Hub and Spoke Network Design for the Inbound Supply Chain

By

Olufemi Oti

B.S. Industrial Engineering
University of Florida, 2006

Submitted to the MIT Sloan School of Management and the Engineering Systems Division in Partial Fulfillment of the Requirements for the Degrees of

Master of Business Administration
and
Master of Science in Engineering Systems

In conjunction with the Leaders for Global Operations Program at the Massachusetts Institute of Technology

June 2013

© 2013 Olufemi Oti. All rights reserved.

ARCHIVES
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
MAY 30 2013
LIBRARIES

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Signature of Author _____
Engineering Systems Division, MIT Sloan School of Management
May 10, 2013

Certified by _____
Karen Zheng, Thesis Supervisor
Assistant Professor of Operations Management, Sloan School of Management

Certified by _____
Chris Caplice, Thesis Supervisor
Senior Lecturer, Engineering Systems Division
Executive Director, Center for Transportation and Logistics

Accepted by _____
Olivier L. de Weck, Chair, Engineering Systems Division
Professor of Aeronautics and Astronautics and Engineering Systems

Accepted by _____
Maura Herson, Director of MIT Sloan MBA Program
MIT Sloan School of Management

This page intentionally left blank.

Hub and Spoke Network Design for the Inbound Supply Chain

By

Olufemi Oti

Submitted to the MIT Sloan School of Management and the Engineering Systems Division on May 10, 2013 in Partial Fulfillment of the Requirements for the Degrees of Master of Business Administration and Master of Science in Engineering Systems

Abstract

Amazon is one of the world's leading retailers. At the core of Amazon's business model is providing consumers with endless selection, and as a result, the large number of vendors used to provide that selection greatly increases the complexity and cost of operating the inbound supply chain. This growth has also created many opportunities for the company to leverage its size and scale to lower transportation costs and improve supply chain flexibility. This project explores implementing load consolidation strategies within the "Hub and Spoke" distribution framework to provide these benefits.

As ~65% of total unit volume from the inbound transportation program managed by Amazon is shipped as costly less-than-truckload (LTL) or small-parcel (SP) freight, there are significant opportunities to use consolidation hubs throughout the inbound network to reduce spend on LTL and SP in favor of more cost effective full truckload (TL) shipments. To evaluate the opportunity and provide the inbound team with a useful strategic planning tool, a comprehensive network optimization model was targeted as a project deliverable. After researching the current state of the inbound transportation network through departmental interviews and visits to carrier hubs and fulfillment centers, key inputs were identified to feed the model.

The mixed integer program solution uses these inputs to minimize total inbound transportation cost for the network subject to expected transit time performance targets by choosing what consolidation hubs and destination lanes freight should be routed to. Using a data-set of shipments originating in the Southwestern geography, an average saving of 13.7% on annual LTL and SP spend was projected by routing 37% of freight volume through consolidation hubs. Results showed freight density as an important driver in savings. In areas with more originating freight, outbound full truckloads can be filled more readily and hence consolidation opportunities can be taken advantage of more often.

This tool and the supporting analyses will help the inbound transportation organization uncover more cost saving opportunities in routing freight through its growing network. In addition to financial cost savings, the strategy will increase supply chain flexibility, reduce environmental impact, and can help increase Amazon's control over the end-to-end inbound transportation network.

Thesis Supervisor: Karen Zheng
Title: Assistant Professor of Operations Management, Sloan School of Management

Thesis Supervisor: Chris Caplice
Title: Senior Lecturer, Engineering Systems Division

This page intentionally left blank.

Acknowledgments

I would like to thank Amazon for sponsoring my internship, for providing the resources that make this thesis possible, and for the opportunity to learn and further develop my skills. I would like to specifically thank the following individuals, who had a profound impact on my work during this internship and project: Bob Flannery, Mark Michener, Akshay Katta, and Bijal Mehta.

I would also like to thank my thesis advisors, Professor Karen Zheng and Professor Chris Caplice, for their advice and support during my internship.

I would like to acknowledge the Leaders for Global Operations Program with special thanks for providing me with such an excellent opportunity. Particularly, I would like to thank Don Rosenfield, Jeff Shao, Davicia Smith, Patty Eames, and Leah Schouten for all of their help in making this possible.

Finally, I would like to thank my family for their assistance and encouragement throughout this journey.

Note on Proprietary Information

In the interest of protecting Amazon's competitive and proprietary information, figures presented throughout this thesis may have been disguised, are solely for the purpose of illustration, and may not represent actual Amazon data.

This page intentionally left blank.

Table of Contents

Table of Figures

Tables

Figures

Equations

1 Project Overview and Background

1.1 Introduction

Amazon is one of the fastest growing retailers on the planet with total sales in 2011 of $48B and growth from 2010 to 2011 of 40%[1]. Important to Amazon's goal of providing a place where "people can find and discover virtually anything they want to buy online" (Amazon 2013b) is providing the wide reaching product selection to support that. The organization strives to sell customers anything from books, to batteries, to high fashion clothing at the most competitive prices. This business strategy, along with Amazon's focus on operational excellence and customer experience has been key to its rapid topline growth but it does create unique challenges for operating such a large and diverse supply chain. Much of the focus within Amazon's Supply Chain organizations has historically rested with managing the outbound supply chain since it more directly impacts the customer experience and is much more costly. This research project focuses its attention on a part of the Amazon supply chain organization which is getting more attention – the inbound transportation organization.

In order to continue to lower costs and improve supply chain speed Amazon must continue to adapt as the size and complexity of its vendor base continues to increase. This project looks towards implementing load consolidation within the "Hub and Spoke" distribution framework as one potential strategy to aid Amazon in its growth and transition. This primarily involves consolidating freight at pooling points closer to the location of vendor origin to take advantage of higher truck utilization over the lengthier portion of transit. Over the course of this research paper the merits of this framework and an optimization model to aid in network design for the load consolidation strategy will be overviewed. Before getting in to those details an introduction to Amazon as a company, the inbound transportation organization and an overview of the general thesis structure will help provide the reader with an appropriate background.

[1] Source: 2010 – 2011 Amazon Annual Reports

1.2 Amazon.com – A Leading Retailer

Amazon.com was created in 1994 by founder Jeff Bezos. It started in Seattle after Bezos, who had been doing research on the internet for hedge fund D.E. Shaw, realized that book sales would be a perfect fit for the e-commerce platform. The website was launched in July of 1995 and by September had achieved sales of $20K per week. Over the years Amazon has increased its product offerings from books, to electronics, to home goods, and has developed its own all-time best seller, the kindle e-reader (Hoover's 2013). The organization's growth over the past four years has been the most staggering - growing by 30% over 2007-2008 and as much as 41% over 2010-2011[2]. Key to Amazon's growth and success over the years has been the strong focus on customer satisfaction and long term thinking. Its mission has been, "To be Earth's most customer-centric company where people can find and discover anything they want to buy online" (Amazon 2013a) and it has followed that mission to great extents knowing that the long term payoff will be significant. The Amazon virtuous cycle, created by Jeff Bezos, is representative of that and is considered the flywheel of the company's operation:

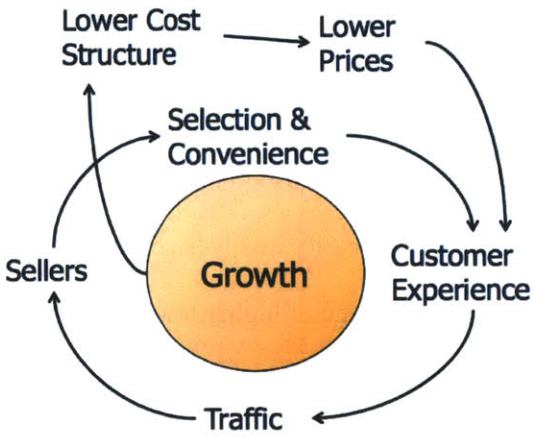


Figure 1 - Amazon Virtuous Cycle[3]

[2] Source: Amazon 2008-2011 Income Statements
[3] Kyle Doherty, "The Power of a Simple Business Model," http://www.kydoh.com/page/2/ (accessed 02/28, 2013).

The diagram depicts a reinforcing feedback loop explaining that by increasing product selection and convenience, customer experience is improved which will improve traffic to the Amazon website and hence improve attractiveness to sellers. This will in turn drive more sellers to Amazon and again more selection and convenience for its consumers. This virtuous cycle has been key to Amazon's growth and success.

1.3 The Inbound Transportation Organization at Amazon

The Inbound Transportation Organization, like many other groups at Amazon, is rapidly increasing in size, in line with the company's rapid growth over the years. The group is responsible for managing the end-to-end transportation considerations for getting products from Amazon's many vendors in to its numerous fulfillment center (FC) locations spread throughout the network.

While historically, a much larger focus has been placed on the Outbound Transportation Organization, a much higher focus has been emerging on improvements to the Inbound Transportation Organization. As retail sales are primarily through the online channel, the need to ship individual shipments to customers (outbound shipments) becomes a primary driver in supply chain cost. Inbound transportation costs can more easily gain economies of scale through use of dense incoming truckloads headed for a few FC locations and hence are a much lower cost in the overall supply chain. As the Amazon supply chain has grown and become more sophisticated over the years, the need to focus on improvements to the Inbound Transportation Organization has become increasingly important.

1.4 Thesis Overview

Over the course of a six-month internship at Amazon's Seattle headquarters I was able to learn about the organization and help work on an exciting opportunity in the supply chain. This thesis will first provide the reader with context and background about the Inbound Transportation Network current state as well as introduction and literature review on the merits of "Hub and Spoke" distribution models. The tools and principles touched on in this literature review then serve as some of the guiding frameworks behind the optimization model developed to help reduce cost and improve performance in the inbound transportation network. In Chapter 3, "The Optimization Model", a discussion on data collection, and a comprehensive overview of the model formulation and design will give the reader a firm understanding of how the model was developed and what inputs feed it. Chapter 4, "Analysis of Model Results" will review results from the pilot run and review a comprehensive sensitivity analysis. Before wrapping up with final remarks in Chapter 6, some of the more qualitative implications of the model and the load consolidation strategy will be reviewed in Chapter 5 as well as next steps and future opportunities for the project.

2 The Inbound Transportation Network

Over the course of this Chapter we will start by reviewing the current state of the inbound transportation network at Amazon. This discussion will familiarize the reader with the basic geographic infrastructure that the network operates within. Next we will continue with a discussion on ground transportation ship-modes commonly used in industry and then conclude with a discussion on Hub and Spoke distribution strategies. These components will provide the necessary context and preparation before discussing the Network Optimization model developed in Chapter 3.

2.1 Current State

An important clarification in understanding the Inbound Transportation network is a distinction between freight that is managed and paid for by Amazon (collect) vs. that paid for by the vendors shipping the freight (pre-paid). This project exclusively deals with collect freight (for the purposes of this thesis let's call it "AmazonPay"). AmazonPay freight is on the rise in the inbound transportation organization and represents an opportunity for the organization to provide more control and reap more benefit from managing the inbound supply chain effectively.

The Amazon North America supply chain is comprised of over thirty FCs dispersed throughout the United States[4]. While some of these locations serve singular or multi-purpose roles within the network, the primary function of the entities is to receive and store product inventory until it is needed to fulfill customer demand. The FCs are generally located to most efficiently serve consumer demand. Locating FCs closer to consumer demand creates a lower outbound cost structure by needing to move the more expensive outbound shipments a shorter distance and also reduces outbound shipment transit time, which directly positively impacts customer experience. Generally, vendors are also positioned in close proximity

[4] Jennifer Dunn, "Locations of Amazon Fulfillment Centers," http://outright.com/blog/locations-of-amazon-fulfillment-centers-2/ (accessed 02/17, 2013).

to the major markets they serve, but the large and geographically disperse nature of Amazon's vendor base increases the challenges and complexity in managing the inbound transportation network.

The inbound transportation organization faces a difficult challenge in finding cost effective transit modes to get inventory from a widely dispersed vendor base in to the relatively small number of FCs serving the network. As we will see later, one of the most significant opportunities to lower cost in the inbound network is to focus on shipments traveling a long distance (i.e. from California to FCs located in the Northeast and Midwest), since these represent a much higher spend than shipments moving a shorter distance. Another important component of understanding the cost and opportunities in the network is familiarity with the different ship-mode options commonly used in transportation networks.

2.2 Transportation Ship-Modes: An Overview

In 2010 the US spent roughly $700 billion on commercial freight movements[5]. In any supply chain there are a number of methods to move freight from point A to point B. Looking at the Amazon Inbound Transportation supply chain we will focus majorly on ship-modes using truck carriers. While imports are a part of the inbound supply chain, this thesis discussion does not focus on import goods and hence ocean freight shipments are not considered. Air freight for that matter, is also a ship-mode more heavily utilized for imports and expedited outbound shipments, so it is also not considered. To that end, this thesis will primarily focus on goods moved by truck with some inclusion of intermodal rail shipments which made up a combined 78% of freight movements (by weight) in 2010[5]. Similarly, at Amazon, these two modes of transit largely dominate the inbound supply chain.

Within the ground transportation ship-modes (trucking, and rail/intermodal) there are primarily four commonly used ship-modes we will discuss throughout this paper: Less-than-Truckload (LTL), Full Truckload (TL), Small Parcel (SP), and Multi-Stop Truckload (MSTL). Each ship-mode has an ideal usage for different needs in the supply chain.

---

[5] Kevin Kirkeby, *Industry Surveys. Transportation: Commercial* (New York, NY: Standard & Poors,[2012]).

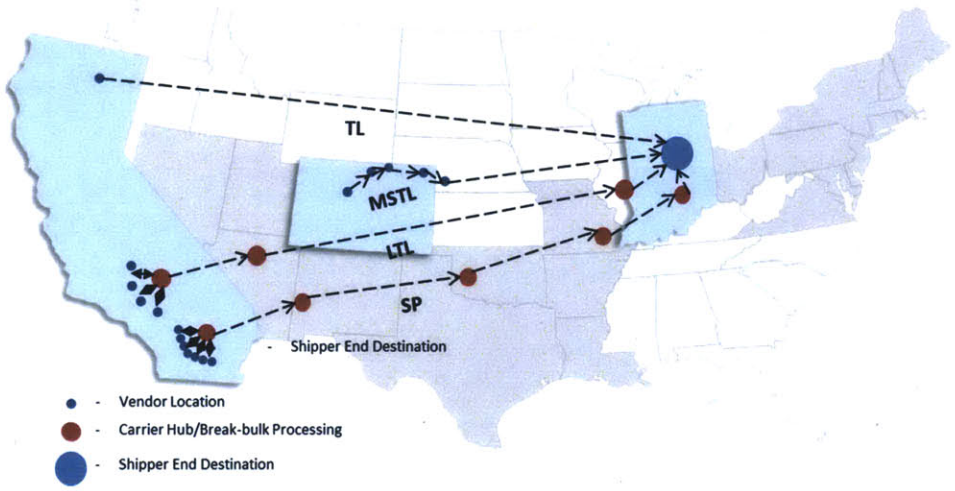


Figure 2 - Ship-Mode Overview

The diagram in figure 3 gives a brief picture of how these modes operate, further detailed diagrams can be found in Appendix A. In general, TL is the most efficient because it takes advantage of high trailer utilization with freight exclusively for the destined shipper; there are also no break-bulk/hub processing operations so there are fewer touch-points in the overall transit path. As one moves down to LTL and SP freight modes, break-bulk/hub operations increase the touch-points in the chain and freight on the trailer is shared across multiple shippers – as a result cost is higher and transit time is increased. The figure below gives a relative cost comparison of moving goods via the three ship-modes. The underlying assumption in the diagram is that you are shipping freight via a full load in each of the relative ship-modes (i.e. a full trailer of LTL freight vs. full trailer of TL freight vs. a full trailer of SP freight).

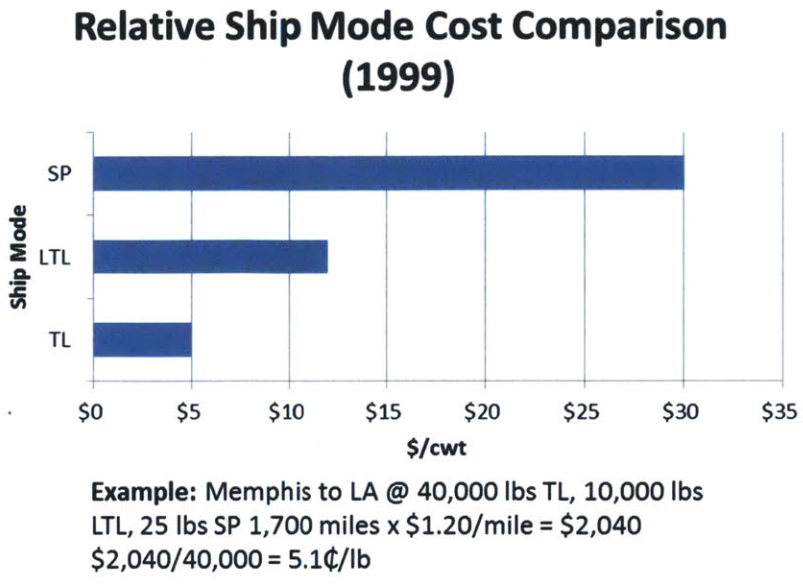


Figure 3 - Ship Mode Cost Comparison[6]

2.2.1 Small Parcel Carriers

The Small Parcel industry is comprised mainly of large companies with the scale capable of offering nationwide services – for example, FedEx, UPS, and the US postal service are some of the bigger carriers domestically. These carriers all operate with large hub and spoke supply chains (a strategy we will discuss in more detail in the next section) to provide rapid delivery and efficient cost throughout their respective supply chains. They are however, the most costly of ground ship-modes and typically are used for shipments of small weight, quantity, and size that are not palletized. In these cases SP is more efficient when there are not sufficient economies of scale to justify ship-modes like LTL and TL, or when it is necessary to pay a premium for expedited shipping. Most companies set an industry policy based off LTL minimum charge expectations that shipments exceeding the 150 pound mark are more cost effective to ship via LTL. In practice however, this static way of looking at when to upgrade from SP to LTL does miss some opportunities for greater savings on shipments that are below this threshold (LMS Logistics).

[6] Paul Huppertz, "Market Changes Require New Supply Chain Thinking," *Transportation & Distribution,* Mar 1999, 1999, 70.

2.2.2 Less-than-Truckload Carriers

LTL carriers are commonly used in supply chains to deliver heavier, bulkier items that do not have enough quantity and volume to justify purchasing an entire truck/trailer (Full Truckload shipping). There are significantly more competitors in the LTL market than SP, and many of the carriers are smaller, regional carriers that do not offer full national shipping. LTL shipments, like SP, are typically combined with freight from multiple customers to make transport more efficient for that carrier. Also like SP transit, there are multiple hubs throughout the carrier network to break and combine freight throughout the shipment route. Since LTL networks are typically smaller in scale and do not have the same resources as larger SP carriers, the break-bulk processing operations do not run 24/7 and hence overall lead-times can be as much as 1-2 days longer on average with a greater lead time variance than SP carriers.

Within the LTL segment carriers are broadly classified as regional or national providers. National carriers have an average length of haul of 850 miles or more and tend to offer full coverage of the domestic US, and parts of Mexico and Canada (Kirkeby 2012). Regional carriers have an average length of haul of 400 to 600 miles and tend to operate in smaller geographic footprints and often specialize in in overnight and second-day services. National LTL carriers have significantly higher overhead costs due to the increased number of terminals and labor force required to operate on a national scale. We will later see that in executing the load consolidation framework, using regional LTL carriers to provide the first leg of transit (pick-up) becomes a distinct cost advantage over using the more expensive national LTL carriers.

2.2.3 Full Truckload Carriers and Multi-Stop Truckload

When a shipper has enough freight coming from one origin location to fill a trailer with high utilization, TL transport becomes the most economical and is often a quicker option for transit. TL shipments are typically used when load size exceeds 10,000 lbs. Like the LTL industry, TL is very fragmented with more than 90% of carriers in the US being classified as small businesses[7]. TL carriers offer the transit time advantage by not needing to stop at multiple hub sortation facilities (break-bulks) throughout the

[7] Hoovers Inc. – Truckload Carriers Industry Overview

route of transit – a carrier will pick up the shipper's cargo from one location and deliver it straight to the end destination with the only stops being for the rest of the driver, or no stops at all in the case of team drivers.

An additional option many shippers are utilizing is intermodal transit which combines one or more ship-modes to take advantage of different supply chain benefits. In this thesis we will only touch on intermodal transit utilizing a TL carrier and the rail network. In this scenario, when there is enough freight to justify it and transit time is not as important, a TL carrier may transfer freight from the truck trailer to the railroad which will carry the freight over the lengthier portion of the transit arc at a much cheaper cost. This provides a great opportunity to save money but at the cost of greater transit time.

Multi-Stop Truckload (MSTL) shipments, in the context of those used in the inbound supply chain, use a multi-pick-up route for multiple suppliers with deliveries destined for one location. In the case that a shipper has multiple vendors (ideally in close proximity to each other) with multiple pieces of freight that do not justify their own truckloads but in combination justify one or more full trailers, a MSTL becomes a good alternative. Overall, these shipments are cheaper than LTL and SP shipments because of high trailer utilization and generally have shorter transit times since there are no hub processing operations between the last pick-up location and the end destination.

This brief overview of transportation ground ship-modes gives a good picture of what types of ship-modes are used throughout this thesis and research discussion. It also provides good context for where the Hub and Spoke distribution framework can have merits in the inbound supply chain. As mentioned earlier, ~65% of AmazonPay freight in the inbound supply chain is shipped using LTL and SP carriers which are more costly and offer lengthier transit times. A strategy that reduces the distance traveled in these modes could provide large savings to the inbound transportation network. In the next section we will discuss the Hub and Spoke distribution framework and what strategies can help Amazon in this regard.

2.3 Discussion: Hub and Spoke Distribution

Delta is commonly known to have pioneered the first Hub and Spoke model in the airline industry back in 1955 (Delta. 2013). The name Hub and Spoke gets its idea from a bicycle wheel where every point on the outer rim is connected to the hub by a single spoke. In application, the methodology allows for any point on the wheel to connect through another by routing through the singular hub on the wheel. For the airline industry, this became a huge operational advantage due to the economies of scale that could be achieved by routing through the hubs. Rather than running multiple low utilized flights from point to point, Delta could run fewer flights to its first hub location in Atlanta, GA and then run fewer highly utilized flights to the final destinations. The diagram in figure 5 depicts this well.

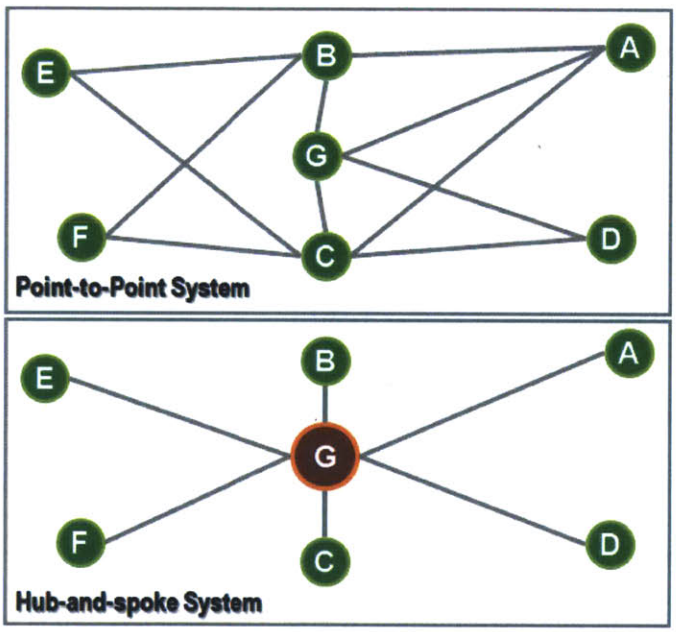


Figure 4 - Hub and Spoke vs. Point to Point Network[8]

In addition, the reduced complexity of the Hub and Spoke model (6 total flight arcs vs. 11 in the point to point model) allows flights to be run more efficiently from the higher utilization in each arc.

[8] J. Coyle, E. Bardi and R. Novack, "Transportation," in *Transportation*, Fourth ed. (New York: West Publishing Company, 1994), 402.

In the early 1970s Frederick Smith, founder of FedEx pioneered the use of the Hub and Spoke in air freight markets (Coyle, Bardi, and Novack 1994, 402). Since FedEx's adoption of the model in the shipping and logistics industry, many logistics providers and shippers have adopted the model to run their distribution networks, including UPS, Wal-Mart, and Lowes for example. The increased number of routing hubs/nodes allow for greater flexibility and responsiveness in the network due to the increased options multiple hubs can offer. Equally important, it also allows for high utilization arcs connecting these hubs to support lower cost structures. Although increasing the number of touch-points in the network should increase transit time and the damage rate from the additional handling, in the ground network strategies we discuss later this effect should be net zero or net positive in most cases. This comes from the fact that while the number of touch-points in the internal network are increased, the number of touches in the break-bulk processing operations of LTL providers are actually reduced – this effect can be observed in Appendix A with the difference in touch-points in LTL delivery and load consolidation ship-modes and will be discussed in more detail in the ensuing sections of Chapter 2. Some of the relevant strategies using the hub and spoke network architecture include small parcel zone skipping and load consolidation which we will now review in further detail.

2.3.1 Small Parcel Zone Skipping

Zone Skipping is a strategy that some small parcel shippers have employed to reduce cost over long distance hauls in their network. The strategy leverages a hub and spoke network and multiple ship-modes to lower cost and improve transit time performance. A good case example of this is from the early 90's with a small parcel consolidator called Small Parcel Service (SPS) that was based in Congers, NY. SPS formed a strategic alliance with a long-haul truck carrier named CRST and used them to perform the long distance haul between SPS's central distribution center and destination hubs before they were finally transported a short distance to the final destinations by UPS. This operation resulted in 10 – 15% cost savings for SPS's customers and in some cases reduced overall transit time by 2 – 3 days, since CRST used team drivers over the long haul portion of the transit (Andel 1992, 34). This clearly provided a cost

and performance improvement to any shippers choosing to use SPS to move its small parcel shipments long distances vs. exclusively using UPS or FedEX. SPS was able to cash in on this opportunity by achieving economies of scale on the full truckloads it filled on CRST's long haul portion of the transit.

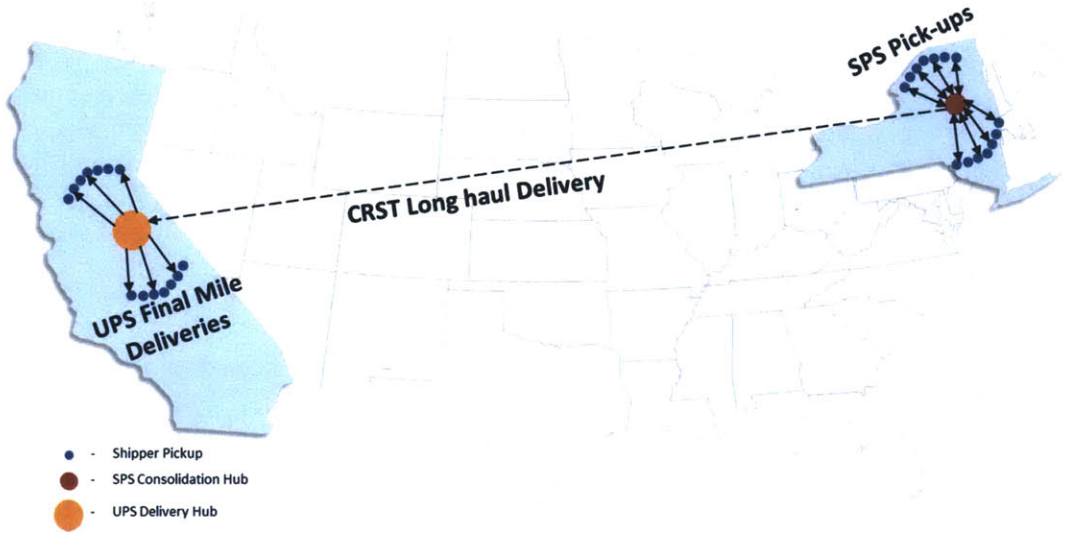


Figure 5 - Zone Skipping Example

While the zone skipping strategy is one for reducing the outbound cost of small parcel shipments, the ideas behind reducing the length of haul traveled by the more expensive ship-mode and gaining economies of scale through high utilization full truck loads are fundamental to the strategies employed in this project.

2.3.2 Load Consolidation

Load consolidation is a broader term than the more specific strategy employed by zone skipping. The term is generic for inbound and outbound operations and the focus of the strategy is to minimize transportation costs and maximize trailer utilization by combining shipments that are produced and used in multiple locations across different times into single vehicle loads (Baykasoglu et al. 2011). The strategy is used across air, ground and rail transport by almost all logistics providers but becomes a very difficult problem to solve within the dynamic context of real business environments.

Load consolidation strategies work off the efficiency of cross-dock operations. Cross-docks are the hub node or "pool point" within the hub and spoke distribution architecture discussed earlier. These hubs serve as the routing and consolidation points in the network and can effectively transfer freight from inbound trailers to outbound trailers without storage needs. Shipments typically will not spend longer than 24 hours in a cross-dock facility and are sometimes transferred in as short as an hour. As of 2004, it was estimated that there were over 10,000 cross-dock facilities throughout the US and Canada, and many retail organizations such as Wal-Mart and Home Depot had adopted these facilities in their logistics operations (Wang 2008). In this paper we will focus on the usage of pre-distribution cross-docks. These cross-docks assume that the destination of the freight to be consolidated is known before entering the cross-dock facility. This is an important distinction from a post-distribution cross-dock which does dynamic routing of shipments once packages arrive at the cross-dock hub. This distinction significantly reduces the complexity of the network optimization problem we will solve in Chapter 3.

2.3.3 Methods for Solving Load Consolidation Problems

There have been many methodologies used in solving load consolidation problems, particularly with respect to third party logistics providers (3PLs). One of the more basic models takes a similar approach to economic-order-quantity (EOQ) models by calculating the minimum amount of weight that should be aggregated to make shipments economical. The economic shipping weight (ESW) is calculated by using the order arrival rate (A), sum of all fixed cost associated with a vehicle load ($\sum F$), expected weight per customer order ($E[w]$) and variable cost of carrying inventory per unit weight per time period (I).

$$ESW = \sqrt{2 * A \sum F * E[w]/I}$$

Equation 1 - Economic Shipping Weight Formula

This simple formulation can be used to solve one piece of the load consolidation puzzle in practice (i.e. what is the economical consolidation weight). The complexity comes in deciding what shipments should be routed from what origins through what hubs to what final destinations. Additional complexity arises

from scheduling changes, human interaction and carrier delivery variability, all of which make it a very dynamic and difficult problem to solve. Generally, ESW models are less suited to solve these problems in practice because they assume consolidated loads are taken from a single origin and shipped to a single destination (Baykasoglu et al. 2011).

An approach developed by Ratliff, Vate and Zhang (Ratliff, Vate, and Zhang 2004) uses a mixed-integer linear program to design a network for load-driven cross-docking systems. In this paper they explore the routing of vehicles through the rail network in Ford Motor Company's North American automobile delivery system. The distinction of a "load-driven" cross-docking system clarifies that delivery vehicles in the network are not dispatched until they achieve a minimum required load (effectively an ESW) rather than in a schedule-driven model that aggregates freight over a defined time period and dispatches it on a fixed schedule regardless of vehicle load minimums. The load-driven approach focuses on maximizing vehicle utilization but does not optimize for an increased service level which schedule-driven models provide. The objective of the mixed-integer linear program is to minimize the average delay time between when a vehicle is produced and when it is delivered. In this case, the motivation for using cross-docks, or "mixing centers" in the context of automobile rail distribution networks, is that consolidating freight allows for the usage of trains with faster transportation times. The model uses two sets of decision variables to achieve this objective: location – number and location of cross-docks and routing – how vehicles should be routed through the identified cross-docks if at all. In the paper the authors explore both single-stage (use of one hub) and multi-stage (use of multiple hubs) approaches to solving the integer program with anywhere from 4,800 to 150,000 variables. An application of the "mixing center" concept with a joint alliance from the UPS logistics group reportedly reduced average transit time at Ford by 25% in 2001(Ratliff, Vate, and Zhang 2004).

Using multi-agent systems has been a growing area of research and study to solve load consolidation problems for large 3PL providers. In these systems an agent is defined as, "a computer system situated in some environment and capable of an autonomous action in this environment in order to meet its design

objectives" (Wooldridge and Jennings 1995, 115-152). Each agent is tasked with the responsibility of optimizing its own set of objectives which could include objectives for specific truck loads, drivers or other system resources. This combined with the socially interactive nature of agents allows for the flexibility of solving multiple smaller consolidation problems rather than large and complex central load consolidation problems. (Baykasoglu et al. 2011) This is very important given the large and dynamic nature of the load consolidation problems 3PL carriers face on a day-to-day basis and allows for quicker and more flexible decision making.

While the multi-agent approach to solving load consolidation problems appears to be an optimal strategy, its complexity along with the resources required to execute it are more suited for the larger transportation networks of more mature 3PL logistics providers. In the next section we'll discuss how such an aggressive strategy may not be the best first and next steps for the Amazon inbound transportation network's potential adoption of load consolidation strategies. As we will see in Chapter 3, the approach we select for solving the load consolidation problem will be similar to the "load-driven" cross-docking solution previously reviewed.

2.4 Amazon Inbound Consolidation: Current State

As mentioned earlier, a large portion of AmazonPay freight for the inbound transportation network is either LTL or SP. There is a significant opportunity within the inbound transportation network to take advantage of hub and spoke distribution strategies like load consolidation to reduce spend on these expensive ship-modes. In recent years, the inbound team has experimented with load consolidation strategies regionally to reduce spend on these higher-cost ship-modes. The simple diagram below well articulates the motivation behind the inbound team's experimentation with the strategy.

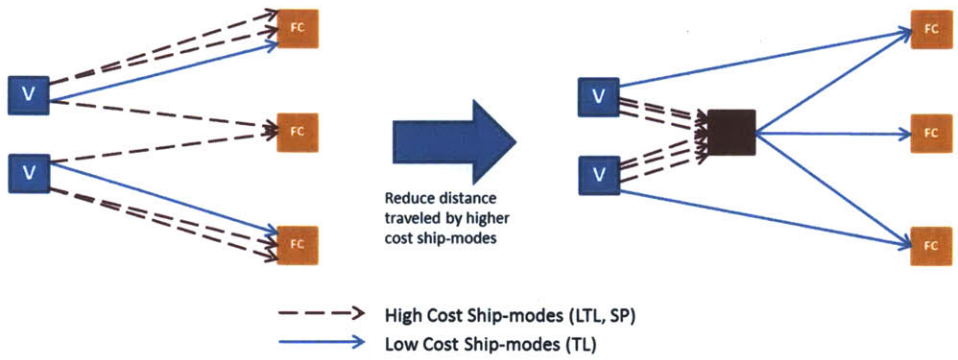


Figure 6 - Consolidation Diagram

By using consolidation hubs to consolidate freight from SP and LTL ship-modes, the inbound team can shift spending to more cost effective TL shipping to lower the overall cost of the transportation program.

Currently the program has largely taken advantage of low hanging fruit opportunities to consolidate. It has primarily focused on long distance moves which offer a greater opportunity to save costs. Further, it has only explored the usage of a limited number of hubs to consolidate freight at. The process of consolidation lane definition has been a largely manual process whereby historical shipping volumes between origin vendor clusters and FCs are rank ordered by freight shipment volume and finally evaluated by potential cost savings opportunity. What the inbound team lacks is a network planning tool that can effectively evaluate all freight movements throughout the network with the information detail to give strategic direction on where to execute the best load consolidation opportunities within the network. The approach we discuss to solving the load consolidation problem in Chapter 3 should help to better bridge the gap between the current state at Amazon and the more sophisticated multi-agent model discussed in section 2.3.

3 The Optimization Model

To support the inbound transportation organization in lowering cost and improving transit time performance, a network planning tool was created to help identify optimal load consolidation opportunities within the inbound network. The tool uses a mixed integer linear program (MILP) to determine optimal freight routing and optimal consolidation hub locations in the inbound AmazonPay network. Over the course of this chapter we will discuss the model's approach and give a detailed overview of the objective function, inputs and constraints governing the model operation. As an outcome of this section we will better understand what the model's goals and operation are before reviewing the results in Chapter 4.

3.1 Optimization Model Approach

The objective of the model is to provide high level planning and identification of load consolidation opportunities within the inbound transportation network. Given the current flows of freight in the network from vendor origin location to FC final destination, the model first calculates the total cost of the current network. Next, given a set of potential consolidation hub locations, the model optimizes freight routing through the network to minimize the total inbound transportation cost. It provides the user with direction by determining which hubs are the most attractive given the unique cost attributes of the hubs and which outbound lanes (path from cross-dock hub to FC final destination) the hub should operate based on optimal truckload utilization. The model also uses transit time on each arc as a lever to restrict or permit load consolidation usage to ensure achievement of transit time performance targets. We will overview the following steps used to create the model:

1. Use vendor origin locations, locations of existing 3PL hubs for Amazon carriers, current/planned locations of Amazon FCs and distance mapping as inputs to create a network infrastructure map.

2. Use historical data to obtain freight shipment traffic by lane (vendor origin cluster to FC destination cluster) by day. Use as basis for current state of flows through the inbound network.

3. Use historical data to estimate carrier rates by ship-mode (SP, LTL and TL). Use linear regressions to predict carrier rates where appropriate.

4. Create expected transit time calculations for all arcs defined in the network based off mileage traveled as basis for transit time measurement in the model.

5. Use inputs to create MILP with the objective to minimize network transportation costs subject to transit time performance targets.

While the intent and focus of this tool was to initially provide a global solution for the full AmazonPay inbound network (national scale), in this thesis we focus on an implementation of the model in the Southwest region of the network which represents 25% of total unit shipment volume. Further, we only review freight that is shipped via LTL or SP with the underlying assumption being that freight shipped via TL and MSTL are already cost optimized. Focusing on the Southwest region gave a representative subset of data that would help to validate the model performance while still being a manageable implementation target given the time available in the internship.

3.2 Network Definition Overview

Before reviewing the details of the objective function and various model constraints and inputs we will start with reviewing the overall network structure. In order to implement the load consolidation strategy we start by looking at a simple one layer network.

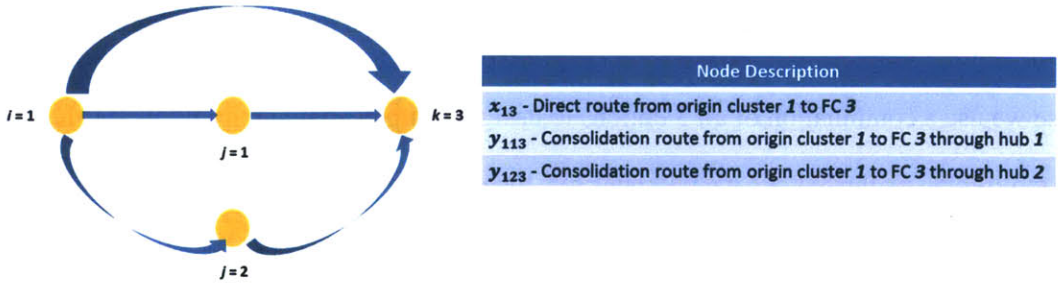


Figure 7 - One Layer Network Overview

This design makes the assumption that freight originating in an origin location *i* only has two options to get to its pre-determined final destination *k*. It can either go directly to its destination via the original ship-mode, x(i,k), or it can use one of the available consolidation hubs to first consolidate freight, y(i,j,k), before entering final destination *k*. All freight entering hubs *j* will be either SP or LTL while all freight leaving hub *j* will be exclusively TL. We do not explore multi-stage routing options where freight could enter 2 or more consolidation hubs before ending at its final destination.

3.2.1 Origin Nodes

Rather than individually define the thousands of vendors in the Amazon network as the origin nodes in the network structure, we take an approach a few levels higher by defining origin nodes by 3-digit zip-codes[9]. For the Southwest region evaluated this reduces the number of origin nodes to 98. The use of 3-digit zip-codes as origin nodes does not lend itself to a model focused on operational execution, but is consistent with the more strategic goals of understanding where load consolidation opportunities exist and how the inbound network should be designed to take advantage of them.

3.2.2 Hub Nodes

The consolidation hubs defined earlier are the cross-dock facilities where load consolidation is executed. There are three approaches to how these hubs could be selected for the model:

1. Use existing Amazon FCs as potential cross-dock hub locations.
2. Use existing Amazon 3PL carrier facilities as potential cross-dock hub locations.
3. Identify optimal hub location based on locations of high origin freight density.

In our case we use a combination of all 3 options to designate 4 potential hub locations for the Southwestern region evaluated. As we will later see, the different parameters in the model could give each of these hub options a distinctive cost profile. In section 5.2, "Who owns the cross-dock?" we will spend more time discussing the pros and cons of each alternative.

[9] See Appendix B for a representative break-down of 3 digit zip-codes in the US.

3.2.3 Destination Nodes

There are currently close to 40 Amazon fulfillment centers throughout the North American network[10]. In a number of cases FCs are located within close proximity of each other (< 70 miles) and will focus on serving the same region with different product groups. There is a basic distinction between FCs that fulfill sortable or non-sortable goods in the Amazon network; sortable items are much smaller in dimension and are often combined as multiple shipments, whereas non-sortable items are larger and not often combined due to size constraints (Wulfraat 2013). To define our destination nodes we first cluster the existing Amazon locations by proximity, with the assumption being that a full truckload destined for an FC cluster could do multi-stop drop-offs to FCs within the cluster in 1 day (<100 miles from each other). We then go a step further and segregate FCs that are sortable vs. non-sortable. This is an important distinction for how freight will flow through the model as we will later see with some of the freight homogeneity assumptions made. The end result of this segregation is that 29 destination clusters are used to define the entire Amazon fulfillment network. See Appendix E for a listing of the destination nodes and FC mappings.

3.2.4 Arcs and Flows

Arcs are defined between each of the *i, j, k* nodes discussed. On each arc a cost and distance are defined. Distance was obtained by using a 3-digit to 3-digit mileage computation based off the PCMILER transportation mapping and routing tool. Cost as we will later discuss in section 3.4.2 is defined as a function of ship-mode and/or mileage. Freight then flows across the arcs in the network in units of weight (pounds). Weight is chosen as the flow unit because of its relevance in freight cost computations. To account for other factors such as cubic volume in truck capacity calculations, hub sortation rates in cost/unit, etc., conversion factors are used throughout the model that take into account Origin-Destination (OD) information as well as FC type (sortable vs. non-sortable) in the computations.

[10] Dunn, *Locations of Amazon Fulfillment Centers*

3.3 MILP Formulation

Just based on the network structure discussed in section 3.2 there are over 28K paths for freight to flow in the modeling of the Southwest region. This translates to a large number of variables and constraints that need to be evaluated, a number far larger than what is supported in basic Excel solver optimization engines. In order to provide a solution with greater flexibility in model definition and results interpretation, AMPL was chosen as the mathematical language to code in with XPRESS as the optimization engine. Over this section, we will familiarize ourselves with the mathematical formulation of the model before getting into the details of the source of the model inputs which will be reviewed in section 3.4.

3.3.1 Objective Function and Decision Variables

The load consolidation strategy we seek to implement works off the basic premise that cost can be saved by consolidating freight early to allow for more economical high utilization trucks to carry freight over the longer distance of the transit. To that end, the optimization model is evaluating two alternatives for freight originating in one origin: consolidate or ship direct.

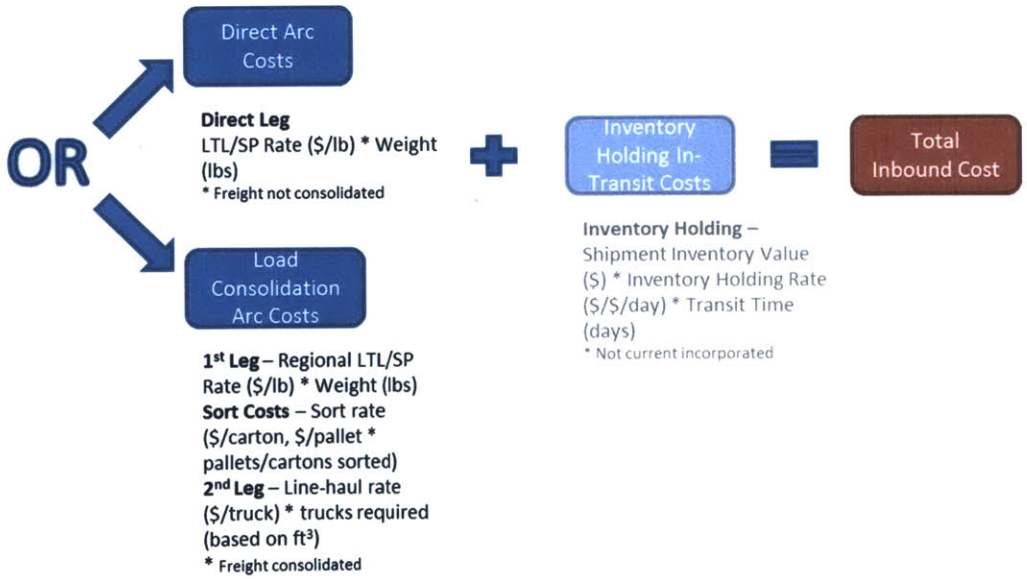


Figure 8 - Total Inbound Transportation Cost Tradeoff

The objective function is to minimize the total inbound transportation cost which is defined as the combination of costs from direct arcs and load consolidation arcs plus any potential increase/decrease in in-transit inventory holding costs from changes in transit time as a result of the new freight routing. While the initial model formulation did include the aspect of in-transit inventory holding cost, a mini pilot run (small scale excel model) showed the impact of this factor to be negligible on total transportation costs given the relatively short lead time of domestic shipments and hence it is not included as a factor in the full model implementation. The mathematical formulation of the objective function is below:

MINIMIZE

$$\sum_{i,k} xl_{i,k} * cL_{i,k} + \sum_{i,j,k} yl_{i,j,k} * \left(cL_{i,j} + \gamma_k * cHL_j\right)$$ — Total **LTL** Costs

$$+ \sum_{i,k} xs_{i,k} * cS_{i,k} + \sum_{i,j,k} ys_{i,j,k} * \left(cS_{i,j} + \beta_k * cHS_j\right)$$ — Total **SP** Costs

$$+ \sum_{j,k} \left(ol_{j,k} * f + \frac{\sum_i yl_{i,j,k} * \tau_k + \sum_i ys_{i,j,k} * \rho_k}{\mu * C}\right) * cT_{j,k}$$ — Total **TL** Costs

$$+ \sum_j o_j * cH_j$$ — One-time cost for using cross-dock hub

Decision Variables (Refer to for all constraints in 3.3.2)

xl = weight of LTL shipments shipped direct
xs = weight of SP shipments shipped direct
yl = weight of LTL shipments shipped to consolidation hub
ys = weight of SP shipments shipped to consolidation hub
ol = open/close outbound lane in cross-dock hub (binary)
o = open/close cross-dock hub (binary)

Data (Refer to for all constraints in 3.3.2)

cL = cost per lb of LTL shipments
cS = cost per lb of SP shipments
cHL = cross-dock cost per pallet for *LTL*
cHS = cross-dock cost per carton for *SP*
γ = lbs to pallets conversion for LTL
β = lbs to cartons conversion for SP
f = truck cost factor (0-1)
τ = lbs to ft^3 for LTL
ρ = lbs to ft^3 for SP
μ = target trailer utilization
C = trailer capacity
cT = cost per outbound truck trailer (load consolidation)
cH = one-time cost of using cross-dock hub

Equation 2 - MILP Objective Function Formulation

This is a single period model formulation, and we define a period at the daily level. In actual execution of the model we will iterate the MILP over multiple periods using AMPL scripts to get an idea for behavior over a full week of demand. The *i,j,k* subscripts refer to the same node definition as described in section 3.2 – *i* represents the origin node, *j* represents a potential cross-dock hub and *k* represents the destination FC cluster.

An additional piece of clarification is needed for the *f* constant in total TL cost definition. As we later discuss in section 3.4.2, TL outbound costs are approximated as a linear function but with a minimum enforcement cost equivalent to the purchase of one full truck. In order to account for this in the cost function, the truck cost factor (*f*) ensures that the total cost of the first truck (fixed and variable component) sums to the cost of one full truck. For cases where the minimum trailer utilization is assumed to equal the max realized trailer utilization the factor is set to 0 (current model operation). If there is a difference between the two (e.g. minimum required utilization is 55% but max realized trailer capacity is up to 70%) the factor is set to still ensure that the fixed and variable components sum to the cost of one truck when the first trailer is filled. The factor, *f*, is initiated to $f = 1 - \frac{\min truck\ capacity}{\max truck\ capacity}$ to achieve this. After the first truck is purchased, the cost function then takes the form of a linear function. Another point of clarity is needed with the conversion factors used throughout the model. In section 3.6 we will later discus the importance of these factors in addressing the freight homogeneity assumptions of the model. Within the model formulation there are 3 sets of decision variables being evaluated:

1. **Freight Routing** – *xl, xs, yl* and *ys* determine how much freight should flow through which arcs in the model (direct vs. through cross-dock). See descriptions under Equation 2 for definition of each variable.
2. **Hub Selection** – *o* determines which cross-dock hubs should be opened to allow for load consolidation to be executed.

3. **Cross-dock Outbound Lane Selection** – *ol* determines for each hub what lanes can be run based on minimum truck utilization and the allowable number of lanes to run per hub

3.3.2 Constraints

The following are the key constraints underlying the optimization model:

1. **Maximum Hubs** – This constraint allows for limiting the maximum number of cross-docks to be used in the model. This is a particularly important consideration in the early stage of evaluating the load consolidation strategy since it is more favorable to start with a smaller number of hubs and gain operational competence before immediately implementing a large number of them throughout the network. This can also help us zero in on the most preferable hub for a given location when we set the maximum equal to one.

$$\sum_j o_j \leq N \,, \forall\, j - \text{Maximum number of hubs}$$

Additional Data

N = Maximum number of active hubs

Equation 3 – MILP: Maximum Hubs Constraint

2. **Maximum Lanes** – A lane, in the context of our model, is defined as the arc from a cross-dock hub (j) to a destination FC (k). Given the size of the particular cross-dock hub (j) we could limit the maximum number of lanes useable out of that hub. This constraint helps to account for available manpower, number of doors useable in the hub as well as other items that limit the number of lanes that could be used out of the hub in practice. In the context of our model, it could be used to prioritize lanes with higher weekly utilization.

$$\sum_k ol_{j,k} \leq L \,, \forall\, j - \text{Maximum number of lanes per hub}$$

Additional Data

L = Maximum number of active lanes in hub j

Equation 4 - MILP: Maximum Lanes Constraint

3. **Maximum Days Pooling** – As we will discuss in further in detail in section 3.4.2, a linear function is used to estimate the cost of a full truck in the TL ship-mode. This constraint performs two functions. First, it enforces that if a consolidation lane is chosen that at least one full truck is purchased; this eliminates the potential for the model to order a ½ trailer TL (impossible since you purchase the whole truck) because it appears economically feasible in the linear model. Second, and along the same lines, it enforces that the truck purchased achieves a minimum economical cubic volume to justify the linear rate we use to approximate the price. The target trailer utilization used in the model is an internal target used in the inbound transportation organization. Further, we insert the flexibility that this minimum cubic volume is achieved over a p day period which gives the effect of allowing freight to "pool" for a number of days before being loaded on a truck. This is a managerial lever that helps to illuminate one potential tradeoff between cost and transit time. In a simple example, shipments coming from one zip code may not have enough freight to economically justify a full trailer via load consolidation in one day but it may have enough freight to justify the trailer every other day. In this case we may be able to justify an additional day of transit time ("pooling time") for the cost savings received in allowing the freight to sit for an additional day and take advantage of load consolidation.

$$\frac{cT_{j,k}*ol_{j,k}*f}{p} + \sum_i \frac{(\rho_k*ys_{i,j,k} + \tau_k*yl_{i,j,k})}{\mu*C} * cT_{j,k} \geq \frac{cT_{j,k}*ol_{j,k}}{p}, \forall\, j,k - \text{Freight Pooling}$$

Additional Data

p = maximum freight pooling days (p>=1, p=1 assumes 0 days of freight pooling)
ρ = SP conversion factor (pounds→ft^3)
τ = LTL conversion factor (pounds→ft^3)
μ = target truck utilization for TL trailers
C = trailer capacity (ft^3)

Equation 5 - MILP: Freight Pooling Constraint

The logical constraint is written to enforce the freight pooling objectives. The left-hand side of the constraint is the TL cost function used in the objective function with one adjustment – the fixed component of cost is adjusted to account for the number of days freight is allowed to pool. The right-hand side of the constraint is the actual cost (as derived from the linear regression) of one full truck

adjusted for the number of days freight is allowed to pool. Ensuring that the cost used in the objective function exceeds the actual calculated cost of one full truck enforces our goal of purchasing at least one full truck when there is no freight pooling. Increasing the hub pooling days, p, then allows us to relax the constraint up to the identified number of pooling days.

4. **Transit Time Performance** – The key factor of performance we seek to observe in this model is transit time, however, our model is in fact optimizing for cost. The transit time constraints in the model allow us to measure and control the performance while still minimizing cost. The simple target we use to do this is to ensure our transit time along a load consolidation arc either meets or exceeds the original transit time along a direct arc by a given factor. This becomes another managerial tool that truly allows for the cost vs. performance tradeoff to be observed. In the current model execution we choose not to make this a hard constraint, it is relaxed. This is done by still allowing the model to proceed even if the transit time target for an arc is not met, but rather throwing a flag to alert the user to which load consolidation arcs do not meet the transit time targets.

$$ys\_t_{i,j,k} \leq xs\_t_{i,k} * (1 + \omega)\,, \forall\, i, j, k \text{ - Transit Time Check (SP)}$$

$$yl\_t_{i,j,k} \leq xl\_t_{i,k} * (1 + \omega)\,, \forall\, i, j, k \text{ – Transit Time Check (LTL)}$$

Additional Data

ys_t = total transit time for consolidation arc (SP)
yl_t = total transit time for consolidation arc (LTL)
ω = % load consolidation arc can exceed or must be lower than the original (direct) transit time arc (can be negative)
xs_t = total transit time for direct arc (SP)
xl_t = total transit time for direct arc (LTL)

Equation 6 - MILP: Transit Time Constraints

These constraints are not included in the mathematical model formulation but are actually included in the pre-solve run script. If the constraint is made hard, all load consolidation arcs that violate the transit time performance constraints are pruned from the model during the pre-solve run script. This can easily be achieved since we calculate expected transit times for all direct and consolidation arcs from the mileage mapping between all arcs before the solver engine is executed.

5. **Model Operational Constraints** – These constraints include the flow conservation constraints and the hub/lane switch constraints. They enforce that the model output is physically feasible.

$$\sum_j ys_{i,j,k} + xs_{i,k} = ds_{i,k}, \forall\, i,k$$ – Flow Conservation Constraint (SP)

$$\sum_j yl_{i,j,k} + xl_{i,k} = dl_{i,k}, \forall\, i,k$$ – Flow Conservation Constraint (LTL)

Additional Data

dl = total shipments (pounds) between origin i and destination k (LTL)
ds = total shipments (pounds) between origin i and destination k (SP)

Equation 7 - MILP: Flow Conservation Constraints

The flow conservation constraints ensure that all shipments originating from vendor cluster i will arrive at destination FC k. Since we do not intend to make any new destination routing decisions from our model, we only need to ensure that the model maintains the same origin-destination routing that was provided from the historical data-set.

$$ys_{i,j,k} \leq ol_{j,k} * ds_{i,k}, \forall\, i,j,k$$ – Lane Switch Constraint (SP)

$$yl_{i,j,k} \leq ol_{j,k} * dl_{i,k}, \forall\, i,j,k$$ – Lane Switch Constraint (LTL)

$$ys_{i,j,k} \leq o_j * ds_{i,k}, \forall\, i,j,k$$ – Hub Switch Constraint (SP)

$$yl_{i,j,k} \leq o_j * dl_{i,k}, \forall\, i,j,k$$ – Hub Switch Constraint (SP)

$$ol_{j,k} \leq o_j, \forall\, j,k$$ – Active hub for active lane

Equation 8 - MILP: Hub and Lane Switch Constraints

The switch constraints ensure that the model cannot choose to use a hub or lane within that hub if the corresponding hub or lane is not active. The model in principle will force the binary variables representing the "switches" for each lane/hub to "1" or "on" if it is cost optimal to do so. If a hub/lane is "0" or off, no freight will be allowed to flow through it.

6. **Hub Sortation Capacities** – Each hub has 3 capacity factors: Pallet cross-dock capacity (for LTL freight), Carton sort capacity (for SP freight), and Total Unit capacity (both LTL and SP). As mentioned earlier, one distinction between SP and LTL freight is that LTL freight is palletized by the

vendor while SP freight is in carton/package form. This has different implications on sortation (cross-dock) capacity as well as load utilization of outbound trailers. LTL freight will be much quicker to sort and the main constraint on sortation capacity will be available floor-space in the cross-dock hub. SP freight will require cartons to be handled individually and hence labor availability and sortation equipment becomes a larger driver. Further, trailers will need to be either fluid loaded (cartons loaded directly in to trailer bed) or palletized prior to loading on the truck. As a result of these distinctions we give each hub one capacity attribute for LTL cross-docks and one for SP cross-docks. Lastly, we use combined total unit capacity as a good proxy for overall hub sortation capacity within the Amazon network. For the current analysis these constraints are relaxed to understand the maximum throughput the facilities would need to support under optimal load consolidation scenarios.

$$\sum_{i,k} \beta_k * ys_{i,j,k} \leq cSP_j \,, \forall\, j \text{ - SP Sort Capacity (Cartons per day)}$$

$$\sum_{i,k} \gamma_k * yl_{i,j,k} \leq cLP_j \,, \forall\, j \text{ – LTL Sort Capacity (Pallets per day)}$$

$$\sum_{i,k} ys_{i,j,k} * \alpha_k + \sum_{i,k} yl_{i,j,k} * \delta_k \leq cTP_j \,, \forall\, j \text{ – Total Unit Sort Capacity (Units per day)}$$

Additional Data

β = Conversion factor (pounds→cartons) for SP
cSP = Carton sort capacity of hub
γ = Conversion factor (pounds → pallets) for LTL
cLP = Pallet (LTL) sort capacity of hub
α = Conversion factor (pounds → units) for SP
cTP = Total unit sort capacity of hub
δ = Conversion factor (pounds → units) for LTL

Equation 9 – MILP: Hub Capacity Constraints

The formulations of the capacity constraints use conversion factors to give meaningful measurement to the capacities being used. The assumptions behind these conversions will be discussed further in the Data Inputs section, 3.4.1.

The objective function, variables and constraints make up the formulation of the MILP. In the next section we will discuss the inputs that feed the model.

3.4 Model Inputs

All data to feed the model is derived from historical cost and metrics tables from the inbound transportation organization. We use historical data to feed this largely static model as we do not expect there to be large variances or changes in the flows (outside of expected growth trajectory) of freight from Amazon's current vendor base and because historical carrier costs are an easily accessible proxy of freight pricing. Since our model's intent is not to provide operational (day-day) routing decisions, but to give high level strategic direction, the choice of historical data should fit our needs well. That being said, there is one major caveat with using this data source. In particular, we use FC receipt level data to represent actual vendor pick-up data. In the data-sets accessed during the internship, pick-up level data did not have granularity to all the specific details (i.e. units, SKU weight, etc.) needed to feed the model, but FC receipt level data did have all required details. As a result, we adjust the receipt data accordingly to account for the fact that FC receipts happen 7 days/week but most carriers do shipment pick-ups 5 days/week. We use a simplifying technique of assuming items received on Sunday were actually shipped on Friday and that items received on Saturday were actually shipped on Wednesday. This gets us to the closest approximation of the pick-up distribution without spreading the receipts from Saturday and Sunday across multiple days. The figure below shows the results from a representative sample week.

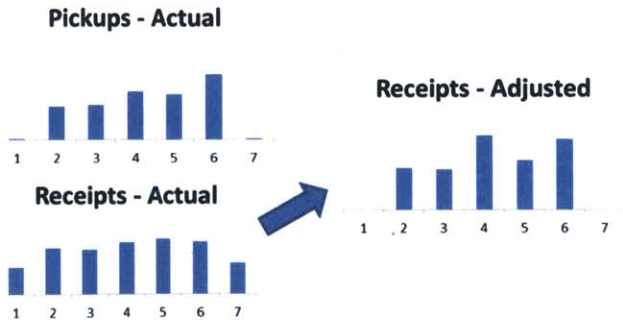


Figure 9 - FC Receipts to Expected Pick-up Date Adjustment

The data-set used in the model is inclusive of freight originating in the Southwest geography. For confidentiality purposes we will not discuss what geographical bound this represents nor discuss specific

details about potential cross-dock hub or vendor locations, but this data-set does represent 25% of all AmazonPay domestic LTL and SP shipping by unit shipment volume.

3.4.1 Demand (Shipments)

Demand is defined in aggregate by the total SKU-weight originating in 3-digit zip node *i* destined for FC final destination *k*. The concept SKU-weight comes from the fact that we are using actual FC receipts at a SKU level to get accurate origin-destination (OD) detail, and we use the SKU's master data to get a weight for each of the items received daily along an arc (*ik*) and then roll the total up as aggregate demand. While demand (*ds* and *dl)* and the associated routing variables (*xl, xs, yl* and *ys*) in the model are defined by SKU-weight, the model incorporates actual units, cartons and SKU-cubic volume shipped along arcs by using various conversion factors in the model.

3.4.2 Freight Rates (Cost)

There are fundamentally 4 different types of cost flowing across the arcs in the model: direct rates (LTL and SP), regional rates (LTL and SP), TL rates, and cross-dock handling rates. Freight rates for LTL and SP are modeled as a simple cost/lb based off the historical average cost between OD pairs. For TL rates we use a linear regression to predict cost per full trailer based off the mileage traveled. Data to feed these rates are calculated from the 2 months preceding model execution. Below, we will review the structure and reasoning behind rate selection by ship-mode.

SP Rates – For direct arc costs, rates are defined in $/lb and are calculated by dividing total charge ($) by total SKU-weight (lbs) shipped at a state-state level and distinguishing between sortable and non-sortable FCs. As a result, all vendor origin-FC cluster pairs inherit the shipping rate of the parent state-state mapping. This state-state mapping of rates gives a cost structure very similar to the zone pricing tables that larger small parcel carriers use where cost is largely a function of two factors – weight and zone. For the first leg of a load consolidation arc (vendor origin → cross-dock hub) we first use the state→FC

cluster mapping, where the cross-dock hub being evaluated is an existing FC, or the historical inner state rate where an existing FC does not exist.

LTL Rates – LTL rates are some of the most complex freight cost structures to be accurately estimated due to the large number of factors that go in the pricing. In addition to OD pair (distance traveled), freight class, origin region, carrier type and discounts are all significant factors that affect an LTL rate. While a more sophisticated rate estimation engine such as that described in "Estimating and benchmarking Less-than-Truckload market rates" (Özkaya et al. 2009) would offer more accurate rate estimates, we choose to go with a simple rate structure for the initial model based on the historical average cost per pound at a state-state mapping level. The state-state mapping structure is used for all direct arcs. For the first leg of travel on load consolidation arcs a user-defined target is currently set to reflect the expected rate structures of regional LTL carriers.

Using regional LTL carriers to do the first leg of transit on load consolidation arcs is where some of the savings from load consolidation come from. In a highly fragmented LTL market-place, smaller regional carriers will offer lower rates to remain competitive in their region vs. the higher rates that larger national LTL carriers like Con-way, Estes and ABF Freight demand. Allowing this rate to be a user-defined target also makes it a managerial lever by allowing the user to find "break-even" regional rate targets that make load consolidation attractive in the region being evaluated.

TL Rates – For TL rates we take a different approach to modeling cost. Unlike SP and LTL ship-modes the primary cost driver in TL shipping is the mileage traveled. Since using TL is effectively purchasing a full trailer, cost is relatively independent of how much weight you load on that trailer. An analytical benchmarking tool called Integrated Freight Market Intelligence (iFMI) by Chainalytics offers a more comprehensive model for predicting market TL rates similar to the tool developed for LTL rate prediction discussed earlier (Chainalytics 2012). In our model, we initiate the TL freight cost in each cross-dock hub→destination FC arc based off a linear regression of TL trailer invoice cost against mileage traveled.

The intercept of the regression represents the minimum charge for a trailer while the coefficient represents the additional cost/mile. The resulting total cost estimates the total full trailer cost between origin and destination pair. The following figure presents the regression result.

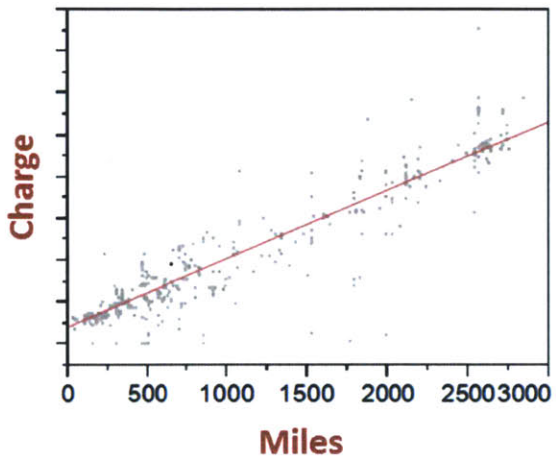


Figure 10 - TL Regression (Charge vs. Miles)[11]

The data-set used is from a 2-month period (May and June of 2012) inclusive of 496 observations consisting of freight traveling throughout the full domestic network. The linear fit has an R^2 of .84 and mileage as a coefficient has a t-ratio of 50.49 suggesting that the variance in prices in the data-set can be largely explained by the miles traveled. In the May-June data-set used, there was a mean absolute error (MAE) of 17% relative to the total average cost in the data-set. Checking the regression against data from July – December of 2012 we similarly see an MAE of 15% relative to the average trailer cost over the 6 month period. While mileage does not fully explain TL costs, this regression approach is sufficient for our model's needs.

Cross-dock Sortation Rates – Cross-dock sortation rates are modeled differently based on whether the originating shipment is LTL or SP. SP freight is assumed to be all carton freight and hence an internal Amazon labor rate for the cross-docking of outbound shipments is benchmarked. This variable rate is a cost/carton, and the pounds→cartons conversion factor handles that conversion. LTL freight is assumed

[11] Actual rates are hidden for confidentiality.

to be all pallet freight and the cross-dock rates are modeled based on quoted prices for regional LTL carriers to handle the cross-dock operation.

3.4.3 Expected Transit Times

Expected transit time is defined as a pure function of mileage. We choose not to use historical transit time data because of the high variation in actual performance on different 3-digit zip → FC Cluster lanes. Using data with this much noise would unintentionally prune or add load consolidation arcs to the model. All of the transit time calculations are handled in the pre-processing routine of the AMPL script file based off the assumptions used in Appendix C. The underlying assumption prevalent in the transit time calculations is that a standard TL driver can legally drive 10 hours a day at an average of 45 MPH (450 miles in one day). Given this we make additional assumptions about added transit time for using SP and LTL ship-modes to account for break-bulk processing and sortation.

Even with a robust model formulation, the results and conclusions of this model are only as good as the data feeding it. While the data sources from Amazon are plentiful and an appropriate amount of time was spent building good inputs, there are many potential improvements for input definition and data recording that we will discuss later in section 5.3, Next Steps and Future Opportunities. Before reviewing the results in Chapter 4, section 3.5 will briefly familiarize us with the outputs defined in the model.

3.5 Model Outputs

Interpreting the results in a managerial context is an important aspect of this project's goals and ambitions. Rather than receiving a one-line read out of "choose hub X to receive total cost savings Y", the model needs to give a concise list of relevant, interpretable information. Interviews with different members of the inbound transportation organization guided the definition of most of these outputs and extensive scripting was then created to post-process the AMPL results into a digestible user form. Broadly speaking, there are 3 categories of outputs: Cost, Routing and Transit Time.

Cost Outputs	Routing Outputs	Transit Time Outputs
Original Cost - $	Outbound Truckloads (HD) - trucks	Transit Time Direct (OD) - days*
Post-Consolidation Cost - $	Total Units Routed by hub (H) – units	Transit Time Cross-Dock (OHD) - day*
Lane Cost (HD)- $	Total Units Received by Hub (H) – units	Weighted Avg. Transit Time Current (D) - days
Effective TL Rate (HD) - $/lb	Hybrid/Consolidation Flows (OHD) – lbs	Weighted Avg. Transit Time Post Run (D) - days
Effective Consolidated LTL Rate (HD) - $/lb	Direct Flows (OD) – lbs	Weighted. Avg. Transit Time Hybrid (D) – days**
Effective Consolidated SP Rate (HD) - $/lb		Weighted. Avg. Transit Time Direct (D) – days**
Total Savings - $		

* Outputs are only displayed if hybrid lane violates transit time constraint| ** Times are separately displayed for SP and LTL freight|*** O,H,D denote output definition at Origin, Hub and Destination level respectively

Table 1 - Model Outputs Definition

The cost outputs give a high level snap-shot of overall model performance but also dive deeper by giving the user effective rates per pound on the active consolidation arcs. Managerially, this is a simple number commonly used in the organization. Most discussions on cost among carrier managers will center on a comparative cost per pound for the relative shipping lane being discussed. The routing outputs give an understanding of where the freight will move throughout the network. Again, high level details comes in the form of how much freight will run on each consolidation lane, or how many units will be routed through which hubs, but more granular details are available as well. The transit time outputs are majorly summarized by FC destination so that the user can quickly and easily see the impact of any routing changes on transit time performance for specific FCs.

The definition of these outputs gives a sound overview of how the results of this model will be interpreted. As we begin to discuss the results of the model run, these create a good base to build off of for sensitivity analysis.

3.6 Model Assumptions and Risks

Throughout this Chapter the model and inputs were explained in detail. A number of key assumptions were touched on and over this section we briefly highlight them as well as their associated risks.

1. **Freight Homogeneity** – Throughout the model freight is assumed to be largely homogeneous. That is, we assume that for a given OD pair that each pound of shipments along that arc for a

given day has the same characteristics. In reality we know that different shipments from different vendors will have different dimensional attributes and thus cannot be treated the same. Ignoring this fact entirely could cause an under/over-estimation of consolidation costs based off skewed trailer utilizations resulting from the differences in freight dimensional attributes. We do our best to get around this assumption by defining the different conversion factors throughout the model. By segregating non-sortable vs. sortable FC destinations we have one layer of heterogeneity since we know that non-sortable shipments exceed a certain dimensional size. Further, by defining our pound-cube conversion factors at the destination FC level, we get some distinction between the cubic dimensional attributes of freight headed for one FC vs. another. That is, perhaps freight headed to FC cluster 1 is typically media (DVDs, CDs, video games) and hence have smaller dimensions per pound while freight headed to FC cluster 2 are high volume, low weight items like diapers and have larger dimensions per pound. The distinction will affect how many pounds of freight can be loaded on to a TL trailer given our trailer utilization targets.

2. **Linear Cost Estimation for TL** – The total cost of cross-dock arcs is largely dependent on the cost of the outbound TL trailer. Rather than using a discrete model that requires trucks to be purchased in one-truck increments we approximate cost by enforcing that at least 1 truck is purchased but thereafter a linear function is used to approximate total cost for the given lane. Enforcing a discrete integer truck constraint would largely impact the runtime of the model. We do, however, note that using the linear approach has a greater tendency to misstate cost in lower volume scenarios than in the higher volume scenarios (greater than 5 trucks per lane) which we typically see in the model.
3. **Static Model Definition –** The model operates statically and assumes no variance in demand (we use absolute values for all inputs). While we do analyze the impact of variability by running the model across multiple days in the week and by using data from peak vs. non-peak periods as we will later discuss in Chapter 4, the potential effects of demand variability are not captured.

Overall, these assumptions present risks; but given the strategic (as opposed to operational) intent of the model, these assumptions should not largely affect the interpretation of our results. We are focused on identifying the best load consolidation opportunities and how to design a network to support it; the current model will help to achieve this goal. In the next Chapter we will review the model's results.

4 Analysis of the Model Results

As previously discussed, the model uses a data-set from the Southwest region of the US, representative of ~25% of AmazonPay LTL and SP freight by units shipped. It's expected that this data-set should represent some of the higher load consolidation opportunities given its location on the west coast where longer haul shipments to the east coast are frequent. Further, in the Southwest there is a 20% higher mix of LTL to SP origin freight volume and a 36% higher unit cube density on average units shipped than in the Northeast. The higher LTL to SP freight mix suggests that shipment sizes on average are higher and the increased cubic density per unit suggests larger bulky items that can make filling a full trailer more likely. This makes sense as many suppliers in the Northeast are book publishers which are high in weight but low in cubic volume. We highlight these differences because while the data-set we review is of sufficient size, the results and load consolidation potential in the Southwest may not reflect the same opportunities in the Northeast or other regions and hence further analysis will be needed to fully vet the total load consolidation opportunities in the inbound network. Over this chapter we will review the results from the model run, test sensitivity of the results to various model assumptions and discuss the managerial impacts of these results and suggestions.

4.1 Results Overview

For the Southwest region the model was initially run over a 4 week data set inclusive of 2 weeks in October and 2 weeks in July of 2012. The results from these 4 runs were fit to an annual unit demand distribution to project savings over the full year. The choice of weeks in July and October highlight the distinction between shipments in the peak vs. off-peak season which is an important attribute of the Amazon business cycle. Over the peak period (October – December) FC receipts are as much as 70% higher than those during off-peak which, as we will discuss later, has significant implications for cross-dock options using existing Amazon facilities.

The initial results were run under the shipment demand for 2012 but we also ran the model under growth assumptions for the 2013 fiscal year, based on expected growth assumptions. We will now review the results from these runs over the 3 sets of outputs discussed – cost, routing and transit time.

4.1.1 Results: Cost and Routing

After fitting results to the annual demand distribution the model projects annual savings of 5.2% and 8.9% of total AmazonPay inbound LTL and SP spend in 2012 and 2013 respectively. In the case of 2012 this comes by routing 24% of freight through 1 of the 4 cross-dock hub options and running lanes to 9 different FC clusters over the 4 week period analyzed. The figures below show the savings detail from week to week over the four weeks analyzed prior to fitting to the annual demand distribution.

2012 - Weekly Model Results

Week	% Savings	Lanes/Week	% Freight Consolidated	Total Weight (Normalized)*
FW28 - Off-Peak - Hi	5.8%	19	25%	0.90
FW30 - Off-Peak - Low	4.8%	19	22%	0.93
FW41 - Peak - Hi	5.6%	25	29%	1.24
FW43 - Peak - Low	4.6%	20	24%	0.94

* Weight is normalized to average over the 4 weeks analyzed *

Table 2 - 2012 Model Results by Week

In addition to choosing months in peak vs. off-peak period, each month targets the lowest and highest volume weeks in that month (designated by "Hi" and "Low" in the above table). The effect of increasing freight density is most apparent in the increased % of freight consolidated in FW41. In comparing FW41 and FW28, it's noted that a roughly 33% increase in volume flowing through the model increases the amount being consolidated by 18%. We see that savings are still slightly higher in FW28 but this is due to the opportunistic nature of some lanes having a larger cost disparity between direct and potential consolidation rates for the given week.

In 2013 42% of freight is routed to 20 FC clusters across 3 of the 4 cross-dock hub options during FW41. Further, savings are increased by 57% over the 2012 FW41 number by increasing the volume flowing through the model.

2013 - Weekly Model Results

Week	% Savings	Lanes/Week	% Freight Consolidated	Total Weight (Normalized)**
FW41 - Peak - Hi	8.8%	57	42%	2.26

** Weight is normalized to the 2012 average weight for the 4 weeks analyzed **

Table 3 - 2013 Model Results FW41

These simple results point to the importance of freight density and volume in defining load consolidation strategies. More lanes and hubs become attractive as there is more volume in the given area to drive higher outbound TL utilization. We also see that 85% of freight routed in 2013 is LTL, consistent with the idea that shipments with higher freight density are more attractive.

These model runs have many of the key constraints relaxed, including cross-dock sortation capacities, maximum number of hubs, and maximum number of lanes. Relaxing these constraints helps to establish an upper bound on the savings potential for the load consolidation initiative and allows us to identify the operational requirements needed to achieve that potential.

From the weekly routing patterns documented in Appendix D it's observed that some of the lanes selected during the identified week did not have enough volume to drive operation on a consistent basis. When bringing these recommendations to the scope of operational implementation, lanes with low utilization may not add value. Running more lanes out of a hub increases overhead, staffing needs and most importantly, in the case of using a potential Amazon cross-dock facility, reduces the available space for other operations. A more conservative recommendation may limit the lanes being run to force that they have at least 50% utilization initially. Utilization is simply defined as the number of days a lane was utilized relative to the 5 days per week the lane could be operating.

Looking at the hubs chosen in the 2012 and 2013 cases, it's noted that only 1 hub is chosen in the lower volume 2012 scenario. Hub-1 is the most attractive option out of the four and this is largely because 70% of unit volume from the Southwest region originates within a 120-mile radius of the hub. This again creates economies of scale by allowing more shipments in the area to utilize load consolidation because of their close proximity to the hub. As unit volume increases in the 2013 scenario, volume surrounding the other hubs achieves the minimum threshold where consolidation can occur at a slightly lower cost than their previous consolidation through Hub-1. While the spread use of 3 hubs rather than 1 hub does provide a slight cost benefit, there is a potential impact to operations from using more hubs as discussed earlier with the low-utilization lanes. Particularly in the case that Amazon-owned cross-dock facilities are used, there will be a loss in operating efficiency from spreading the cross-dock operation across multiple hubs with lower volume. This may not be a concern in the long-run as Amazon continues to grow and scale, but in the short-term increasing use of this program, the added simplicity of using one cross-dock hub for the identified region could provide a greater benefit.

From Appendix D, load consolidation is seen as most attractive for freight destined for the CHA-N and CVG-S clusters (by total trucks routed). Adopting load consolidation for these clusters generates effective load consolidated rates that are 17% and 9% lower respectively than historical direct rates out of the region. In both cases there is enough volume to warrant at least 1 truck/day when including small parcel freight. One point to note is that in order to take advantage of running lanes into CVG some care needs to be taken in negotiating effective contracts with TL carriers to cover the potential for increased deadhead costs into the Kentucky-Ohio area. It is often difficult for TL carriers to find loads exiting the region and hence shippers sending freight to the area are forced to pay "deadhead" costs to cover the under-utilized trucker's return trips. In some cases historical invoice costs for freight loads entering the CVG clusters were up to double the cost predicted by the linear TL regression function.

4.1.2 Results: Transit Time (Performance)

One of our initial hypotheses and expectations was that overall transit time (TT) could be reduced by employing load consolidation strategies in the inbound network. Principally this would come from eliminating a number of the break-bulk processing touch-points that are used in SP and LTL ship-modes. While the consolidation and cross-dock activities do increase freight handling and processing time up-front, the thought is that much of that time could be made up for on the second leg of the transit by using the dedicated TL driver.

Using the 2013 run as reference we actually see transit time to most FCs increase marginally. Table 2 gives the break-down by lane. LTL transit remains relatively unaffected by the routing of 37% of its freight through consolidation. SP freight on the other hand is more adversely affected since direct transit times for SP freight are lower.

%Increase/Decrease in Expected TT		
Destination	LTL	SP
ONT-S	31%	42%
AVP-S	4%	16%
SEA-N	4%	35%
RNO-S	3%	29%
RIC-S	1%	13%
CAE-S	1%	14%
RIC-N	1%	-
CHA-S	0%	15%
PHL-S	0%	12%
PHX-S	0%	28%
SDF-S	0%	15%
BNA-S	0%	15%
CHA-N	-1%	14%
GSP-N	-1%	-
CVG-S	-1%	14%
IND-S	-2%	15%

** Numbers highlighted in red represent an increase in TT, green represents a decrease in TT**

Table 4 - Transit Time Performance of Model

Taking a deeper look at this however, the larger increases in TT (for both SP and LTL freight) come from routing freight to FC clusters that are relatively close to the origin locations. The ONT, SEA, PHX and RNO clusters are all located in the west/southwest part of the country and hence are close to the points of

vendor origin. In these cases the direct TL arc does not get a chance to make up time over the SP and LTL break-bulk operations because the overall transit arc is a much shorter distance. This hints to the notion that adding cross-dock operations to short legs of transit may be inefficient for transit time performance. This is also backed from a cost standpoint because the model highly favors shipping along longer arcs (i.e. clusters located in the NE and MW) than the closer options located on the same coast.

Overall, the results from the initial runs of the model are positive. The cost savings range of 5.2% to 8.9% positively impacts the bottom line. While the reduced transit time performance does highlight a potential negative impact from the strategy, there are some actions that could offset the transit time impact. One of the more immediate alternatives would be to use team drivers on the outbound consolidation leg rather than standard single drivers; team drivers provide 24 hour driving coverage but at an increased cost. Further, working with the retail team to order goods 1 or 2 days earlier in these cases would eliminate the impact of potential increased transit time on product availability. The caveat, however, would be an increase in the Amazon cash cycle since the goods would effectively be in Amazon's possession for a longer period of time before being sold. Further, it may pose a greater difficulty for the retail team to forecast demand for an additional 1 – 2 days in the future. In the next section we will review opportunities to increase the savings impact of the load consolidation strategy by looking at opportunities to consolidate freight at the vendor level before pick-up.

4.2 Vendor Level Consolidation Opportunities

In addition to savings from the increased economies of scale on outbound trailer loads from consolidation, additional savings are expected on the first leg prior to entering the cross-dock hub. As a result of these opportunities we estimate an effective cost reduction of 20% in the first leg of consolidation (LTL and SP pick-up → Cross-dock hub). The result of this change translates to an increase in savings over 2013 to 13.7% in LTL and SP spend (up from 8.9%). Over this section we'll review 2 strategies that can help achieve these savings while section 4.3.3 will show the quantitative impact of these savings on total load consolidation costs.

4.2.1 SP Shipment Upgrades

In the case of small parcel shipments, the vendor level consolidation opportunity most readily appears in the potential to upgrade shipments from SP to LTL and hence take advantage of a cheaper first leg consolidation rate. A simple example to consider is the case with 1 vendor who has multiple orders destined to multiple FCs that would be routed through the same cross-dock hub. In this case, if the combined weight of those shipments exceeds the 150 pound mark they could likely take advantage of cheaper rates through LTL pick-ups. In one example, this opportunity was explored across three eastbound lanes (PHL, CHA and IND) originating out of the Southwest region. The resulting analysis suggested that a 34% saving in first leg costs could be achieved if the identified shipments were upgraded to LTL from SP and the actions below could be carried out:

1. That the vendor is capable of palletizing freight and is willing to combine multiple orders to different FCs into one freight bill. Support would be needed by the retail team to initiate negotiation with the vendors to execute this action.

2. That the necessary software changes can be made to club shipments initiated from one vendor to multiple FCs into one freight bill.

3. That the hub being used to cross dock is capable of breaking and sorting potential mixed destination pallets to the appropriate lane. If using a 3PL hub to perform this operation, there will likely be an increased variable handling cost at the hub.

Based on this example we use a conservative 20% saving on SP rates in the more aggressive model run.

4.2.2 Order Frequency Alignment and Optimization

Another opportunity to consolidate at the vendor level manifests itself in order frequency alignment. The thesis of Chong Keat NG, LGO '12 (Ng Keat 2012), discusses a number of freight consolidation strategies implementable in inbound supply chains. One principal area of focus and finding was on the savings potential from reducing the shipment frequency of orders by increasing coordination with

vendors. The result would lower the number of shipments and increase opportunities to consolidate freight at the vendor level before pick-up. In our case, aligning order frequency across multiple FCs (for shipments that would be destined for the same cross-dock hub) would add incremental savings. For LTL carriers this consolidation at vendor level could result in upgrades to MSTL shipments and potentially full TL shipments. The savings potential quantified for LTL shipments in this analysis ranged from 10 – 30% and hence the mid-point, a 20% saving rate, on LTL first leg rates was used to account for this opportunity in the more aggressive model run. In the next section we will take a look at how changes like the above savings on first leg rates affect the model in the sensitivity analysis.

4.3 Sensitivity Analysis

Much of the benefit from this optimization tool comes from the ability to quickly manipulate important parameters for sensitivity analysis. From a managerial standpoint this can help identify actionable target areas for the organization to improve cost or performance in the network. The table below outlines some of the key parameters that can be manipulated in the model.

Parameter Name	Description
spsave	Parameter to adjust the SP rate for 1st leg of consolidation by identified percentage (0-1)
ltlsave	Parameter to adjust the LTL rate for 1st leg of consolidation by identified percentage (0-1)
tlsave	Parameter to adjust the cost/mile of 2nd leg (TL) for consolidation by identified percentage (0-1)
max_pool	Parameter to adjust maximum pooling time (in days).
maxH	Parameter to control maximum number of hubs activated in network.
maxL	Parameter to control maximum number of lanes a given hub can use.
h_cost_s, h_cost_l	The sortation cost/carton (small parcel) and cost/pallet (LTL) to cross-dock at a given hub.
truckut	Target trailer utilization for outbound TL (0 – 100%); maximum trailer assumed to be 3960 ft^3
time	Transit time performance target - % that consolidation transit time can exceed direct transit time for comparable arcs.
sp_vol_g, ltl_vol_g	Demand growth factor for all origin nodes (quick manipulation of volume without updating actual demand inputs)

Table 5 - Key Sensitivity Parameters

While there are a number of potential scenarios to review, we will focus on 3 key ones in this section: inclusion/removal of SP freight in consolidation, the usage of hub pooling days and savings on carrier rates (LTL, SP and TL). For all sensitivity analysis we look at one week of data (FW41) under 2012 demand assumptions.

4.3.1 Effect of Small Parcel Freight

LTL freight is the most natural and least demanding option to implement load consolidation immediately within the inbound network. Since freight is already palletized, there is minimal space and infrastructure needed to run the cross-dock operation. If, for example, an Amazon-owned cross-dock hub were used, the only needs would be dock space and minimal labor to perform the pallet transfers from LTL to TL trailer. Consolidating SP freight, on the other hand, is a bit more involved – freight would need to be sorted and fluid loaded in its carton form on to the trailer which is much more time consuming and requires more resources. If the cartons need to be palletized prior to load, this would again require even more dock space and time to perform the operation. Hence, it is tempting to initially leave SP freight out of consolidation, especially since items shipped by SP are typically of smaller weight and dimensions and are thus less efficient at filling trailers. In our model however, we see that when restricting to only consolidate LTL freight, overall savings are reduced by 40%.

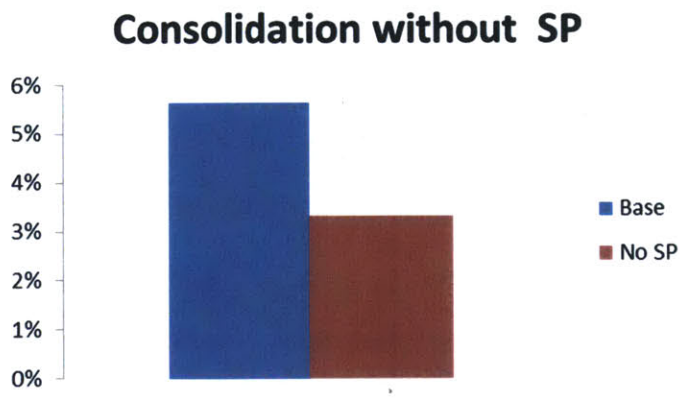


Figure 11 - Load Consolidation without SP Freight

While SP consolidation constitutes only 15% of consolidated volume in the base scenario, it has a dramatic impact on savings because of its significance in driving the second leg TL utilization. When SP freight is not consolidated, 7 fewer lanes are activated and the effective total LTL consolidation rate (1st leg LTL rate + hub sortation rate + effective 2nd leg TL rate) is increased by 7%. This highlights that including SP freight in load consolidation is a big driver in providing incremental savings for the program.

4.3.2 Effect of Freight Pooling

Truck utilization on the cross-dock hub outbound lane is a large factor in driving savings. By allowing freight to pool for an additional day, more lanes become economical to run due to the increase in TL utilization. Under the current model assumptions, adding an additional day of freight pooling saves an additional 58%/week (up to 8.9% from 5.6%) by opening 4 additional lanes. This action of course impacts transit time by an additional day so steps would need to be taken to mitigate its impact on customer experience. Updating transit time expectations with the retail organization on the affected load consolidation lanes would be necessary. Further, managing daily trailer utilization for the identified lanes would be critical in keeping cost expectations in line. Running under-utilized trucks negatively impacts the cost of consolidation transit and this becomes a larger risk with the lower volume lanes that benefit from freight pooling the most.

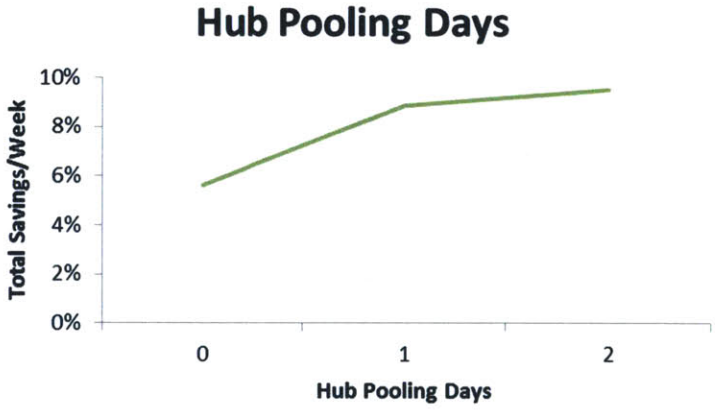


Figure 12 - Effect of Freight Pooling on Load Consolidation

It's also clear from the chart that there are diminishing returns to using this strategy. Allowing freight to pool for a second day, only provides an incremental savings of 9%/week (up 9.5% from 8.9%). The biggest bang comes from allowing freight to pool for just 1 day.

4.3.3 Effect of Freight Rates

In this sensitivity analysis we quantitatively look at the effect of rate savings similar to those discussed in section 4.2. We start with the base scenario in FW41 of 2012 assuming there are no savings on any legs of

transit and then adjust the savings on each leg to see the impact. While a key aspect of load consolidation savings is lowering costs on the more expensive first leg of transit, the results below suggest that once the freight is within reasonable proximity to the cross-dock hub, the additional savings are highest by reducing costs in the outbound TL portion of the shipment. This makes sense because for longer haul shipments (i.e. California → CVG), the TL leg represents over 50% of the total freight cost in load consolidation. To test the effect of these savings in the model, we look at reducing SP rate, LTL rate and TL rate by 5%, 10% and 20%.

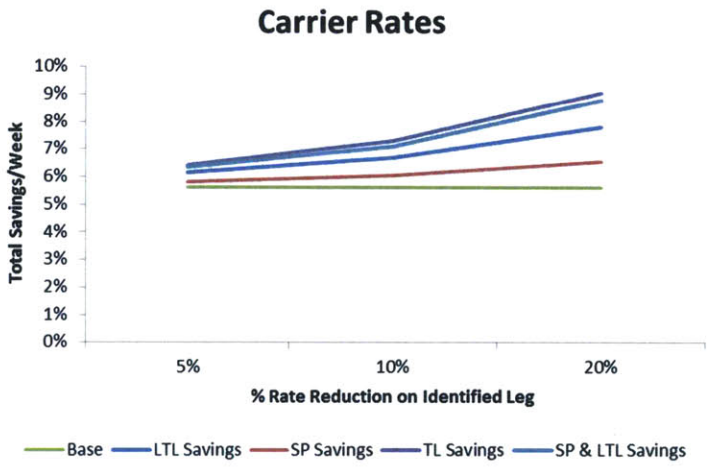


Figure 13 - Effect of Carrier Rates on Load Consolidation Savings

The purple line showing the effect on savings from a reduction in the TL rate reports the highest total savings rate/week throughout the 5 – 20% band while saving 20% on both LTL and SP legs (turquoise line) is a close second. It's also noted that saving only on the SP leg of transit (red line) has the lowest effect throughout the band – we should expect this since only 15% of freight of consolidated freight in the initial model results was SP.

In order to achieve these savings there are a few strategies Amazon can employ. For savings on the TL leg of transportation, one example would be using intermodal/rail vs. a standard truckload ship-mode. For transit over the 1,300-mile range, the intermodal ship-mode begins to offer lower pricing than standard

truckload[12]. To take advantage of this option, more attention would need to be focused on managing the scheduling complexity associated with this ship-mode – there can be a lot more variance in rail schedules and transit times are often longer. Another way to reduce costs in the TL leg of transport is to improve trailer utilization. Since the full trailer is being purchased, any improvements that drive greater utilization and allow more products to be packed on the same truck will yield additional savings. While many trailer utilization projects in Amazon have focused on increasing trailer efficiency for the outbound organization (savings are much higher because more is spent here), similar initiatives can be considered to improve trailer utilization in the inbound organization. Currently, much of this effort is left up to the vendor, who does not have much incentive to spend additional time packing an efficient truck.

The cost of the first inbound leg can be lowered by reducing the number of pick-ups and freight bills through aligning order frequencies, and clubbing shipments together destined for the same hub. We discussed a number of opportunities in this regard in section 4.2. The 20% savings we estimate on LTL and SP of legs of transit as a result of these actions can drive up to a 56% increase in total load consolidation savings (up to 8.8% from 5.6%).

4.3.4 Summary of Sensitivity Analysis

Using the sensitivity analysis as a tool, managers can focus on initiatives that have the greatest cost impact to the network. From the above discussion, we make a few preliminary conclusions:

1. Including SP freight in the potential load consolidation strategy is important – there is a 40% reduction in network savings from *not* including SP freight in the scenario executed.
2. Allowing freight to pool on some lanes has a large impact on cost savings – we see a 58% increase in savings by allowing freight to pool for 1 day. However, it is important to understand the resulting transit time increase in conjunction with this cost saving.

[12] W. Plunkett, G. Taylor and J. English, *Intermodal Profitability Analysis in the Truckload and LTL Industries* (Fayetteville, AR: Mack-Blackwell Transportation Center,[1998]).

3. Targeting cost savings initiatives on the outbound hub lane may yield the largest savings – there are a number of strategies that could help save on the shipping rates for all legs of transit, however, focusing on the TL leg which constitutes a large portion of total transit costs would yield the most benefit.

This sensitivity analysis concludes the discussion of the model's results. We observe that there is significant cost-saving potential from implementing the load consolidation strategy within the regional inbound network evaluated. The sensitivity analysis also helps to target some areas of focus that could increase the savings impact of this strategy further. In the next chapter we will discuss the qualitative impact of the load consolidation strategy

5 Qualitative Implications of Load Consolidation

In this chapter we will discuss some of the more qualitative implications of the load consolidation strategy. Further, we'll discuss some of the pros and cons of utilizing Amazon-owned hubs to perform the required cross-dock operation vs. using 3PL carrier hubs. Finally, we will conclude by addressing potential improvements to the model and methodology as well as future opportunities with this project and the inbound transportation organization.

5.1 Additional Benefits

5.1.1 Environmental Impact

Companies have become increasingly aware of the importance of green supply chains in business today. They are realizing that by reducing their carbon footprint, they not only reduce negative impact on the environment but also see significant cost savings in doing so. A case in point is Wal-Mart, one of world's largest retailers, who undertook a large scale project in its supply chain organization to reduce packaging usage from their suppliers by 5%. This project resulted in a 667,000 m^3 reduction in CO_2 emissions and consequently an annual saving of $3.4 billion (Hoffman 2007).

As we have seen from our model analysis, load consolidation represents a significant opportunity to save on transportation and logistics costs. There is also a significant environmental impact of this strategy. By consolidating, fewer LTL and SP trucks will be utilized, replaced by higher utilization TL shipments. Since the overall transportation miles are reduced and the ton-miles per vehicle per year are increased, there will be a significant reduction in CO_2 emissions. In a paper by Ali Ülkü the author discusses a new approach to computing potential savings in CO_2 emissions by using a discrete-time load consolidation model (Ali Ülkü 2012, 438-446). In a world more and more focused on environmental sensibility, focusing on strategies that reduce carbon emissions can yield benefit in more ways than one. Customers respond well to organizations that are focused on greening their supply chains and this often translates to increased customer goodwill.

5.1.2 Supply Chain Flexibility

While the hub and spoke architecture does add nodes to the existing inbound transportation network, there are additional benefits that can come from these nodes. In the case of Wal-Mart, a pioneer of using cross-docking in retail, these hubs are the life-blood of the distribution network. Through efficiently enabled operations, they allow Wal-Mart to maintain a low cost of inbound logistics – in 1993 this was estimated to be 3.7% of discount store sales vs. 4.8% for its direct competitors (Bradley and Ghemawat 2002, 6-7). In the context of the load consolidation strategy, Amazon differs from Wal-Mart in that cross-dock hubs would primarily be needed to consolidate freight and save costs, as opposed to serving as true distribution centers. However, utilizing cross-dock hubs throughout the network could also enable the opportunity of re-routing freight real-time based on demand needs at different FCs. Further, in scenarios where Amazon-owned FCs are used as cross-dock hubs, the option of shipping directly from the hub to consumers for high-demand items could turn out to be a competitive advantage. Without load consolidation, the in-transit goods on an LTL or SP truck would have no option to be re-routed. Adding an additional touch-point in the network could provide this flexibility and lends itself to a more flexible supply chain. Nevertheless, we remark that additional analysis is needed to fully assess the operational implications of such a strategy.

5.2 Who owns the cross-dock?

We discuss earlier that the cross-dock hubs could be modeled as Amazon-owned and operated facilities or those of 3PL logistics providers. In our model analysis, we make no distinction between the cost models of each. However, in practice, there can be different benefits of using each alternative. Leveraging the existing hubs of the current 3PL provider base gives the benefit of speed and scale. If contracts are initiated with LTL or SP providers to perform the cross-dock operation, the hub network of these 3PLs could be quickly utilized with no investment costs or extended time to do so. This would allow Amazon to quickly reap some of the savings of load consolidation without a significant impact on their current

operations. The negative side of this approach is that Amazon is paying for the service of cross-docking, similar to the small parcel zone skipping case in section 4.2.1.

In the case of utilizing an Amazon-owned hub for cross-docking, there are a number of benefits. One benefit is cost sharing in the growing base of FCs being built – there is an abundant and growing pool of available hub options. If cross-dock operations are to occur in these hubs, the capital and overhead cost of the cross-dock operations could be shared with running the standard distribution operations. This also has the added benefit of improving supply chain flexibility as discussed in the previous section. Further, running the cross-dock operation as an internal process provides more managerial control over the end-to-end supply chain. If relying on 3PL carriers for both the pick-up of freight and the sortation of that freight at cross-dock facilities, there is significant dependency on the competency of these 3PL providers. Scheduling would also be more seamless internally and again, this creates an easier opportunity for Amazon to adjust and be flexible to the needs in the supply chain.

While having the cross-dock operation internal to Amazon does provide significant benefit in terms of control, flexibility, and cost, the impact to operations is an important aspect to consider. Fulfillment centers are primarily built to "fulfill", i.e. to ship products to customers. Operations in FCs, similar to operations at the brick and mortar retail locations of Wal-Mart, are optimized to receive goods and distribute those goods to the customer base. When looking at available FC capacity during peak periods, dock doors are at a premium because they are highly utilized to handle those functions. Adding a cross-dock operation to the mix would increase operational complexity and in some cases require an increase in dock space. In the case of cross-docking non-palletized shipments, this would likely require capital investment in the form of new sortation equipment to handle the cross-dock sortation. The take-away here is that appropriate operational planning would be needed to ensure that the core operations of a fulfillment center are not interrupted by the added operations of cross-docking.

While leveraging 3PL logistics providers to perform the load consolidation opportunity does find its benefit in limiting capital expenditure and speeding up implementation, internalizing the operation may be a better long-term strategy. As Amazon continues its growth, its vendor base will continue to grow, and larger scale consolidation opportunities will exist. Further, its supply of FCs to use as future cross-dock locations will continue to grow. By bringing this operation in-house, the organization would be less reliant on the performance of 3PL providers and would be able to better harness the operational excellence of its core business model to perform consolidation in a more cost effective way while also increasing the flexibility of its supply chain.

5.3 Next Steps and Future Opportunities

This project studies the potential impact of load consolidation strategies within a large inbound supply chain. The model developed is a proof of concept for the Southwest region and demonstrates realizable savings targets for the program to achieve. In order to take full advantage of this strategy, there are a number of subsequent steps to follow-up with and opportunities that could be further explored in future projects.

1. **Use the model to analyze other regions and the full network** – The process of preparing and consolidating data as inputs into the model is lengthy and time consuming. This drove the decision to examine the load consolidation strategy only for the Southwest region under the available time constraint. To truly understand the potential of load consolidation in the domestic network, it is necessary to analyze other regions – particularly regions with different freight characteristics than those coming from the Southwest. Further, a fully comprehensive domestic model, could give better guidance and strategic direction to the question of cross-dock hub location in the context of the growing inbound network. One of the key deliverables at the conclusion of this project is to ensure appropriate knowledge transfer to other logistics engineers within the organization to allow for this next stage of development and analysis to occur seamlessly.

2. **Identify operational and technical requirements needed for cross-docking** – As discussed earlier, there are three primary ways to implement this strategy at Amazon: using existing FC locations as cross-dock hubs, building new cross-dock hubs, or using 3PL carriers hubs to perform the needed cross-docking. In each case, appropriate planning needs to take place for the solution to be scalable. For example, shipping destination labels would need to be updated to reflect intermediate consolidation hubs as a stage in the shipping cycle. Further, sortation planning for small parcel freight needs to be addressed as more manpower and sortation equipment are needed to support the consolidation of carton-level shipments from multiple vendors. Wang (Wang 2008) discusses many of the operational strategies that modern single-stage cross-docks utilize to help master operational efficiency.
3. **Consider implementation of lanes recommended by the model** – The model recommends as many as 14 load consolidation lanes to be implemented out of various hub locations in the Southwest. After reviewing these opportunities against the operational and technical requirements, Amazon should consider onboarding the new lanes.
4. **Consider predictive model for freight rate estimation** – The model currently uses a flat rate structure to provide the cost for SP and LTL prices based on historical averages. While this does have an advantage in deriving first-cut managerial implications (as it is easy to quickly update the rate of a particular lane to do sensitivity analysis), it does limit the accuracy of data to locations that Amazon has previously shipped to/from. It also forces us to make assumptions that the cost on a 3-digit zip arc inherits that of the parent state-state historical averages. The full network map built in to the model includes distance mapping between all nodes which is only currently used to estimate costs of the TL leg in our linear regression. Using a more predictive regression model similar to that used by Özkaya (Özkaya et al. 2009) to predict LTL rates would help to smooth out some of the inconsistencies in the historical data-sets and provide a more objective estimate of arc costs based on hub locations since miles traveled is an important factor in the prediction.

5. **Assess impact of demand variability on model** – With the limitations of this model's structure inbound demand was only evaluated over a discrete 4 week period. To better understand performance in a more holistic and dynamic context, the effect of demand variability should be incorporated. Using data from the off-peak vs. peak business cycles, the variance in model performance and lane selection was noticeable. Pairing this model with a simulation under varying demand conditions would be an opportunity to better assess the impact of demand variation on the model's recommendations.

Continuing with these next steps and opportunities will increase the effectiveness of the network modeling tool and better prepare the inbound transportation organization to evaluate and fully adopt load consolidation as a useful strategy in its operations. In the next chapter we will conclude with a few closing remarks.

6 Conclusion

Amazon's growth over the past years has provided significant benefit for its top line but has also created the opportunity to explore numerous cost savings opportunities for its bottom line. The inbound transportation organization's operations are one of many stages of the supply chain affected by this growth, and it must adapt to new strategies to remain competitive. Over the course of this research project, I specifically study the utilization of the "hub and spoke" distribution framework as an opportunity for the inbound supply chain to lower cost and improve transit time performance. Load consolidation is pinpointed as an important strategy within that framework to convert the currently heavily-used LTL and SP ship-modes in the inbound organization to more efficient TL ship-modes.

By developing a network optimization model, I am able to quantify the transit cost savings of this strategy in a region representative of 25% of the AmazonPay LTL and SP shipment volume. The model shows a potential saving of 13.7% from load consolidation. In addition, the model can be used as a tool to gain strategic insights into where hubs should be placed and how consolidation freight should be routed throughout the network. Further, it finds its usefulness as a managerial tool to guide decision making based off the sensitivity analysis of important model parameters. Qualitatively we also see that load consolidation can offer an opportunity to increase supply chain flexibility and reduce carbon footprint.

Over the course of this project I have had the opportunity to learn a tremendous amount about Amazon as well as deepen my knowledge in supply chain design and analysis. I sincerely enjoyed this opportunity and appreciated the help and support I received from everyone involved in making this a reality.

7 Bibliography

Ali Ülkü, M. 2012. "Dare to Care: Shipment Consolidation Reduces Not Only Costs, but also Environmental Damage." *Int. J. Production Economics* 139: 438-446.

Amazon. "Amazon: Company Facts.", accessed 02/28, 2013, http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-factSheet.

———. "Amazon: Overview.", accessed 05/01, 2013, http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-Mediakit.

Andel, Tom. 1992. "Parcel Shippers are Put to the Test." *Transportation & Distribution*, Apr 1992, 34.

Baykasoglu, Adil, Vahit Kaplanoglu, Rizvan Erol, and Cenk Sahin. 2011. *A Multi-Agent Framework for Load Consolidation in Logistics*: The Free Library: Transport.

Bradley, Stephen and Pankaj Ghemawat. 2002. "Wal*Mart Stores, Inc." *Harvard Business Review* 9-794 (024): 6-7.

Chainalytics. "Chainalytics Unveils Tool for Analyzing Freight Markets.", accessed 05/05, 2013, http://www.supplychainbrain.com/content/general-scm/quality-metrics/single-article-page/article/chainalytics-unveils-tool-for-analyzing-freight-markets/.

Coyle, J., E. Bardi, and R. Novack. 1994. "Transportation." In *Transportation*. Fourth ed., 402. New York: West Publishing Company.

Delta. "Delta History.", accessed 02/06, 2013, http://news.delta.com/index.php?s=18&cat=39.

Doherty, Kyle. "The Power of a Simple Business Model.", accessed 02/28, 2013, http://www.kydoh.com/page/2/.

Dunn, Jennifer. "Locations of Amazon Fulfillment Centers.", accessed 02/17, 2013, http://outright.com/blog/locations-of-amazon-fulfillment-centers-2/.

Hoffman, W. 2007. "Who's Carbon Free?" *Traffic World*, October 22.

Hoover's. 2013. *Hoover's Inc.: Amazon.Com, Inc. Profile*: Hoover's.

Kirkeby, Kevin. 2012. *Industry Surveys. Transportation: Commercial*. New York, NY: Standard & Poors.

LMS Logistics. *Ten Best Practices for Motor Freight Management: An LMS White Paper*. St. Louis, MO: LMS Logistics.

Ng Keat, Chong. 2012. "Inbound Supply Chain Optimization and Process Improvement."Massachusetts Institute of Technology.

Özkaya, Evren, Pınar Keskinocak, Roshan Joseph, and Ryan Weight. 2009. "Estimating and Benchmarking Less-than-Truckload Market Rates." *Transportation Research* E.

Paul Huppertz. 1999. "Market Changes Require New Supply Chain Thinking." *Transportation & Distribution*, Mar 1999, 70-74.

Plunkett, W., G. Taylor, and J. English. 1998. *Intermodal Profitability Analysis in the Truckload and LTL Industries*. Fayetteville, AR: Mack-Blackwell Transportation Center.

Ratliff, H. Donald, John Vate, and Mei Zhang. 2004. ***Network Design for Load-Driven Cross-Docking Systems***. Atlanta, GA: Georgia Institute of Technology.

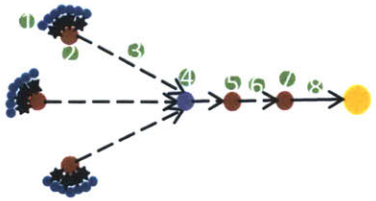
Wang, Jiana-Fu. 2008. "Operational Strategies for Single-Stage Crossdocks." Ph.D., University of California, Irvine.

Wooldridge, M. and N. R. Jennings. 1995. "Intelligent Agents: Theory and Practice." *The Knowledge Engineering Review* 10 (2): 115-152.

Wulfraat, Marc. "MWPVL International Supply Chain Experience: Amazon.Com Distribution Network.", accessed 02/24, 2013, http://www.mwpvl.com/html/amazon_com.html.

Appendix A – Ship-mode Diagrams

LTL Delivery

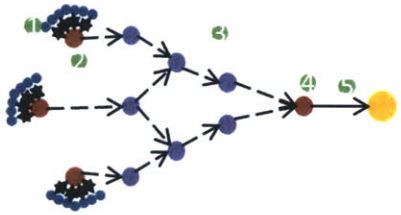


1. Local carrier hubs pick-up at vendors after morning deliveries
2. Amazon **pallet** shipment cross-docked on to linehaul trucks
3. Linehaul leg of route
4. There can be an arbitrary # of stops & additional break-bulks on linehaul leg (most likely case is 0)
5. Linehaul leg to regional carrier hub. Step eliminated for regional LTL carriers (nationals will do full)
6. Delivery from regional to local hub
7. Truck is broken down and sorted to truck for Amazon FC delivery
8. Carrier morning delivers to FC

TL Delivery

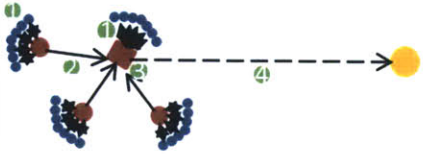
1. Local carrier does **pallet** pick-up at 1 vendor
2. Carrier delivers strait to Amazon FC with no stops

SP Delivery



1. Local parcel trucks do pick-ups at multiple vendor locations
2. Amazon **cartons** are sorted and organized on to 1st leg shipment truck
3. There are an arbitrary # of stops and additional sortations across linehaul route (higher than LTL linehaul route)
4. Amazon cartons are sorted to local parcel delivery truck
5. Amazon cartons are delivered at local FC

Load Consolidation Delivery

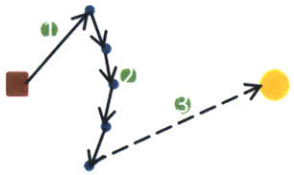


1. Local pick-ups are done at various LTL regional hubs close to vendor
2. Regional hubs deliver Amazon **pallets** to local 3PL consolidation hub (ignore for pickups originating at consolidation hub)
3. Amazon pallets are cross-docked to trucks for consolidation lanes
4. TL carrier completes linehaul delivery leg without stops to Amazon FC

Legend

●	Vendor Pick-up
●	Carrier Hub (Pick-up/Drop-off)
●	Carrier Hub (Interim) – Arbitrary # of points
■	Carrier Hub (Consolidation/Linehaul)
●	Amazon Fulfillment Center
←→	Pick-up & Return Leg
→	Pick-up / Delivery Leg
- - →	Linehaul Leg

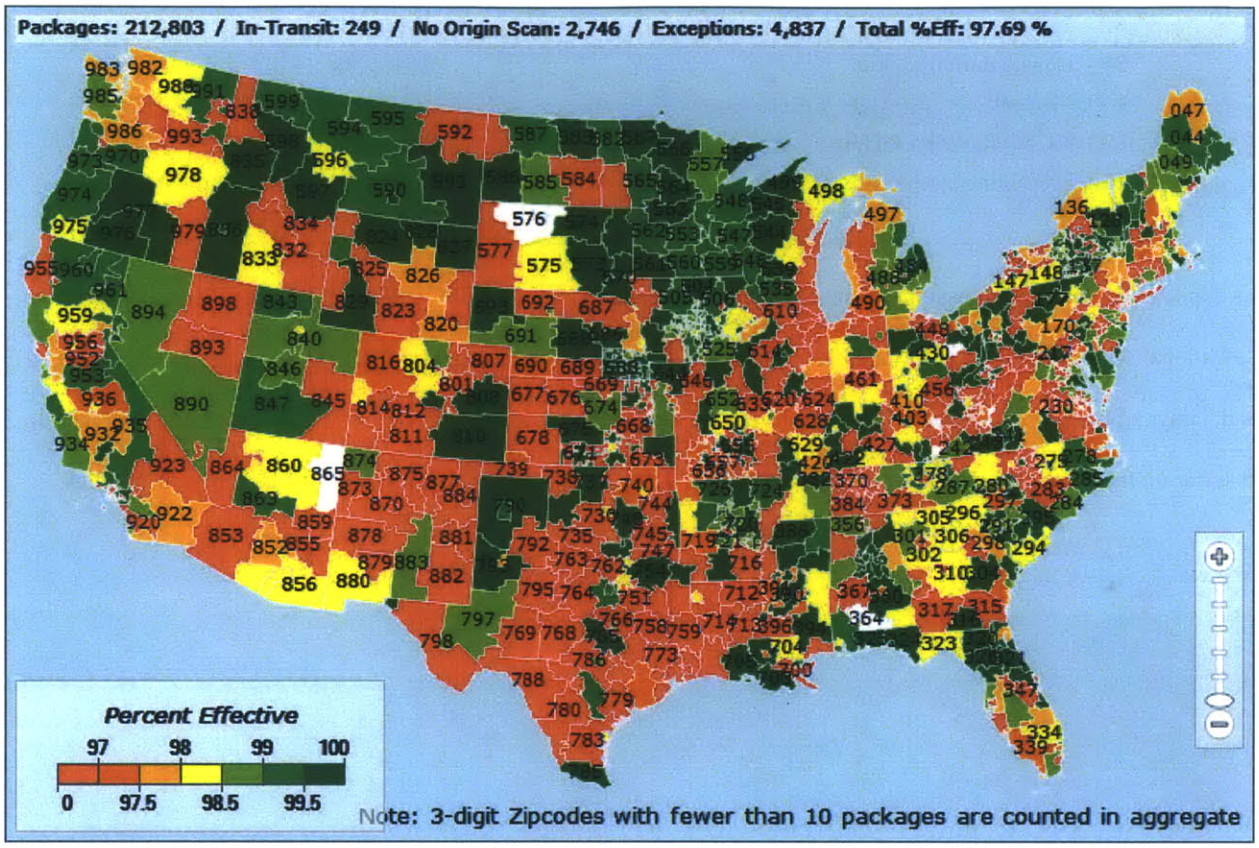
Milk-run



1. Carrier does **pallet** pick-up at first vendor for milk-run route
2. Carrier continues to do pick-ups at next legs of milk-run route (routes are static or dynamic dependent on program)
3. Carrier does final delivery to Amazon FC(s)

* This process includes SMR & DSP deliveries. DSP deliveries differ in that shipment drop-offs will be done as well.

Appendix B – 3 Digit Zip-Code Overview



http://www.parcelview.com/images/map1-lg.jpg

Appendix C – Expected Transit Time Assumptions

Transit Mode	<300 Miles	300+ Miles
SP – Direct	2.5	(Miles-300)/450 + 2
SP – Consolidation Pickup	1.5	(Miles-300)/450 + 1.5
LTL - Direct	2.5	(Miles-300)/450 + 3
LTL – Consolidation Pickup	1.5	(Miles-300)/450 + 1.5
Hub Processing	.5	.5
TL	1	(Miles-300)/450 + 1

This expectation calculation assumes a standard solo driver for TL can get 300 miles on the first day of pick-up and an additional 450 miles every day thereafter. For SP and LTL direct arcs the assumption is that there is an additional 1 and 2 days of transit for SP and LTL respectively in break-bulk processing. For SP and LTL consolidation pick-up and sortation the total expected times are 2 days plus additional expected transit time for distance over 300 miles.

Appendix D – Projected Lane Utilizations for 2013

Lane Utilization (FW28 & FW30)

	% Lane Active	Total Trucks (Normalized)
HUB-1		
IND-S	100%	1.85
PHX-S	100%	2.45
CHA-N	100%	3.31
SDF-S	100%	1.95
CVG-S	100%	2.15
CHA-S	80%	1.76
CAE-S	60%	0.77
RNO-S	60%	0.60
ABE-S	50%	0.82
AVP-S	40%	0.56
TUL-S	20%	0.28
HUB-2		
CHA-N	70%	0.76
SDF-S	60%	0.60
CVG-S	30%	0.30
CHA-S	20%	0.20
IND-S	20%	0.20
IND-N	20%	0.20
HUB-3		
SEA-N	40%	0.71
PHL-N	40%	0.41
ABE-S	10%	0.10
Grand Total		**1.00**

Lane Utilization (FW41 & FW43)

	% Lane Active	Total Trucks (Normalized)
HUB-1		
GSP-N	100%	3.09
RIC-N	100%	2.12
CVG-S	100%	4.38
SDF-S	100%	1.77
PHX-S	100%	2.47
BNA-S	80%	1.99
RIC-S	70%	1.18
CHA-N	70%	3.08
CHA-S	70%	1.27
IND-S	60%	0.93
AVP-S	50%	0.72
RNO-S	50%	0.49
ABE-S	40%	0.55
CAE-S	20%	0.28
PHL-S	10%	0.11
TUL-S	10%	0.12
HUB-2		
CHA-N	50%	0.58
GSP-N	40%	0.39
CVG-S	40%	0.39
SDF-S	30%	0.29
RIC-N	20%	0.20
IND-N	20%	0.20
CHA-S	20%	0.26
AVP-S	10%	0.12
BNA-S	10%	0.10
HUB-3		
SEA-N	50%	0.69
ONT-S	10%	0.10
PHL-N	10%	0.13
Grand Total		**1.00**

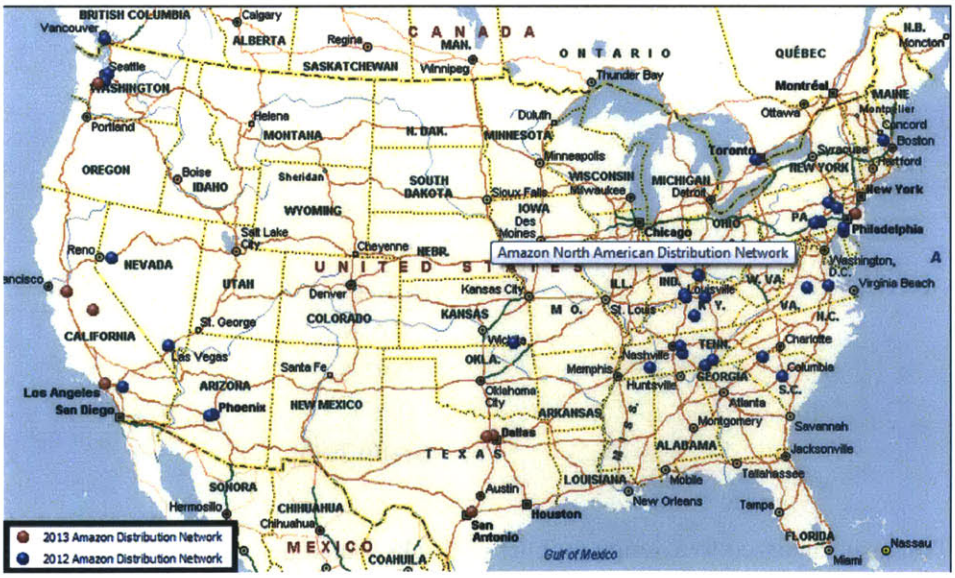
** Total Trucks normalized to average number of trucks routed per lane per 2 week period **

Appendix E – Fulfillment Center Cluster Mapping (Destination Nodes)

Region	Cluster	FC	Region	Cluster	FC	Region	Cluster	FC	Region	Cluster	FC	Region	Cluster	FC
MW	IND-N	IND2	NE	ABE-S	ABE1	NW	SEA-N	BFI1	SE	BNA-N	BNA2	SW	LAS-S	LAS2
		IND5			ABE2		SEA-S	SEA6		BNA-S	BNA1		ONT-S	ONT2
	IND-S	IND1			ABE3						BNA3		PHX-N	PHX5
		IND3		AVP-S	AVP1					BOS-S	BOS1			PHX7
		IND4		BWI-S	BWI1					CAE-S	CAE1		PHX-S	PHX3
	SDF-S	SDF8		PHL-N	PHL4					CHA-N	CHA2			PHX6
	TUL-S	TUL1			PHL5					CHA-S	CHA1		RNO-N	RNO2
					PHL6					CVG-N	CVG2		RNO-S	RNO1
				PHL-S	PHL1					CVG-S	CVG1			
					PHL3						CVG3			
					PHL7						CVG5			
				RIC-N	RIC1					GSP-N	GSP1			
				RIC-S	RIC2					SDF-N	SDF2			
										SDF-S	LEX1			
											SDF1			
											SDF4			
											SDF6			

****All FC locations and name conventions courtesy of Wulfraat, 2013[14] ****

The Amazon naming convention for fulfillment centers is consistent with the name of major airports in the domestic US. A suffix of 'N' represents a non-sortable cluster grouping while a suffix of 'S' represents a sortable cluster grouping. The map below gives an overview of these current locations[13]. Regions are arbitrarily defined for project organizational purposes and do not reflect internal Amazon segmentation.



[13] Marc Wulfraat, "MWPVL International Supply Chain Experience: Amazon.Com Distribution Network," http://www.mwpvl.com/html/amazon_com.html (accessed 02/24, 2013).