Analytics Driving Supply Chain Segmentation for Lenovo

by
Luiz Gosling
Master of Science, Logistics and Supply Chain Management, PUC-Rio, Brazil, 2014
Bachelor of Engineering, Industrial Engineering, PUC-Rio, Brazil, 2011
and
Javier Urrutia
Master of Engineering, Operations Research, Universidad de Concepcion, Chile, 2014
Bachelor of Engineering, Commercial Engineering, Universidad de Concepcion, Chile, 2011

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2019

Signature of Author: _____

Department of Supply Chain Management
May 10, 2019

Signature of Author: _____

Department of Supply Chain Management
May 10, 2019

Certified by: _____

Dr. Maria Jesus Saenz
Executive Director, Supply Chain Management Blended Program
Capstone Advisor

Accepted by: _____

Dr. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

Client-Oriented Supply Chain Segmentation for High-Tech Hardware
by
Luiz Gosling
and
Javier Urrutia
Submitted to the Program in Supply Chain Management
on May 10, 2019 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

Abstract: Although segmentation strategies and their benefits are common topics in the academic literature, two realities contrast in the realm of supply chain management: While a vast number of companies still perceive operations as cost-centers, only a few have adopted profit-driver approaches. The second group embraced segmentation to deploy end-to-end supply chain strategies that clearly match the segments' requirements to the company's capabilities, adding sustainable value in the process.

Lenovo Data Center Group (DCG), sponsor of this project, proposed an assessment of their Hyperscale Business Unit's supply chain and how they could better serve their client portfolio. With that aim in mind, this capstone had three goals: (i) review current frameworks present in the supply chain segmentation literature; (ii) identify DCG's key client and product groups in its portfolio through quantitative methods; and (iii) propose specific supply chain policy guidelines for each identified group, creating baselines for change.

We applied a clustering model to process DCG's sales and operations records to not only identify clusters of clients and products, but also quantify their differences in terms of supply chain. Data was preprocessed and then ran through an EFA for dimensionality reduction without losing traceability for insights and discussions. Two dimensions were identified: Importance and Complexity. k-Means clustering algorithm was then applied on the resulting dataset, and four client-product clusters were identified. With results in hand, a workshop was conducted with experts from DCG to explore a policy-cluster framework and understand how policies are set according to products' and clients' nature.

Our analysis shows that four segments exist within DCG's portfolio and operating segmented supply chains is likely to positively affect performance. A workshop identified guidelines behind distinct supply chain policies for each cluster and provided a framework for segmented strategies' design, thus helping managers rethink supply chains to better fit given segments. Moreover, the framework enables a data-driven approach to segmentation, which can be deepened with further analysis – for example, micro-segmentation of clients. To this end, we believe that our work has significant practical implications for Lenovo DCG.

Keywords: supply chain, segmentation, policy, high-tech, cluster, k-Means, EFA

Capstone Advisor: Dr. Maria Jesus Saenz
Title: Executive Director, Supply Chain Management Blended Program

Table of Contents

List of Figures

List of Tables

1   Introduction

Segmentation strategies and their benefits are common topics in the academic literature. Applications are offered across multiple business functions such as corporate strategy, operations, marketing, and product management, often built upon the works of Porter (1980). Researchers and practitioners have acknowledged the value of segmentation of markets, products and/or suppliers, to drive value creation and process alignment accordingly (Sanchez Rodriguez, Harris, & Manson, 2015; Protopappa-Sieke & Thonemann, 2017).

In the realm of supply chain management, however, two realities contrast: While a vast number of companies still perceive operations as a cost-center, few have adopted a profit-driver approach. The second group has embraced segmentation to deploy end-to-end supply chain strategies with sustainable outcomes and value addition for clients and shareholders (Protopappa-Sieke & Thonemann, 2017). This value addition fundamentally means that those companies align operations with customer requirements so that different customers are served with differentiated supply chain strategies (Protopappa-Sieke & Thonemann, 2017). Because of this contrast, while there is an abundance of literature on corporate strategy and the benefits of business segmentation strategies, studies focused on how the supply chain can support these strategies and help achieve business goals have gained more interest and recognition only in recent years.

The objective of this research project is to support our corporate partner, Lenovo DCG, to analytically assess their supply chain capabilities to better serve their clients' requirements by proposing individual segmentation policies through clustering. Lenovo DCG is a global player in the data server industry, acting as a global hardware manufacturing and system integrator. This industry is expected to expand at a fast pace in the coming years, mainly driven by the rise of cloud computing applications and decentralization of computational infrastructure across the globe (Lenovo DCG, 2018a).

Large-scale cloud, search engine and social media providers (commonly called *Hyperscale* or *Cloud* providers in the high-tech industry) have become increasingly focused on optimizing their capital expenditure and in reducing hardware supply risk in their data centers (Lenovo DCG, 2018b). At the same time, those companies have to balance the expected fast-paced innovation in cutting-edge technologies and provide superior customer experience. As a result, players such as Lenovo DCG act as specific solution providers, offering (i) Hardware Design, (ii) Procurement, and (iii) System Integrator options for their clients. The first two functions are more straightforward as they encompass tasks commonly seen in other industries. The third function – systems integrator – is the most unique in this case. In short, a system integrator (SI) acts as a middle man between a client and their suppliers and vendors, while also acting as a vendor itself. As such, a SI may provide physical logistics solutions such as procurement, transportation, tracking, and customs clearance, and at the same time design and model the client's solutions for them (i.e. data centers' infrastructure and software capabilities), including the selection and management of vendors across the globe.

In order to adequately understand Lenovo DCG's current supply chain design, its practices and policies, and the added value a segmented supply chain strategy could provide, a series of meetings and interviews were conducted with key stakeholders. These meetings were called "initial immersion". A semi-structured kick-off meeting at Lenovo DCG's headquarters allowed an open-question dialog with a broad range of senior supply chain leaders, which provided valuable insights and allowed for a clear project design (Figure 1 illustrates the leaders' positions and roles under the Vice President of Global Supply Chain as well as the team members involved in the project). Weekly follow up meetings were carried out to discuss progress and answer more specific inquiries as the project moved forward.

For their data server portfolio, Lenovo DCG has historically deployed novel supply chain strategies for their products and clients. The company understands, though, that a "one size fits all" approach fails to address clients' and products' individual needs in a more precise manner and more specific approaches

already occur in practice, even if not formally addressed. All things considered, Lenovo DCG is potentially limiting its ability to deliver superior perceived value for bespoke clients and, at the same time, higher efficiency for cost-sensitive ones.
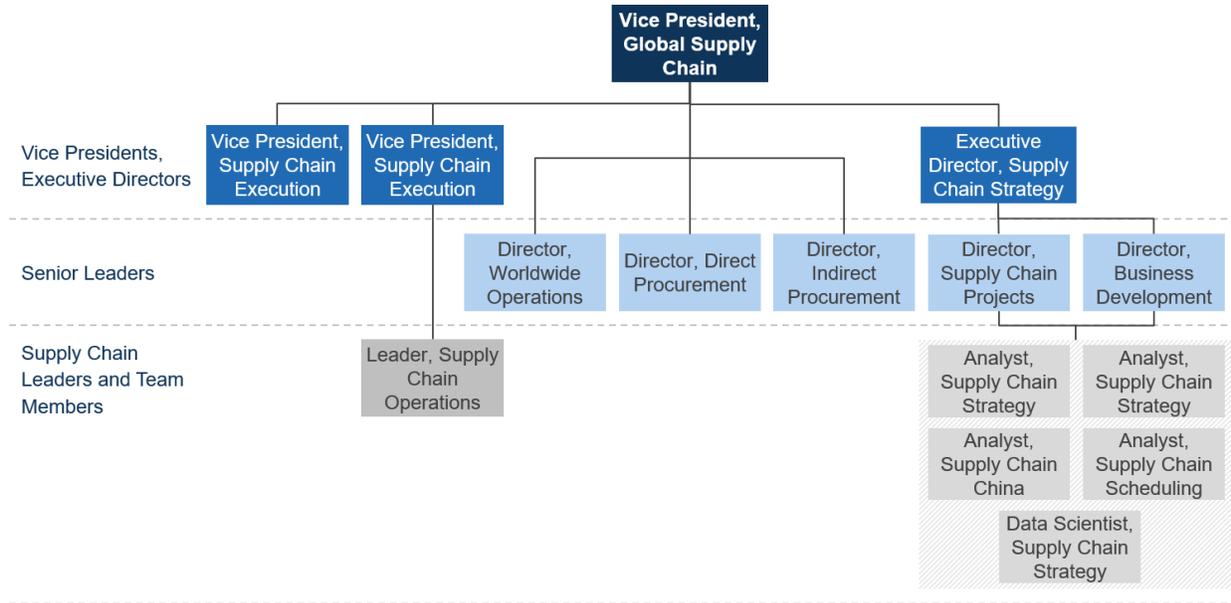
Figure 1: Lenovo DCG supply chain leaders' organizational chart (adapted from Lenovo DCG)

As recent projections from Lenovo DCG show twofold growth potential by 2020 (Lenovo DCG, 2018c), developing superior supply chain strategies becomes paramount. Such strategies should consider the necessary capabilities to ensure differentiated service levels for each client segment, thus ensuring the desired growth the company foresees. In this context, we propose the following research questions (RQ):

RQ1. How homogeneous is Lenovo DCG's client-product portfolio, in terms of supply chain?

RQ2. Are there measurable similarities and differences among items in the firm's portfolio?

RQ3. Can the firm's portfolio be segmented in groups based on its supply chain characteristics?

RQ4. What are the current policies guiding the supply chain execution of each identified group and what are the similarities and differences among them?

RQ5. Could segmented supply chain policies be beneficial to ensuring alignment between supply-chain execution and service performance at Lenovo DCG?

The goals of this capstone project are trifold: (i) to review current frameworks present in the supply chain segmentation literature; (ii) to identify Lenovo DCG's key customer and product groups in its portfolio through quantitative scientific methods; and (iii) to propose individual supply chain policy guidelines for each identified group, creating baselines for future discussion.

This capstone is organized as follows: Chapter 2 provides a review of existing literature on supply chain segmentation. Chapter 3 addresses the methodology used to provide the answers to Lenovo DCG questions and deliver the desired outputs. Chapter 4 explores and discusses the results of the segmentation exercise at Lenovo DCG. Finally, Chapter 5 concludes the work with a summary of the findings, the authors' assessment of this work's managerial implications for Lenovo DCG as well as its limitations, and a proposed avenue for Lenovo DCG and future academic research.

2   Literature Review

This chapter reviews the existing literature on supply chain segmentation, the major research construct of this study, and identifies the necessary stages for the segmentation exercise. This selected body of knowledge serves as the foundation for the development and application of this project's quantitative research model.

The key objective of this chapter is to answer five questions: (i) what are the established frameworks present in the literature capable of guiding supply chain segmentation initiatives?; (ii) which of the established frameworks can be applied to Lenovo DCG?; (iii) what are the qualitative and quantitative segmentation drivers present in the identified frameworks?, (iv) what are the stages towards the adequate application of segmentation frameworks?; and (v) what are the segmented supply chain policies and strategies proposed by the frameworks?

Answering these questions is paramount to determining which type of segmentation is better suited for Lenovo DCG and which variables must be measured to develop a segmented strategy for the company.

2.1.   Supply Chain Segmentation

Supply chain segmentation approaches have been present in the literature since the 1980's (Alicke & Forsting, 2017; Protopappa-Sieke, 2017). However, despite being contemporary with works on corporate strategy such as those of Porter (1980), tailored supply chain approaches have seldom been considered strategic differentiation enablers.

Essentially, supply chain segmentation describes a method for designing customized strategies according to selected value drivers (Childerhouse, Aitken, & Towill, 2002; Walters, 2006). This customization is built upon the definition of unique segments, which consist of a set of products, clients, or regions. For each

segment, a unique supply chain strategy is derived that fits the products' or customers' needs according to the identified drivers (Godsell, Harrison, Emberson, & Storey, 2006).

Callinan (2018) argues that current market trends have caused demand fulfilment to become an increasingly challenging task as customers' expectations evolved significantly and became more dynamic. The article urges supply chain managers to realize the untapped potential of providing value through supply chain as customer experience becomes a key differentiator, beyond price and product. To succeed in this scenario, companies must understand the different expectations of different segments, what drives their experience, and then align their value proposition and resources planning accordingly to better serve those top tier customers while not overserving the lower tiers with adequate flexibility (Sanchez, 1997; Agarwal, Shankar, & Tiwari, 2006).

In this current market scenario, of highly customizable products/services and increasing customer expectations; big data has enabled segmentation depth or granularity to advance beyond traditional segmentation based on few factors and agglomerative approaches (e.g.: the classical 2-by-2 matrix of sales volume and profitability, at a customer group level). Computational power and machine learning techniques have broadened the spectrum of analysis. Companies now can track individual purchase behavior, identifying and predicting behavior for micro-segments or even advancing towards a completely personalized, one-to-one marketing approach (Payne & Frow, 2006; Rubin, 1997; Sanders, 2016) as shown in Figure 2.
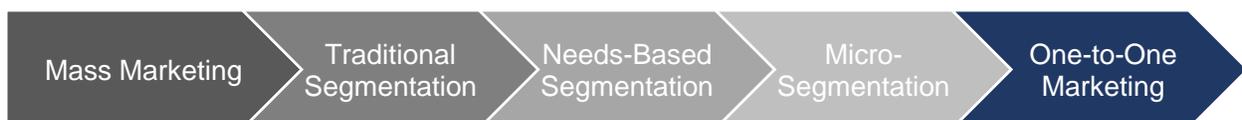
Figure 2: Segmentation continuum (adapted from Rubin,1997)

Beyond marketing applications, this approach serves to evaluate and adjust end-to-end supply chains. The level of granularity will depend on the industry and company maturity on the use of these techniques, but

current capabilities allow for a deep granular and frequent analysis, proving valuable insight, to identify and react quickly, towards changes in market/customer trends.

For end-to-end supply chain purposes, companies may adopt distinct approaches when segmenting, three approaches can be identified as guidelines: Market-Driven, Product-Driven and Hybrid Segmentation (Alicke & Forsting, 2017). Approach selection will rely on previously identified drivers, but the end result, which their research calls 'supply chain strategy', will be a set of activities that define a supply chain design (sourcing, manufacturing and distribution guidelines) tailored per the segment's needs.

For Lenovo DCG, supply chain segmentation exercises focused on identifying key-segments of products that could guide the design of segment-specific supply chain policies. These policies consist of sets of rules that drive and/or restrict operations, namely: Sourcing; Inventory; Production; Fulfillment; and Customer Service. Figure 3 illustrates the five policies and provides an overview of Lenovo DCG's supply chain echelons.
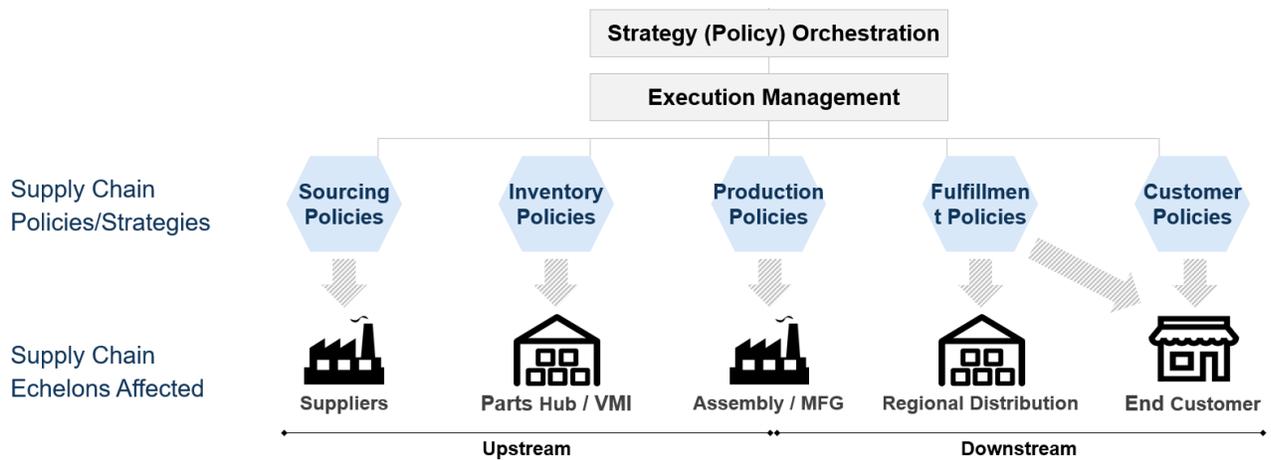
Figure 3: Lenovo DCG's supply chain echelons and policies (adapted from Lenovo DCG)

2.1.1. Market-driven Segmentation

Market-driven segmentation frameworks develop supply chain segments according to external factors, such as region, client, or market landscape. One of the first market-driven segmentation frameworks was proposed by (Hill, 2000), who focused on differentiated manufacturing capabilities to better service different customer groups. This framework included two key-drivers groups: the supplier's prerequisites to be able to service a customer and the differentiating factors that drive the customer's final decision. Hill, (1985) called these groups Order Qualifiers (OQ) and Order Winners (OW). The groups may include cost, reliability, availability, and lead-time, for example. The framework proposes differentiated manufacturing strategies that meet each segment's specific requirements, rather than one single response to market.

Later works developed the market-driven segmentation approach to include other business capabilities beyond manufacturing.(J. L. Gattorna, Chorn, & Day, 1991) proposed customer policies that further reinforce the segmented strategies, and Gattorna and Walters (1996) argue that in order to add value, there needs to be alignment between such policies and the proposed strategy. The authors recommend segmenting the market according to customer buying behavior, including demand volatility and price sensitivity. The proposed supply chain strategies range from fully flexible, agile and lean supply chains, to continuous replenishment (Christopher & Gattorna, 2005). Walters (2006) proposed the alignment of supply chain processes and market orientation further by introducing a set of customer value drivers.

Lastly, Hjort et al. (2013) provide a rather precise segmentation exercise focused on the e-commerce business, recommending not only a customer segmentation based on buying and returning behaviors, but also the development of differentiated reverse logistics capabilities.

In short, market-segmenting a supply chain can lead to valuable insights as well as superior performance (Skipworth, Godsell, Wong, Saghiri, & Julien, 2015). That said, market requirements should not be focused

as the sole factor affecting supply chain decisions, as they are not the only dimension that brings different requirements.

2.1.2. Product-driven Segmentation

In contrast to the market-driven segmentation approach, the product-driven approach focuses on internal product information, sales and performance records, and demand characteristics (Protopappa-Sieke & Thonemann, 2017). The product-driven approach was initiated by Fisher (1997) through its theory of functional *versus* innovative products. The author suggests that functional products have stable demand and a longer product life cycle, lower contribution margins when compared to the rest of the firm's portfolio, and low variety. Innovative products, on the other hand, are defined by varying demand, shorter life cycles, higher contribution margins within the firm's portfolio, and a much higher variety. The proposed theory matches different supply chains to the different product types, which ensures a physically efficient service for functional products (focusing on reducing costs) and a responsive supply chain for innovative products (focusing on flexibility and agility in responding to unstable demand).

Modern variations of Fisher's (1997) product segmentation approach are found in literature and managed to develop the original model by incorporating not only demand uncertainty but also supply uncertainty to a given product set (Lee, 2002), replenish lead-time (Christopher & Towill, 2002; Peck, Christopher, & Towill, 2006), required agility (Christopher & Towill, 2000), among other characteristics.

In short, product-driven segmentation frameworks aim at segmenting a firm's product portfolio according to different product, demand, and supply factors. These factors not only include product characteristics and demand data (level and volatility), but also supply chain aspects such as lead-times, risk, and footprint complexity. Moreover, while recognizing the factors that differentiate a firm's portfolio is paramount for an adequate supply chain strategy design, the product-driven approach is limited by not considering external factors covered in the market-driven orientation.

2.1.3. Hybrid Segmentation

In this third approach, both market and product factors are used jointly to segment products and define supply chain strategies that better fits the firm's internal and external operational requirements. Overall, the hybrid segmentation approach provides a more holistic strategy, which considers both internal and external factors when deciding between efficiency and agility in a firm's operations.

Lastly, although the academic studies recognize the importance of identifying the market's needs and responding to them, many remain theoretical and do not explain how the supply chain strategy could be operationalized in practice. Works like the ones described in this section are fine examples of practice-focused research.

By identifying the value drivers and applying the approaches described above, companies are able to identify which characteristics define each of their serviced segments and then allocate products/customers within those groups. Best practices suggest segments definition will follow four principles (Protopappa-Sieke & Thonemann, 2017):

- **Heterogeneous segments:** Within segments, strategic elements should have significant differences. Optimal supply chain approaches will be unique.
- **Exclusivity:** Segments should be mutually exclusive. An attribute should, ideally, not be present in multiple segments.
- **Practicality:** The segmentation exercise must provide actionable insights that can be understood by the organization – it should not become a source of additional complexity.
- **Holistic supply chain:** Segments should be understood by stakeholders throughout the whole supply chain, from purchasing to distribution. These functions must equally understand the segment requirements.

One important aspect of any segmentation exercise is that it must always provide value, either from improved customer service levels or in the form of cost reductions (Godsell, Diefenbach, Clemmow,

Towill, & Christopher, 2011). With SKUs or customers assigned to each of the segments, companies must then simulate scenarios to evaluate the performance of new supply chain strategies and policies.

2.2. Supply Chain Segmentation Best Practices

Successful implementations of segmentation strategies have been well documented. Likewise, there is an abundance of literature covering approaches to markets with particular conditions (e.g. high volatility of sales, or long lead times from suppliers), all of which serve as a valuable basis that connects the segmentation exercise to external factors.

From earlier publications such as Fisher's framework for supply chain strategy identification (Fisher, 1997), academia has provided research on how to identify or manage for agile, lean or flexible supply chain, or on how to tackle seasonality, volatility or market disruptions.

Some researchers have focused on particular supply-chain strategies or market characteristics, and how to address these. The responsive supply chain strategy case was analyzed by Roh, Hong, and Min (2014), providing perspectives on how successful responsiveness can be achieved. The flexibility and agility required by the fashion industry was analyzed by Chan, Ngai, and Moon (2017) to determine the key supply chain elements successful garments manufacturer presented.

Others have provided comparative analysis of different supply chain strategies. Goldsby, Griffis, & Roath, (2006), took the more traditional strategies from Fisher (lean, agile, and the hybrid leagile) and evaluated the implications on both physical networks and a series of supply chain metrics of these approaches.

It is precisely this type of analysis that is required for the segmentation exercise. Lee (2002) argues that a single approach might not be enough for modern supply chains. In volatile market conditions, companies would benefit from a portfolio approach to supply chain design (Olavson, Lee, and DeNyse, 2010), where each segment will be addressed by tailored strategies, each with different performance target metric, and

the simulations and evaluations will enable that match to occur. In Lenovo DCG's case, this can be translated into potential micro segmentations after the preliminary round, which can lead to specifically tailored strategies for the firm's portfolio across the globe.

More than just traditional supply chain KPIs such as cost or lead-time, researches have also provided tools that incorporate into the design and evaluation of supply chain other strategic objectives such as social responsibility, environmental impact, or resilience (Sáenz, Revilla, & Acero, 2018). These tools allow for a better alignment between supply chain strategies and qualitative factors of the overarching strategy. This capstone project aims to do just that: apply quantitative methods in assessing Lenovo DCG's portfolio characteristics to provide analytical support to the decision makers in the firm. Ultimately, as in many supply chain decisions, there is no single right answer when it comes to policy design and strategy definition, and the scientific rigor of quantitative methods are a powerful asset in kindling strategic discussions with solid foundation.

2.3. Clustering Algorithms

Grouping of objects is required for various purposes in different areas of engineering, science and technology, business, medical science and our daily life. For example, picture a group of passengers waiting to board a flight at the airport. Passengers are often classified by their frequent flyer status, which represent their score on some predetermined metric (number of accumulated miles over the last twelve months, for instance). When premium passengers arrive at the boarding gate, they are queued in a specific line, separated from the non-premium passengers. The passengers grouped for that status will be treated accordingly while passengers not belonging to that group should be handled differently. The airline company cares that its ground staff treats their loyal premium members in a differentiated manner in order to ensure future revenues from this group, thus labeling premium individuals is very important. Whenever we find a labeled individual, we place it into the group with same label. It is a simple task when labels are given in advance. However, on many occasions, no labelling information is available in advance and objects

need to be grouped based on some characteristics. Both of these in- stances represent a wide range of problems occurring in analysis of data. In generic terms, these cases are dealt under the scope of classification ((Duda, Hart, & Stork, 2012); Farid, Zhang, Rahman, Hossain, & Strachan, 2014). Precisely, the first case when the label of an object is given in advance is called **supervised classification** whereas the case when the class label is not tagged to an object in advance is called **unsupervised classification**.

In this project, clustering is an unsupervised way of segmenting Lenovo DCG's portfolio. In short, this activity consists of analytically grouping sets of data through the use of measurable characteristics that ensure: (i) similarity among points in the same group; and (ii) differences between points in different groups (Dhanachandra, Manglem, & Chanu, 2015; Aryuni, Madyatmadja, & Miranda, 2018). We chose *k-Means* among several different clustering algorithms (see Figure 4) as the approach for product segmentation due its non-hierarchical clustering approach and its good convergence, as defended by Vora & Oza (2013). Moreover, the authors argue that *k-Means* clustering is simple, flexible, can be understood and implemented easily. Kalyani, Ramudu, & Reddy (2018) points out that, in terms of clustering structure, clustering algorithms can be categorized as: (i) Exclusive-clustering; (ii) Overlapping-clustering; (iii) Hierarchical-clustering; (iv) Probabilistic-clustering; and the proposed *k-Means* algorithm fits into the first group.

In hierarchical clustering methods, clusters are formed by iteratively dividing the dataset using top-down or bottom up approaches. There are two forms of hierarchical methods: agglomerative and divisive clustering (Murtagh, 1983). Agglomerative algorithms follow a bottom-up approach, which builds up clusters starting with single object and then merging these atomic clusters into larger and larger clusters, until all of the objects are finally lying in a single cluster or otherwise until certain termination conditions are satisfied. Divisive algorithms follow a top-down approach, which breaks up a larger cluster containing all objects into smaller clusters on each step, until each object is in a cluster of its own or until a certain conditions is satisfied (Saxena et al., 2017).
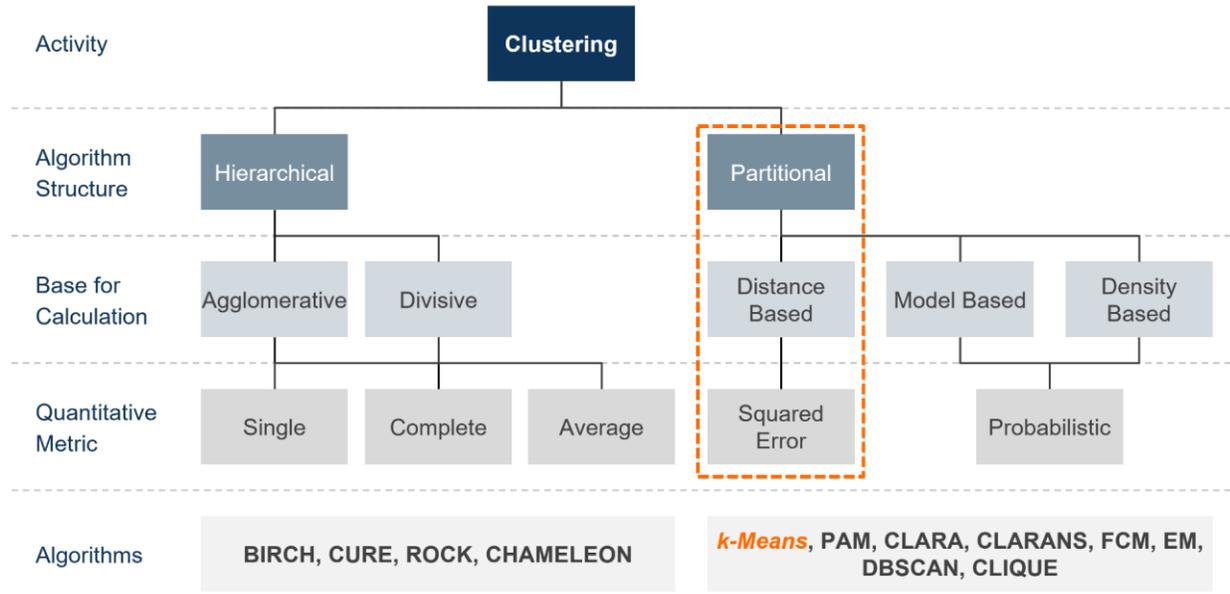
Figure 4: Taxonomy of clustering approaches (adapted from Fraley & Raftery (2010))

Partitional methods are opposite to hierarchical clustering. On those methods, data is assigned into k-clusters without any hierarchical structure by optimizing some criterion function (Lam & Wunsch, 2014). The most commonly used criterion is the Euclidean distance, which finds the minimum distance between points with each of the available clusters and assigning the point to the cluster. The algorithms studied in this category include: *k-Means*, PAM (Kaufman & Rousseeuw, 1990) CLARA (Kaufman & Rousseeuw, 1990), CLARAN (Ng, Raymond T., 2002), Fuzzy c-means, DBSCAN, among others. This study first explored two methods for clustering: (i) agglomerative hierarchical clustering, and (ii) *k-Means*. Both hierarchical clustering and *k-Means* methods were applied in thus study, and given its wide adoption (Kulis & Jordan, 2011), approach simplicity, and seamless handling of the data available, *k-Means* was chosen as preferred algorithm for this capstone. Moreover, hierarchical methods are usually applied not for result improvement or insights, but rather to form a preliminary intuition on the number of clusters present in a given dataset. Those methods are often applied together in data science studies (Lam & Wunsch, 2014).

Regarding the calculation involved, the *k-Means* algorithm consists of a heuristic unsupervised machine-learning that classifies data points into a specified number of clusters (k) based on the distance between

points and cluster centers. Through multiple iterations (or rounds), data points are assigned the cluster with

the center closest to them, with centers being recalculated to reflect the assigned points' centroid. These

steps are repeated until no changes are observed in the classification, as shown in Figure 5. The algorithm

follows a basic as well as simple approach to characterize a given dataset via a specific number of clusters

settled *a priori* by the user. The general goal of the algorithm is to minimize the total distance between each

cluster centroid and its points. Mathematically, the squared error function is defined as the objective

function for the unsupervised machine-learning as shown in Figure 6.

1	2	3	4
Locate k arbitrary points into the dataset space. These points represent the initial set of centroids.	Allocate every data point to a set which consists of a closest centroid.	When the entire dataset has been allocated, re-calculate the positions of the k centroids.	Repeat step 2 as well as 3 until centroids suffer no additional change. The dataset will be classified into clusters.

Figure 5: k-Means steps (elaborated by authors)

$$SSE = Z = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|$$

Figure 6: k-Means objective equation (distance function)

In the SSE equation above, Z is the objective function to be minimized, k is the number of clusters, n is the

number of cases (or products, in Lenovo DCG's case), x_i is the i-iteration case, and c_j is the centroid for

cluster j. The subtraction observed on the equation is often referred to as distance function.

Two downside risks are identified when adopting k-Means: (i) The cluster result sensitivity to the initial

clusters centers selection, meaning that user inputs may sharply affect the final clustering results in some

cases; and (ii) The algorithm's lack of optimization for the number of clusters, but rather a way to test

different values as set by the user. Additionally, there are modern, improved versions of k-Means that make

use of more advanced statistics for better precision or flexibility (Kulis & Jordan, 2011). Nevertheless, the

k-Means method is widely adopted and serves sufficiently well the intended objectives of this study.

2.4. Insights from the Literature Review

Specifically, regarding the segmentation approach to data, we believe that the observed Lenovo DCG situation is best addressed by the adoption of a hybrid model. Per the discussion with key executives (discussed in detail on the following chapters), there are clearly differentiated product segments, a feature that usually points towards a product-driven segmentation. However, there are significant differences within clients within the same product family in terms of expectations, strategic value for Lenovo DCG, and their negotiating power with clients, thus making the case for a more comprehensive framework through a hybrid approach. Finally, Swan, Pal, & Lippert (2009) fundamentally argue that segmentation brings rewards to companies that perform operations that are tailored to specific channel needs and costs. The benefits vary across industries, since the focus of segmentation is quite different depending on the industry. The authors point out that in high tech, where it is common to see short product lifecycles and unpredictable trends and component availability, segmentation typically improves service levels by 5 to 10 percent while reducing inventory levels by 15 to 20 percent. The focused allocation of inventory successfully frees up working capital that can be reinvested in software and hardware capital expenditures. Figure 7 illustrates the general idea of focusing a multi-capable supply chain into multiple chains serving specific segments of suppliers and/or clients.
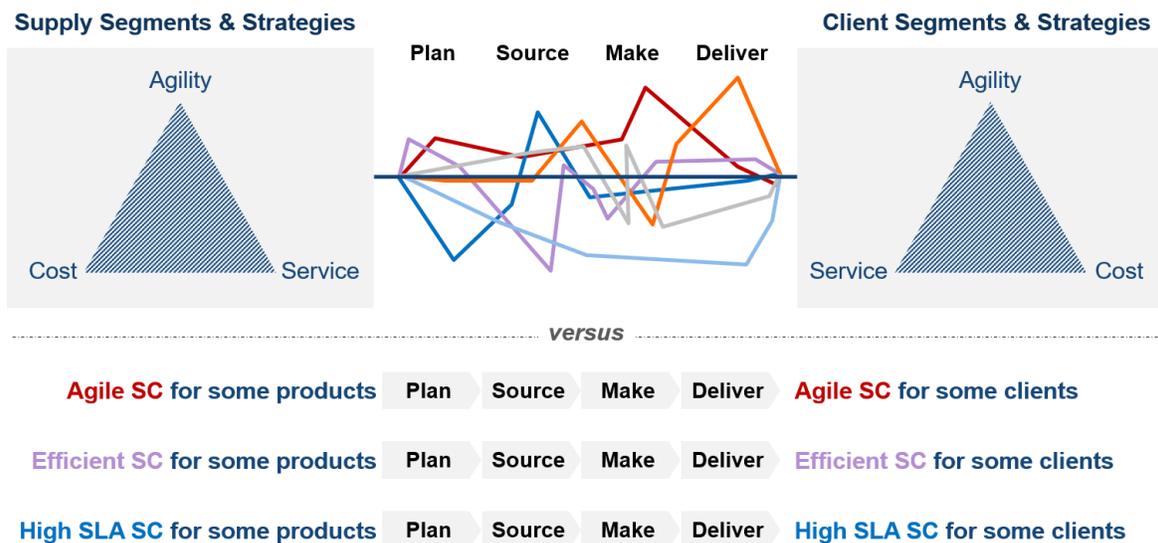
Figure 7: Single versus segmented supply chain approach (adapted from GT Nexus white paper)

The following chapter explores in detail the methodology applied in this project, including the application of the aforementioned framework and unsupervised machine-learning. By exploring Lenovo DCG's sales records and supply chain data, a hybrid segmentation analysis was conducted and included both product sales and supply-chain execution information. With access to a group of variables able to characterize Lenovo DCG's portfolio, several clustering analyses were built, through the use of *k-Means*. Resulting clusters were then contrasted for qualitative analysis and specific policies discussion, which served as a guideline for the future implementation of the segmented strategy.

3   Methodology

This chapter explores the methodology followed to answer the research questions proposed in Chapter 1. Table 1 outlines this process by breaking it down into five parts. The main objective of the chapter is to cover the tools used and the measures taken to achieve the project objectives.

Table 1: Project Phases Overview (elaborated by authors)

#	Phase	Focus	Main Activities
1	**Literature Review**	▪ Identification of SCM strategies, key-SCM drivers, objectives identification and mapping	▪ Literature search for previous work covering supply chain strategies and segmentation
2	**Company and Sector Immersion**	▪ Internal and external insights gathering	▪ Interviews with executives and key-people from the sponsor company ▪ Technical visits to the sponsor company headquarters and interactions to better understand the firm's portfolio and operations ▪ Weekly meetings with project team from DCG
		▪ Development of Research Questions	▪ Questions development (R1, R2, R3, R4, R5) ▪ Identification of hypothesis for each question ▪ Analysis for testing each hypothesis
3	**Data Gathering**	▪ Collection of necessary data to perform the listed analysis	▪ Data requests for sales, inventory, client, channel, product, and delivery ▪ Identification of variables able to describe Lenovo DCG's client and product portfolio
4	**Analysis**	▪ Provide quantitative insights for the key-questions ▪ Provide structured responses for each hypothesis	▪ Product dataset exploration ▪ Selection of descriptive variables ▪ Exploratory Factor Analysis (EFA) on the descriptive variables ▪ Hierarchical and *k-Means* clustering ▪ Feedback and client validation of clusters
		▪ Provide guideline for adequate policy discussion and design	▪ Track back clusters components to identify key-variables and build policy matrix
5	**Research Insights**	▪ Provide quantitative insight on the segmentation ▪ Policy-cluster framework for indicating targets of future work	▪ Workshop: Qualitative discussions with project team (MIT and DCG) for policy selection and future redesign for each cluster ▪ Feedback and client validation of guidelines for policy redesign ▪ Suggest future work to improve practical results, as well as additional data collection

3.1.   Part One: Literature Review

The first part of this process was an extensive literature review on supply chain strategy, supply chain segmentation, and segmentation methods covered in Chapter 2. Supply chain design articles on particular supply chain requirements such agile, flexible, or lean were extensively consulted. Quantitative methods for supply chain segmentation, such as clustering algorithms, were explored in order to define a preferred method for Lenovo DCG. Lastly, white papers from consulting firms and reports on industry practices were also reviewed.

3.2.   Part Two: Company and Sector Immersion

The first step in conducting the practical study was an immersion into the Lenovo DCG business. For such, technical visits to headquarters enabled the authors to conduct interviews with executives and key-people from the internal project team, as well as subject-matter experts able to provide insights on the company's sales, operations, product development, planning, and overall strategy for different segments and geographies.

After an initial immersion, weekly meetings ensured the internal project team from Lenovo DCG was kept up to date with the analyses and results as they evolved. Moreover, the participation of the internal project team was fundamental in exploring the questions and raised issues that would eventually take form as this project's research questions.

3.3.   Part Three: Data Gathering and Analysis

In our analyses, sales and operational data from the past three years was collected from the sponsor's ERP system. A few areas of information were not adequate for the intended application and thus were not fully used during the research, namely: (i) inventory position evolution, given its high level of granularity, which did not fit the client-product level of the work; (ii) manufacturing and shipping capacity was limited to lead-times and network information (such as origins and destinations for each product sold) because costs,

volumes, and capacity data were not available on the client-product level; client information was restricted to internal records data, including descriptions, industry, and business unit, while cross-information usually associated to SRM (e.g. Share of Wallet) was not available. Moreover, the data gathered covered North American operations and did not include sales from countries such as China or European offices. That said, the data collected was paramount in providing factual quantitative information on the customer-segmentation exercise, including sales trends, macro inventory levels, geographical footprint, channels performance, product characteristics, transportation modes, and the performance of deliveries on time.

The three-year horizon was set by Lenovo DCG as the maximum available timespan for the data requested, in particular the Hyperscale Business Unit. As the data center division has been growing rapidly since its establishment four years ago, increasing amounts of information have been recorded and more data generated. The three years constitute, therefore, a sufficiently large period for analyzing Lenovo DCG's servers' operations and this project's main focus: the hyperscale segment.

Regarding the data files, data was stored in individual files for each year and then merged into one single dataset. From this dataset, 54 features were identified and used to characterize Lenovo DCG Hyperscale Business Unit's portfolio of 111 products and 142 client-product pairs. Of those 54 features, 13 were identified as key variables for describing the firm's portfolio, as shown in Figure 8. It is important to mention that most of those variables were not provided in the original dataset, but rather calculated using the existing data as KPIs, aggregated summary statistics, or treated as categorical variables (i.e. Distribution Channel). Additionally, a "product" is a unique Material Description from the dataset, while a "client" is a standardized description based on the Client Name from the sales records.

Figure 8: The 13 key variables used in characterizing the firm's portfolio's supply chain

Lastly, quantitative data was normalized before its application on EFA and subsequent clustering. This was particularly important, given k-Means' sensitivity to distance and thus to the scale of the selected features (e.g. Sales in Dollars and Total LT are described in units with different order of magnitude, which distorts their weight on clustering analyses when distance is set as the target factor for group formation on the geometrical space). Normalization tackles that issue by transforming the data in each feature to a range between 0 and 1.

3.4.   Part Four: Data Analysis

Three methods were applied to analyze the constructed dataset and gain the quantitative insights necessary for answering the project's research questions: Exploratory Factor Analysis (EFA), hierarchical clustering, and *k-Means*.

Conversations with different leaders at Lenovo DCG provided insight into the most pressing issues for the firm's supply chain. From the company's perspective, (i) inventory position, (ii) profitability, and (iii) lead time pose higher risks of operational disruption and profitability loss, while from the customer size, (iv) cost and (v) lead time were described as drivers of inefficiency. Unfortunately, it was not possible to obtain

granular data on average inventory levels, profitability, or costs, as mentioned in section 3.3. Therefore, the initial approach towards the segmentation exercise considered only Total Lead Time and Average Delivery on Time (%) (DOT%) as target metrics for cluster performance, with all other features initially considered descriptors, or explanatory variables.

This approach was subsequently refined. Regardless of the statistical validity of some of these concepts to explain the data correlations in EFA, not all features should be used in the cluster analysis. For example, some of the mentioned features are not exactly valid descriptors of the products and their supply chains, but rather a result of Lenovo DCG's policies. Sources of Supply is an example: its value depends on the company's decision to single source (assemble servers in one single manufacturing site) or manufacture in multiple sites, closer to demand. There is no clear scale when such variable is applied in a quantitative analysis, as a higher value (multi-sourcing) is not necessarily better than the single sourcing option. Such examples illustrate the trade-offs commonly seen in supply chain decisions and thus were not considered valid for application in the capstone.

The original list of 13 features was thus reduced to 10, as shown in Figure 9: Sales in Dollars, Sales in Units, Demand Uncertainty, Order Frequency, Average Order Size, Distribution Channel, Total Lead Time, Manufacturing Lead Time, Destinations, and Delivery on Time (%). Total Lead Time and Delivery on Time (%) were kept as target metrics.
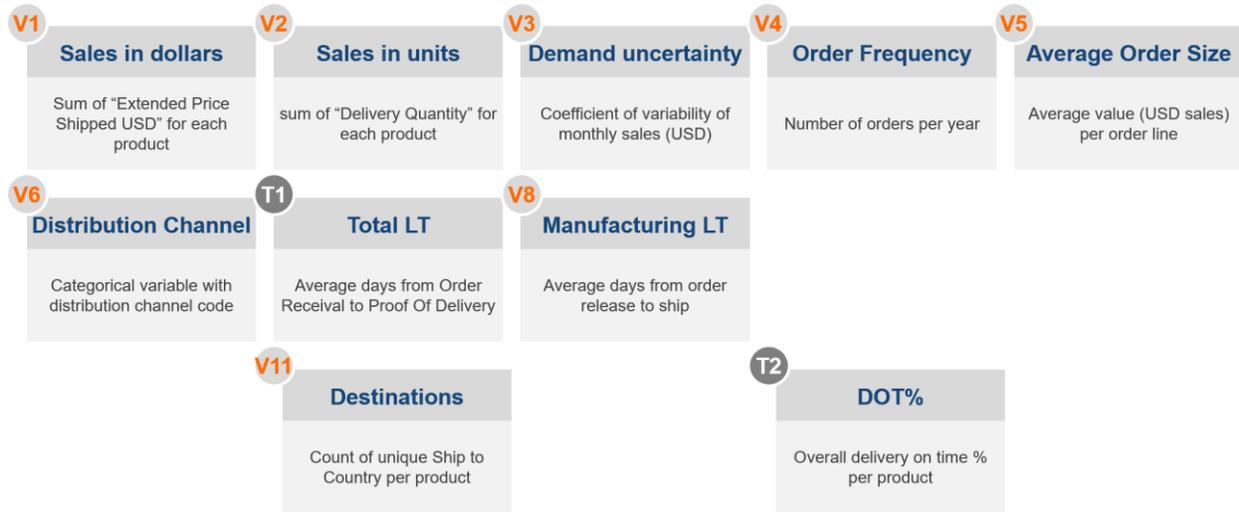
Figure 9: The 10 key variables used in characterizing the portfolio after the initial cut. The target metrics T1 and T2 are set aside for cluster comparison after the model outcome (elaborated by authors)

3.4.1. Exploratory Factor Analysis (EFA)

From the final list of features in Figure 9, some common factors appeared inherent at a glance. A group of features relate to sales volume and revenue, while others relate to complexity in the supply chain. Hence, an EFA was performed to reduce the number of features, from their normalized values per client-product group. As expected from the variable set, two inherent constructs appear, demonstrated by eigenvalues superior to 1 only for up to two factors. The scree plot in Figure 10 helps visualize the relative importance of adding new factors.

Additionally, an arbitrary, yet commonly used, cut-off value of 0.4 was set for the feature loading and applied to all factors. This led to the exclusion of Distribution Channel and Manufacturing Lead-time from the list of features, leaving 8 features for the clustering exercise.
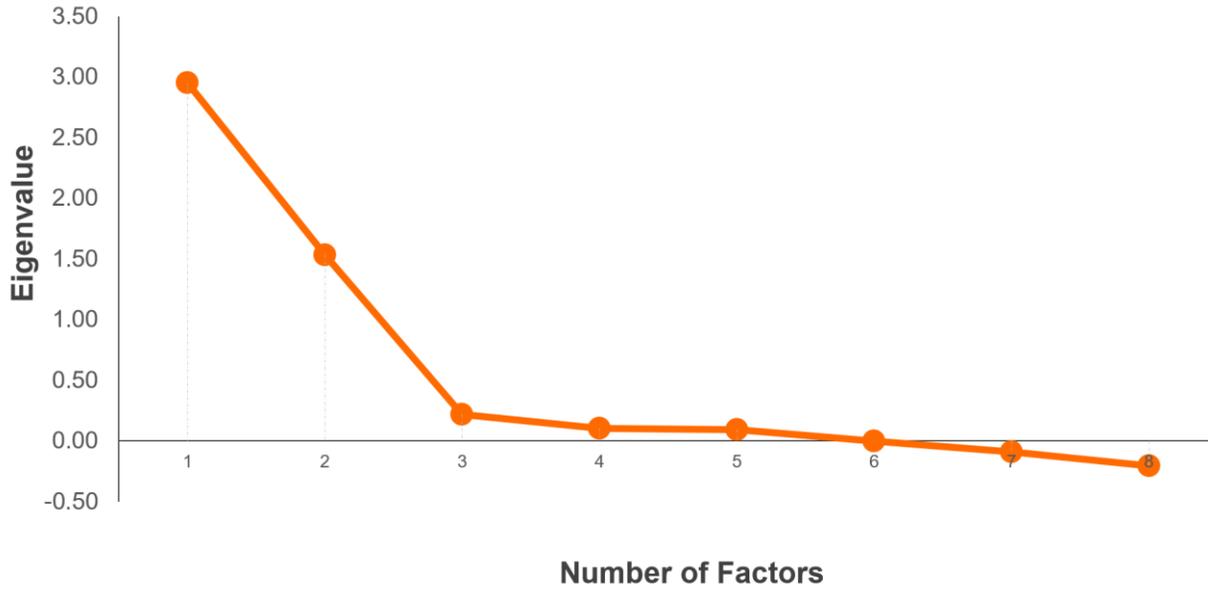
Figure 10: Scree plot of Eigenvalues after EFA (elaborated by authors)

The resulting factors (groups of features) were named **Importance** and **Complexity**, per their composition. Importance included Sales Volume and Revenue features, while Complexity was composed of Demand Uncertainty and Number of Destinations (countries). Table 2 compiles the rotated factor loadings per factor.

Table 2: Load factors for each variable on EFA (elaborated by authors)

Variable	Factor 1 (Importance)	Factor 2 (Complexity)
Sales in Units	0.992	
Sales in Dollars	0.989	
Order Frequency	0.617	
Average Order Size	0.416	
Destinations		0.874
Demand Uncertainty		-0.843

The negative sign on Demand Uncertainty for the Complexity factor on Table 2 was unexpected, as a higher coefficient of variability of demand – and thus more unpredictable purchase pattern – would intuitively lead to higher supply chain complexity. When analyzing the underlying data for Destinations and Demand Uncertainty, it became clear that most products with a high value for Destinations had a more frequent ordering pattern and thus less demand uncertainty, while products with only a single destination

tended to be purchased very unfrequently, presenting high demand uncertainty. Thus, Destinations and Demand Uncertainty present opposing directions in the factor loading for Complexity.

3.4.2. k-Means Clustering

From the EFA analysis, a score for Complexity and Importance was determined for each of the client-product data points. The initial distribution of data was plotted into the two-dimension scatter plot shown in Figure 11.
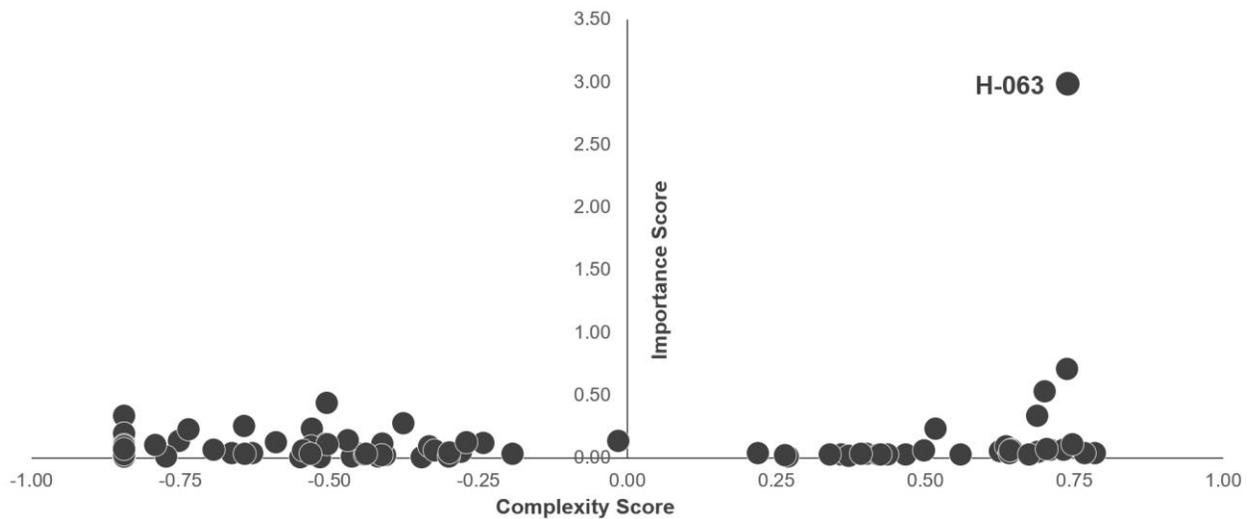
Figure 11: Distribution of client-product pairs on the Complexity-Importance space: Product H-063 is a clear outlier due to its far higher Importance Score (elaborated by authors)

Product 063 for client H, evidently an outlier, was previously identified as such by Lenovo DCG during the initial interviews, stressing the need to address its supply chain in a particular context. Cluster analysis was done on this dataset both including and excluding this outlier. The results were very similar, meaning the product is likely to be clustered on its own cluster and does not affect the grouping of other data points. Also, the optimal number of clusters *k* does not seem to be affected by H-063's presence on the dataset, revealing a rather independent profile in terms of its supply chain when compared to the rest of the portfolio.

For the purpose of the cluster exercise, we adopted the solution (list of client-product points) with all pairs, therefore including H-063. However, the data point was momentarily withdrawn from the dataset for running cluster-mean-analysis and ANOVA, as its statistical calculations are not feasible with only a single point in a cluster.

Considering the remaining client-product records, *k-Means* algorithm was performed for *k* ranging from 2 to 12 clusters. The elbow curve for the filtered data shown on Figure 12 suggests 3 to 4 clusters as the ideal number of clusters, consistent with the all-record analysis. Also, this result is aligned with the hierarchical method.
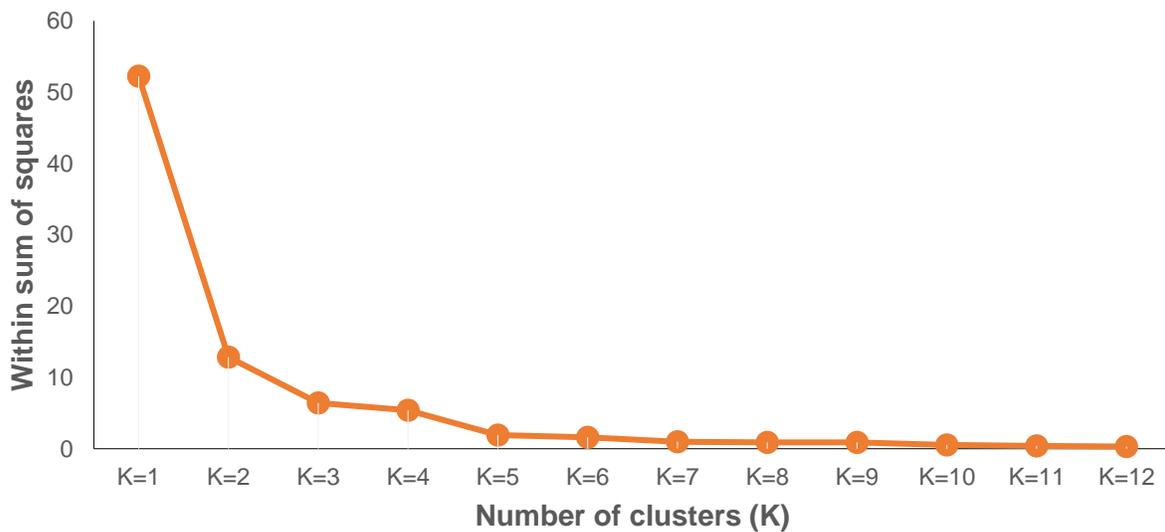
Figure 12: Elbow curve for k-Means without the outlier point H-063 (elaborated by authors)

With results in hand, the composition of each cluster was analyzed for an initial understanding and assessment of results. Including H-063 as Cluster 1, Appendix A shows the Client-Product composition for different *k*. Table 3, extracted from Appendix A, shows compositions for k=4.

Table 3: Cluster composition for k=4 with outlier point H-063 included as Cluster 1.

k	Cluster 1	Cluster 2	Cluster 3	Cluster 4
4	1 product System-level product	29 products Options and spares for Client H	109 products All other clients-products pairs	3 products Client H, one product type

Lastly, Figure 13 shows the graphical representation of the four clusters on the dimensional space, while Figure 14 filters out the outlier point H-063 to explore the other three clusters without the scale distortion.
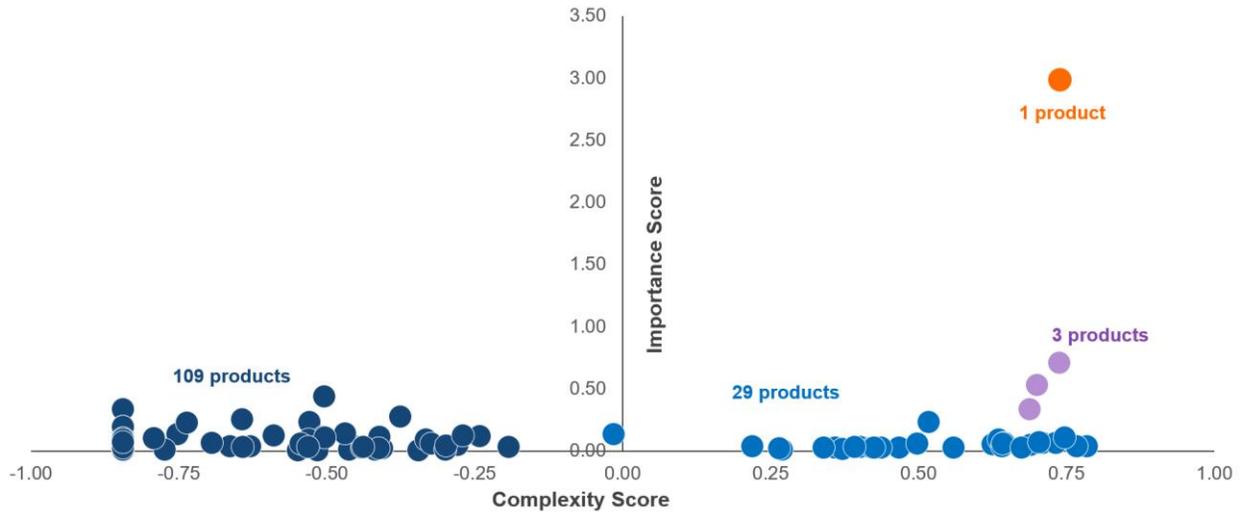
Figure 13: segmentation of data points on the Complexity-Importance space (elaborated by authors)
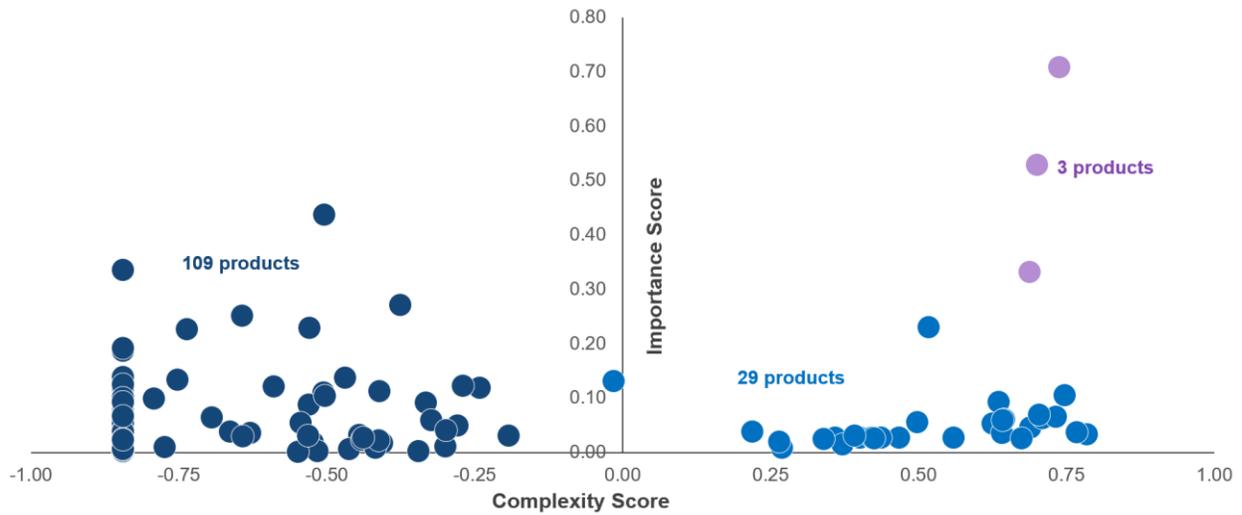
Figure 14: segmentation of data points on the Complexity-Importance space ex-outlier point H-063.

The clear revelation of the analysis is that Client H tends to be separated from all other clients, particularly from by the difference in the Complexity factor. Increasing the number of clusters, the more relevant Importance factor differences become, but this goes to the detriment of cluster heterogeneity.

A means comparison was conducted via ANOVA and Tukey analysis to assess inter-cluster heterogeneity. This combination has been suggested as a good approach to perform configurational analyses (Flynn, Huo, & Zhao, 2010; Revilla, Saenz, & Knoppen, 2013; Cao, Hou, Li, & Zhao, 2015). Removing Cluster 1 from the analysis (given its single data point), the means comparison reveals that for Importance, Clusters 2, 3 and 4 are all significantly different. Regarding the Complexity dimension, we found that Clusters 2 and 4 (H client only) show a not significant means difference at a 0.05 level. Figure 15 and Figure 16 summarize the statistical results for these analyses.

Policies	Cluster 1	Cluster 2	Cluster 3	Cluster 4	ANOVA F, (Sig)
Cluster Description	System-level product	Options and spares for Client H	All other clients	Client H, one product type	
Cluster Size (members)	1	29	109	3	
Avg Importance Score	2.979	0.051	0.058	0.523	561.291 (0.000)
Importance Position	High	Low	Low	Median	
Avg Complexity Score	0.740	0.515	-0.716	0.710	358.085 (0.000)
Complexity Position	High	Median	Low	High	
Target Metric 1 Total Lead Time	47.437	30.943	26.606	46.787	
Target Metric 2 Manufacturing Lead Time	23.373	11.455	9.955	22.784	
Target Metric 3 Delivery on Time %	0.884	0.748	0.763	0.890	

*Figure 15: Complexity and Importance among clusters.*

Dimensions	(I) Cluster ID	(J) Cluster ID	Mean Difference (I-J)	Significance
Importance	2	3	0.095	0.000
		4	-1.490	0.000
	3	2	-0.095	0.000
		4	-1.585	0.000
	4	2	1.490	0.000
		3	1.585	0.000
Complexity	2	3	1.220	0.000
		4	-0.189	0.229
	3	2	-1.220	0.000
		4	-1.408	0.000
	4	2	0.189	0.229
		3	1.408	0.000

Figure 16: Tukey-Test, Pairwise comparison of clusters, excluding cluster 1.

Despite the similitude of two clusters in the Complexity dimension, this solution was presented to Lenovo DCG for discussion. Particularly, we wanted to explore if the differences in the Importance dimension for Client H's clusters are in fact reflected in the unique supply chain requirements and demand signals for those clusters. Also, Figure 16 indicates very similar levels of performance when the target metrics are observed. This evidences the need for additional target metrics (e.g. Cost Efficiency, as discussed in section 3.4), which would enable further differentiation between clusters and an a deeper, more accurate level of segmentation. The reason this is important is that it is clear by the analyses conducted in this capstone that the four clusters are fundamentally different, yet the performance metrics available do not capture differences in execution.

3.5.   Part Five: Research Insights

With the clustering results in hand, it becomes clear that Lenovo DCG's portfolio differs in multiple aspects related to its supply chain. However, that result has limited value to Lenovo DCG's operations if the gap between quantitative insights and the supply chain practice is not bridged. With that in mind, understanding the practical insights behind the clustering outcomes and enabling the internal discussion of supply-chain policies for each cluster is paramount in addressing the issues raised by the subject-matter experts and senior executives early in the project.

Moreover, such a discussion would be able to build a guideline for future work, clearly indicating areas of focus for Lenovo DCG and clarifying how the proposed clusters differ not only in execution results, but also in policies and strategies. In order to do that, a workshop was designed and conducted between authors and a handpicked team of experts from Lenovo DCG. This guideline was designed by the authors with the intention of creating a tangible resource for internal discussion and was shaped as a framework that associated supply chain policies with each identified cluster. This matrix-shaped framework was proposed to the team as a way to differentiate clusters through their policies, given that – in the current practice – supply chain policies are not defined for each cluster suggested by this capstone. Policies were grouped

into five categories, based on the initial business immersion in Lenovo DCG and the weekly meetings with the company's internal project team: Sourcing Policies, Inventory Policies, Production Policies, Fulfillment Policies, and Policies to Customers. Figure 17 illustrates the framework and provides an example of what the populated matrix could look like after the workshop.

The workshop took place in two parts, with two days in between them. The first consisted of a four-hour meeting aimed at (i) explaining the results and getting feedback and business insights from Lenovo DCG's team in order to validate the results; (ii) proposing the policy framework to the team, shaping it according to their vision, and populating the matrix based on current Lenovo DCG's practices in order to identify similarities and differences among clusters, policy-wise. Also, discussions on how policies could be changed in the future to better suit the clusters' perceived requirements were explored qualitatively. These feedbacks and insights eventually guided a few additional tasks that were incorporated into the study during the 2-day gap between workshop parts and helped shape the final version of the policy-cluster matrix.

Policies	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Sourcing Policies	Very important	Not as important	Somewhat important	Very important
Inventory Policies	Very important	Somewhat important	Somewhat important	Not as important
Production Policies	Not so important	Not as important	Very important	Not as important
Fulfillment Policies	Not as important	Somewhat important	Not as important	Very important
Policies to Customers	Somewhat important	Very important	Not as important	Not as important

Very important: policy fundamentally defines the cluster's supply chain.
Somewhat important: policy affects the cluster's supply chain, but if changed would not necessarily dismantle the cluster
Not as important: policy does not directly affect the cluster's supply chain.

Figure 17: Policy-cluster matrix with exemplified values for policies' weights (elaborated by authors).

The second part consisted of a two-hour meeting to go over the final matrix, quantitative results from additional analyses, and validate the product and its application at Lenovo DCG. Lastly, the study's limitations and important remarks regarding the analytical methodology were posed to the sponsoring company, and recommendations were made on how to further improve this study in the future.

4   Results and Discussions

4.1.   Key Insights

Building on the existing research on supply chain segmentation, our research contributes to the understanding and application of this concept at Lenovo DCG. This project applied data covering three years of sales records, as well as operational data and inferred statistical information to identify homogeneous client-product clusters among the firm's rapid growing division of Hyperscale computing.

Overall, the cluster analysis along with ANOVA and Tukey tests were able to confirm four clusters an appropriate number of clusters among the firm's hundreds of client-product pairs, based on the two dimensions, Complexity and Importance. Among those four clusters, one is highlighted as a strong outlier (H-063 data point), and all analyses were re-run without it, with no significant change in results. For all other client-product pairs, the same cluster composition resulted with or without the outlier.

Additional clusters could be seen in the analyses for k-values higher than 4, yet we felt that the analytical benefits of such a decision would not compensate for the added complexity of having extra clusters, as evidenced by the elbow chart illustrated in Figure 12. After all, while the k-Means algorithm accepts any number of k clusters smaller than the number of observations (n), the final delivery of this work and the research questions raised at the beginning of the study become increasingly complex to be addressed if the number of groups is too high. There is – as is often the case in supply chain – a recommended level of granularity in every analysis and discussion to ensure practical application of the quantitative results.

That said, this study found that Lenovo DCG's portfolio is not homogeneous when it comes to supply chain practices or policies. As a final illustration of such differences, Figure 18 shows the policy-cluster matrix filled with the validated weights and summarized insights from the workshop. In this matrix, each cluster has its guiding-policies identified and discussed and insights on what could be an area of focus for future adjustment, discussion, or further analytical work were necessary. While the matrix presents a skimmed

version of the content discussed in the workshop, complete notes covering these topics were shared among the project members. Also, perhaps the most interesting highlight of Figure 18 is how well it shows the different weight (or significance) of policies for each cluster. While Cluster 1 is strongly defined and affected by Sourcing Policies, Cluster 2 is mostly driven and regulated by Inventory Policies, given the nature of its products. Other examples of quantitative and qualitative similarities and differences among clusters are present not only in this matrix, but also throughout the analyses in this study. They allow us to affirm that the company could segment its supply chain execution based on predefined clusters. This could likely improve adherence to policies that are often conflicting between clusters, improve service levels, and prevent inefficiencies.

Cluster Description	System-level product	Options and spares for Client H	All other clients	Client H, one product type
Cluster Size	1	29	109	3
Policies	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
Sourcing Policies	Very important, but mostly dictated by client's requirements (BoM, source of supply, cost) with the minority of parts procured directly by Lenovo DCG.	Somewhat important, products are mainly commodities in comparison to Cluster 1, all single-sourced from other continents (Client H)	Very important, similar products as C1, but entirely managed by Lenovo, which acts as a traditional supplier, not a systems integrator for its clients as in Cluster 1.	
Inventory Policies	Somewhat important, but mostly dictated by client's requirements. Cash flow-wise, a few commodities receive hyper focus: SSD, CPU, memory (80% of BoM value).	Very important, the company purchases to stock based on forecast	Somewhat important, mostly focused on scale, since Lenovo is in control of the design, BoM, sourcing, and inventory (what, where, and how much).	**Same as Cluster 1** Products in this cluster are procured, stored, assembled, and shipped to order after demand is awarded.
Production Policies	Not as important, no specific rules that directly affect the cluster	Not as important, components not manufactured by Lenovo	**Heterogeneous Cluster**: Includes several clients and different supply chain policies, thus execution is not homogeneous within the group. A few larger customers could be separated from this cluster given Lenovo's focus on sales growth (share of wallet). Also, some clients have a more domestic supply chain in their respective geographies, which generates particularities. Further segmentation could help us better understand these points.	Supply chain guidelines are pre-arranged by Client H and thus Cluster 4 is subject to similar supply chain policies and execution restrictions as Cluster 1.
Fulfillment Policies	Somewhat important, as demand is awarded and timeline is designed in reverse. This policy affects all tasks.	Not as important by itself, blends with Inventory Policies		
Policies to Customers	Very important, with higher flexibility found in relationships with larger clients.	Items are mostly provided as *off-the-shelf* replacement products for servers, supply chain service is not as flexible as Cluster 1, with price becoming an important client decision factor.		

Very important: policy fundamentally defines the cluster's supply chain.
Somewhat important: policy affects the cluster's supply chain, but if changed would not necessarily dismantle the cluster
Not as important: policy does not directly affect the cluster's supply chain.

Figure 18: Policy-cluster matrix with insights and weights from the workshop (elaborated by authors)

With that in mind, this study addresses research questions RQ1, RQ2, RQ3, RQ4, and RQ5 and supports the hypothesis that there is not one unique way in which supply chains could be managed in order to adequately align execution with product-client characteristics (Sanchez et al., 2015). Figure 19 illustrates the steps taken in order to address such questions.


Dataset (n=142, p=6)	EFA (d=2)	K-Means UML (k=4)	Policies for Clusters
List of all selected client-product pairs with the variables ("columns") that describe them.	Combining variables to explain products with fewer dimensions, only the significant ones.	Based on the identified dimensions, explore similar product clusters.	Discussion and insights on how clusters perform and to which policies they are subject to.

Figure 19: Summary of project steps to address the research questions (elaborated by authors)

Additionally, this project acts as a platform model that can be further used as more data becomes available for Lenovo DCG. As previously discussed, data limitations prevented key variables from being included, and we believe that Profitability and Inventory information could significantly improve the model capacity to provide insightful results. The four clusters solution presented did not show statistically significant differences (at 0.05) for Total Lead Time or Average Delivery on Time (%), yet Lenovo DCG acknowledges that Average Inventory on Hand and Profitability differences are perceived among clusters (particularly between clusters 1 and 3). This serves as a motivating factor for developing multiple supply chain approaches. Despite not being able to show this analytically, the company does find great value in providing a methodological approach, with multivariate description of Client-Product pairs, that presents them with the opportunity to segment and align their supply chain strategy.

4.2. Future Lenovo Work Development

As for future development and improvements, the company embraces these results as a powerful momentum generator to engage across the organization. As a starting point it serves to request complementary key variables, that have already been identified, such as Profitability and Inventory

Position. In Lenovo's pursue for a "Customer Driven" supply chain strategy, this exercise goes beyond a classification exercise via Importance or Complexity dimensions. At a cluster level, a deeper discussion on "Customer Experience" would be engaged, to identify the critical expectations from the customer.

Through different methodologies Lenovo DCG would assess what are the critical expectations for a cluster of Client-Product such as Lead Time, Customer Delivery Reliability, Demand Visibility, Customer Data Transparency, Service Quality (e.g. delays at commissioning), or a mixture of multiple. Paired with Cost or Profitability data, the combination with Customer Experience allows for an assessment of the overall Cost Efficiency of the supply chain, quantifying the cost of execution and comparing it to the delivered Customer Experience. One important remark is that Customer Experience is a qualitative and subjective concept, that will need to be defined, in our vision, by Lenovo DCG's team for each cluster of clients. A more complex, accurate, and complete set of expectations are paramount in exploring the tradeoffs between service level for customers and costs for the company.

The establishment of such Cost Efficiency metric and the Supply Chain characteristics mentioned above would allow Lenovo DCG to periodically asses the performance of their approach (clusters definition and supply chain policies design) to market segments. This is often referred to as an Evaluation Loop, and consists in periodic reruns of the model with comparisons between outputs in order to identify which levers (supply chain policies, in this case) better ensure target metrics optimization (e.g. maximization of Customer Experience, maximization of Cost Efficiency, or minimization of Total Lead Time). Changing policies and remeasuring the new environment – through either data collection or simulation – creates a formal, robust process that evidences the individual impacts of such changes. When an Evaluation Loop is considered for this segmentation exercise, a methodological approach can be developed for Customer Driven Supply Chain, as illustrated in Figure 20. The iterative process should be run at a predetermined frequency (e.g. monthly, quarterly or annually), or against a disruption in the industry, such as entry/exit of clients or competitors, tariff changes or changes in manufacturing costs – note that this analysis could be

run with a simulation approach i.e. evaluate (simulate) Cost Efficiency as the "Total Expected Supply Chain Cost" or "Profitability" for the desired level of Customer Experience.
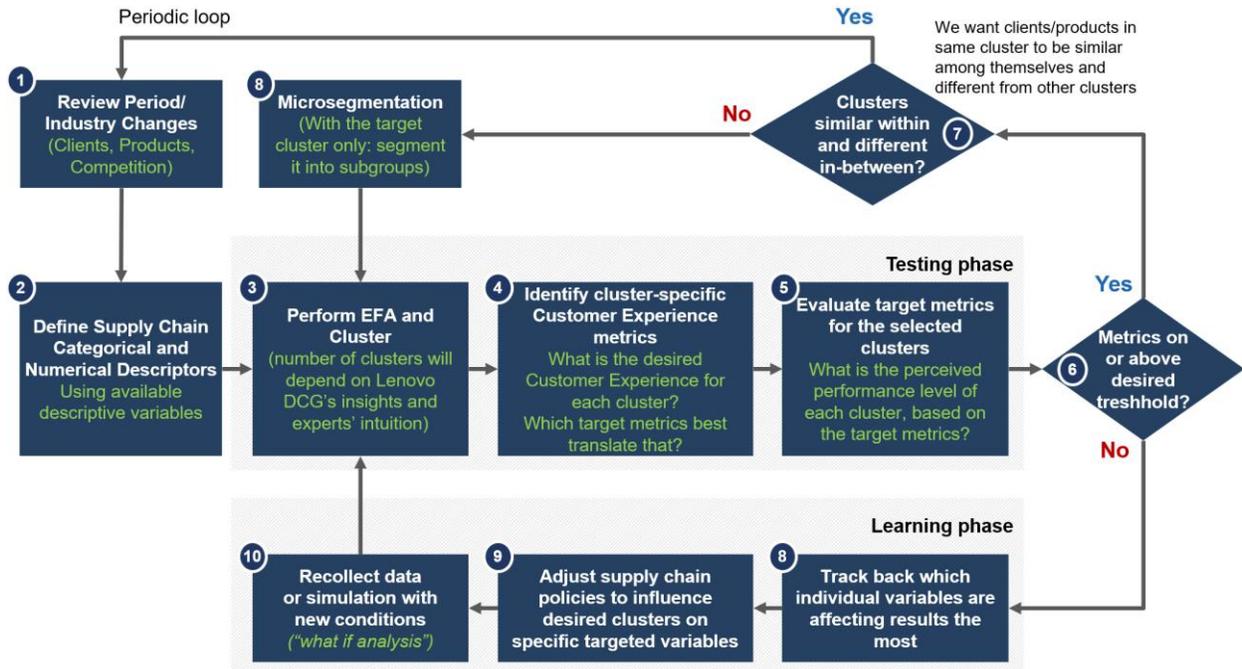

Figure 20: Implementation for Customer Driven Supply Chain Policies with an Evaluation Loop

(elaborated by authors)

5   Conclusions

5.1.   Final Remarks

The scope of this project was to provide analytical support to Lenovo DCG on segmenting their supply chain to better service clients and improve bottom line results. As the company's Hyperscale division grows rapidly, senior management believes that additional levels of capability design will be paramount in addressing the challenges that come with such evolution. Based on this belief, the main objective of this research project was to answer two questions: first, what are potential groups of products among the firm's portfolio and how do those groups differ? Second, what are individual policies that could adequately serve each of those groups?

We approached this challenge through a clustering exercise and found that four different client-product groups can be identified from among the 142 points that compose Hyperscale's 2018 portfolio. Those four clusters should be treated as individual, independent groups and their policies should be defined individually based on the client-product nature. Cluster 1 is an outlier and is composed of only one product, 063, supplied to one single client, Client H. Cluster 2 is a group of 29 components (tier-2 and/or tier-3 Bill of Material (BoM) goods) that are also supplied exclusively to Client H. Cluster 3 has 109 products and could be further segmented in order to define accurate policies for more segregated groups; a few large clients in the cluster, for example, have a clear share-of-wallet expansion target for the foreseeable future and that should be taken into account, among other things. Cluster 4 represents three items supplied to Client H with demand and supply chain particularities that differentiate them from Cluster 2.

Our quantitative research was improved by a qualitative discussion with subject-matter experts from Lenovo DCG that explored how different supply chain policies affect each identified cluster. This allowed us to define the current state of supply chain policies for each of the four clusters, which we believe will guide future discussions within the firm to adequate parts of their supply chain to a more segmented reality.

In such reality, clusters with specific needs (e.g. lower lead times, lower costs, high flexibility, high visibility, etc.) are handled with capabilities (e.g. procurement pooling, inventory planning, transportation consolidation, among others) that adequately match their requirements.

The goals of this capstone project were trifold: (i) to review current frameworks present in the supply chain segmentation literature; (ii) to identify Lenovo DCG's key customer and product groups in its portfolio through quantitative scientific methods; and (iii) to propose individual supply chain policy guidelines for each identified group, creating baselines for future discussion. We believe this capstone has successfully fulfilled its objectives and has contributed to advancing knowledge of supply chain segmentation by providing empirical evidence on its benefits in the high-tech industry: both the existence of clusters within Lenovo DCG's portfolio as well as clear differences in supply chain policies. We hope that further studies will build upon our results to produce relevant knowledge on this increasingly important field and that Lenovo DCG works towards continuously improving their supply chain to better serve their clients and provide sustainable value to their shareholders.

5.2.   Opportunities for Future Research

Although our study contributes to both academic audience and supply chain practitioners by aiming at bridging the gap between quantitative clustering and supply chain segmentation, objectives were targeted at providing analytical value to Lenovo DCG. That said, additional research is recommended to fully understand the potential segmentation in Lenovo DCG's portfolio and its operational and financial impacts. Such impacts should be calculated and verified with real operational data after the implementation of new policies, permitting the development of the feedback loop illustrated on Figure 20. Our research scratches the surface on a number of techniques and supply chain topics, and we see significant opportunity to dive deeper in quantitative research, data collection, and analysis.

Further availability of data would enrich the analysis. There are elements in the descriptive features that could be added to asses, for example: the product Value-Density, the Integration or Digitalization of Information (relationship with client), the product Manufacturing Complexity, or the scarcity of supplies. This could add a third or more dimensions, that would accurately segment, from a Supply Chain perspective, the Client-Product pairs.

As discuss on Chapter 4, from a performance criterion for evaluation of clusters, the Efficiency of the supply chain policies could be assessed as the Cost or Profitability to reach a certain level of Customer Experience. Furthermore, this would enable a periodic and methodological approach for organizing and optimizing a customer-driven supply chain.

Finally, It would be important to analyze how the potential segments among other business units at Lenovo DCG compare to Hyperscale in terms of policies. It is possible that different business units (e.g. DCI, HPC/AI, IoT, SDI, etc.) share similar customer requirements and policies and, thus, supply chain capabilities. Synergies would then be beneficial for the firm and its clients and would likely enhance the existing relationships between business units. This would expose our conclusions to additional angles and strengthen the empirical validity of this research.

References

Agarwal, A., Shankar, R., & Tiwari, M. K. (2006). Modeling the metrics of lean, agile and leagile supply chain: An ANP-based approach. *European Journal of Operational Research*, *173*(1), 211–225. https://doi.org/10.1016/j.ejor.2004.12.005

Alicke, K., & Forsting, M. (2017). Supply Chain Segmentation Scientific Frameworks. In M. Protopappa-Sieke & U. W. Thonemann (Eds.), *Supply Chain Segmentation* (pp. 5–13). https://doi.org/10.1007/978-3-319-54133-4_2

Aryuni, M., Madyatmadja, E. D., & Miranda, E. (2018). Customer Segmentation in XYZ Bank Using K-Means and K-Medoids Clustering. *2018 International Conference on Information Management and Technology (ICIMTech)*, 412–416. https://doi.org/10.1109/ICIMTech.2018.8528086

Callinan, L. (2018). *How to Leverage Logistics to Deliver Great Customer Experiences* (No. G00375720; p. 11). Gartner.

Cao, Z., Hou, B., Li, Y., & Zhao, X. (2015). The impact of organizational culture on supply chain integration: a contingency and configuration approach. *Supply Chain Management: An International Journal*, *20*(1), 24–41.

Chan, A. T. L., Ngai, E. W. T., & Moon, K. K. L. (2017). The effects of strategic and manufacturing flexibilities and supply chain agility on firm performance in the fashion industry. *European Journal of Operational Research*, *259*(2), 486–499. https://doi.org/10.1016/j.ejor.2016.11.006

Childerhouse, P., Aitken, J., & Towill, D. R. (2002). Analysis and design of focused demand chains. *Journal of Operations Management*, *20*(6), 675–689. https://doi.org/10.1016/S0272-6963(02)00034-7

Christopher, M., & Gattorna, J. (2005). Supply chain cost management and value-based pricing. *Industrial Marketing Management*, *34*(2), 115–121. https://doi.org/10.1016/j.indmarman.2004.07.016

Christopher, M., & Towill, D. (2000). *Marrying the lean and agile paradigms*.

Christopher, M., & Towill, D. R. (2002). Developing Market Specific Supply Chain Strategies. *The International Journal of Logistics Management*, *13*(1), 1–14. https://doi.org/10.1108/09574090210806324

Dhanachandra, N., Manglem, K., & Chanu, Y. J. (2015). Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm. *Eleventh International Conference on Communication Networks, ICCN 2015, August 21-23, 2015, Bangalore, India Eleventh International Conference on Data Mining and Warehousing, ICDMW 2015, August 21-23, 2015, Bangalore, India Eleventh International Conference on Image and Signal Processing, ICISP 2015, August 21-23, 2015, Bangalore, India*, *54*, 764–771. https://doi.org/10.1016/j.procs.2015.06.090

Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern Classification*. Retrieved from https://books.google.com/books?id=Br33IRC3PkQC

Farid, D. M., Zhang, L., Rahman, C. M., Hossain, M. A., & Strachan, R. (2014). Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. *Expert Systems with Applications*, *41*(4, Part 2), 1937–1946. https://doi.org/10.1016/j.eswa.2013.08.089

Fisher, M. L. (1997). What is the right supply chain for your product. *Harvard Business Review*, *75*(2), 105–116.

Flynn, B., Huo, B., & Zhao, X. (2010). The Impact of Supply Chain Integration on Performance: A Contingency and Configuration Approach. *Journal of Operations Management - J OPER MANAG*, *28*, 58–71. https://doi.org/10.1016/j.jom.2009.06.001

Gattorna, J. L., Chorn, N. H., & Day, A. (1991). Pathways to customers: reducing complexity in the logistics pipeline. *International Journal of Physical Distribution & Logistics Management*.

Gattorna, John L., & Walters, D. (1996). *Managing the supply chain : a strategic perspective.* Basingstoke: Macmillan Business.

Godsell, J., Diefenbach, T., Clemmow, C., Towill, D., & Christopher, M. (2011). Enabling supply chain segmentation through demand profiling. *International Journal of Physical Distribution & Logistics Management*, *41*(3), 296–314. https://doi.org/10.1108/09600031111123804

Godsell, J., Harrison, A., Emberson, C., & Storey, J. (2006). Customer responsive supply chain strategy: An unnatural act? *International Journal of Logistics Research and Applications*, *9*(1), 47–56. https://doi.org/10.1080/13675560500534664

Goldsby, T. J., Griffis, S. E., & Roath, A. S. (2006). MODELING LEAN, AGILE, AND LEAGILE SUPPLY CHAIN STRATEGIES. *Journal of Business Logistics*, *27*(1), 57–80. https://doi.org/10.1002/j.2158-1592.2006.tb00241.x

Hill, T. (1985). *Manufacturing strategy. Text and cases.* London: Macmillan.

Hill, T. (2000). *Manufacturing strategy: text and cases*. Irwin McGraw Hill.

Hjort, K., Lantz, B., Ericsson, D., & Gattorna, J. (2013). Customer segmentation based on buying and returning behaviour. *International Journal of Physical Distribution & Logistics Management*.

Kalyani, C., Ramudu, K., & Reddy, G. R. (2018). A review on optimized K-means and FCM clustering techniques for biomedical image segmentation using level set formulation. *Biomedical Research*, *29*(20). https://doi.org/10.4066/biomedicalresearch.29-18-1052

Kaufman, L., & Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction To Cluster Analysis*. https://doi.org/10.2307/2532178

Kulis, B., & Jordan, M. I. (2011). Revisiting k-means: New Algorithms via Bayesian Nonparametrics. *ArXiv:1111.0352 [Cs, Stat]*. Retrieved from http://arxiv.org/abs/1111.0352

Lam, D., & Wunsch, D. C. (2014). Chapter 20 - Clustering. In P. S. R. Diniz, J. A. K. Suykens, R. Chellappa, & S. Theodoridis (Eds.), *Academic Press Library in Signal Processing* (Vol. 1, pp. 1115–1149). https://doi.org/10.1016/B978-0-12-396502-8.00020-6

Lee, H. L. (2002). *Aligning Supply Chain Strategies with Product Uncertainties*. *44*(3), 16.

Murtagh, F. (1983). A Survey of Recent Advances in Hierarchical Clustering Algorithms. *The Computer Journal*, *26*(4), 354–359. https://doi.org/10.1093/comjnl/26.4.354

Ng, Raymond T., H., Jiawei. (2002). CLARANS: a method for clustering objects for spatial data mining. *IEEE Transactions on Knowledge and Data Engineering*, *14*(5), 1003–1016. https://doi.org/10.1109/TKDE.2002.1033770

Olavson, T., Lee, H., & DeNyse, G. (2010). A Portfolio Approach to Supply Chain Design. *Supply Chain Management Review*, *14*(4), 20–27.

Payne, A., & Frow, P. (2006). Customer Relationship Management: from Strategy to Implementation. *Journal of Marketing Management*, *22*(1–2), 135–168. https://doi.org/10.1362/026725706776022272

Peck, H., Christopher, M., & Towill, D. (2006). A taxonomy for selecting global supply chain strategies. *The International Journal of Logistics Management*, *17*(2), 277–287. https://doi.org/10.1108/09574090610689998

Porter, M. (1980). *Competitive Strategy: Techniques for analyzing industries and competitors*. New York: Free Press.

Protopappa-Sieke, M., & Thonemann, U. W. (Eds.). (2017). *Supply Chain Segmentation: Best-in-Class Cases, Practical Insights and Foundations* (1st ed.). Springer International Publishing.

Revilla, E., Saenz, M. J., & Knoppen, D. (2013). Towards an empirical typology of buyer-supplier relationships based on absorptive capacity. *International Journal of Production Research*, *51*(10), 2935–2951.

Roh, J., Hong, P., & Min, H. (2014). Implementation of a responsive supply chain strategy in global complexity: The case of manufacturing firms. *International Journal of Production Economics*, *147*, 198–210. https://doi.org/10.1016/j.ijpe.2013.04.013

Rubin, M. W. (1997). Creating Customer-Oriented Companies. *Prism*, *4*, 5–27. Arthur D. Little.

Sáenz, M. J., Revilla, E., & Acero, B. (2018). Aligning supply chain design for boosting resilience. *Business Horizons*, *61*(3), 443–452. https://doi.org/10.1016/j.bushor.2018.01.009

Sanchez, R. (1997). Preparing for an Uncertain Future: Managing Organizations for Strategic Flexibility. *International Studies of Management & Organization*, *27*(2), 71–94.

Sanchez Rodriguez, V., Harris, I., & Manson, R. (2015). Horizontal logistics collaboration for enhanced supply chain performance: and international retail perspective. *Supply Chain Management*, *20*(6), 631–647.

Sanders, N. R. (2016). How to Use Big Data to Drive Your Supply Chain. *California Management Review*, *58*(3), 26–48. https://doi.org/10.1525/cmr.2016.58.3.26

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., … Lin, C.-T. (2017). A review of clustering techniques and developments. *Neurocomputing*, *267*, 664–681. https://doi.org/10.1016/j.neucom.2017.06.053

Skipworth, H., Godsell, J., Wong, C. Y., Saghiri, S., & Julien, D. (2015). Supply chain alignment for improved business performance: an empirical study. *Supply Chain Management: An International Journal*, *20*(5), 511–533. https://doi.org/10.1108/SCM-06-2014-0188

Swan, D., Pal, S., & Lippert, M. (2009). Finding the perfect fit. *CSCMP's Supply Chain Quarterly*, (4). Retrieved from https://www.supplychainquarterly.com/print/scq200904segmentation/

Vora, P., & Oza, B. A. (2013). *A Survey on K-mean Clustering and Particle Swarm Optimization*.

Walters, D. (2006). Effectiveness and efficiency: the role of demand chain management. *The International Journal of Logistics Management*, *17*(1), 75–94. https://doi.org/10.1108/09574090610663446

Appendix A

K-Means cluster composition, for 2 to 6 clusters (elaborated by authors)

# Clusters	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
2	**1 Client-Product** H-063	**109 Client-Product** All other H products, and client-product records	---	---	---	---
3	**1 Client-Product** H-063	**32 Client-Product** 31 products of client H 1 Product client B	**109 Client-Product** 20 H-Client products, all other client-product records	---	---	---
4	**1 Client-Product** H-063	**29 products** H-Client, low complexity products	**109 Client-Product** 20 H-Client products, all other client-product records	**3 Client-Product** H-Client, high complexity products	---	---
5	**1 Client-Product** H-063	**28 Client-Product** H-Client, low complexity products	**79 Client-Product** 13 H-Client products, 66 other client-product records	**3 Client-Product** H-Client products, high complexity products	**31 Client-Product** 7 H-Client products, 24 other client-product records	---
6	**1 Client-Product** H-063	**28 Client-Product** H-Client, low complexity products	**78 Client-Product** 13 H-Client products, 62 other client-product records	**3 Client-Product** H-Client, high complexity products	**27 Client-Product** other client-product records	**5 Client-Product** other client-product records