Utilizing Data and Data Analytics to Improve Supply Chain Performance

A literature review

Bachelor Thesis
Management - Data Science

Name
Pascal H. Aaldering

ANR
606686

Delivery date
December 9th, 2015

Faculty
Tilburg School of Economics and Management

Supervisor
prof.dr.ir. H.A.M. Daniels

Word count
7572

Abstract

In the current business environments, big data and data analytics are seen as the next revolution in management, business, and competitive differentiation. Firms that rely on the effective management of supply chains can gain significant gains in performance if data analysis techniques are correctly integrated into business processes, and utilized in an effective way. Research shows that data analytics can have an impact on supply chain performance and on business performance in general. In this thesis, different types of data and data analysis techniques are discussed regarding the area of supply chain management to see whether these techniques can have a substantial contribution to the performance of the supply chain. Through reviewing literature, the term supply chain performance is broken down into four main supply chain processes; Plan, Source, Make, and Deliver. For each of these processes, analysis techniques are discussed and analyzed to see how they can provide a contribution to improving the performance of the supply chain.

Keywords: data analytics, supply chain performance, big data, data integration.

Table of Contents

1 Introduction

The current business environment is evolving into a state where the analysis of obtained data can significantly help the decision-making process and improve business performance. As McAfee & Brynjolfsson (2012) observe, "...companies in the top third of their industry in the use of data-driven decision making were on average, 5% more productive and 6% more profitable than their competitors."

One of the areas of business where the analysis of (big) data can have a significant impact on decision making, process design and performance, is the area of supply chain management. The competitive environment is changing as well, and companies need to adapt to this change. Fawcett & Waller (2014) emphasize this statement with the following: "Because Darwin's warning rings true -that is, only the adaptable survive- more purposively adaptable supply chain design and governance is needed." Research shows that the use of business analytics can improve supply chain performance (Trkman, McCormack, Oliviera and Ladeira, 2010). The purpose of this thesis is to gain insight in how companies can effectively integrate (big) data and data analysis techniques into their business processes to improve supply chain performance.

1.1 Research questions

To be able to provide an answer to how companies can effectively utilize data analytics to improve supply chain performance, several sub questions will need to be answered. First, it is important to have clear definitions of the proposed variables, therefore the first sub question is: *How are data, big data, data analytics, a supply chain and supply chain performance defined?*

After having defined the variables of interest, focus will be on the different types of data that are generated throughout the supply chain process to gain better insight in different decision making processes. This leads to the following research question: *What are the different types of (big) data that are involved with the management of supply chains?*

When the different types of data are categorized, it is important to understand how these data from possible different sources are analyzed to create insight in the supply chain process. To address this matter, the following research question is constructed: *Which data analysis techniques can be used to effectively analyze data and gain insight in the supply chain process?*

If the collected data are from different sources, integration of these varying types of data into decision making processes is important for companies to have effective decision-making units that will improve future performance of the supply chain and the business process as a whole. This leads to the following research question: *How can data from multiple sources be integrated into a decision making process?*

Answering the above mentioned questions will lead to better understanding of the supply chain processes and the data that is integrated within these processes. The most important goal of this research process is to gain better understanding of the ability of data to improve supply chain performance. The following research question is constructed to address this vital aspect: *How can analyzed data improve supply chain performance?*

1.2 Research design

To evaluate the opportunities created by the use and analysis of big data, a literature review will be performed. To get access to relevant literature, the WorldCat.org database that is linked to Tilburg University and Google Scholar will be the main applications that are used. Keywords for relevant articles will be: big data, data analysis, data analytics, data science, business analytics, data integration, supply chain (management), business processes and decision making. This thesis will answer the research questions in the above mentioned order, therefore it will be structured as follows: First, existing literature will be reviewed to define the proposed variables 'data', 'big data', 'data analytics', 'supply chain' and 'supply chain performance'. Following this, a review of the various types of data involved in the area of supply chain management will be conducted to report types of data that are relevant and important for managing supply chains.

After having determined the definitions of the variables and the types of data involved, focus will be on literature that discusses data analysis techniques that can improve the level of insight within the supply chain process to be able to provide a summary of data analysis techniques that can have a significant impact on the supply chain process. When having determined relevant data analysis techniques, focus will be on the integration of data into decision making processes and business processes. Conclusively, information about data, data-analysis techniques and business processes will be combined to provide an overview of how firms can ultimately use these aspects to improve their performance for the main supply chain processes.

2 Defining data, big data, data analytics, a supply chain and supply chain performance

To be able to understand and comment on the effect that data, big data and data analytics can have on the performance of a supply chain, it is important to have a clear definition of the terms 'data', 'big data', 'data analytics', 'supply chain' and 'supply chain performance'. In this section, these variables will be defined by reviewing existing literature.

2.1 Data and big data

The term 'data' is a very common and understandable construct, it usually consists of a set of quantitative or qualitative variables that have a certain value. In itself, data is nothing more than a largely unstructured representation of a certain situation. The term 'big data' refers to large sets of data that are simply too big to handle for regular database management systems. Chen, Chiang and Storey (2012) state that "More recently big data and big data analytics have been used to describe the data sets and analytical techniques in applications that are so large (from terabytes to exabytes) and complex (from sensor to social media data) that they require advanced and unique data storage, management, analysis and visualization technologies." These large sets of unstructured and complex data contain information that can possibly improve the performance of vital business processes.

In a 2001 research report, Doug Laney discussed challenges and opportunities regarding the management of data. Laney stated that data management challenges existed along three dimensions: volume, velocity and variety (The 3V model). *Volume* refers to the (growing) amount of data that is available. *Velocity* stands for the speed at which new data are generated (getting close to real-time information) or for the frequency of data delivery (Russom, 2011). *Variety* refers to the different types of data that are available from different sources, making the data (partially) unstructured. These three elements of data and data management describe the basic dimensions in which (big) data exists, but do not capture the most important aspect that businesses in today's competitive environment face when dealing with (big) data, namely the capturing of value from these data.

A report from IDC (2011) defines big data in a way that captures the essence of harnessing new technologies to improve business performance: "Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis." This definition introduces the fourth V that can turn the 3V model into a 4V model: *Value.* As mentioned before, data in itself is nothing more than a set of variables that represent a certain situation. These data in itself have no value. Therefore it is of crucial importance to effectively analyze obtained data to get results that will have business value and can impact business performance.

An even more thorough and complete definition of big data is a definition where "big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective and innovative forms of information processing that enable enhanced insight, decision making, and process automation" (Gartner IT Glossary, 2013). This definition describes 'big data' as a new technology that demands cost-effective and innovative forms of information processing to gain better insight into various processes. One of the important goals of this thesis is to show how (big) data analytics can enable this enhanced insight and decision making for the specific area of supply chain management, with the ultimate goal of improving the performance of the supply chain.

2.2 Data analytics

Once data is collected, the process of analyzing the data can start. Data in itself can be seen as a large collective of answers to certain questions. The challenge that arises here is to ask the right questions so that the data will be transformed from a meaningless set of variables to meaningful information that can have an impact on a business process. In a 2013 paper, Vasant Dhar, professor at the Stern School of Business and director of the Center for Digital Economy Research, captured this challenge by stating that in a situation where the computer taunts us by saying "If you only knew what question to ask me, I would give you some very interesting answers based on the data!", we often don't know what question to ask. This statement underlines the importance of having the right data analysis techniques to transform raw data into meaningful information. This transformation can be compared to the process of manufacturing in general; raw materials (data) are obtained for further processing, these raw materials then enter the process where value is added (data analytics); the raw materials are getting transformed into end products (useful information).

More specifically, IBM states that big data analytics is the "use of advanced techniques against very large, diverse data sets that include different types such as structured/unstructured and streaming/batch, and different sizes from terabytes to zetabytes (one zettabyte is 1000^7 bytes) and the analyzing of this big data allows analysts, researchers and business users to make better and faster decisions using data that was previously inaccessible or unusable." The main message that can be distilled from this definition is that (big) data analytics allows professionals to make better and faster decisions by using data and data analysis techniques.

In a business perspective, the analyzing of data and using the information that is created throughout this process to support decision making is often referred to as *business intelligence (BI).* According to Sanders (2014), the combination of analytics and big data can together turn information into intelligence. Once business intelligence is created, it can have powerful impact on business performance if acted upon correctly.

2.3 *Defining a supply chain, supply chain management and supply chain performance*

To investigate the effect that the use of data analytics can have on the performance of a supply chain, the term 'supply chain performance' has to be broken down into concrete variables. To do this, a well-constructed definition of the term 'supply chain' is important since it forms the foundation of understanding the management of supply chains. Mentzer, DeWitt, Keebler, Min, Nix and Smith (2001) have performed a literature review on the definitions of a supply chain and supply chain management. They defined a supply chain as "a set of three or more entities (organizations or individuals) directly involved in the upstream and downstream flows of products, services, finances, and/or information from a source to a customer." Using this definition of a supply chain, the researchers (Mentzer et al., 2001) also constructed a definition of the management of supply chains (SCM) and described it as "the systemic, strategic coordination of the traditional business functions and the tactics across these business functions within a particular company and across businesses within the supply chain, for the purposes of improving the long-term performance of the individual companies and the supply chain as a whole."

This definition describes supply chain management by focusing on improving the long-term performance of a supply chain and the individual companies involved. Even though this definition was published in 2001, it still holds truth for companies today, mostly because business functions and the tactics used within these business functions can almost always be improved to create potential additional value for specific business functions or departments.

The performance of a supply chain is harder to define, since it encompasses a variety of key indicators that reflect the supply chain performance. Researchers (Gunasekaran, Patel & McGaughey, 2004) have constructed a framework for supply chain performance measurement that distinguishes three different processes that can be measured; strategic, tactical and operational processes. These processes contain a wide range of activities that describe specific elements of supply chain management which again can be divided into four categories: planning, sourcing, make/assemble and delivery. This subdivision into four main categories can be traced back to the Supply Chain Operations Reference model (SCOR): a model that was developed in 1996 by PRTM, a management consulting firm which is now part of PricewaterhouseCoopers LLP (PwC). According to the APICS Supply Chain Council, the SCOR model is the world's leading supply chain framework, linking business processes, performance metrics, practices and people skills into a unified structure. Zhou, Benton, Schilling and Milligan (2011) have conducted an extensive literature review on the SCOR model and describe the model as follows:

Level 1 contains the scope and content of the supply chain; the main supply chain processes: *Planning*; the process of balancing demand and supply to develop policies that meet sourcing, delivery and production requirements. *Sourcing*: the process of connecting manufacturers and suppliers to meet (planned) demand. *Make*: the process of transforming raw materials or products into finished goods to meet (planned) demand. *Deliver*: the process that provides finished goods and services to meet (planned) demand, and *Return*: the process of returning products or receiving returned products. Since the first four versions of the SCOR model did not contain the Return process, Zhou et al. (2011) stated that this process was not as mature as the other four processes and decided to focus on the four remaining processes that were widely adopted by practitioners. Level 2 describes the core process configuration and the type of the supply chain and Level 3 specifies the best practices of each process within the supply chain.

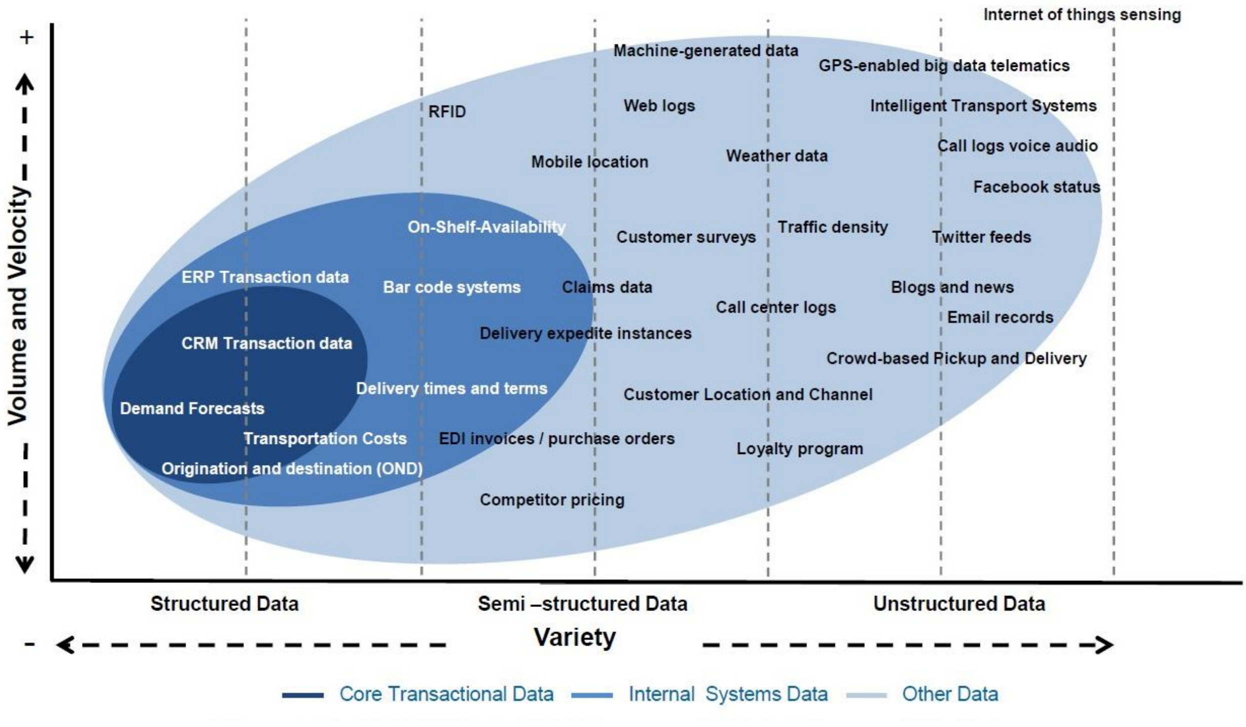
The SCOR model is an extensive supply chain performance analysis tool and it will not be used in this thesis to focus on deconstructing supply chain processes to evaluate them individually, but it is merely used to provide a solid foundation of understanding how supply chain performance is measured. To provide a clear overview of the opportunities and challenges that are involved with the use of data analysis technologies, data analysis techniques will be discussed and will be assigned to one or more of the four main supply chain processes as to how these techniques may affect these processes.

3 Types of data involved with the management of supply chains

Since companies involved in a supply chain do not only have a flow of raw materials and/or finished goods, but also a flow of information, a lot of data is created throughout the entire supply chain process. The main reason behind the enormous flow of information is the fact that accessing and storing data is becoming less expensive. Dhar (2013) states that most of this information is stored because businesses generally associate data with a positive option value. This is an understandable phenomenon because somewhere within these large unstructured sets of data, there may exist crucial information about certain business processes that can have a positive effect on business performance. Therefore, companies may assume that it is safer to store potentially useless data than to disregard potential useful data. This method of storing of large amounts of potentially useless data will increase the total costs of data storage and must be considered thoroughly.

Rozados & Tjahjono (2014) have analyzed current trends in Supply Chain Management and have identified 52 sources of big data across the supply chain. These types of data were categorized in Structured-, Semi-structured-, and Unstructured data and were scaled to the level of Volume and Velocity the data possessed. The main sources of data are shown in Figure 1.

Figure 1. SCM Data Volume and Velocity vs. Variety (Rozados & Tjahjono, 2014)



As shown in the figure, the core transactional data is mostly structured data whereas other (non-internal) data is often unstructured. The fact that only 8 out of the 52 data sources identified by the researchers belonged to the structured core transactional data that have a moderate level of volume and velocity leaving the number of unstructured data relatively large, shows an asymmetry in data sources; large amounts of data (high volume) that are generated at an increasingly fast pace (high velocity) are mostly unstructured of nature, whereas smaller amounts of data (lower volume) that are generated at a slower rate (moderate velocity), are mostly structured of nature. The three types of data are defined below.

3.1 Structured data

Structured data can be seen as the 'traditional' form of data. Chen et al. (2012) state that structured data is mostly apparent in BI&A 1.0 (Business Intelligence and Analytics 1.0), referring to companies that collect data through various legacy systems, and often store this data in commercial relational database management systems (RDBMS). As seen in Figure 1, examples of structured data within the supply chain are CRM Transaction data, Demand forecasts, ERP transaction data, Transportation costs and more. A common characteristic of these sources of data is the fact that they can be stored and managed in a RDBMS (relational database management system). In the IT Glossary of Gartner, the definition of a RDBMS is described as follows: "It is a DBMS in which the database is organized and accessed according to the relationships between data items. In a relational database, relationships between data items are expressed by means of tables."

Transportation costs are a good example of structured data. The costs for transportation of products or raw materials can easily be tracked in a database. They can also be linked to for instance weather reports, to see if there is a relationship between weather reports and delivery time; heavy rainfall could for example increase delivery time and thus the transportation costs. This proposed relationship between variables contains a structured variable on one hand (transportation costs) and a highly unstructured variable on the other hand (weather reports).

3.2 Semi-structured data

Semi structured data can be seen as data that has an irregular, implicit, and partial structure (Abiteboul, 1997). The irregularity of this data refers to the fact that the dataset often consists of heterogeneous or incomplete elements. In some places, information may be structured differently than in other places; if a firm, for instance, has international partners, currencies among systems may vary, which will lead to a semi-structured database if these two different types of data are merged. The data is also implicit, meaning that the structure is not obviously apparent on first sight. The partial structure of data speaks for itself; parts of the dataset may lack structure or have a very sketchy structure (unstructured text).

A good example of semi-structured data are the RFID (Radio Frequency Identification) chip and Barcode scanning systems, because although these sensors contain valuable information, when this information will be placed in a RDBMS, it will not have the structure or the scheme of the other elements in the database. Besides barcodes, products can contain an RFID chip that stores all the information. One of the biggest advantages of this chip is the fact that it does not require Line-of-Sight Scanning, whereas barcodes do require strict line of sight scanning (Michael & McCathie, 2005).

3.3 Unstructured data

Unstructured data can be seen as data that does not follow a specific format and does not reside in a traditional database. Losee (2006) states that unstructured data, such as text or images, contain information but contain no explicit structuring information, such as tags. A good example, also relating to the field of supply chain management, is social media content. Within these social networks, a variety of valuable information can be found, but it is not as simple as entering a search query in a database management system. Data is not labelled or tagged with an identifier, which makes the collection of possible useful data significantly harder than collecting structured data.

4 Data analysis techniques to gain insight in the supply chain process

Analyzing data sounds like an easy and straightforward process where the input is the collected data sample and the output is the information that can be acted upon. In reality, it is not that simple. As Lavalle, Lesser, Schockley, Hopkins and Kruschwitz (2011) state, "Information must become easier to understand and act upon", emphasizing that even if data has been analyzed, it still needs to be interpreted by managers and professionals so it can be transformed into action.

Before analytics can be performed, Lustig, Dietrich, Johnson and Dziekan (2010) state that businesses need to treat data and information as a strategic asset. Once companies recognize the power of data and information, the process of business analytics can begin. The analysis of structured data was categorized into three different categories by Lustig et al. (2010):

4.1 Descriptive analysis

Descriptive analysis is a set of technologies and processes that use data to analyze and understand business performance. This technique is often used to gain insight in past processes to approach the future with more confidence. Descriptive analysis usually answers the question of "what happened?" by deriving information from data.

Souza (2014) describes descriptive analysis and states that "real time information about the location and quantities of goods in the supply chain provides managers with tools to make adjustments to delivery schedules, place replenishment orders, place emergency orders, change transportation modes, and so forth." This type of information is very useful for managers and policy makers since it provides an overview of a situation by answering the "what is happening?" question. Integrating this type of information-processing technologies into business processes will help keeping track of everything that is happening within certain links of the supply chain.

4.2 Predictive analysis

Predictive analysis is a technique that analyzes real time and historical data to make predictions in the form of probabilities about future events (Rozados & Tjahjono, 2014). The basis of predictive analysis can be found in statistical and mathematical methods. One of the most important aspects of predictive analysis is the ability to forecast demand.

In a 2012 big data report, Lora Cecere (founder and CEO of Supply Chain Insights LLC), states that (supply chain) leaders are turning to new forms of predictive analysis to map multiple 'ifs' to multiple 'thens' through learning systems. Dhar (2013) states that "machine learning is characterized by statistical induction aimed at generating robust predictive models." To clarify the previous two statements; predictive analysis can be used to formulate multiple 'if A, then B' situations to improve responsibility and flexibility of the firm. Responsibility and flexibility are improved because companies will already have made predictions about demand fluctuations for example, and may have already constructed different policies for certain scenarios. This way, the policies only need to be put into practice when a predicted situation occurs. Cecere (2012) also states that the combination of new forms of pattern recognition, optimization and learning systems is improving the ability for the organization to improve the response.

Lustig et al. (2010) have classified predictive analysis into six categories. This classification is displayed in Table 1 with their accompanying descriptions. Possible applications of the techniques to improve supply chain performance, and to which main supply chain process these techniques might belong, are included in Table 2.

4.2.1 Predictive analysis techniques & application to SCM

Categories - Lustig et al. (2010)	**Descriptions - Lustig et al. (2010)**
Data mining	What data is correlated with other data?
Pattern recognition and alerts	When should I take action to correct or adjust a process or piece of equipment?
Monte Carlo simulation	What could happen? (A Monte Carlo simulation is a simulation technique where the process is repeated several times with different starting conditions. The result of these simulations is a distribution function that shows all possible outcomes.)
Forecasting	What if these trends continue?
Root cause analysis	Why did something happen? Trying to identify the root causes or problems of a phenomenon.
Predictive modeling	What will happen next if?

Table 1 – Analysis techniques and descriptions by Lustig et al. (2010)

Analysis technique	Possible application to improve supply chain performance	Supply Chain Process: Plan/Source/Make/Deliver/Return
Data mining/Pattern recognition and alerts	Discover demand patterns to anticipate future demand levels (linked with forecasting)	Plan/Source/Make/Deliver/Return – Discovering patterns in large data sets can lead to more efficient planning. For each supply chain process, patterns can be recognized that may provide insight or show weaknesses in the process.
Monte Carlo simulation	Evaluating risk by creating models that show assumed distributions of selected values. Wu & Olson (2008) used a Monte Carlo simulation for evaluating risk relating to different ordering plans.	Plan/Source/Make/Deliver – since a Monte Carlo simulation can give an assumed distribution of future events, it can be applied to any process to evaluate risk or show possible outcomes, given the right starting conditions.
Forecasting/Predictive modeling	Forecast demands using CPFR-Collaborative Planning, Forecasting and Replenishment systems.	Plan – Forecasting and predictive modeling belong to the area of planning since this process focuses on future demand levels so that policy can be adapted and adjusted to meet the requirements that are expected. Make (predictive modeling) – Responding to customer sentiment, supplier evaluation, questionnaires or industry standards so that production can be altered to meet customer demands.
Root cause analysis	Analyzing the management of recalls to prevent them. (Kumar & Schmitz, 2011)	Return – Products that get returned because of flaws in design or products that do not 'work' are sent back to the manufacturer, therefore root cause analysis can for instance be used in the area of the Return process. Deliver – Higher than average delivery costs on for instance a specific route may have an underlying cause, such as the number of traffic lights or crowded highways.

Table 2 – Possible applications to Supply Chain Management and placement in the main supply chain processes

4.3 Prescriptive analysis

Prescriptive analysis uses predictions based on data to inform and suggest proposed sets of actions that can serve to take advantage or to avoid on a particular outcome (Rozados & Tjahjono, 2014). Some of the prescriptive analysis techniques overlap with predictive analysis techniques because to be able to 'prescript' a future course of actions, there first must be a prediction of the situation in the future, for instance by using a Monte Carlo simulation where all possible outcomes are reviewed. This way, a proposed set of actions for the future can be constructed.

5 Integrating data from different sources into a decision making process

As early as in 1974, Galbraith discussed the design of organizations through an information processing view. He argued that uncertainty required more information processing to achieve a higher performance level; “the greater the task uncertainty, the greater the amount of information that must be processed among decision-makers during task execution in order to achieve a given level of performance.” He also stated that uncertainty limits the ability of an organization to preplan or make decisions in advance of their execution. This uncertainty can be reduced by firms through better planning and coordination.

One of the distinctive aspects of data and especially that of big data, is the variety of data. All this data can be transformed into information, if it is integrated successfully into the process of decision making and management. If the integration of systems and data coming from automation and simplification of processes inside the enterprise is performed skillfully, it will allow to support decision making (Kościelniak & Puto, 2015).

5.1 Data integration

When dealing with large volumes of data, the datasets are often not from the same source or of the same type. Therefore, integrating data from multiple sources into a unified (decision making) process is of significant importance for companies dealing with large amounts of data.

An example of a data integration process is proposed by Berry (2012). She identified four phases that describe the integration process from searching for information to analyzing data and creating value out of it:

Phase	**Description**
Search and navigate	Find information within any system or source
Consolidate	Combine relevant information from multiple sources
Correlate	Identify information relationships within disparate types of data from different systems
Analyze	Draw conclusions and uncover themes within data, as well as information relationships

Table 3- Process that can turn data into information (Berry, 2012)

The goal of the process mentioned above is to be able to draw conclusions and uncover information relationships and themes within the selected and consolidated data. The crucial factor in this process is being able to consolidate different sources of data into a unified (decision-making) process.

To do this efficiently, firms might consider implementing a Metadata Layer into the companies’ system environment. In the e-book *“4 Steps to Leveraging Supply Chain Data Integration for Actionable Business Intelligence”* by Take – Supply Chain (the supply chain division of Take Solutions Inc.; a global technology solutions and service provider), it is mentioned that data mapping using a metadata layer (Metadata is ‘data about data’, or ‘information about information’) provides the technical means to assure that the data from all systems is optimally integrated and accurate. This might sound like an ERP (Enterprise Resource Planning) system, but Take- Supply Chain states that “that strategy is no longer effective, nor will it help you stay agile and competitive in today’s networked economy.” Take – Supply Chain also states that this metadata layer “lowers total cost of ownership through centrally storing and managing all information about data sources, content, business rules, and access authorizations in a single metadata layer.” This way, firms will have an accurate overview of all data in the systems that were once disparate and disconnected, but are now accessed through the metadata layer.

5.2 The role of data scientists

Another important aspect of integrating data into a decision making process to improve supply chain performance, is selecting the right source(s) of data. One way to make the selection of data easier, is to hire a good data scientist. An important reason for hiring a good data scientist is the fact that they identify rich data sources, join them with other, potentially incomplete data sources, and clean the resulting set (Davenport & Patil, 2012). The most important quality of a data scientist that can have an impact on data analytics and the results that come out of it, is the ability to identify rich data sources and joining them with other, potentially incomplete data sets.

Kowalczyk & Buxmann (2014) performed a case study regarding big data and information processing in organizational decision processes and have stated that a combination of data-centric mechanisms and organizational mechanisms is needed for effective integration of analytic capabilities with domain-specific knowledge. Therefore, to successfully identify rich data sources, a data scientist not only needs well developed quantitative skills, but he also needs to have knowledge about the domain to identify potential problem- or opportunity areas. This domain can vary, as business environments differ per company.

Another important aspect of data and data-driven-decision making is data-analytical thinking. This skill is important, not just for the data scientist, but throughout the entire organization. The reason behind this is that managers and employees in other functional areas will only get the best from the company’s data science resources if they have some basic understanding of the fundamental principles (Provost & Fawcett, 2013).

Putting this in line with improving the performance of the supply chain, companies that rely on effective supply chain management need to hire data scientists that know how the supply chain operates and thus also have knowledge about how the performance of a supply chain is measured. Besides knowledge about the supply chain operations, data-analytic thinking also needs to be implemented in the organization culture and within functional departments. Only this way, data will effectively be selected from different sources to construct a broad view of the performance of the supply chain.

6 Improving supply chain performance by using analyzed data

After having discussed data analysis techniques and possibilities regarding the integration of data into decision making processes, the most important part remains untouched; actually improving the performance of the supply chain by using the analyzed data. To deconstruct the performance of the supply chain once again, for each of the four main processes in the supply chain (Plan, Source, Make and Deliver), concrete ways of improving these processes will be discussed here.

Research has already shown that the use of data analytics can have an impact on these main supply chain processes. Souza (2014) has categorized possible applications for data analytics for each process. The main results of this categorization are shown in Table 4 below.

Supply chain process	**Application of data analytics within this process**
Plan	**Independent demand**: plan for inventory safety stocks using predictive analysis techniques such as time series methods (moving average/exponential smoothing/autoregressive models) where time is the only predictor of demand. **Production planning**: the S&OP (Sales and Operations Process) uses aggregate demand forecasts to establish aggregate production rates, aggregate levels of inventories and workforce levels. This is mainly done to derive the Master Production Schedule (MPS) that shows periodic production quantities. (related to the Make-process) **Data mining**: precedes the use of causal forecasting techniques by finding appropriate demand drivers for a product that can be used in further analysis.
Source	**Strategic sourcing:** using an AHP (Analytic Hierarchic Process). This decomposes a complex problem -such as selecting a supplier from a large group of possible suppliers- into more easily comprehendible problems that can be analyzed individually. **Tactical sourcing**: prescriptive analytic techniques (such as game theory) are used to determine auction rules (Auctions where firms can buy parts, materials or services). A combination of statistics and game theory can also be used to prescribe more sophisticated contracts that can improve retailers' product availability.
Make	**Strategic:** using a MILP (Mixed Integer Linear Program) to determine the optimal location and capacity of plants, DCs (distribution centers) and retailers. Data that is needed for this 'program' to function include for instance: aggregate product demands at each retailer, plant capacities, unit shipping costs between two locations, and the annual fixed cost of a DC at each potential location. To achieve more accurate forecasting, this problem can be formulated in a more complex way by including more variables. **Tactical:** minimize costs by using decision variables that include the amount to produce for each product category using regular time, overtime and subcontracting. The number of workers to be hired or to be laid off can also be included. Additional variables may include inventory cost, wages, hiring costs and layoff costs. **Operational:** using MRP (Materials Requirements Planning) to schedule manufacturing plans. An objective function can be used that minimizes the maximum completion time across all jobs. The basic idea of the function is: there are N jobs to be scheduled and M different resources, how can these effectively be scheduled when processing time, due date and the priority of the jobs are known. (Workforce/labor scheduling can also be seen as a similar problem) Some ERP systems have scheduling modules, like the Applied Planning and Optimization module in SAP.

Deliver (and Return)	**Strategic**: Fleet planning, which refers to the dynamic acquisition and divestiture of vehicles for delivery, to meet the delivery demand or the returns demand. Coca Cola Enterprises started replacing some of its diesel fleet vehicles to diesel electric hybrid vehicles. Decision variables are fuel costs, usage based deterioration (depreciation of assets) and seasonal demand. **Tactical:** using a multi-commodity network flow model to maximize efficiency and minimize costs. This represents a network of nodes (factories/retail locations/DCs). Each arc (connection between two nodes) represents a way of shipping (air/rail/carrier) and a given capacity (Truckload/Less-than-truckload). Constraints include capacity at a given arc, at a time period, or at a node. Data that is needed includes shipping costs per arc, supply forecasts for each node (provided by the S&OP), demand forecast at sink nodes (retail locations) and arc capacities. **Operational:** using a VRP (Vehicle Routing Problem) that optimizes the sequence of node visitation on a route. This can be done for a delivery truck, a return truck or both. This classic model determines in what sequence nodes should be visited, ending at the same starting point. Possible goals of using VRP are lowering transportation costs, create more effective planning for route optimization, and increasing customer satisfaction by optimizing visitation sequences.

Table 4 - Application of data analytics per supply chain process (Souza, 2014)

As shown in the table above, there are several ways to improve the main supply chain processes on strategic, tactical and operational levels. To see if there are more methods of improving these processes, other literature is reviewed below to see if it can confirm or deny these methods, or if additional techniques can be used that will help improve the processes and eventually the performance.

6.1 Plan

As showed in Table 2 and Table 4, predictive analysis/predictive modeling is the driving factor behind the planning process within a supply chain. Creating more efficient forecasting and planning processes will lead to better fulfilment of demands in the supply chain, which will increase the performance (even if it is marginally) of every organization involved in the process. The reason that more than one company in the supply chain can benefit from better planning- and forecasting processes is the fact that better planning and forecasting reduces gaps between supply and demand, which will eventually reduce the bullwhip effect: small fluctuations high up the supply chain can create significantly larger fluctuations further down the supply chain. Sharing customer demand information across the supply chain can also significantly reduce, but not completely eliminate the bullwhip effect (Ouyang, 2007).

Improving the planning process will not only have an impact on the accuracy of forecasting and will help reduce the bullwhip effect, but it also has a significant positive influence on the Source, Make and Deliver process (Zhou et al. 2011).

6.2 Source

An important aspect of sourcing is the ranking of suppliers in a supplier-selection situation. Souza (2014) proposed the AHP-method (Analytical Hierarchic Process) to select valuable suppliers. Another method of selecting suppliers is the DEA (Data Envelopment Analysis) method. This method focuses on the efficiency of a decision alternative (in this case a supplier). The alternatives (suppliers) are evaluated on input (cost) and output (benefit) criteria. How efficient an alternative is, is defined as the ratio of the weighed sum of its outputs (the performance of the supplier) to the weighed sum of its inputs (the cost of using a supplier). For each supplier, the DEA method finds the most favorable set of weights (the set of weights that maximizes the supplier's efficiency rating), without making its own or any other supplier's rating greater than 1 (Boer, Labro & Morlacchi, 2001).

Thus, by performing a DEA analysis, firms can get insight in the efficiency of possible suppliers. Talluri & DeCampos (2013) mention that a high level of efficiency is indicative of a firm that has certain practices and capabilities in place to be able to respond effectively to a buying company's current and possible future needs.

6.3 Make

Besides using Linear Program modules to optimize location, capacity, and for instance flow of materials, and using MRP systems to schedule manufacturing activities, there is an additional way of improving the efficiency and productivity of manufacturing. Li, Song and Huang (2015) have used a Scientific Workflow Management System (SWfMS) based on a cloud service platform for manufacturing big data analytics. (A SWfMS system is a system that defines, creates and manages the execution of workflows (Liu, Valduriez, Pacitti & Mattoso 2015)). They state that such a system has a significant advantage because it can provide easy access to complicated computing technologies without knowing the underlying knowledge. The system consists of an Infrastructure Layer, a Management Layer, a Service Layer and an Application Layer, in that order.

The Infrastructure Layer is the layer that facilitates the network, either through hardware and local resources, or through a cloud environment. The Management Layer is "the bridge of physical resources and workflow execution." In this layer, the scientific workflow scheduling module is the essential part that decides which scheduling algorithm has to be used and it provides the specific scheduling strategies. The Service Layer receives the requirements of the upper layer and the scheduling strategies from the lower layer to optimally facilitate the workflow system. The Application Layer is the layer that shows the workflow design and can create visualizations and presentations of the analyzed data. It can show simplified visualizations of complex experiments and data processing techniques that are used to analyze manufacturing big data. These visualizations are beneficial to possible scientific discoveries that can be made within the analyzed data.

Li et al. (2015) also state that when heterogeneous resources within systems are used (different types of data), an efficient workflow scheduling algorithm is important to perform analytics that will give meaning to the outcomes.

6.4 Deliver

Even though analytics can have an effect on the delivery process -as shown in Table 4-, Trkman et al. (2010) have conducted research to see which of the relationships between supply chain process analytics (plan/source/make/deliver) and supply chain performance have the biggest effect. The results concluded that the use of analytics in the Deliver area has the smallest effect, whereas the use of analytics in the Make area had the biggest effect on supply chain performance.

An explanation for this is also given by the researchers; they state that it might be possible due to the fact that the delivery process is often outsourced and that the decisions often take place at the end of the entire process, where their effect may be limited. This is a fair assumption to make, because even if the delivery process is optimized in the best way possible by using data, the other main processes (plan/source/make) also need to be streamlined and efficient to gain significant performance gains.

7 Conclusion & Implications

In Chapter 2, definitions of the variables of interest were given. In Chapter 3, the different types of data involved with the management of supply chains were discussed. In Chapter 4, data analysis techniques that can have an impact on supply chain performance are mentioned and broken down. In Chapter 5, the integration of data from different sources into a decision making process is discussed and in Chapter 6, applications and possible utilizations of data analysis techniques to improve the supply chain performance are mentioned and described for each separate supply chain process.

The purpose of this thesis, as mentioned in the introduction, was to gain more insight in how companies could effectively integrate and utilize data and data analysis techniques to improve supply chain performance. It has become clear that supply chain management teams face large volumes of different types of data all throughout the supply chain. To extract (business) value out of this data, data analysis techniques have to be used with a thorough and goal-focused mindset that requires analytical thinking within organizational units to effectively recognize and analyze possible problem areas to ultimately improve business processes. Hiring competent data scientists that have knowledge about quantitative analysis skills, as well as functional knowledge of the business domain, can simplify this process.

Besides finding solutions that have an internal impact for the company, firms that rely on effective supply chain management need to cooperate with upstream and downstream links within the supply chain process to realize the biggest gains in performance. Sharing customer demand data and forecast analyses can prepare the ‘next link’ in the supply chain for future demand fluctuations, which will lead to performance gains for individual companies in the supply chain, as well as performance gains for the supply chain as a whole.

Supply chain management is just one of the many business areas where data analytics can improve and possibly transform the business environment. The reason that this topic is especially relevant for the SCM area, is the fact that more than one company is involved in the entire process of conducting business, making it very attractive to look at possible new and innovative ways to significantly improve the performance. Another important aspect is the digitalization of society. Every day more and more data is being generated and is more easily becoming available for companies, but the biggest factor within this topic is to act upon this data. Adopting data analysis techniques into business processes together with competent human resources to analyze the growing amount of data will not only lead to better insight in the business, but will also help to make decisions based on a more solid understanding of the functioning of business processes, which can lead to an increase of the overall performance of the company.

8 Recommendations for future research

This thesis was written from an overview standpoint to try and point out the most significant contributions of data and data analytics to the improvement of supply chain performance. This method allowed the research to not get too specific for certain elements of the main supply chain processes. It also is one of the disadvantages of this thesis; data analysis and data integration solutions that are specifically designed for each of the supply chain processes are not discussed in detail because this would create a very in-depth, possibly very technical review of certain solutions, and would not capture the general application and value of these solutions. Future research could also include conducting empirical research in certain business environments to better map the requirements and difficulties of utilizing data analytics for a specific business domain.

References

Abiteboul, S. (1997). *Querying semi-structured data.* Springer.

Berry, D. (2012). *Unstructured Data: Challenge or Asset?* Retrieved from http://www.zdnet.com/article/unstructured-data-challenge-or-asset/

Boer, L., Labro, E., & Morlacchi, P. (2001). A review of methods supporting supplier selection. *European Journal of Purchasing & Supply Management, 7*(2), 75-89.

Cecere, L. (2012). *Big Data: Go Big or Go Home?* Supply Chain Insights LLC.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly, 36*(4), 1165-1188.

Data Mining for Supply Chain Management in Complex Networks. (n.d.). Retrieved from http://what-when-how.com/information-science-and-technology/data-mining-for-supply-chain-management-in-complex-networks/

Davenport, T. H., & Patil, D. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review, 90*(10), 70-76.

Dhar, V. (2013). Data Science and Prediction. *Communications of the ACM, 56*(12), 64-73.

Galbraith, J. R. (1974). Organization Design: An Information Processing View. *Interfaces, 4*, 28-36.

Gantz, J., & Reinsel, D. (2011). Extracting Value from Chaos. *IDC iView*, 1-12.

Gartner. (2013). *IT Glossary*. Retrieved from http://www.gartner.com/it-glossary

Gunasekaran, A., Patel, C., & McGaughey, R. E. (2004). A framework for supply chain performance measurement. *International Journal of Production Economics, 87*(3), 333-347.

IBM. (n.d.). *What is Big Data Analytics*. Retrieved from http://www-01.ibm.com/software/data/infosphere/hadoop/what-is-big-data-analytics.html

Kościelniak, H., & Puto, A. (2015). BIG DATA in decision making processes of enterprises. *Procedia Computer Science, 65*, 1052-1058.

Kowalczyk, M., & Buxmann, P. (2014). Big Data and Information Processing in Organizational Decision Processes. *Business & Information Systems Engineering: The International Journal of WIRTSCHAFTSINFORMATIK, 6*(5), 267-278.

Kumar, S., & Schmitz, S. (2011). Managing recalls in a consumer product supply chain - root cause analysis and measures to mitigate risks. *International Journal of Production Research, 49*(1), 235-253.

Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity and Variety.* Application Delivery Strategies. META Group.

Lavalle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review, 52*(2), 21-31.

Li, X., Song, J., & Huang, B. (2015). A scientific workflow management system architecture and its scheduling based on cloud service platform for manufacturing big data analytics. *The International Journal of Advanced Manufacturing Technology*(8).

Liu, J., Valduriez, P., Pacitti, E., & Mattoso, M. (2015). A Survey of Data-Intensive Scientific Workflow Management. *Journal of Grid Computing*.

Losee, R. M. (2006). Browsing mixed structured and unstructured data. *Information Processing and Management, 42*(2), 440-452.

Lustig, I., Dietrich, B., Johnson, C., & Dziekan, C. (2010). *The Analytics Journey.* Institute for Operations Research and the Management Sciences.

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*, 60-68.

Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining Supply Chain Management. *Journal of Business Logistics, 22*(2), 1-25.

Michael, K., & McCathie, L. (2005). The pros and cons of RFID in supply chain management. *Proceedings of the International Conference on Mobile Business*, (pp. 623-629).

Ouyang, Y. (2007). The effect of information sharing on supply chain stability and the bullwhip effect. *European Journal of Operational Research, 182*(3), 1107-1121.

Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data, 1*(1), 51-59.

Rozados, I. V., & Tjahjono, B. (2014). Big Data Analytics in Supply Chain Management: Trends and Related Research. *6th International Conference on Operations and Supply Chain Management.* Bali.

Russom, P. (2011). *The Three V's of Big Data Analytics*. Retrieved from http://TDWI.org.

Sanders, N. R. (2014). *Big Data Driven Supply Chain Management: A Framework for Implementing Analytics and Turning Information into Intelligence* (1st ed.). Pearson, NJ.

SCOR framework. (n.d.). Retrieved from http://www.apics.org/sites/apics-supply-chain-council/frameworks/scor

Souza, G. C. (2014). Supply Chain Analytics. *Business Horizons, 57*, 595-605.

Take Solutions, Inc. (2013). *4 Steps to Leveraging Supply Chain Data Integration for Actionable Business Intelligence.*

Talluri, S., DeCampos, H. A., & Hult, G. T. (2013). Supplier Rationalization: A Sourcing Decision Model. *Decision Sciences, 44*(1), 57-86.

Trkman, P., McCormack, K., Oliviera, M. P., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems, 49*(3), 318-327.

Waller, M. A., & Fawcett, S. E. (2014). Supply Chain Game Changers—Mega, Nano, and Virtual Trends—And Forces That Impede Supply Chain Design (i.e., Building a Winning Team). *Journal of Business Logistics, 35*(3), 157-164.

Wu, D., & Olson, D. L. (2008). Supply chain risk, simulation and vendor selection. *International Journal of Production Economics, 114*(2), 646-655.

Zhou, H., Benton Jr., W. C., Schilling, D. A., & Milligan, G. W. (2011). Supply Chain Integration and the SCOR Model. *Journal of Business Logistics, 32*(4), 332-344.