77%		-49.38%	-16.79%
Total Impact			-29.76%			-66.17%

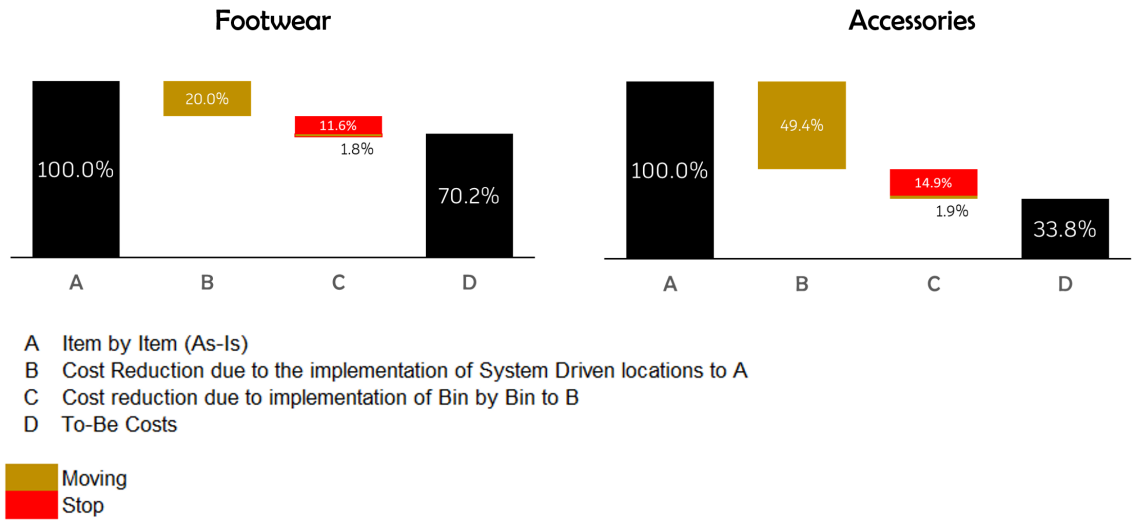


Figure 6.3: Preliminary results for the categories "footwear" and "accessories" - graph.

These results, which refer exclusively to the labor-time spent on the Put-Away waves, are also presented on the two graphs shown in Figure 6.3. For each graph, dedicated to each category, the first and last columns represent, respectively, the As-Is and To-Be total time spent. Each column in the middle represents the impact of the corresponding operational change with the color, brown or red, distinguishing whether the impact was on the "Moving Time" or "Time Stopped", respectively. All values are expressed as percentage of the "Total Time" of the As-Is scenario.

6.3 Discussion

For the footwear category, only by applying the first change (from A to B), original locations to system-driven locations, an economy of 19.99% of the time spent was achieved. By changing from Item-by-Item to Bin-by-Bin, an extra 9.79% reduction was attained. From B to C, the time spent travelling, which might be expected not to change considerably, increased slightly. Overall, Put-Away operations in the To-Be scenario, for this category and restricted to the area under study, are expected to take 29.76% less time.

As far as accessories are concerned, the gains were considerably higher. The time economy due to the first change (A-B) was 49.38%. One reason for this result is that, the area of the category accessories considered is closer to the Put-Away area, where full trolleys are queued. As a result the paths followed are shorter leading to a smaller basis for the percentage value. This result may also be justified based on the fact that, for the accessories category, it might be harder for Material Handlers, without taking out the bin from the shelf, to check where is there available space. By applying the second change (B-C), the time-labor spent was reduced 16.79% further. In comparison to the footwear category, the improvement due to the second change was bigger because each bin for accessories can store up to five Items, while for the footwear category, only three are possible. Overall, the simulation model forecasts a reduction of 66.17% of the time spent in Put-Away operations for the accessories area studied.

Considering that the estimated time savings for the restricted area studied, which represent 5% to 10% of the overall Warehouse volume, are fairly representative of the whole Warehouse and taking into account average labor costs, annual savings well above six figures are projected. The proposed operational changes are thus very positive for improving the overall costs of this Warehouse.

7 Conclusion

The main objective of this work was to create a model of the Venray Warehouse and its operations, which accurately incorporate a high level of detail, to support continuous warehouse analysis, improvement, and optimization. A business case, intended to be further explored until final conclusions are drawn, was used to guide the model's development by supporting the choice of the details to prioritize. After this business case is closed, the same model can be used for the coming business cases which will rely on details already modeled and on new ones that might be necessary to incorporate. This is not only true for business cases related to the Venray Warehouse, but also for other business cases involving other warehouses, considering that, though the scenario is different, most of the processes might be the same.

In fact, as a consequence of this project, the methodology for analysis and design of operations followed by the company has evolved towards relying more on simulation. This methodology will be key in the time coming due to the increasing complexity of operations and the introduction of automation.

The concept of Digital Twin (DT) does not have a consensual definition and is seen by many as a buzzword. However, the developed model, incorporating several features often associated with DT, offers an additional use case that may contribute to the ongoing pursuit of a universal definition for this concept.

The methodology followed was aligned with the degree of novelty of this project for both the organization and the author. The chosen methodology also took into consideration that what is feasible or not might not be completely clear at the beginning. For that matter, the pilot model was crucial, not only allowed to determine that the main model features were feasible, but also provided valuable insights into the optimal implementation approach. The agile nature of the plan followed also provided a good deal of flexibility to cope with obstacles that appeared down the road. One interesting problem that appeared very late in the development of the project was the Java class limit for constants that was reached only by defining about half of the required nodes for floor zero. A solution to this problem was found soon after. Yet, the importance of always having a working model that is refined through an iterative process was evidenced as this kind of obstacle could completely prevent the project from being developed further.

Through the process of developing this project, a number of limitations become evident. One of them is that, though people carry out their work according to the processes designed, these designs don't specify every detail of that work. One example of that are the locations that the

Material Handlers chose to store items, is left to the worker criterion within certain restrictions. And, naturally, such a thing is not easily caught by a simulation model where people inevitably follow a much more mechanic behavior.

Another important limitation is related to the capacity of the bins and trolleys. For instance, accessory items might have a wide range of sizes, shapes, and weights. Still, this work considered that each bin would have a capacity for five items regardless of the Stock Keeping Units (SKU) involved, which is the average value according to the organization's documentation. Nevertheless, not considering its variation may impact certain simulation results.

Regarding the main tool used for the development of the model, if, on the one hand, it allows for fast modeling by providing artifacts that suit the needs of the problem under study, it is also limited in terms of the options for versions management and importing/reusing artifacts between projects. As the size and complexity of the model increase, this aspect may require some attention.

Computational resources also represent a key limitation for this project. However, for such projects, it is expected and also desirable that a limited computational budget does exist. In the same way that the model must be developed iteratively, the available resources shall also increase iteratively as their need and suitability are demonstrated and justified.

In the near future, the features that are waiting for appropriate computational resources are expected to be implemented. The most notable one is the full activation of all storage positions in the Pick Tower. Another must-have feature expected to be implemented soon is the data splitting for calibration, along with the assessment of the model's accuracy according to the model's intended purpose. Still related to calibration, for some complex cases, it might be important to implement distribution functions instead of fixed values for the parameters.

As far as business cases are concerned, within the scope of the first one, the next steps will be to consider more data representing longer time spans. Instead of one day or one month, it would be important to simulate several months straight, to better catch the consequences of the proposed changes over time. Another important measure would be to always compare the real metrics from the process's implemented changes and compare with what the model forecast. This is the ultimate evaluation on "unseen data" and provides a definitive answer to the accuracy of the model guiding thus its continuous development.

Regarding more technical aspects, it is worth considering changing the framework for future model versions so the model is instantiated from a main Java application. This change may be useful for applying machine learning techniques where the same simulation has to be run several times. Another useful modification would be to set up socket connections. These connections can substitute the Pyconnector currently in use, thus allowing for the model to be restarted without killing the Python environment. This change also opens doors for connection with other applications allowing, for instance, elaborate dashboards for real-time data visualization more sophisticated than those built in Anylogic.

As the model is expected to grow considerably in size and complexity, adding unit tests assuring the proper function of key features may provide an important support for large scale development, in terms of model size and teamwork.

Bibliography

Adam, J. (2022, November 2). *What is agile software development?* Retrieved March 17, 2023, from https://kruschecompany.com/agile-software-development/

Allen, B. D. (2021). *Digital twins and living models at nasa*. NASA. Retrieved April 12, 2023, from https://ntrs.nasa.gov/citations/20210023699

Anylogic. (2023). *Anylogic simulation software is used in these industries*. Retrieved March 7, 2023, from https://www.anylogic.com/

Banks, J., Carson, J. S., Nelson, B. L., & Nicol, D. M. (2005). *Discrete-event system simulation* (4th ed.). Pearson Practice Hall.

Barricelli, B. R., Casiraghi, E., & Fogli, D. (2019). A survey on digital twin: Definitions, characteristics, applications, and design implications. *IEEE Access*, *7*, 167653–167671. https://doi.org/10.1109/access.2019.2953499

Bartholdi, J. J., & Hackman, S. T. (2019). *Warehouse & distribution science*. Georgia Institute of Technology.

Bazaz, S. M., Lohtander, M., & Varis, J. (2019). 5-dimensional definition for a manufacturing digital twin. *Procedia Manufacturing*, *38*, 1705–1712. https://doi.org/10.1016/j.promfg.2020.01.107

Borshchev, A., & Grigoryev, I. (n.d.). *The big book of simulation modeling - multimethod modeling with anylogic 8*. AnyLogic North America. Retrieved April 20, 2023, from https://www.anylogic.com/resources/books/big-book-of-simulation-modeling/

Božić, V. (2023). Methodology of system dynamics. https://doi.org/10.13140/RG.2.2.27204.71049

Camarotti, A. (2020, December 10). *Forbes de Camarotti #8: José Neves, fundador e CEO da Farfetch*. Retrieved May 21, 2023, from https://www.youtube.com/watch?v=fxvRa7X_wfs

Chung, C. A. (2004). *Simulation modeling handbook - a practical approach*. CRC Press LLC.

Edwin E. Aldrin, J., Collins, M., Charles Conrad, J., Cortright, E. M., Gilruth, R. R., Chrisopher C. Kraft, J., Lovell, J. A., Low, G. M., Mueller, G. E., Newell, H. E., Petrone, R. A., Philips, S. C., Schmitt, H. H., Alan B. Shepard, J., Sherrod, R., von Braun, W., & Webb, J. E. (n.d.). *Apollo expeditions to the moon* (Vol. 13.1). NASA Langley Research Center.

EMR. (2023). *Global logistics market outlook*. Expert Market Research. Retrieved April 23, 2023, from https://www.expertmarketresearch.com/reports/logistics-market#:~:text=The%20global%20logistics%20market%20size,USD%2014.37%20trillion%20by%202028

Farfetch. (2023a). *About farfetch*. Retrieved April 10, 2023, from https://www.farfetchinvestors.com/our-company/about-farfetch/default.aspx

Farfetch. (2023b). *Farfetch 2022 annual report*. Retrieved April 8, 2023, from https://s22.q4cdn.com/426100162/files/doc_financials/2022/ar/2022-Annual-Report-(1).pdf

Farfetch. (2023c). Farfetch onboarding material - operations.

Ghiani, G., Laport, G., & Musmanno, R. (2004). *Introduction to logistics systems planning and control* (S. Ross & R. Weber, Eds.). John Wiley & Sons, Ltd.

Grieves, M. (2015). Digital twin: Manufacturing excellence through virtual factory replicatoin.

Langevin, A., & Riopel, D. (2005). *Logistics systems: Design and optimization*. Springer Science + Business Media, Inc.

Law, A. M. (2015). *Simulation modeling and analysis* (5th ed.). McGraw-Hill Education.

Liu, M., Fang, S., Dong, H., & Xu, C. (2021). Review of digital twin about concepts, technologies, and industrial applications. *Journal of Manufacturing Systems*, *58*, 346–361. https://doi.org/10.1016/j.jmsy.2020.06.017

Liu, X., Jiang, D., Tao, B., Xiang, F., Jiang, G., Sun, Y., Kong, J., & Li, G. (2023). A systematic review of digital twin about physical entities, virtual models, twin data, and applications. *Advanced Engineering Informatics*, *55*, 101876. https://doi.org/10.1016/j.aei.2023.101876

Macal, C. M., & North, M. J. (2009). Agent-based modeling and simulation. *Proceedings of the 2009 Winter Simulation Conference (WSC)*. https://doi.org/10.1109/wsc.2009.5429318

Madni, A., Madni, C., & Lucero, S. (2019). Leveraging digital twin technology in model-based systems engineering. *Systems*, *7*(1), 7. https://doi.org/10.3390/systems7010007

Negócios. (2018, September 21). *Farfetch consegue quase 900 milhões de dólares com o ipo*. Jornal de Negócios. Retrieved April 15, 2023, from https://www.jornaldenegocios.pt/mercados/bolsa/detalhe/farfetch-consegue-quase-900-milhoes-de-dolares-com-o-ipo

Rushton, A., Croucher, P., & Baker, P. (2014). *The handbook of logistics and distribution management: Understanding the supply chain* (5th ed.). Kogan Page.

Sharma, A., Kosasih, E., Zhang, J., Brintrup, A., & Calinescu, A. (2022). Digital twins: State of the art theory and practice, challenges, and open research questions. *Journal of Industrial Information Integration*, *30*, 100383. https://doi.org/10.1016/j.jii.2022.100383

A Model Artifacts

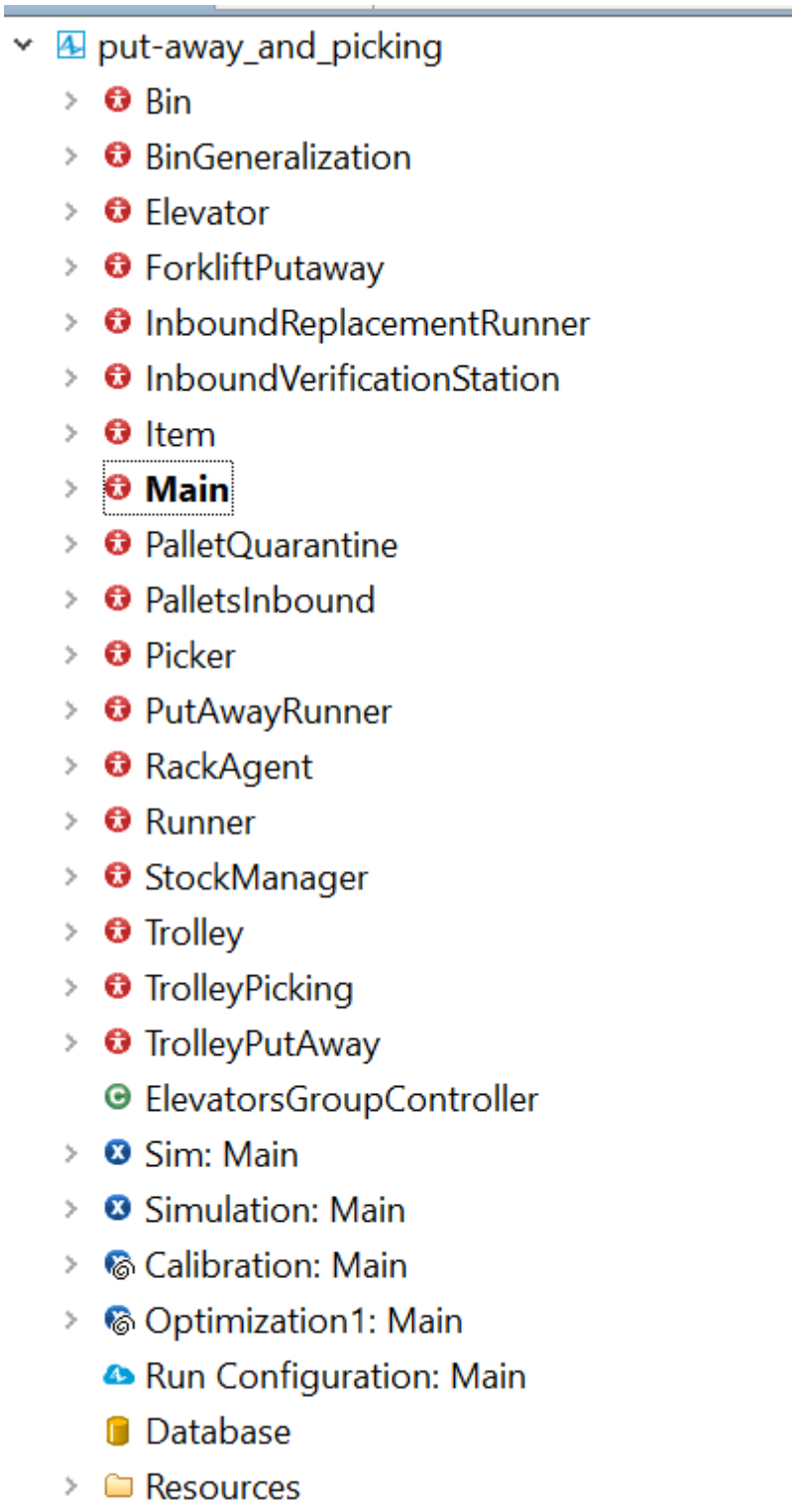


Figure A.1: List of Agents

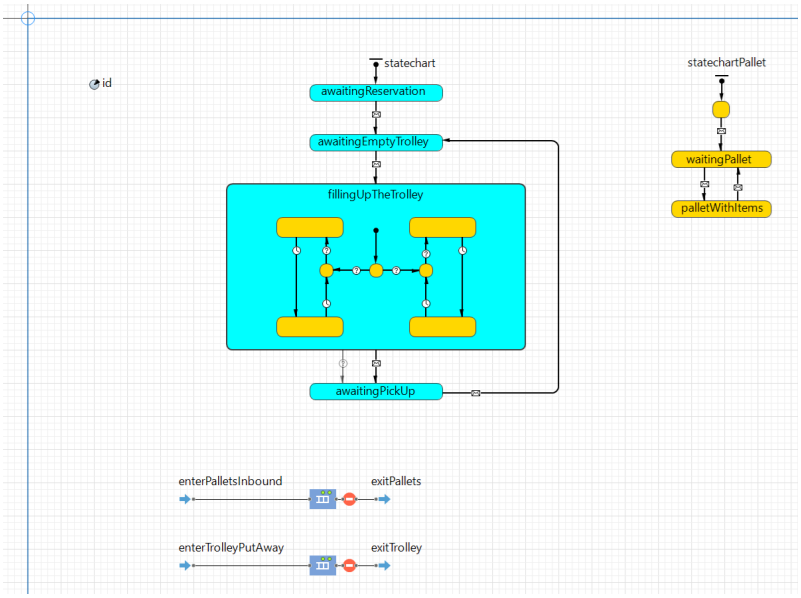


Figure A.2: Verification Station

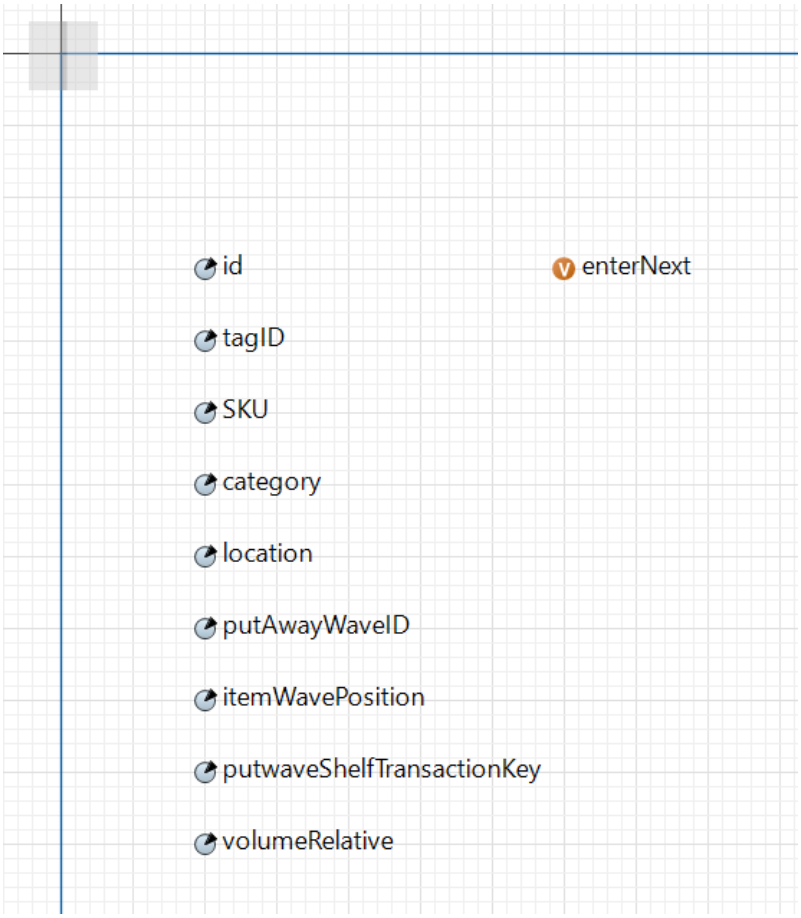


Figure A.3: Item

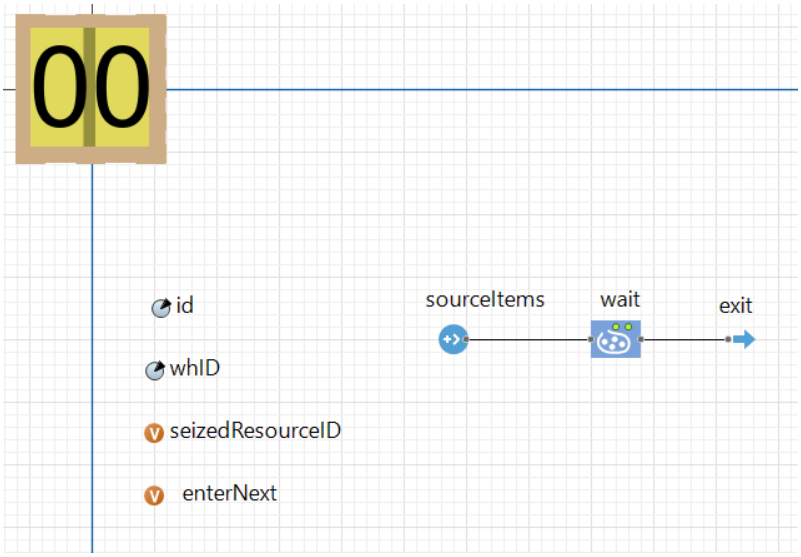


Figure A.4: Pallets Inbound

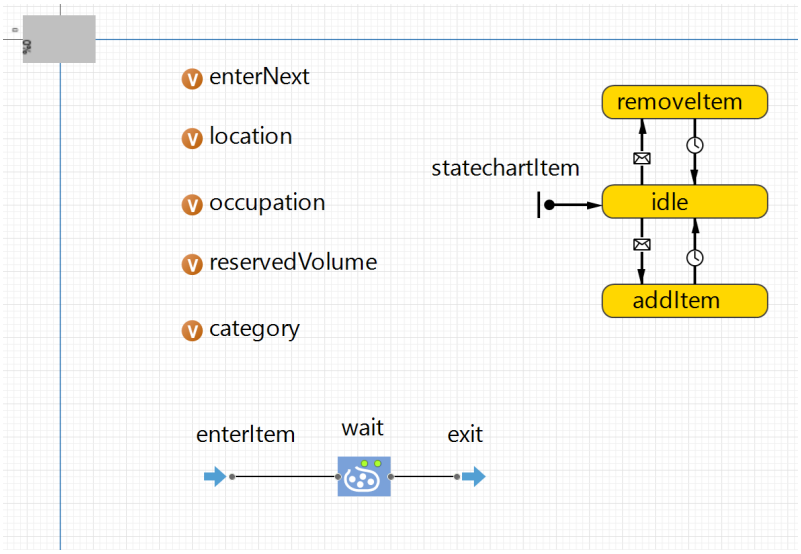


Figure A.5: Bin

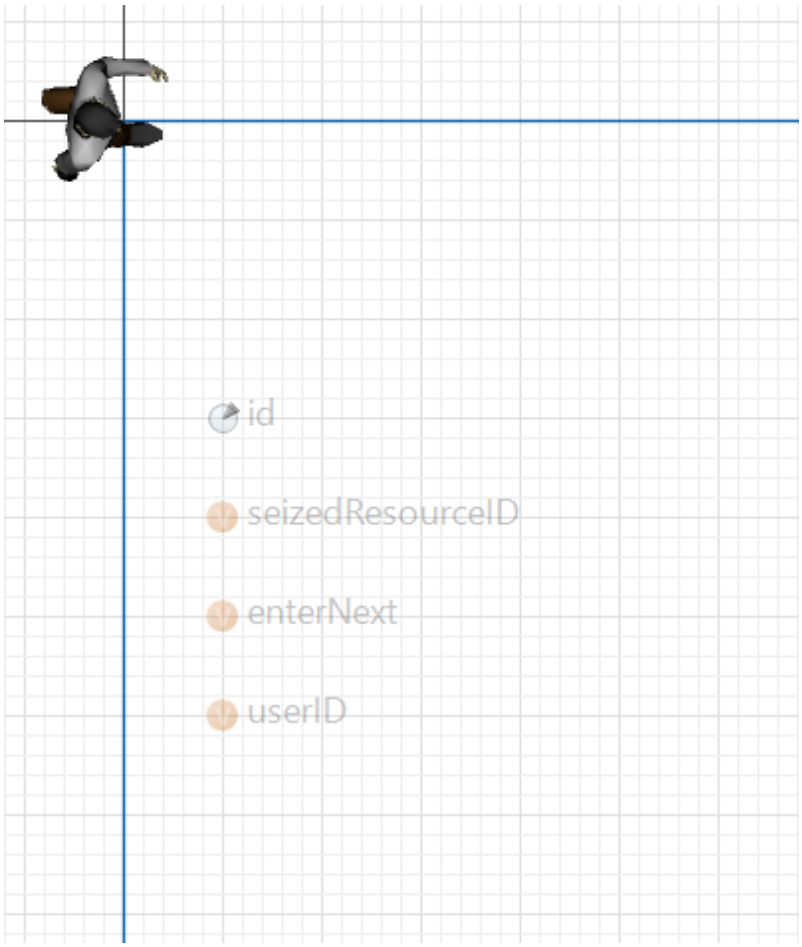


Figure A.6: Picker

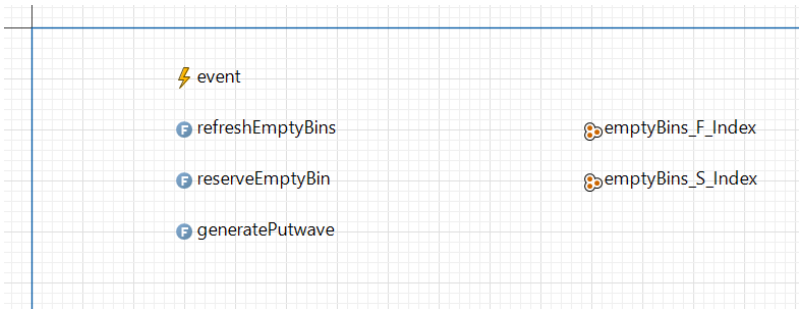


Figure A.7: Stock Manager (agent)

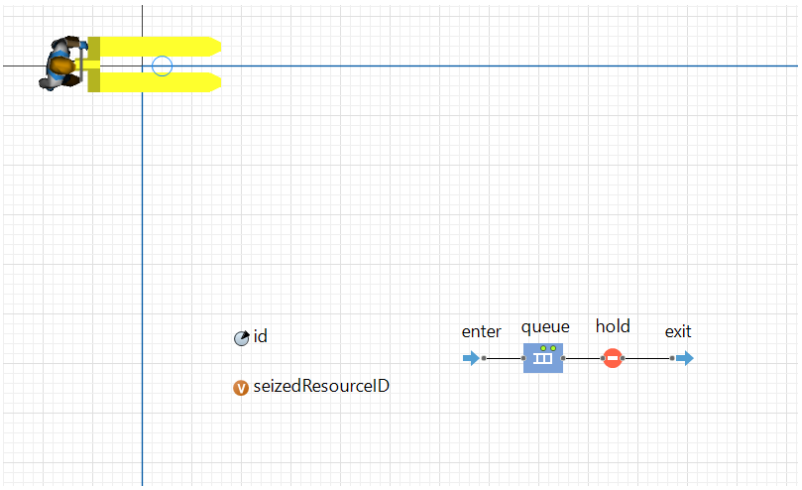


Figure A.8: Forklift

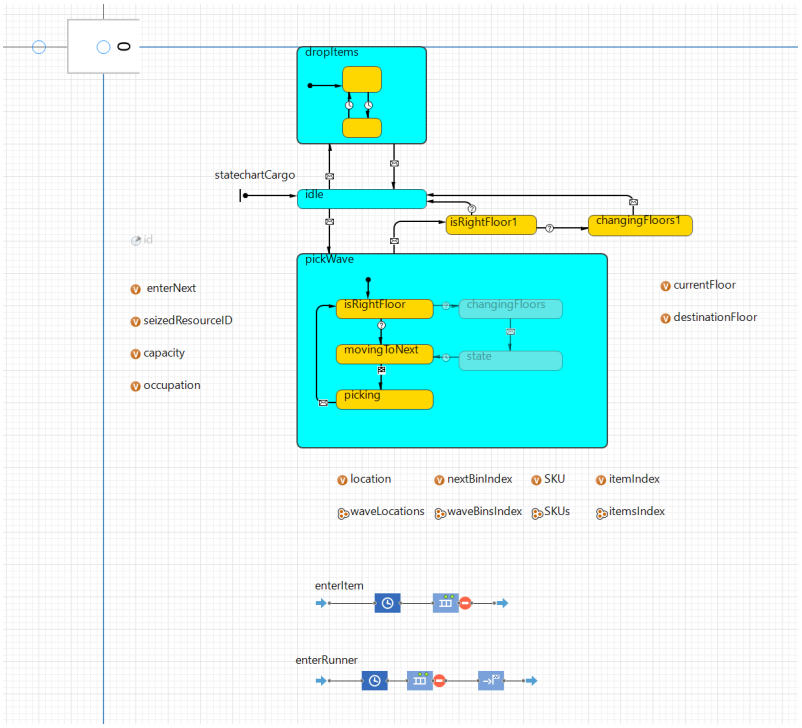


Figure A.9: Trolley Picking

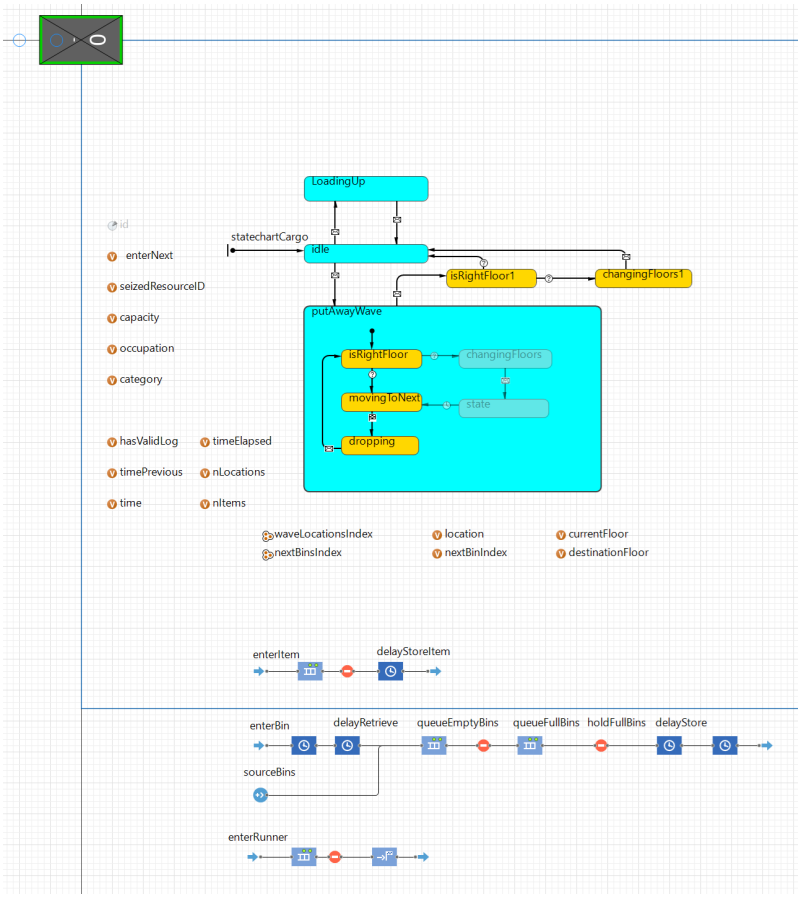


Figure A.10: Trolley Putaway

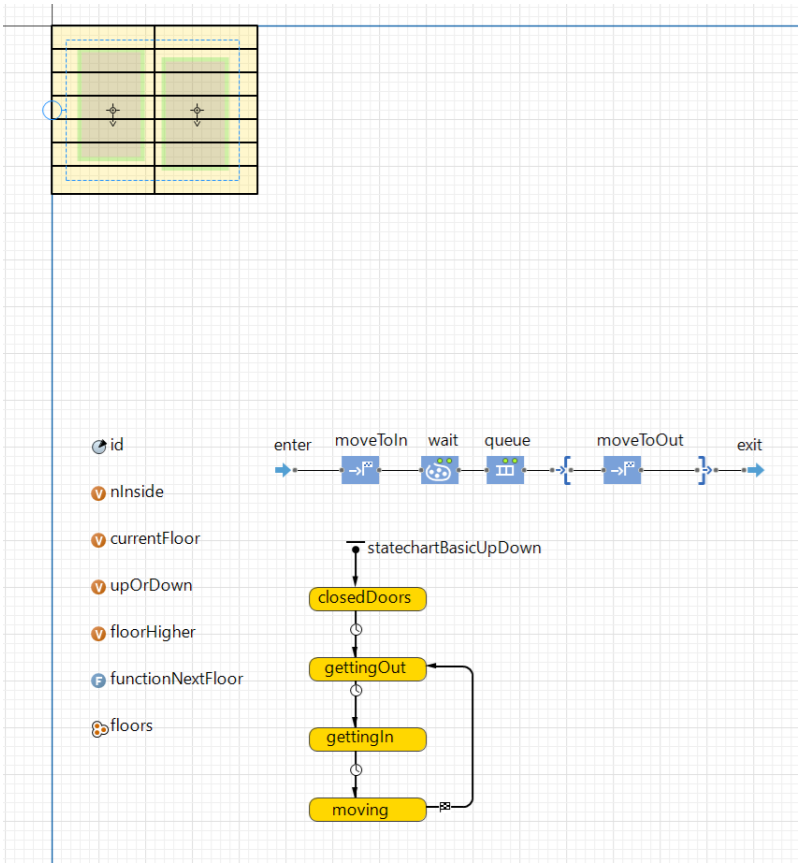


Figure A.11: Elevator

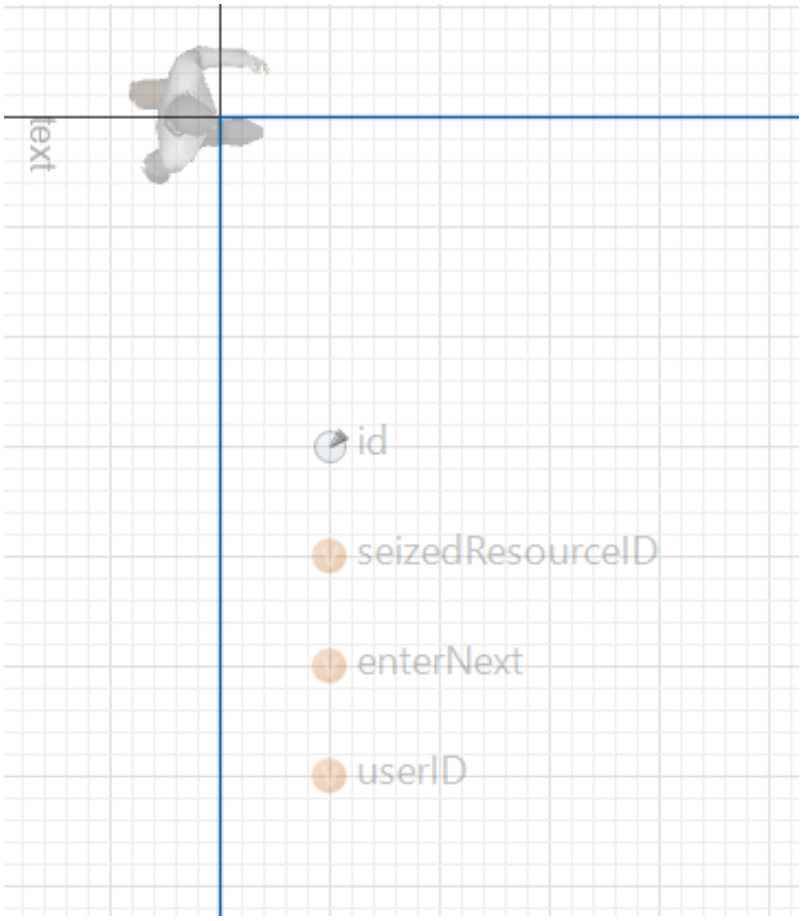


Figure A.12: Replacement Runner