Master's Thesis

A Reference Model for Big Data Analysis in Shipbuilding Industry

Yonghyeok Lee

Department of Management Engineering

Graduate School of UNIST

2017

A Reference Model for Big Data Analysis in Shipbuilding Industry

Yonghyeok Lee

Department of Management Engineering

Graduate School of UNIST

A Reference Model for Big Data Analysis in Shipbuilding Industry

A thesis/dissertation
submitted to the Graduate School of UNIST
in partial fulfillment of the
requirements for the degree of
Master of Science

Yonghyeok Lee

01. 06. 2017

Approved by

Advisor
Changyong Lee

ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

A Reference Model for Big Data Analysis in Shipbuilding Industry

Yonghyeok Lee

This certifies that the thesis/dissertation of Yonghyeok Lee is approved.

01.06.2017

Signature

Advisor: Changyong Lee

Signature

Minseok Song

Signature

Marco Comuzzi

ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

ABSTRACT

Global shipbuilding industry has gone through a tough time due to the reduction of shipping order quantities and shipbuilding tonnages since the global financial crisis of 2008. To overcome the challenges, big data analysis is expected to be an effective solution to increase the practical efficiency in the shipbuilding industry. After an organization applies big data analysis, benefits, such as better aimed marketing, more straightforward straight-forward business insights, client based segmentation, and recognition of sales and market chances, are anticipated. In the future, the key for competitiveness is finding an appropriate way of applying big data analysis. Although numerous studies for big data analysis are conducted, the studies tend to focus on the technical aspect of analyzing data including method, algorithm, and architecture. Therefore, it is required to study how to applying the analysis technique in the practice, specifically shipbuilding industry in this study. In this thesis, the reference model for big data analysis in shipbuilding industry is developed. The proposed reference model provides the big data analysis guideline according to four phases such as contract, design, production, and service. They are categorized based on value chain of shipbuilding industry. Each phase consists of three levels of big data analysis, e.g., category of analysis, analysis method, and detailed algorithm. Moreover, the importance of the analysis method is determined in order to increase the applicability of the reference model. To verify the validation of the model, experts of the shipbuilding industry consulted the model it is consulted by the experts of the shipbuilding industry.

Keywords: Reference Model, Big Data Analysis, Shipbuilding Industry

Contents

List of Figures

List of Tables

I. Introduction

1.1. Research Background

The global shipbuilding industry has gone through a tough time due to the reduction of ship order quantities and shipbuilding tonnages since the global financial crisis of 2008. The current economical context is pushing companies to produce the product with less cost and better quality (Aramja, Kamach, & Chafik, 2015). In addition, the companies are required to do it faster and most of all in a cost effective manner. Furthermore, three major Korean shipbuilders recorded the substantial amount of loss, over seven trillion won for the third quarter in 2015, even though Korea is one of the top countries in the shipbuilding industry. Due to globalization, new competitors bring new tools and approaches to the market (Younus, Hu, Yong, & Yuqing, 2009). Although China's shipbuilding industry has emerged lately as the new competitor in the market, it has experienced rapid growth with effective cost advantage. Most orders in the market were placed for bulk carriers and tankers sectors in which China has competitiveness on a global scale (Jiang, Bastiansen, & Strandenes, 2013). Now facing increasingly fierce competition and hostile environments, it is crucial for shipbuilders to find the next-generation growth of opportunities and make innovations in internal and external environments as one of the solutions, the big data analysis is promising.

The big data analysis is a data analysis based on big data. The data analysis is responsible for finding the hidden patterns, rules, and information from the data. Most researchers in this field use the term to describe how they refine the "ground" (i.e., raw data) into the "gold nugget" (i.e., information or knowledge) (Tsai, Lai, Chao, & Vasilakos, 2015). Big data refers to the explosion of available information. It is driven by the massive amounts of very high-dimensional or unstructured data, which are continuously being produced and stored with much cheaper cost than they used to be (Fan, Han, & Liu, 2014). These data are generated from online transactions, emails, videos, audios, images, click streams, logs, posts, search queries, health records, social networking interactions, science data, sensors, and mobile phones and their applications (Sagiroglu & Sinanc, 2013). The analysis of big data mainly involves analytical methods for traditional data and big data, analytical architecture for big data, and software used for mining and analysis of big data (M. Chen, Mao, Zhang, & Leung, 2014). After an organization applies some form of big data analysis, benefits, such as better aim marketing, more straightforward business insights, client based segmentation, and recognition of sales and market chances, can be anticipated. Specifically, the potential of big data analysis in the manufacturing industry is evaluated to make a positive impact, such as improved demand forecasting, supply chain planning, sales support, developed production operations, and web search based applications (M. Chen et al., 2014; Manyika et al., 2011).

The existing studies of the big data analysis focused on the way of analyzing big data such as method, algorithm, and architecture. M. Chen et al. (2014) introduced the methods, architectures and tools for the big data analysis. In addition, Gandomi and Haider (2015) studied the analytic methods used for big data. Especially, the analytics related to unstructured data are described into various areas of text analytics, audio analytics, video analytics, and social media analytics. Zhang, Liu, Wang, and Gruen (2007) described the steps for planning the statistical analysis involving genome-wide data. H. Chen, Chiang, and Storey (2012) illustrated the application, data, analytics, and impacts of data analysis in domains such as science and technology, smart health and wellbeing, security and public safety, and others. The types of technology for analysis are introduced like RDBMS, data warehousing, data mining, and others. Moreover, the processing of big data for analysis was studied. Sandryhaila and Moura (2014) proposed the framework to cope with large-scale data for the analysis based on the discrete signal processing (DSP). Even though the techniques for big data analysis are important, there are only a few studies providing practical guidelines on how to apply analysis techniques in specific industries. The target areas for the analysis, which were studied in the literature, are broad like the whole industry and the effect of the analysis was conceptually described. Although the organization wants to analyze the data with the purpose of extracting useful values and providing suggestions or decisions, there is a problem on analyzing the data because they do not understand what value can be drawn by analyzing the data and how to apply it. The various potential values of different levels can be generated through the analysis of datasets depending on the applied areas in the industry. Therefore, it is required to provide a practical guideline for big data analysis specific to each industry, in this case, the shipbuilding industry.

This thesis proposes a reference model for big data analysis in the shipbuilding industry. The reference model is developed based on the value chain of shipbuilding industry. The value chain offers the overview of the area that big data analysis is applied to create the value. It is necessary to develop the reference model that considers the characteristics of the industry to offer the effective guidance. It provides the big data analysis guideline according to four phases, such as contract, design, production, and service. They are categorized according to the value chain of the shipbuilding industry. Each phase consists of three levels of data analysis, i.e., category of analysis, analysis method, and detailed algorithms. The level of guidance needs to be concrete and detailed for industry personnel to properly understand and utilize the big data analysis results. At the level of analysis method, the concrete guidance will be provided, such as data, detailed algorithms, and analysis result.

This paper is organized as follows. Section 2 starts with discussing related works. Then, Section 3 describes the value chain of the shipbuilding industry. Section 4 introduces the reference model developed in this study. The analysis methods by each phase will be explained. Section 5 presents the result of the interview with domain experts, which is used to validate the reference model. Section 6 has a discussion and Section 7 concludes the thesis.

1.2. Research Method

The reference model development discussed in this article was realized within a four-phase research method in Figure 1 derived from the method for reference model development (Ahlemann, 2009; Schütte, 1998). The research method is comprised of:

(1) *Problem Definition*. The research objective was defined and the problem domain specified as documented in the first section of this paper.

(2) *Development*. The second phase consists of two different activities:

(2a) *Literature Review*. Research conducted by other authors and organizations involving the shipbuilding industry, for example, data analysis in shipbuilding industry and the methods of applying big data analysis was also taken into consideration.

(2b) *Development of A Reference Model*. The initial development of a reference model was based on the knowledge obtained from the literature review.

(3) *Evaluation*. The objective of this phase was to validate, refine, and stabilize the initial reference model development.

(3a) *Interviews with Domain Experts*. The interview was conducted with experts in the shipbuilding industry with the objective of gathering further empirical evidence for the reference model in order to validate it. Possible improvements were discussed during the interview. The reference model would then be refined or redesigned if the interview results indicated that this was necessary. It was then concluded that the domain experts had reached a consensus on the reference model’s propositions.

(3b) *Refinement of a Reference Model*. The experience gained from the interview was also used to refine the reference model.

(4) *Documentation*. The documentation of the reference model contains a description of the research method, the finalized reference model, and the interview results for empirical evidence.

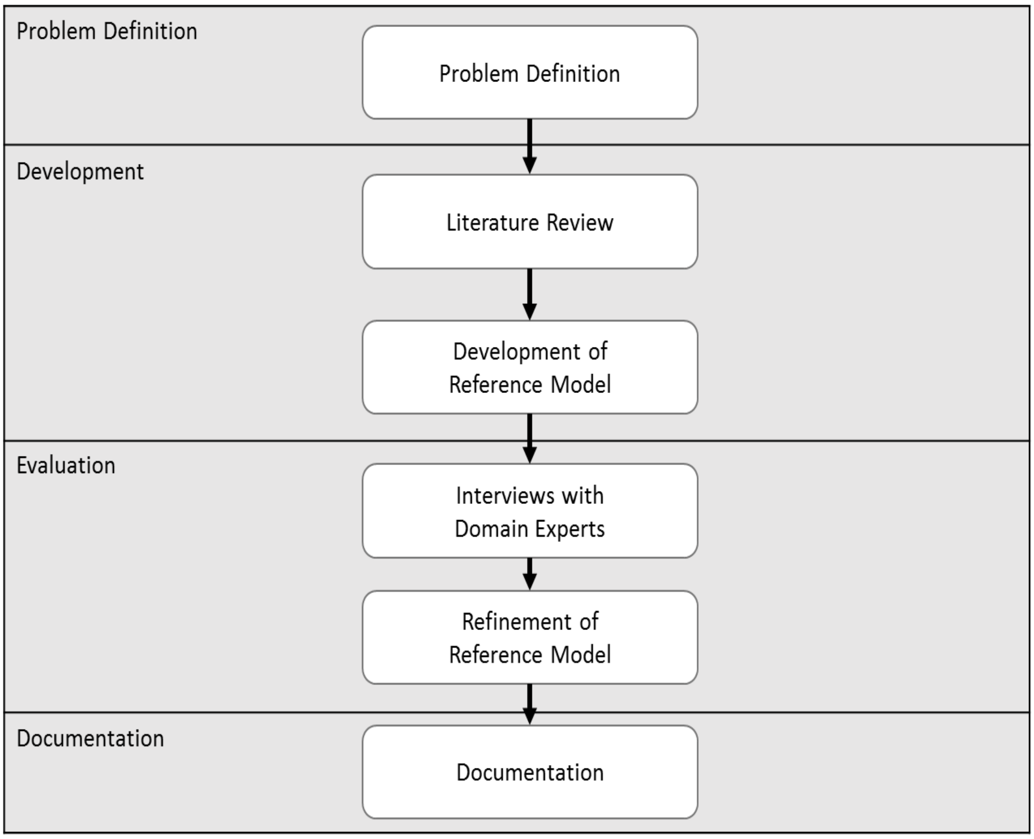

Figure 1. Research Method

II. Related Works

2.1. Big Data Analysis in Manufacturing Industries

Nowadays data is collected from all aspects in business, but the large volume of data would be of little use if it has not been effectively analyzed so that insightful information can be extracted and applied in the decision-making and business processes. Davenport (2006) argued that it is vital for business to compete on analytics, which means "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions". In addition, big data analysis can be regarded as a sub-process in the overall process of insight extraction from big data and used as competitive differentiator to the company (Gandomi & Haider, 2015; LaValle, Hopkins, Lesser, Shockley, & Kruschwitz, 2010). There are four main goals of big data analysis: (1) to develop effective methods that can accurately predict the future observations; (2) to gain insight in to the relationship between the features and response for scientific purposes; (3) to explore the hidden structures of each subpopulation of the data, which is traditionally not feasible and might even be treated as ‘outliers’ when the sample size is small; (4) to extract important common features across many subpopulations even when there are large individual variations (Fan et al., 2014; Fan & Lv, 2008).

Big data refers to the data that is unable to be handled and processed by most current information systems or methods due to the explosion of available information (Fisher, DeLine, Czerwinski, & Drucker, 2012). In addition, Laney (2001) presented a well-known definition (also called 3Vs) to explain what is the “big” data: volume, velocity, and variety. These data are generated from online transactions, emails, videos, audios, images, click streams, logs, posts, search queries, health records, social networking interactions, science data, sensors and mobile phones and their applications (Sagiroglu & Sinanc, 2013).

Manufacturers have tremendous potential to generate value from the use of large datasets, integrating data across the extended enterprise and applying advanced analytical techniques to raise their productivity both by increasing efficiency and improving the quality of their products (Manyika et al., 2011). Big data analysis is applied in the manufacturing industry based on the value chain. The value chain provides an overview of the business area for the company or the industry. Manyika et al. (2011) describes the expected effect of big data with the value chain of manufacturing industry. The value chain is composed of research and development, supply chain management, production, marketing and sales, and after sales service. Sowar and Gromley (2011) proposed the analysis in the steel industry. They mentioned that the value chain is the important in that it is the starting point for finding the advantage of the initial data analysis. Additionally, Patil and Giffi (2015) suggested the applicable analysis based on the value chain in the automotive industry. The analytics were described on each value chain and the cross value chain. However, the level of description for the analysis is not in detail and only enumerates the possible analytics. Therefore, the value chain based approach is used in the thesis to develop the reference model for the data analysis. The analysis will be explained specifically.

2.2. Big Data Analysis in Shipbuilding Industry

Big data links ships, ports, ship inspection institutions, repair dockyards, and equipment suppliers, so it is expected to strengthen the competitiveness in the shipbuilding industry and coexistence between small, medium, and large-sized companies. Recently, with the intention of remote ship monitoring for better services for shipping customers, shipbuilders started to adopt new sensor technology by installing different sensors for different components on board and transmit data using satellite communications to land-based service centers, e.g., Health Monitoring System (HEMOS) by Rolls Royce Marine (Hao et al., 2015). After a while, builder companies realized that the collected sensor data could also help improve ships' maintenance and future design. In addition, from the aspect of regulators, big data analysis on sensor enabled operation data can improve energy efficiency and environmental performance, safety verification and assessment, the monitoring of accidents and environment risks, and help regulators introduce more quantified regulations for the administration of ships and seas.

As importance of the big data analysis increase in the shipbuilding industry is, a few studies for using big data analysis in shipbuilding industry were conducted. In the maritime domain, H. Hwang, Kim, Shin, Song, and Nam (2016) defined the big data as the meaningful information generated by the navigation and communication equipment from the many ships on the ocean. Plus, they developed a vessel traffic display and statistic system based on AIS messages from the many vessels of maritime. Seunghee Oh and Lee (2015) proposed accident prediction mechanism using maritime big data, which comes from vessel traffic control system. However, these studies limited the sources of data and certain problems. There are more data that can be utilized. Big data analysis is necessary in shipbuilding industry in that it has a lot of opportunities. Furthermore, there is no practical guideline on how to apply data analysis technique in shipbuilding industry. The reference model should include overall big data analyses and provide the guidance for the analysis in shipbuilding industry.

2.3. Reference Model

There is no mutual understanding of the term “reference model” (Fettke & Loos, 2004). Generally, one can distinguish between approaches that regard models as direct representations of reality and approaches that follow a constructive paradigm. The latter regard a model as a development of one or various modelers. In accordance with this, a reference model is defined as an “information model used for supporting the development of other models” (Thomas, 2006). The use of reference models is frequently based on the expectation that they can

- accelerate the development of information systems (a time aspect),
- reduce the associated costs (a cost aspect),
- help to communicate innovative ideas and best practices (a quality aspect), and
- reduce the risk of failure (a risk aspect) (Ahlemann, 2009).

Although widely accepted in business informatics, the term reference model is not always applied. The terms “standard model,” “framework” or “architecture” are frequently used as synonyms. However, the term, reference model, is used to share the information, not for the development of other models in this thesis. The reference model in the study is expected to accelerate the application of big data analysis, to reduce the associated costs, to help to communicate the ideas for applying the data analysis, and reduce the trial and error for the data analysis.

III. Value Chain of Shipbuilding Industry

The shipbuilding value chain in Figure 2 is comprised of four major phases, i.e., contract, design, production, and service. The value chain is reorganized from the shipbuilding process of Korea Shipbuilding Industry Cooperative (KBIZ, 2012). The contract phase includes the phases of sales activities and marketing. The design phase includes the phase of design and project management. The production phase includes the activities required to build the ship such as assembly, marking, cutting, and others. Finally, the service phase includes maintenance, repair, technical training, and customer support. Comparing to other phases related to traditional values of shipbuilding, the service phase is expected as a new business model using huge amount of data accumulated for several decades.

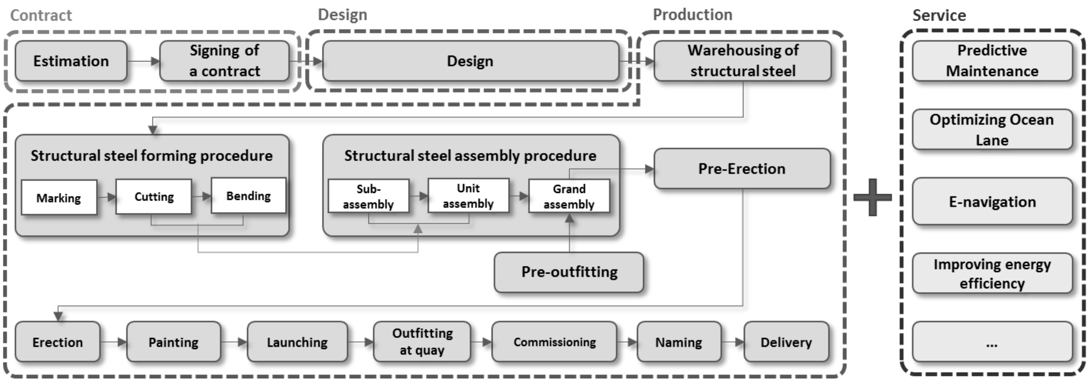

Figure 2. Vale Chain of Shipbuilding Industry

Each component of the value chain is described as following:

Estimation. In order to obtain an order, shipbuilding company proposes the cost for the contract. Specification and cost of a ship are determined generally depending on requirements from ship-owners or clients.

Signing of contract. After discussions about specifications and cost of the ship, ship-owners and the shipbuilding company makes contracts for shipbuilding.

Design. Design process consists of four stages: initial design, basic design, detail design and production design. Initial design is a process for setting a concept of the ship and specification of vital materials. Basic design is a process for determining detailed specifications of the ship and making drawings. Detail design is a process for verification of results from basic design and making drawings for fabrication stage. Production design is a process for making drawings of each part of the ship and setting process of forming and assembling.

Warehousing of structural steel. After purchasing structural steels from steel mill companies, the shipbuilding company stacks structural steels in a stockyard. The company cares about the method to take out necessary steels on each day and try to prevent the damage of steel during the storage.

Pre-treatment. Before steel cutting process, several preliminary processes are conducted to structural steels. For example, 'shot blasting' is a process to eliminate mill scales and debris on surfaces of structural steels by hitting surfaces with small iron beads or sand. 'Primer painting' is a process for painting steels before processes of forming and assembling to prevent rusting.

Structural steel forming process. For making components of blocks before assembly process, several work processes are done to structural steels. Although the forming process is composed of marking, cutting, and bending, each process has distinguishing properties compared to a structural steel assembly process.

Marking. Signs for work on the next stages are marked on steels. These signs are for gas cutting, bending forming and assembling.

Cutting. For making parts of components, structural steels are cut as drawings of blocks. As methods of cutting, there are gas cutting, plasma cutting, laser cutting and edge milling.

Bending. For making curved plates of blocks, structural steels are bent. As methods of bending, there are two categories: cold working and hot working. Cold working processes (Bending roller, Die-less forming, Universal Press) are for making steels as developable surfaces. Hot working processes (Line heating) are for making steels as undevelopable surfaces.

Structural steel assembly process. After making parts of components, parts are assembled as small components and blocks through assembly processes. Structural steel assembly process is composed of sub-assembly, unit-assembly, and grand-assembly. In a stage of *Sub-assembly*, small components composing of blocks are assembled. Due to these components are small and have light-weights, sub-assembly processes can be automated with a conveyor system. In the stage of *Unit assembly* and *Grand assembly*, small components and parts become blocks. Difference between unit assembly and grand assembly is the size of assembled blocks. Blocks assembled at grand assembly process are bigger than those from unit assembly process. Some blocks assembled at unit assembly process make up the blocks at grand assembly process, and others become the blocks of the ship directly.

Pre-erection. Before erection (assembly of blocks) in the dock, some blocks are assembled in pre-erection yard (between grand assembly shops and the dock).

Pre-outfitting. Before doing outfitting processes at the quay, some outfitting processes are done during pre-erection processes. Pre-outfitting processes reduce man-hours for outfitting processes and prevent reworks of following processes, especially painting processes.

Erection. Blocks from unit assembly, grand assembly and after pre-erection are assembled in the dock.

Painting. For preventing rusting, damages from waves and wind and sticking of seaweeds, painting processes are done to the ship. Painting processes are done during forming and assembly processes as well as erection processes.

Launching. If the ship (or the part of ship) is estimated to have enough buoyancy, the ship is transferred to the water. There are two kinds of launching processes: partial ship launching and full ship launching.

Outfitting at quay. Outfitting and finishing works like correction of deformations are done at quay after launching.

Commissioning. After finishing all production processes and before shipbuilding company delivers the ship to the owner, the ship is sailed on the sea and gets tested whether the ship satisfies the specifications like strength, velocities at specific drafts and stability as designed.

Naming. After finishing all production processes, the ceremony for naming the ship is hosted.

Delivery. Completed ship is sailed off to the owner.

IV. Development of A Reference Model for Big Data Analysis in Shipbuilding Industry

4.1. Overview

A reference model for big data analysis has been developed based on the literature survey and refined considering the feedback from the interview. The model is expected to work as the starting point for applying big data analysis to shipbuilding industry. It provides the applicable data analysis to shipbuilding industry according to the four phases of value chain in Figure 3, e.g., contract, design, production, and service. Each phase has categories of analysis related to tasks on each value chain. However, the categories of analysis on the production phase of the reference model are not accordance with the activities on the production phase of the value chain such as painting, erection, and others. The categories are more likely to depend on the division of department in the general manufacturing industry. It is because understanding of the applicable area for the data analysis is increased in terms of the practitioner. In the literature, the data analysis is utilized according to such categorization. The analysis composition for each phase will be explained in the following section.

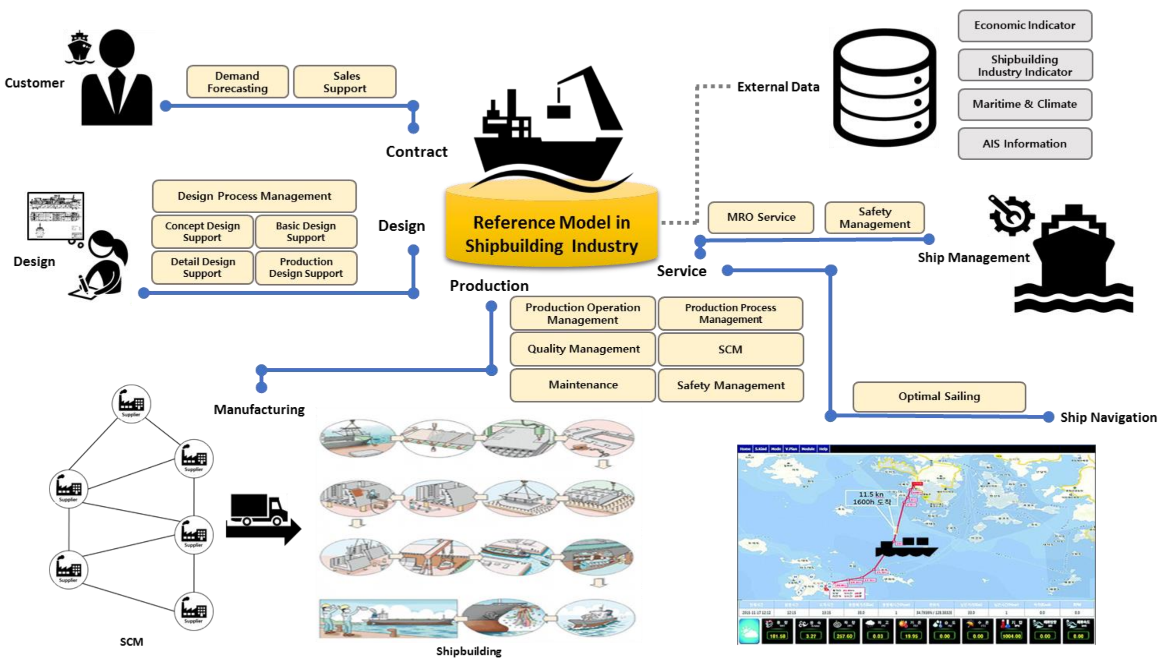

Figure 3. Overview of A Reference Model

The reference model is composed of four layers in Figure 4, which are phase, category of analysis, analysis method, and detailed algorithm. There is a study for the big data analysis (H. Chen et al., 2012). This study provides application, data, analytics, and impact according to each area. Such an arrangement for the big data analysis was used in the thesis. In the reference model, data, detailed algorithms, results are provided at the level of analysis method. The category of analysis is equal to the level of applications in the literature. Also, the phase is compared to each area in the previous study. In addition, the reference model was tried to be specified as possible in that a specific industry was studied in our study. The phase is categorized according to the value chain of the shipbuilding industry. Each phase is composed of three levels of data analysis, i.e., category of analysis, analysis method, and detailed algorithm. Category of analysis is involved in main tasks or the issues in each phase. Considering the category, the analysis method supports to achieve the purpose of it. The analysis method is the major layer to provide the practical guideline for data analysis. It is described with data, detailed algorithm, analysis result, and related technology. The detailed algorithm is the applied techniques to analyze data. Other alternative algorithm having the equal purpose of analysis can replace the algorithm.

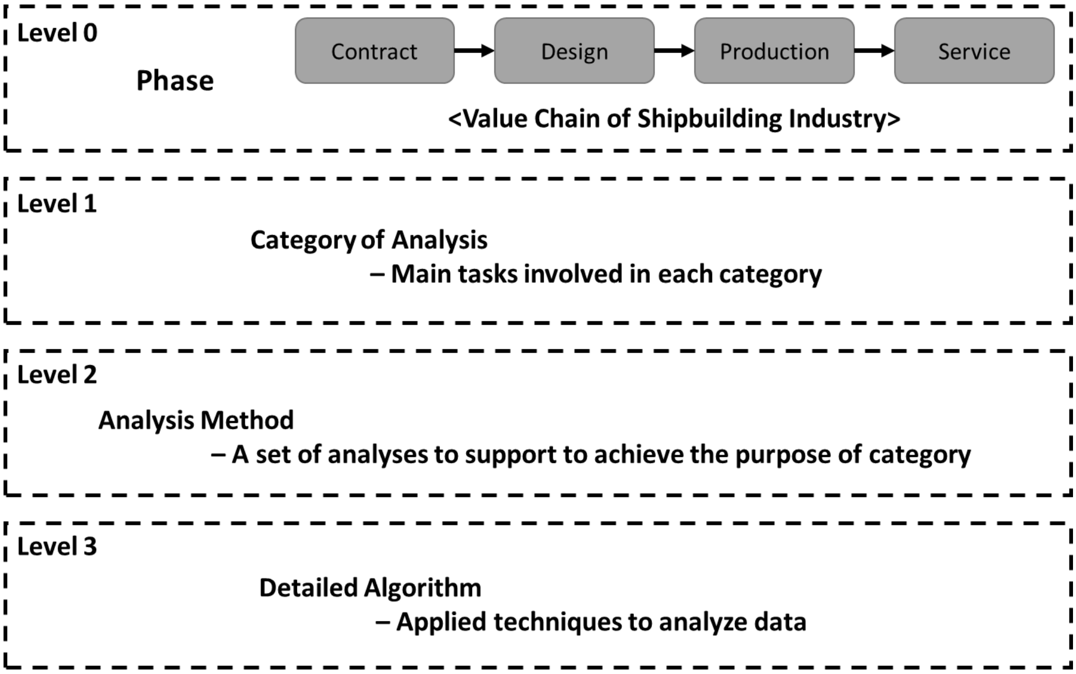

Figure 4. Layers of A Reference Model

4.2. Data Analysis in Contract Phase

4.2.1. Contract Phase

The shipbuilding company highly depends on the demand of ship owners as shipbuilding is an order-made production industry. Moreover, the shipbuilding company is vulnerable to the market estimation for obtaining the ship order. As the competition in the global market increases due to emerging countries like China and the recession of shipbuilding industry holds out, major shipbuilders in the global market try to overcome these situations such as holding a conference to discuss about the prospect of new order market and countermeasure for the crisis. Big data is expected to be used in the contract phase of shipbuilding industry. It would help to forecast the demand in the market and support sales activity. Daewoo Shipbuilding & Marine Engineering (DSME) carried forward the development of forecasting model for a new demand of a ship in 2015. The model forecasts the new demand of a ship is based on the ship's owner data, economic indicators, and ship sailing data. It complements the character of the shipbuilding industry in which it is hard to create new demand. Shipbuilding is a traditional order-made production industry that it is difficult for manufacturers to create new ship demands, unlike other manufacturing industries. DSME has tried to analyze data on cargo volumes, macro-economic indicators, and shipping-related indicators, and to preemptively discover new ship types and technology that will lead the market. Considering a sharp reduction in the ship order, demand forecasting of ship order and supporting sales activity are important to increase the amount of orders received. Therefore, the analyses to support these activities are described in the following section.

4.2.2. Data Analysis Description

The analysis composition in the contract phase is in Figure 5. There are two major analyses on contract in the shipbuilding industry: demand forecasting and sales support. Demand forecasting is analyzed in detail: technology trend and promising technology analysis, and potential customer estimation. Sales support is further analyzed in three methods: estimation of initial production cost, estimation of initial production period, and estimation of earnings and success rate of a ship order. The analysis method descriptions are summarized in Table 1. Each analysis method is described as the following.

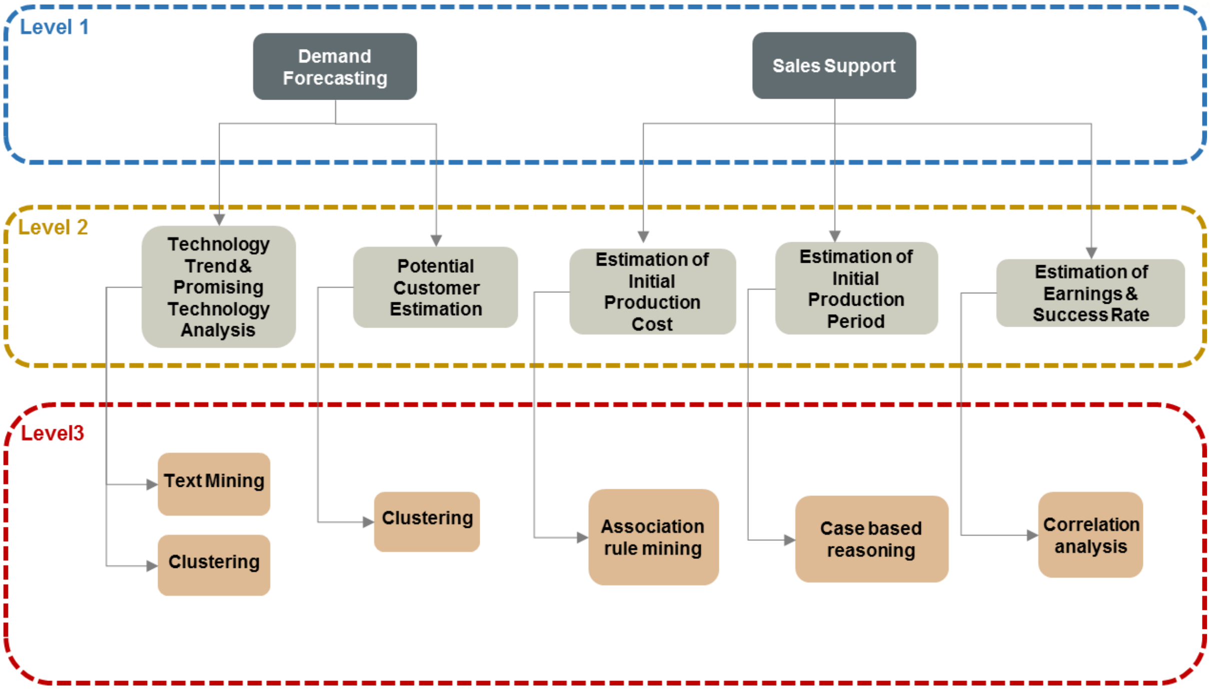

Figure 5. Analysis Composition in Contract Phase

· *Technology Trend & Promising Technology Analysis*

As the ship and related technology develops, technology which is applicable to construct a new ship becomes various. The company is able to utilize technology in shipbuilding and to predict prospective technology in the future by analyzing technology trend and promising technology. The ship can be guessed based on the change of a technology trend. For example, if technology related to energy saving for the ship stands out, the company may expect that the demand of green ship will be increased. Unstructured data including the contents of ship technology such as ship technology related patent and paper is used for this analysis. In terms of detailed algorithm, text mining and clustering are used for the analysis. The trend of ship technology is verified from the change of primary key words by time as the result of applying text mining to the unstructured data related to ship technology. The importance of the keyword is evaluated based on the term frequency. Moreover, the result is visualized as word cloud. With text mining results, clustering is used to make a group composed of key words including affinity. Clustering results are visualized as cluster dendrogram. The cluster is adjusted by the distance between clusters. Finally, technology trend related to the ship and promising technology is derived as the result of the analysis method (P. Kim, Hong, & Koh, 2014; S. Kim, Nam, & Sun, 2016; Min, Kim, & Ji, 2014; W. Park & Hwang, 2016).

· *Potential Customer Estimation*

Company in order-made production industry produces the product which is customized by customer requirements. It needs to manage customers to increase customer satisfaction and to obtain new order in the future. Potential customer estimation supports the company to prepare for demand by finding out the characteristics of the customers expected to make a new order. Classification information of customer groups is helpful to establish a marketing strategy considering customer property. Effective customer management will be conducted based on it. Customer (i.e., shipping company) information is used for this analysis. In terms of detailed algorithm, clustering is applied to the data. Applying clustering algorithm to the information of the current customer, including size of fleet, region, status of ship order, and others, customers are clustered depending on various standard in the information (S. Kim et al., 2016). Each cluster has the property for making a ship order. For example, customers are clustered by the size of the ship. Moreover, customers are grouped by the probability of marking a ship order in each cluster. In the end, clusters of the potential customer by standard are generated as the result of the analysis method.

· *Estimation of Initial Production Cost*

It is crucial to rapidly deal with new order because the order quantity is not stable in order-made production industry. At the phase of contract, initial suggestion of the necessary cost involved in ship production that is suitable with the requirement of ship owner is significant. The expert has conducted existing estimation of the production cost. After the estimation, including the requirement of ship order, is requested, the expert examines the information and estimates the cost based on similar records with the information. The estimation result may be distorted by a subjective viewpoint on the expert and take a long time because it is conducted by each person. It is available to estimate the initial production cost more quickly and accurately through the estimation of the initial production cost based on the data. The requirement of the order and the historical record data including specification of ship, type and shape of ship, and component information, are used for this analysis. In terms of detailed algorithm, association rule mining is applied to the data (Y. Kim, Park, & Oh, 2005). This technique derives the associative relation among the primary specification of ship and the parts information corresponding to the specification. Given the required specification of the order, the most similar case is found. Finally, the parts used in the case are found out as the result of the analysis method. In addition, the initial production cost is estimated by using the unit cost of the parts to support sales activity for obtaining a ship order.

- *Estimation of Initial Production Period*

Contract in the shipbuilding industry is concluded with the negotiation depending on the specification of the ship. As the existing calculation for the production period of the ship relies on the expert's experience, the estimated period is not accurate. The accuracy of the estimated production period could be increased by using the historical record data. The data includes required specifications of ship order, size, and type, duration of process activity, and others. A more accurate estimation of the production period is available through estimation of initial production period. Moreover, it prevents excessive reduction of the production period to get the order by calculating reasonable production period using the data. In terms of the detailed algorithm, case based reasoning is applied to the data for the analysis (K. Oh & Park, 2005). Case based reasoning predicts the result of new case based on the results of the historical cases (Watson, 2001). It measures the similarity between the new case and historical cases. Then, it selects the case which is the most similar with the new case. The cases having similar specification to the requirement is searched on the historical data. Therefore, the process activity and the duration of it is verified on a similar shipbuilding case as the result of the analysis method. The overall production period is estimated using the result. It will be used as the estimated production period in the sales of contract.

- *Estimation of Earnings & Success Rate of Ship Order*

When the shipbuilding contract is awarded on the way of competitive bids in shipbuilding industry, shipbuilders often tend to put up with the unfavorable bid conditions to obtain the order. Accordingly, it needs to consider the earning rate of getting the order as well as the success of that. Although the shipbuilders currently try to propose the favorable bid conditions based on the similar precedent, it takes a long time and the result may not be precise because the process of finding similar precedent is conducted manually. By analyzing the historical bid information such as region, order type, shipping company, earning rate and others, the reasonable bid condition will be derived. In terms of detailed algorithms, correlation analysis is applied to the data (Sohn, 2011). Correlation between the bid condition and success & earning rate is derived by the analysis using the data. The result will be used to make the bid condition which maximize the profit for winning the order.

Table 1. Description of Analysis Method for Contract Phase

Demand Forecasting				
Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Technology Trend & Promising Technology Analysis	· Ship technology related patent · Ship technology related paper · Unstructured data including contents of ship technology	· Text mining · Clustering	· Technology trend · Promising technology	(Cho & Kim, 2011; P. Kim et al., 2014; S. Kim et al., 2016; Min et al., 2014; W. Park & Hwang, 2016; Song, Park, Jung, & Song, 2013)
Potential Customer Estimation	· Customer (Shipping company) information (e.g., size of fleet, region, status of ship order)	· Clustering	· Potential customer by type	(S. Kim et al., 2016)
Sales Support				
Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Estimation of Initial Production Cost	· Historical performance data (e.g., specification of ship, component information) · Required specification of ship order	· Association rule mining	· Information of component used in ship having similar specification to order	(Y. Kim et al., 2005)
Estimation of Initial Production Period	· Historical performance data (e.g., size and type of ship, duration of process activity)	· Case based reasoning	· Duration of process activity in similar ship	(K. Oh & Park, 2005)
Estimation of Earnings & Success Rate of Ship Order	· Historical bid information (e.g., region, order type, shipping company, earning rate)	· Correlation analysis	· Relationship between bid condition and success rate · Relationship between bid condition and earning rate	(Sohn, 2011)

4.3. Data Analysis in Design Phase

4.3.1. Design Phase

The design in shipbuilding industry consists of four stages: initial design, basic design, detail design and production design (J. Oh, Park, Kim, & Jung, 2014). At the stage of the initial design, a concept of ship satisfying the requirement of the ship owner and the primary specification is determined. The basic design describes the process of determining the detailed specifications of the ship such as arrangement of equipment including navigation devices, housing facilities, electrical devices, and others. Moreover, the overall construction drawings are made. The drawings made at the stage of basic design is specified in detail to be understandable to the worker on the production field at the stage of detailed design. In addition, the results of basic design are verified on the detailed design. Lastly, a shop drawing of each part of the ship is made and procedures of forming and assembling are planned at the stage of production design. More than 12 months are averagely sent on the phase of design in shipbuilding process. Moreover, there are unnecessary cycle of tasks because the frequent change of design is general in shipbuilding industry and it causes to carry out the task related to the change. These rework and irrelevant cycle of the tasks need to be handled in terms of the whole process of the design. The big data analysis is able to support design phase by solving some issues on each stage of design. As the result of the analysis, it is expected to reduce the additional time for working and increase the efficiency. Considering these issues on the design, the analyses are proposed in the next section.

4.3.2. Analysis Description

The analysis composition in design phase is in Figure 6. There are five major analyses on design in the shipbuilding industry. They are initial design support, basic design support, detail design support, production design support, and design process management. These are developed considering the characteristic of design in the shipbuilding industry. Firstly, initial design support has an analysis method, inference of similar case of ship. Secondly, the basic design support is composed of two analysis methods. They are forecasting of ship stability and forecasting of the fatigue lifetime of ship materials. Thirdly, there is an analysis method, forecasting of material requirement, in detail design support. Production design support has two analysis methods, which are estimation of duration in production process and optimization of task allocation in production design. At last, there are four analysis methods on design process management. They are visualization of design process, performance analysis of design process, diagnosis & visualization of cooperation network of design, and knowledge & information analysis of design. The analysis method descriptions are summarized in Table 2. Each analysis method is described as the followings.

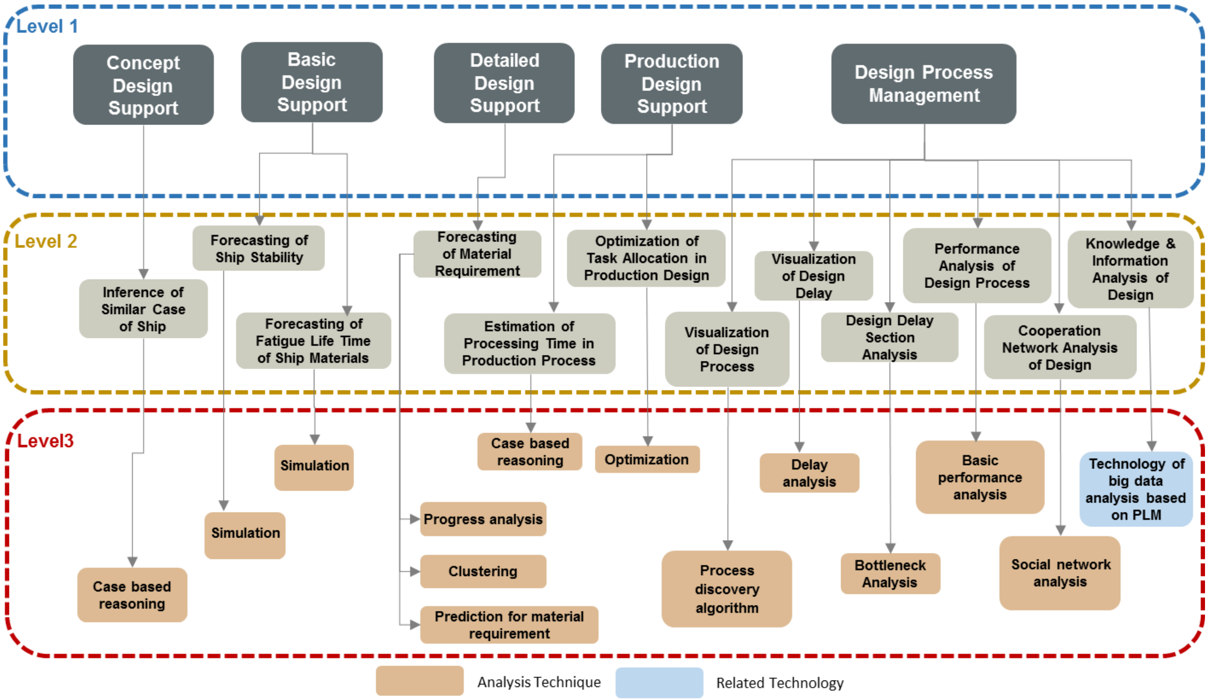

Figure 6. Analysis Composition in Design Phase

· *Inference of Similar Case of Ship*

Initial design stage in shipbuilding industry draws the optimal concept of ship satisfying the requirement of ship owner such as classification of freight, carrying capacity, size, speed, and expected route. The concept of ship is generally drawn depending on the experience of the designer. However, the empirical approach to derive the concept of ship is not reliable because the subjectivity of the designer may affect the process of drawing the concept. It needs to determine the basic performance of ship and the primary specification using the historical ship design data (Delatte & Butler, 2003). The data includes the type of ship, dead weight capacity, speed, the requirement of ship owner, and others. In terms of detailed algorithm, case based-reasoning is applied to the data. Case-based reasoning predicts the result of a new case based on the results of the historical cases (Watson, 2001). It measures the similarity between a new case, which is the requirement of ship owner, and the historical cases. If the importance of each requirement is available, it will be considered as the weight when measuring the similarity. Then, it selects the case, which is the most similar with new case. As the result of this analysis method, the historical shipbuilding case, which is the most similar to the requirement, is selected. Based on the selected case, the value of design is referenced for the initial design.

- *Forecasting of Ship Stability*

Current regulation related to the maritime safety suggests the standard for the stability such as the limitation of capacity depending on size and type of ship. However, the regulation is insufficient because it does not cover the stability considering the property of ship and sea route. Furthermore, the alternative of design is needed to increase the stability by evaluating the damaged stability considering the characteristic of the ship and sea route. It is assessed whether the design result is satisfied with the standard for the stability through forecasting of ship stability (D. Lee, Choi, Park, Kang, & Lim, 2009). Such analysis uses the data such as historical ship accident data (e.g., accident type, cause, extent of damage), current regulations related maritime safety (e.g., maximized allowance for damage), sea route data (e.g., wave height, wave), ship data (e.g., type, size), and others. In terms of detailed algorithm, simulation is applied to the data. Before implement simulation, the scenario of damaging ship is generated from the historical ship accident data. Simulation is conducted with setting actual wave state as the variable to predict the damaged stability. It is decided that the designed ship has the damaged stability accord with the standard if the measured damaged stability is under the maximized allowance of damage. In the case of exceeding the allowance, the alternative will be designed to correspond to the standard.

- *Forecasting of Fatigue Lifetime of Ship Materials*

Fatigue lifetime refers to the number of repetitions or the time until the structure is destroyed when loads are repeatedly applied to the structure. The ship's fatigue lifetime needs to be considered in the design process as the standards of ship design that can be operated in extreme environments such as the North Sea are recently tightened. Materials with the high fatigue lifetime would be derived through foresting of fatigue lifetime of ship materials. The data such as fatigue lifetime of ship materials and marine and climate data of sea route. Regression and simulation are applied to the data for the analysis. The value for the forecasting model of fatigue lifetime is calculated using the regression (C. Yoo et al., 2012). Moreover, the variables are randomly generated for the simulation. The effect of each variable to the fatigue lifetime is analyzed as the distribution of fatigue lifetime is derived from random variables. As the result of this analysis method, the distribution of fatigue lifetime is created, which considers the actual sea route. The distribution describes the accumulated probability that materials are destroyed before the applying of weight is repeated in N times, when N is the number of a cycle (Yoon & Zhang, 2008). Such a result supports the decision making on selecting the materials on the stage of basic design.

- *Forecasting of Material Requirement*

In the case of custom-made materials used in shipbuilding, the required quantity should be predicted and materials would be ordered in advance before actual use. Material requirement needs to be predicted and ordered at the stage of the design so that materials can be procured before shipbuilding. The order for the materials when they are needed causes the delay of work because the custom-made materials takes a long time for the production. The material requirement is predicted through the analysis using the data. Data of ship and shipbuilding such as material and production plan is used for the analysis. In terms of detailed algorithm, the property of material requirement is derived by analyzing how the trend of material usage changes as time goes on. The ships are clustered based on the property and the standard for classification of the ship is reestablished. The forecasting model of material requirement predicts material requirement by the type of ship and the quality.

- *Estimation of Duration in Production Process*

The way of operating the actual production process is planned at the stage of production design. Production planning is conducted depending on the experience of the worker on the actual work-site operations. Reasonable production planning is important based on the proper evidence because loss in the profit is caused by the compensation for the delay that actual work is not implemented according to the plan. The data driven production planning will be conducted through the estimation of duration in production process. The data is historical performance data such as size of ship, type of ship, duration of each process activity, and required specification of the ship owner. In terms of the detailed algorithm, case based reasoning is applied for the analysis. The most similar case to the requirement of ship order is found from the historical record data by using cased based reasoning (Watson, 2001). The duration of each production process of the selected case is used to make a production plan for shipbuilding of a new order (K. Oh & Park, 2005). As the result of this analysis method, the similar ship and the duration of each process activity of it are generated from the similar case to the requirement of the ship order. This information is used as the basis for the production planning.

- *Optimization of Task Allocation in Production Design*

Tasks of production design requires the longest working hours among the stages of design. In addition, it directly or indirectly affects the productivity improvement because it is related to the production of a ship. Production design tasks have a precedence relationship due to the process based work of it and the following task can start after the precedent task ends. The production design for a specific block usually takes about one month because a worker is responsible for production design for a specific block. The current allocation of the tasks is determined by the situation of the field relying on the experience and knowledge of the production design manager. It is necessary to consider experience, ability and past performance of the worker to allocate effectively tasks. The data driven approach is able to consider them. The effective allocation of production design tasks is implemented through optimization of task allocation in production design (Son & Kim, 2012). The data is the task data of production design such as level of difficulty of task and standard duration of task. Furthermore, worker data for task is used such as the experience of design tasks, performance, and others. In terms of detailed algorithm, Hungarian algorithm of optimization is applied to the data. The algorithm is generally used to solve the optimal allocation problem given tasks and workers which are one or more. As the result of the analysis method, the production design tasks are allocated to the workers when the total sum of duration is minimized.

- *Visualization of Design Process*

During the design of the ship, the complex stream of design works appears because there are many works transferring due to the changes of the design. Due to the complex stream of the design work, it is difficult to understand the relationship among the design works clearly. By visualizing the stream of design works and establishing the design process model based on the event log of the design work, it is possible to understand the relationship among the design works. The data needed is the event log of design: for example, caseID, activityID, timestamp, and originatorID. The algorithm for analysis is process discovery algorithm. By using process discovery algorithm (ex. α-algorithm, Heuristic Mining, Fuzzy Mining, and Genetic Mining), it is possible to set and visualize the process model. The result of the analysis is the process model that treats each design work as an activity.

- *Visualization of Design Delay*

The stage of designing the ship takes over twelve months on average. If the delay occurs on the design stage, it would affect the schedule of shipbuilding. Due to the delays on the design process occurring in practice, it is needed to understand the period that delay occurs, and the stream of the works. By visualizing the delays on the design stage, it is possible to know the current state of the delay on each design process, and the cause of the delay. The data for analysis is the event log of design (ex. caseID, activityID, timestamp, and originatorID). Delay analysis method is applied to the data, as the method for analysis. From comparison and analysis between the plans and actual data, it is possible to describe the current state of the delay on each design work according to the period of the delay. If certain work is delayed over the certain period, the work can be judged as the cause of the delay on the next design stage, and the target to reduce the delay. From the analysis, it is possible to visualize the number of delaying days, and understand the current state of the delay on design works according to the period of the delay.

- *Design Delay Section Analysis*

When a new order of change comes on the stage of designing the ship, the delay on the design works occurs frequently because some of the completed works should be reworked. Due to the design processes being in progress simultaneously by many workers, and transferring the works happen frequently, it is needed to diagnose the delay in the viewpoint of the process. The data to analyze is an event log of design, such as caseID, activityID, timestamp, and originatorID. There are two methods to analyze the data. The first one is process animation. The progress of the works can be animated based on the process model generated by the event log data. The bottleneck point can be found by visualizing the speed of the progress on each design work, and the workload concentrated in a certain section. The other one is performance analysis with process model. With performance analysis with process model, it is possible to know the section and the work that consume long time on the process model. The result of the analyses is the bottleneck point on the design process.

- *Performance Analysis of Design Process*

For effectiveness of the design process, it is needed to analyze the performance of the design process according to the performance on each work and worker, and the workload on the design process. By analyzing performance of the design process on each case, work, and worker, according to the frequency of the work and the time consumed, it is possible to diagnose the current state of the process. The data for analysis is the event log of design (ex. caseID, activityID, timestamp, and originatorID). The method to apply to data is basic performance analysis. By analyzing performance of the design process on each case, work, and worker, it is possible to diagnose the current state of the process based on the frequency of the work and the time consumed. By calculating working time and waiting time based on the event log, the simple statistics like minimum, maximum, and average value can be calculated. In addition, the basic performance analysis visualizes the result as graphs like bar chart, pie chart, meter chart, and time chart. The result from the analysis is the performance graph in various viewpoints like cases, works, and workers.

- *Cooperation Network Analysis of Design*

However, many workers design the ship together during the stage of the ship design so transferring works among the workers happens frequently. For improving the efficiency by managing the overall design process, it is needed to understand and manage the work stream among the workers. By establishing the cooperation network among the workers designing, it is possible to understand the primary work stream among the workers, and the primary workers. The data needed is an event log of design: for example, caseID, activityID, timestamp, and originatorID. The method for analysis is social network analysis. Social network analysis shows the relationship among the workers according to works transferred. The node meaning the worker, who has high frequency of performing the work, appears big. It is possible to visualize the primary relationship above the standard, according to the frequency of the relationship among the workers. From the analysis, the social network among the workers who transfer their works to each other is visualized.

- *Knowledge & Information Analysis of Design*

The error of the design found at producing/installing/operating/decommissioning, occurs safety problems and the economic loss. However, it is difficult to find inefficiency and errors on the design process with manpower. If the inefficiency and the errors are prevented in advance by using big data analysis, it is possible to improve the competitiveness of production. By analyzing the data accumulated on PLM (Product Lifecycle Management, a system for controlling the data and information of the ship and offshore platform), and finding and resolving the problems in advance, it is needed to support the decision making for designers and managers to take action effectively. The related technology is a technology of big data analysis based on PLM. When the high-quality data analyses technology, big data analysis to PLM is supported. To realize the technology, five things are needed: I) a multi-dimensional data model for analyzing PLM data; II) a system of online analytical processing and visualization for analyzing PLM multi-dimensional data; III) a model and an engine of data mining for analyzing PLM; IV) PLM data structure integrated with internet of technology things; V) a technique of big data analysis for the internet of things. By analyzing the data accumulated on PLM, for example, component masters, the structure of the product, and the design information on drawings, it is possible to extract patterns of designing and non-standardized design knowledge. By analyzing the data accumulated on PLM, and understanding the various errors, it is also possible to prevent the safety problems and the economic loss in advance when the errors are found in designing, operating, and decommissioning.

Table 2. Description of Analysis Method for Design Phase

Concept Design Support				
Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Inference of Similar Case of Ship	· Historical ship design data (e.g., type of ship, speed, dead weight capacity) · Requirement of ship owner	· Case based reasoning	· Similar case	(Delatte & Butler, 2003; Watson, 2001)
Basic Design Support				
Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Forecasting of Ship Stability	· Historical ship accident data (e.g., accident type, cause, extent of damage) · Current regulations related maritime safety (e.g., maximized allowance for damage) · Sea route data (e.g., wave height, wave) · Ship data (e.g., type, size)	· Simulation	· Assessing stability being suitable to standard	(D. Lee et al., 2009)
Forecasting of Fatigue Life Time of Ship Materials	· Fatigue life time of ship materials · Marine and climate data of sea route	· Regression · Simulation	· Probability distribution for fatigue life time of materials	(C. Yoo et al., 2012; Yoon & Zhang, 2008)
Detail Design Support				
Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Forecasting of Material Requirement	· Data of ship and shipbuilding (e.g., material, production plan)	· Progress analysis · Clustering · Prediction for material requirement	· Predictive value for material requirement	(ECMiner, 2015)
Production Design Support				
Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Estimation of Duration in Production Process	· Historical performance data (e.g., size of ship, type of ship, duration of process activity) · Required specification of ship owner	· Case based reasoning	· Duration of process activity for similar ship	(K. Oh & Park, 2005; Watson, 2001)

Optimization of Task Allocation in Production Design	· Task data of production design (e.g., level of difficulty of task, standard duration of task) · Worker data for tasks	· Optimization	· Optimized allocation of tasks	(Son & Kim, 2012)
Design Process Management				
Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Visualization of Design Process	· Event log of design (e.g., caseID, activityID, timestamp, originatorID)	· Process discovery algorithm	· Process model	(S. Kim et al., 2016)
Visualization of Design Delay	· Event log of design (e.g., caseID, activityID, timestamp, originatorID)	· Delay analysis	· Current state of design delay	(S. Kim et al., 2016)
Design Delay Section Analysis	· Event log of design (e.g., caseID, activityID, timestamp, originatorID)	· Process Animation · Performance Analysis with Process Model	· Delay section and task	(S. Kim et al., 2016)
Performance Analysis of Design Process	· Event log of design (e.g., caseID, activityID, timestamp, originatorID)	· Basic performance analysis	· Performance of case/task/worke r	(S. Kim et al., 2016)
Cooperation Network Analysis of Design	· Event log of design (e.g., caseID, activityID, timestamp, originatorID)	· Social network analysis	· Social network	(S. Kim et al., 2016)
Knowledge & Information Analysis of Design	· Related technology: technology of big data analysis based on PLM			(KISTEP, 2015)

4.4. Data Analysis in Production Phase

4.4.1. Production Phase

Shipbuilding industry has three properties: labor-intensive, capital-intensive, and technology-integrated. Shipbuilding industry is the industry of comprehensive fabrications; ships are constructed with numerous materials and components in the extensive scale, compared to products from other manufacturing industries. Simply, the shipbuilding process can be described as cutting huge steels, welding them, and making the structures of the ship. These processes are complicate works and demand lots of labor. In addition, for shipbuilding, the shipbuilding company should get fabrication yards equipped with huge facilities like building berths, docks and cranes and secure funds for management. Due to the market of shipbuilding being worldwide and a single market, the competitiveness of cost and technology is very important as well. Works of shipbuilding need combination of various types of technology like shipbuilding engineering, mechanical engineering and electronic engineering. The competitiveness of the shipbuilding company is, therefore, determined by the types of technology for design and production.

The procedures of shipbuilding are conducted through forming of various materials and components, sub-assembly and grand assembly so smooth procurements for materials and components and swift construction of components, piping and blocks are the primary factors for delaying the period of delivery. Compensation for delaying makes profitability of shipbuilding deteriorate. In order to manage supply chains effectively and installing pipes on right time, the needs for application of big data analysis exist. Although automation of processes in shipbuilding is inadequate compared to those of other manufacturing industries, the quality and quantities of data collected during construction of the ship are expected to improve because of the improvement of the concerned types of technology and reduction of cost for introducing the sensors. Considering the current circumstances of shipbuilding industry mentioned above, the importance of the application of big data analysis is expected to increase gradually.

With the application of big data analysis to improve efficiency of shipbuilding procedures including improvement of the types of technology related to shipbuilding, management of supply chains, installation of pipes on the right time, etc., the competitiveness of shipbuilding industry appears to be strengthened.

4.4.2. Data Analysis Description

The analysis composition in production phase shows in Figure 7. There are six major analyses on production in shipbuilding industry: production process management, production operation management, maintenance, quality management, safety management, and supply chain management. These are developed by considering of the characteristic of production in the shipbuilding industry.

Firstly, production process managements have visualization of production process, visualization of production delay, production delay section analysis, diagnosis of abnormal workflow in production process, performance analysis of production process, production workload analysis, and connection analysis of the worker (facility). Production operation management is composed of two analysis methods: feedback analysis of the worker, and inspection of changes to the production process. Thirdly, maintenance has abnormality detection of manufacturing facility, planning for maintenance, schedule planning for maintenance, failure type analysis, risk analysis of failure, and deficiency analysis. Quality management includes optimization of welding quality. Safety management has two analysis methods, safety analysis of the worker, and environment analysis of the shipyard accident. Lastly, there are seven analysis methods on supply chain management: forecasting of supply chain lead-time, diagnosis of abnormal workflow of supply chain, performance analysis of supply chain process, visualization of delay time of supply chain, delay section analysis of supply chain, supplier workload analysis, and logistics optimization. The analysis method descriptions are summarized in Table 3 ~ 5.

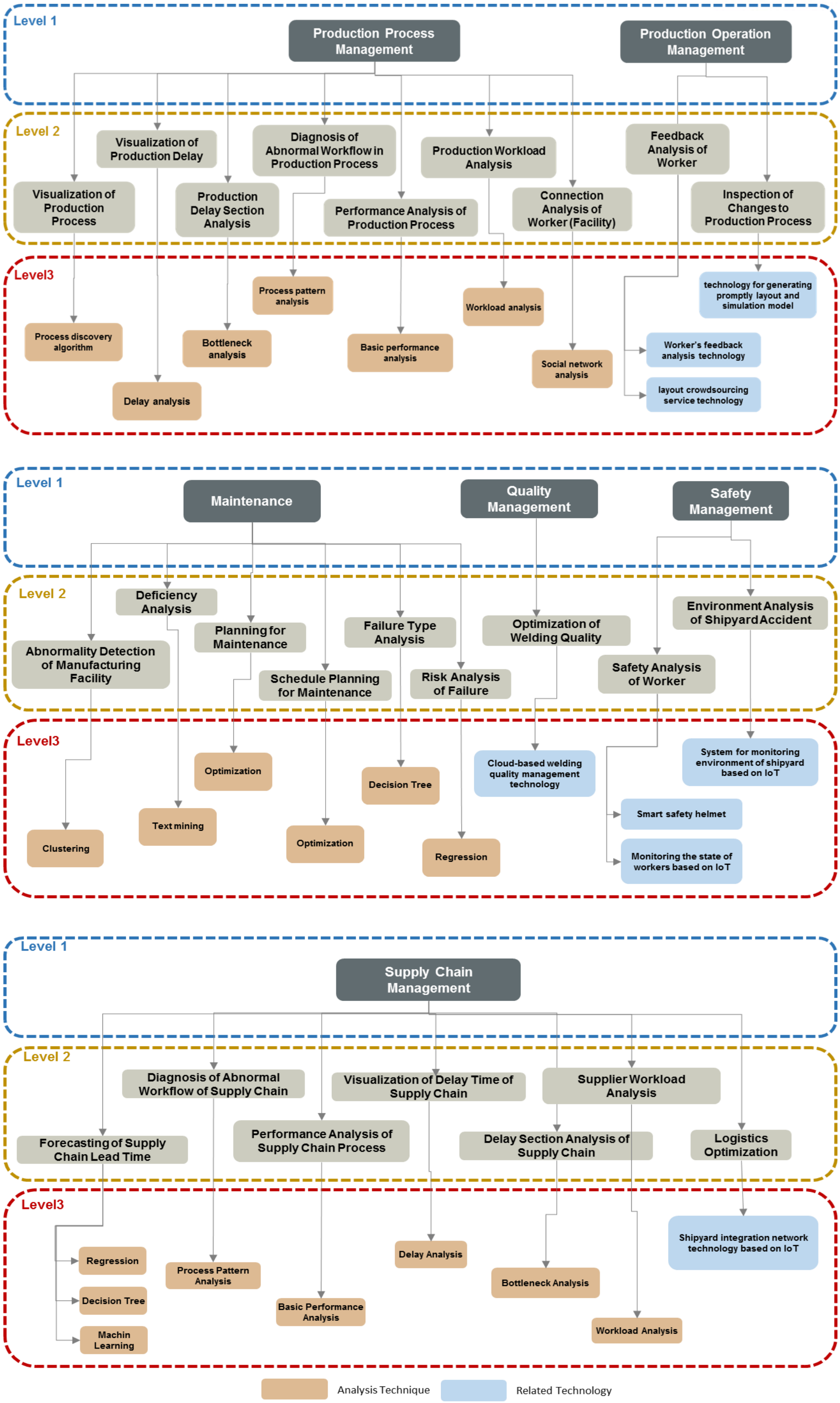

Figure 7. Analysis Composition in Production Phase

- *Visualization of Production Workflow*

Due to the enormous projects of shipbuilding industry that are in progress simultaneously, production workflows become extremely complex. Although delaying the period of delivery happens, there are difficulties to comprehend production processes accurately. By deriving production process model, however, it is possible to visualize complex production workflows with a realistic description based on data (S. Kim et al., 2016). The data includes production planned and actual data: for example, blockID, taskID, planned start/end time, actual start/end time, and originator ID. For detailed algorithm, process discovery algorithm is applied to the data. Production planned and actual data should be converted to MXML form to apply to process discovery algorithm. Process model can be derived by using process discovery algorithm such as a-Algorithm, Heuristic Mining, Fuzzy Mining, and Genetic Mining. As the result of the analysis, production workflow can be visualized as a process model that regards each procedure of workflow as an activity.

- *Visualization of Production Delay*

In the field of ship and offshore platform, it is important to construct and deliver them within the fixed period. The scale is so extensive and the knowledge and experiences about construction do not accumulate especially for offshore platform. For these reasons, delaying the period occurs frequently and causes significant cost loss. By visualizing production delays, it is possible to monitor current state of production delay and diagnose cause of delay (S. Kim et al., 2016). The data of production plan and actual data (ex. blockID, taskID, plan start/end time, actual start/end time, and originator ID) are needed, and algorithm that is applied to data is delay analysis. With comparison analysis between plan and actual data, the current state of production delay can be categorized from the period of the delay. The activity causing the long time delay or a series of delays to following the process can be selected to be handled first to solve the delay. From this analysis, the number of days of delaying on each process can be reported in real time and current state of delay can be checked along the period of delay.

· *Production Delay Section Analysis*

Production delay can cause another delay on the following process, and by chain reaction, become the cause of delay on the period of the overall shipbuilding schedule. In a fabrication yard, although the delay on each process occurs, it is difficult to diagnose which delay of process causes overall delay of shipbuilding schedule because the scale of production processes is extensive. By analyzing the production delay section, it is possible to discover the delay on production processes and derive measures for improvement from the analysis (S. Kim et al., 2016). The data needed are production planned and actual data such as blockID, taskID, planned start/end time, actual start/end time, and originator ID. There are two algorithms to be applied to the data: process animation and performance analysis with process model. Firstly, process animation describes current state of process as the animation on the process model, based on production planned and actual data. By visualizing the speed of the process that is ongoing and work amount concentrated at specific section, the section where bottleneck phenomenon occurs can be found. On the other hand, performance analysis with process model can show the section or work that consumes the time abnormally or relatively more. This analysis can show the process model with classification by the time consumed, using the visual effect like colors. As the result of the analyses, bottleneck point of process model and the work that causes overall delay can be detected.

· *Diagnosis of Abnormal Workflow in Production Process*

Purpose of this analysis is managing workflow of primary process and abnormal patterns of process by analyzing and understanding the patterns of the process that are primary or abnormal. It is possible to understand primary or abnormal patterns by analyzing each case`s time consumed and events occurred (S. Kim et al., 2016). From this understanding, management of processes can be improved. The data for the analysis are production planned and actual data: for example, blockID, taskID, planned start/end time, actual start/end time and originator ID. The detailed algorithm applied is a dotted chart. Dotted chart visualizes distribution of events from various viewpoints like cases of processes, works, and workers. By analyzing activities that can be batched and time-consuming or cases which is the outlier in the viewpoint of the number of events, it provides concrete clues for analyzing the production process with more details. From the analysis based on dotted chart, it is possible to understand various patterns of processes.

· *Performance Analysis of Production Process*

It is required to diagnose the current states for efficiencies of production processes by analyzing the works of production processes, performance and the amount of works for each worker. From the analysis of performance, performance of frequencies and time consumed in the processes can be visualized along each work, worker and case (S. Kim et al., 2016). With related performance analyses like analysis of bottleneck points and process pattern analysis, it is possible to evaluate the performance of work. From the evaluation, effectiveness of production process monitoring can be improved and the result of evaluation becomes the index for understanding the causes of delaying in detailed and managing the performance. The required data for the evaluation include production planned and actual data, such as blockID, taskID, planned start/end time, actual start/end time and originator ID. The detailed algorithm applied to data is basic performance analysis. Basic performance analysis measures performances from various viewpoints like works, workers, and cases based on frequencies and time consumption. From the data given, working time and waiting time can be measured and it is possible to compile statistics like maximum, minimum, and average value. In addition, analysis provides the visualization of result with bar charts, pie charts, meter charts, and time charts. The result of analysis is various visualized graphs based on frequencies and time of each case, work, and workers.

· *Production Workload Analysis*

Delaying the period of delivery is the primary issue in the field of ship and offshore platform. It is essential to prevent delays and manage production workload effectively by analyzing production workload. From understanding current states of workload on each worker and shop of production, it is possible to take action for preventing production delaying preemptively, like dispersing workload occurring delays, and allocating additional workers (S. Kim et al., 2016). The data for the analysis are production planned and actual data (ex. blockID, taskID, planned start/end time, actual start/end time, and originator ID). The algorithm applied to data is workload analysis. Workload analysis shows current states of production workloads on each worker and shop by comparing and analyzing plans and actual data. It is possible to understand the present workload compared to plans, and the performances of works on each worker and shop. Form the analysis, the current state of workload on each worker and ship could be understood.

- *Connection Analysis of Worker (Facility)*

Due to the production of ship and offshore platform conducted in enormous scale, making the production process effective is important. Managing the workflow by each process and workflow between workers (facilities) that conduct each process causes effectiveness of overall production process. By establishing the network between the workers (facilities) conducting each process, it is possible to figure out the workflow between the workers, and the primary workers (S. Kim et al., 2016). The data for analyzing are production planned and actual data, like blockID, taskID, planned start/end time, actual start/end time, and originator ID. The detailed algorithm for the data is Social network analysis. Social network analysis establishes the relation between workers (facilities) based on the workflow of production processes and works. Workers (facilities) with higher frequency of conducting works show relatively a larger network node. It is possible to visualize critical workflows between workers (facilities), which have higher frequencies than certain value. The result of analysis is the social network between workers (facilities).

- *Feedback Analysis of Worker*

Due to the swift discussion system between workers, engineers, and managers not being established, improvement of production using knowledge of workers who have abundant field experiences is limited to production field. Most effective way to improve the current problems about field and production processes of constructing ship and offshore platform is reflecting opinions of workers, engineers, and managers widely, which is crowdsourcing. Crowdsourcing is the technology that analyzes and standardizes the knowledge of workers to improve production processes effectively and swiftly. In the poor field environment like the shipbuilding company, especially, the layout crowdsourcing service technology can be a measure to collect opinions of various workers effectively and a platform of the technology that can improve problems of the work field, production process, and design stages systemically. This technology includes development of the service platform and the user interface supporting the shipbuilding company`s crowdsourcing. By collecting the opinions from workers, the shipbuilding company can diagnose and improve the ineffectiveness of the existing work processes. Crowdsourcing also improves the efficiency and safety of work processes through the supporting system for workers. In addition, through communication between workers using smart equipment based on the types of information technology, it is possible to optimize the environment of the work field, and reduce the cost. There are three things to conduct feedback analysis for workers. First, the technique for analyzing non-standardized data should be developed. Second, domain ontology should be established. Lastly, the system to analyze the feedback of workers` non-standardized knowledge and information should be established. Through feedback analysis of workers, knowledge of field workers can be applied to production processed, and production processed can be improved.

- *Inspection of Changes to Production Process*

To generate simulating models for supporting swift verification of changes of shipbuilding processes, too many times, costs, and endeavors are needed. When a change of certain information occurs, existing works of generating layouts and simulating models require too much workload to modify the contents related to the change. It is needed to develop a support technology that can make the work of generating and modifying the layout and simulating model swift. By reflecting and verifying the changes of designs and productions swiftly through analyzing verification of process changes, it is possible to reduce the costs and improve the productivity. Technology for generating layout and simulating model is based on swift reflection and verification of changes of design and production. This technology includes six contents: I) designing models of neutral information, which supports generating models for simulation; II) research for optimizing plans for operating production field; III) designing modules for generating layouts automatically; IV) designing modules for generating models of neutral information; V) programs to automatically generate layouts and models for simulation based on neutral information; VI) establishing the system generating layouts and models for simulation swiftly. By reflecting and verifying changes through the layouts and models for simulation, it is possible to apply the changes to the production field swiftly. It is also possible to operate and manage the production field properly, like optimizing the plans for operating production field. In addition, the time and cost consumed for generating simulating models, to verify production process that changes, are applied to can be reduced.

Table 3. Description of Analysis Method for Production Phase (1)

Production Process Management				
Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Visualization of Production Process	· Production planned and actual data (e.g., blockID, taskID, planned start/end time, actual start/end time, originator ID)	· Process discovery algorithm	· Production process model	(S. Kim et al., 2016)
Visualization of Production Delay	· Production planned and actual data (e.g., blockID, taskID, planned start/end time, actual start/end time, originator ID)	· Delay analysis	· Current state of production delay	(S. Kim et al., 2016)
Production Delay Section Analysis	· Production planned and actual data (e.g., blockID, taskID, planned start/end time, actual start/end time, originator ID)	· Process Animation · Performance analysis with process model	· Delay section and tasks	(S. Kim et al., 2016)
Diagnosis of Abnormal Workflow in Production Process	· Production planned and actual data (e.g., blockID, taskID, planned start/end time, actual start/end time, originator ID)	· Dotted chart	· Process pattern · Process characteristic	(S. Kim et al., 2016)
Performance Analysis of Production Process	· Production planned and actual data (e.g., blockID, taskID, planned start/end time, actual start/end time, originator ID)	· Basic performance analysis	· Performance of case/task/worker	(S. Kim et al., 2016)
Production Workload Analysis	· Production planned and actual data (e.g., blockID, taskID, planned start/end time, actual start/end time, originator ID)	· Workload analysis	· Current state of workload of worker/department	(S. Kim et al., 2016; M. Park et al., 2015)
Connection Analysis of Worker (Facility)	· Production planned and actual data (e.g., blockID, taskID, planned start/end time, actual start/end time, originator ID)	· Social network analysis	· Social network of worker/facility	(S. Kim et al., 2016)
Production Operation Management				

Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Feedback Analysis of Worker	· Related technology: worker's feedback analysis technology, layout crowdsourcing service technology			(KISTEP, 2015)
Inspection of Changes to Production Process	· Related technology: technology for generating promptly layout and simulation model			(KISTEP, 2015)

· *Abnormality Detection of Manufacturing Facility*

In the field of ship and offshore platform, many enormous projects are in progress simultaneously. One delay of the production process can cause delaying in the next production processes and delivery, as a serial action. Analyzing the data about variables of facilities after completion of each production process in real time and finding the abnormality of the process supplement the quality of production in early stages. In addition, it is possible to minimize the delay of production processes by swift actions. By diagnosing the abnormality of production processes, whether the abnormality of manufacturing facilities happens or not, can be diagnosed in real time. The data for diagnose is facilities data including historical and real-time stream data. The detailed algorithm applied to the data is k-Nearest Neighbor (kNN) algorithm. kNN algorithm is suitable to deal with stream data because this algorithm can predict the class of result with simple calculation of distance, as well as collection of the data. kNN algorithm is most simple and relatively accurate algorithm among algorithms for classification (Kwak & Kim, 2013). The algorithm would define the characteristic space as the Euclidean space having dimensions as the number of facility variables comprising the data based on stream data. Data for production processes are defined as vectors on the characteristic space. A group means a set of data that have similar values of all variables. When classifying the normal groups and abnormal groups based on historical data, and new data is given, the abnormality of manufacturing facility can be detected as the group is classified. The result of algorithm shows whether manufacturing facilities have abnormality or not.

- *Planning for Maintenance*

When there is a problem and a stop of operation happens to the manufacturing facilities, the delay of production processes occurs and causes problems to the overall schedule. Planning for maintenance is for planning to minimize the cost considering the overall product lifecycle. From planning, it is possible to make the optimized plans for maintenance and replacement considering the maintenance, the time during the trouble of facilities, and the cost of replacement (Kalagnanam, Lee, Ide, & Han, 2015). The data needed is recorded data of asset use, budget for maintenance, hazard functions, and other constraints like resources. The methods applied to data are mixed integer linear and nonlinear program. The input data for analyses is data of functions about the rate of the problem. This is based on the analysis of the problem's degree predicting the lifespan of assets from the data of asset use and information about the conditions. The constraints are information about assets, data of the function of the rate of the problem, and budget for maintenance and resources. The result from the analyses plans for maintenance with minimum cost on each facility.

- *Schedule Planning for Maintenance*

As same as the reason mentioned in planning for maintenance, schedule planning for maintenance is needed. Schedule planning for maintenance establishes operating plans for a week or a month, and fulfills the demand of maintenance with minimum cost by establishing optimized schedules for maintenance and inspection, and path plans for resources like human, equipment, and material. It can also be applied to the factories having various assets (Kalagnanam et al., 2015). The data needed is for a period of schedule, asset information for scheduling (ex. location, and distance), order of work, and other constraints like worker type, and part inventory. The method applied is the mixed integer linear program. Mixed integer linear program would find the optimized solution that the number of missions done is maximized, and the moving distance and overall consumed time are minimized within the given period. The results from the program are schedule planning and paths with minimized cost per unit of a shop.

- *Failure Type Analysis*

The purpose of failure type analysis is predicting the possible problems of assets and facilities in the future and taking actions for preventing these problems. The problems of assets and facilities during the production processes affect negative influences to the quality of production, cost of production, and safety. Failure type analysis would find the signs of these problems, detect these problems based on the data in the early stage, and take actions swiftly (W. Hwang, Kim, Jang, Hong, & Han, 2008). The data is the record data of facility failure, such as year of failure occurrence, place, and type of facility and failure. The detailed algorithm for analysis is a decision tree. Through this method, failure types are defined based on the data of facility failure. Properties and contents of each variable also are defined. Variables include the date, time and location of occurrence, and each variable has properties and values like year, date and location. By applying decision tree algorithm based on variables for analyzing failure types, the condition of failure occurrence on each type can be derived. The result of the analysis is the condition of failure occurrence on each type.

- *Risk Analysis of Failure*

Risk analysis of failure estimates expectation of lifespan of assets are based on the record data of asset failure. Expectation of lifespan of assets are determined by record data and properties of assets, and primary stress factor about environment and operation. Generally, models for failure risk expresses remaining lifespan of assets as a function comprising factors of environments and operation, and properties of assets. Remaining lifespan is calculated based on the record data of assets, as a probability that assets operate without failures during certain period in the future. The methods of analysis consist of parametric technique, partial parametric technique, and non-parametric technique. In the case of risk analysis of failure, the effects of variables to remaining lifespan of assets would be analyzed by estimating indexes like functions of survive and functions of failure rate, and comparing them (Kalagnanam et al., 2015). The data needed are record data of maintenance, operation record data of manufacturing facility, and sensing data of manufacturing facility. The method applied to the data is Cox Proportional Hazard Regression. Cox Proportional Hazard Regression calculates the initial values of probability for survive from the record data of maintenance, and the initial values are corrected by variables of production processes. The result of analysis is the life expectancy of the assets.

· *Deficiency Analysis*

Reports of inspection in the field of ship and offshore platform are documents describing deficiencies and supplement points of offshore structures. Analyzing inspection reports improves the production processes of the structures, prevents additional inspections, and finally prevents the delays of the period of delivery and reduces non-operation expenses. However, unstructured data in the reports disturb understanding properties of deficiencies and effectiveness and management. For this reason, it is needed to analyze the production issues to classify types of deficiencies about the results of inspection. Through deficiency analysis based on data, field experts can understand the deficiency types of the offshore structures. It is also possible to reduce a gap between data analyses and actual usages. The data is inspection text log data like inspection reports (S. Lee et al., 2014). There are two stages for analysis: text mining and clustering. Text mining conducts pre-treatment to text data of inspection reports, like tokenizing, PoS tagging, removing stop-word, stemming, pruning, and weighting. It calculates indexes like term-frequency (TM) and document-frequency (DF) on each document, and the indexes are utilized on the stage of clustering. By analyzing the interrelationship and patterns of trend, it is possible to find the issues like components related to deficiency and types of deficiency. Clustering categorizes the documents that have a high similarity of text composition, by comparing the documents. Usually, clustering is conducted by using a self-organizing map. The result of analysis is the conceptual diagram showing deficiency types of offshore structures. From the diagram, it is possible to understand the cause of deficiency and related issues.

· *Optimization of Welding Quality*

When shipbuilding, accuracy of steel cutting and welding deformation during pre-assembly process affects performance and strength of the structure. Modification and correction to deformation can cause unnecessary man-hour work, and reduce productivity. Quality of welding is determined by the skills of welding operators. Problems due to deviation of welding quality can be resolved by securing a suitable welding quality following the standard welding procedure. Optimized conditions for welding can be suggested by researching the way to optimizing welding quality, observing the quality of welding, and analyzing causes of the problems. Through these procedures, welding quality can be improved. Cloud-based welding quality managing technique suggests optimized welding conditions according to the way to weld and the members for welding. It also includes techniques for preserving equipment and managing welding process based on the information technology. Contents of the technique are like these: I) developing the management system for optimized welding based on the analysis of variables about welding; II) collecting and analyzing welding data during production processes; III) developing techniques for preserving and managing welding processes with types of information and communication technology; IV) managing the record of welding equipment and preserving the equipment with wire-wireless networks. Cloud-based welding quality managing technique assures the quality of welding by monitoring welding quality and analyzing a cause and effect. In addition, the technique reduces unnecessary man-hours and improves productivity by managing welding processes with optimized welding conditions which are suggested.

· *Safety Analysis of the Worker*

Shipbuilding process has massive-scaled components and high complexity. Plus, dangerous situations such as workers being exposed to risk elements in a work environment happen frequently. In addition, it is necessary to manage a health state of workers in the shipbuilding industry systemically because the intensity of works in the shipbuilding industry is more strenuous than that in other industries. It is possible to induce workers to do their work with safety by examining their health state in advance, finding signs of abnormality of health state, and adjusting the intensity of the works. From the analysis of safety about workers in the shipyard that has many risk elements, dealing with hazard situations and managing workers` health state can induce prevention of disasters. There are two examples for safety analysis. One is smart safety helmet technique. Smart safety helmet technique supports the countermeasure to disasters in the shipyard with safety. The technique includes designing and realization of smart safety helmet`s software and hardware, and integrated development of smart safety helmet. Smart safety helmet alerts the workers to emergencies, and help them to evacuate from the place that needs to be evacuated from. In addition, when workers get in trouble, safety measures can be operated swiftly as the information and location of workers transferred. The other one is technique for monitoring the state of workers based on the internet of things. It is for preventing the emergencies due to the health problem of workers, by monitoring workers` health state and finding abnormal signs. There are three things to realize this technique: I) a technique finding abnormal signs from workers` health state, II) a program monitoring the health state of workers and finding abnormal signs, III) a system monitoring workers` health state based on the internet of things. By checking the state of workers and adjusting the intensity of works, safety works can be induced. In addition, industrial accidents in the shipyard can be prevented by managing workers` health state systemically.

· *Environment Analysis of a Shipyard Accident*

As the amount of constructing high-value vessels and offshore platforms that are constructed through complex production processes increased, safety accidents also occur more frequently. However, the system for managing the safety systemically to prevent disasters has not been established. By analyzing the environment of shipyard accidents, it is possible to realize the work environment that assures workers` safety. One of the techniques relevant to environment analysis of shipyard accident is the system for monitoring environment of shipyard based on the internet of things. This system supports assurance of workers` safety. It can operate as analysis system for big data of environment for a shipyard accident and the central control system. There are three things to realize in the system: I) a technique to monitor emergencies, II) models to analyze big data of the environment of shipyard accidents, III) a system for analyzing environment around the shipyard and controlling. By establishing the system, it is possible to prevent possible disasters by monitoring and analyzing safety accidents and disasters in the shipyard. In addition, safety for workers in shipyard becomes guaranteed by controlling and preparing possible disasters.

Table 4. Description of Analysis Method for Production Phase (2)

Maintenance				
Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Abnormality Detection of Manufacturing Facility	· Facilities data (e.g., historical and real-time stream data)	· Clustering	· Abnormality detection	(Kwak & Kim, 2013)
Planning for Maintenance	· Record data of asset use · Budget for maintenance · Hazard Function · Other constraints (e.g., resources)	· Optimization	· Maintenance plan for an asset	(Kalagnanam et al., 2015)
Schedule Planning for Maintenance	· Period of schedule · Asset information for scheduling (e.g., location, distance) · Order of work · Other constraints (e.g., worker type, parts inventory)	· Optimization	· Maintenance plan for an factory	(Kalagnanam et al., 2015)
Failure Type Analysis	· Record data of facility failure (e.g., year of failure occurrence, place, type of facility and failure)	· Decision tree	· Conditions causing failure by failure types	(W. Hwang et al., 2008; S. Lee et al., 2014)
Risk Analysis of Failure	· Record data of maintenance · Operation record data of manufacturing facility · Sensing data of manufacturing facility	· Regression	· Life expectancy of an asset	(Kalagnanam et al., 2015)
Deficiency Analysis	· Inspection text log data (e.g., inspection report)	· Text mining · clustering	· Conceptual model of deficiency classification of maritime structure	(S. Lee et al., 2014)

Quality Management				
Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Optimization of Welding Quality	· Related technology: cloud-based welding quality management technology			(KISTEP, 2015)
Safety Management				
Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Safety Analysis of Worker	· Related technology: smart safety helmet, monitoring the state of workers based on IoT			(KISTEP, 2015)
Environment Analysis of Shipyard Accident	· Related technology: system for monitoring environment of shipyard based on IoT			(KISTEP, 2015)

· *Forecasting of Supply Chain Leads Time*

There are a lot of outfitting products comprising the ship and offshore platform. There are also various supply routes according to supply chain process and suppliers. Recently, cost loss has occurred in offshore platform projects because of insufficient consideration of various outfitting products and supply routes. For example, the increase of unnecessary stocking of products because of early supply, and delaying of the supply. To resolve these problems, prediction for lead-time of supply during the specification of outfitting products and supply routes are needed. By analyzing the data of supply chain of outfitting products, it is possible to find the factors affecting lead-time and predict lead-time of each supply process. There are two categories of data needed: one is supply chain (ex. supply quantity, and timestamp of procedure completion), and the other one is plumbing materials data (ex. length, weight, diameter, and the number of component pipes). There are four methods to analyze the data. The first one is multiple linear regression models. This method is applied when there are multiple explanatory variables. There are two assumptions: I) the distribution of the variables is the normal distribution of which the average value of errors is zero and the variance is constant, II) the variables are independent to each other. The second method is partial least squares (PLS) regression. By considering covariance between the variables, PLS regression provides more effective prediction than other regression methods do. The model of PLS regression is made by using correlation between explanatory variables and period of production process. Thirdly, boosted decision tree regression can be applied. This method is a multiple additive regression trees algorithm that stochastic gradient boosting method is applied to. Due to it being the method that the technique of boosting ensemble is applied to regression tree, the accuracy of the analysis is high. The last method is artificial neural network method. Artificial neural network consists of nodes that copy biological neurons. Based on weighted sum of input values, estimation and classification are conducted by converting calculated valued form combination functions to output values. Artificial neural network method usually provides a good performance about the problem of prediction and estimation when input and output variables are classified well. The result of the analyses is the effective supply plan based on a predicted lead-time of each supply for process.

- *Diagnosis of Abnormal Workflow of Supply Chain*

By understanding the primary and abnormal workflows in supply chain processes from analysis of process patterns, it is needed to manage the primary and abnormal workflow patterns. In the viewpoints of time consumed and the number of events on each case, it is possible to find primary and abnormal patterns, and improve the effectiveness of managing supply chains. The data for analysis are supply chain planned and actual data: including supplierID, taskID, planned start/end time, actual start/end time, and originator ID. The method for analyzing the data is a dotted chart. A dotted chart visualizes the distribution of events simply, in the various viewpoints like cases, works, and workers of processes. By understanding works that can be batched, cases with long time consumption and the causes of the long period of time, and the abnormal cases in the viewpoint of the number of the events, dotted chart provides the clues for analyzing supply chains in more detail. From the dotted chart, it is possible to understand the patterns and characteristics of various supply chains.

- *Performance Analysis of Supply Chain Process*

It is needed to diagnose the current state for effectiveness of supply chains by understanding the work of the supply processes, and performance and workload for each worker. By analyzing the performance, the performance of frequencies and time consumed in the viewpoint of process can be quantified according to each work, worker, and cases. In addition, from the performance analysis associated with bottleneck point analysis and process pattern analysis, it is possible to evaluate the performance on each supplier, increase the effectiveness of monitoring supply chains, and utilize the result as the index for understanding detailed causes of delays and controlling the performance. The data needed is supply chain planned and actual data (ex. supplierID, taskID, planned start/end time, actual start/end time, and originator ID). The method applied to the data is basic performance analysis. Basic performance analysis evaluates the performance with various viewpoints like works, workers, and the cases based on frequencies and time. By measuring waiting time and working time based on the given data, it is possible to calculate simple statistics like minimum, maximum, and the average value. The method provides visualization of the result with various graphs: bar charts, pie charts, meter charts, and time charts. The result from the analysis is the performance of suppliers in the aspects of cases, works, and workers, based on frequencies and time as indexes, with various visualized graphs.

- *Visualization of Delay Time of Supply Chain*

In the field of ship and offshore platform, it is important to construct and deliver the ship and offshore platform within the period. Each block consists of various materials and components. If only one part lacks, assembly of the block cannot be completed and the serial problems of delaying occurs. By visualizing the delay on the supply chain, it is possible to understand the current state of delays on supplying the materials and components, and diagnose the cause of the delay. The data to analyze are supply chain planned and actual data, such as supplierID, taskID, planned start/end time, actual start/end time, and originator ID. The method analyzing the data is delay analysis. Delay analysis classifies the current state of delay on each supplier according to the period of delay, by comparison and analysis between the supply chain planned and actual data. It can be utilized as the index for managing the suppliers, and the evidentiary materials for requiring explanation and action about delay to each supplier. From the analysis, it is possible to visualize the days of delaying according to each supply chain, and the current state of delaying according to the period of delaying.

- *Delay Section Analysis of Supply Chain*

Delays of period occurred on supply processes cause delay of overall production process. Due to there being supply chain processes with a massive scale, it is hard to check the current state of delay point in supply chains although delay on each process occurs. By analyzing the delay point on supply chains, it is possible to find the delay of period on supply processes and derive the improvement plan. The data for analyses are supply chain planned and actual data (for example, supplierID, taskID, planned start/end time, actual start/end time, and originator ID). There are two methods to analyze the data: process animation and performance analysis with process model. Process animation shows the current state of supplying the components and materials on supply chains in the process model, based on supply chain planned and actual data, with animation. It is possible to find the bottleneck point on the supply chain by visualizing the progressing speed on each process and the concentrated workload on specific section. The second method, performance analysis with process model, shows the section (work) that consume more time relatively or abnormally on process model. From analyses, it is possible to find the bottleneck point and the process in delay on the supply chain.

- *Supplier Workload Analysis*

Due to the delay of the period of delivery being the primary issue in the field of ship and offshore platform, it is needed to prevent the occurrence of delay and manage supply workload effectively by analyzing the supplier workload. With current workload states on each supplier, it is possible to prevent delaying of supply chains in advance by distributing the overload of suppliers and adding the new supplier. The data to analyze are supply chain planned and actual data (ex. supplierID, taskID, planned start/end time, actual start/end time, and originator ID). The method to analyze the data is workload analysis. Workload analysis shows the current state of workload on each supplier by comparison and analysis between supply chain planned and actual data. It is possible to know the current state of supplier workload compared to plans, and evaluate the supply performance on each supply chain. From the analysis, the current state of workload on each supply chain can be figured out.

- *Logistics Optimization*

The basic factor for shipbuilding is managing the location of and operating the blocks of the ship and transporters. It is important to establish the automatic system managing the blocks and transporters based on the network, rather than the handwork system. It is needed to develop the network and the algorithm, which are optimized to the shipyard environment that is massive-scaled and a lot of metal scattered. By optimizing and monitoring the distribution in the shipyard in real time from logistics optimization, it is expected to increase productivity and reduce the cost related to the human resources, transporters, and stockyard. Shipyard-Integration-Network-Technology based on internet of things is the example of logistics optimization. It is for managing distribution transferring and production processes in real time by monitoring materials, blocks and transporter in the shipyard. It is the information-integrated technology connecting facilities within and outside the shops in the shipyard for collecting and sharing the overall information in the form of a pyramid. The contents of the technology are as follows: I) monitoring the environment of shipyard and controlling the location of materials in real time; II) improving the efficiency of work and preventing damages to blocks by using communication system technology based on group communication; III) establishing the digital map of shipyard; IV) assuring the safety of human and improving the efficiency of work by detecting the location of the flagmen. By applying this technology to the shipyard, it is possible to manage the location of the blocks of the ship and transporters, and operate the shipyard properly.

Table 5. Description of Analysis Method for Production Phase (3)

Supply Chain Management				
Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Forecasting of Supply Chain Lead Time	· Supply chain data (e.g., supply quantity, timestamp of procedure completion) · Plumbing materials data (e.g., length, weight, diameter, number of component pipes)	· Regression · Machine learning	· Predicted supply lead time by procedure	(Ham, Lee, & Woo, 2016)
Diagnosis of Abnormal Workflow of Supply Chain	· Supply chain planned and actual data (e.g., supplierID, taskID, planned start/end time, actual start/end time, originator ID)	· Dotted chart	· Process pattern · Process characteristic	(S. Kim et al., 2016)
Performance Analysis of Supply Chain Process	· Supply chain planned and actual data (e.g., supplierID, taskID, planned start/end time, actual start/end time, originator ID)	· Basic performance analysis	· Performance of supplier	(S. Kim et al., 2016)
Visualization of Delay Time of Supply Chain	· Supply chain planned and actual data (e.g., supplierID, taskID, planned start/end time, actual start/end time, originator ID)	· Delay analysis	· Current state of supply delay	(S. Kim et al., 2016)

Delay Section Analysis of Supply Chain	· Supply chain planned and actual data (e.g., supplierID, taskID, planned start/end time, actual start/end time, originator ID)	· Process animation · Performance analysis with process model	· Delay section and task	(S. Kim et al., 2016)
Supplier Workload Analysis	· Supply chain planned and actual data (e.g., supplierID, taskID, planned start/end time, actual start/end time, originator ID)	· Workload analysis	· Current state of workload of supplier	(S. Kim et al., 2016; M. Park et al., 2015)
Logistics Optimization	· Related technology: Shipyard integration network technology based on IoT			(KISTEP, 2015)

4.5. Data Analysis in Service Phase

4.5.1. Service Phase

The principal profit comes from shipbuilding of the new order in the shipbuilding industry. As a result of the property of order-made production industry, the shipbuilding industry is seriously influenced by a market situation. Global shipbuilding industry faces difficult business environment as the quantity of ship order has been sharply reduced due to economic recession and falling oil prices. Big data as new growth engine is expected to contribute to develop a business model of service, which creates a new value in shipbuilding industry. Service business of the shipbuilding industry is the initial stage. Service bushiness using big data is MRO (Maintenance, Repair & Operation) service for ship and navigation on optimal route. DSME pushed forward the development of service business model using big data and ICT (Information & Communication Technology) of MRO service for the sailing ship in 2015. MRO services is a business model that provides maintenance, repair, and operation services by using data on ship equipment conditions, classification inspection schedules, and equipment suppliers. Moreover, big data will be used for optimal navigation for energy efficiency and safety. Especially, the high added value is expected from service business for MRO service as a huge amount of various parts and materials are used in the shipbuilding industry comparing to other manufacturing industries. There are five types of MRO service for a ship (DSME, 2016). They are supply of consumable components, supply of necessity/spare part, supply of equipment and tool, supply of equipment and tool with repair service, and supply of ship inspection and repair service.

4.5.2. Data Analysis Description

The analysis composition in contract phase is in Figure 8. There are three major analyses on service in shipbuilding industry: MRO service, sailing on optimal route, and safety management. MRO service has three analysis methods which potential customer estimation, optimal MRO package analysis, and condition diagnosis analysis of ship equipment. Optimal sailing has two analysis methods, which are optimal sailing for energy efficiency and optimal sailing for safety. Finally, safety management has accident forecasting as an analysis method. The analysis method descriptions are summarized in Table 6. Each analysis method is described as the followings.

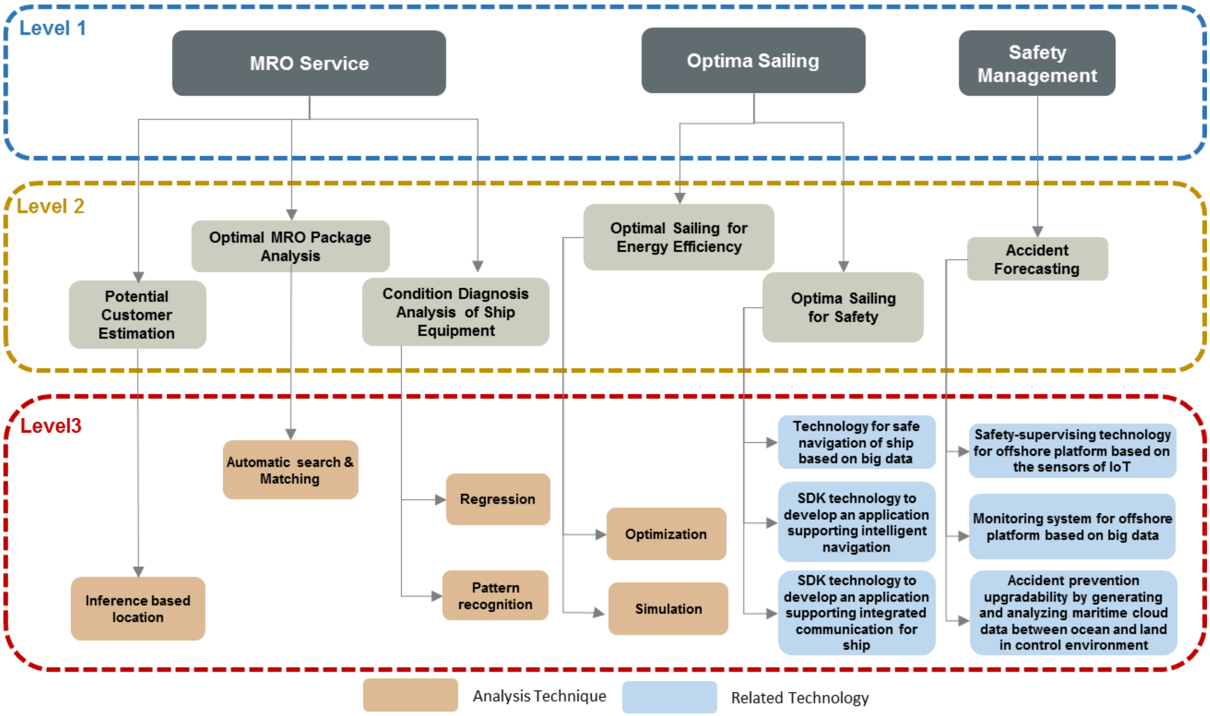

Figure 8. Analysis Composition of Service Phase

· *Potential Customer Estimation*

The most of MRO services are provided on the port because the sailing ship should enter the near port when the technical problem occurs on the ship. In the case that a certain port is the base for offering MRO service, the ship entering the port is regarded as the customer for the service. The potential customer is assumed from the ships entering the port before the ship arrives at the port through the potential customer estimation based on the data (DSME, 2016; S. Kim et al., 2016). The data includes AIS data (e.g., location of ship, speed, and course), ship classification data (e.g., schedule of classification survey), ship equipment data (e.g., type of equipment and usage period), and departure and arrival of port data. In terms of detailed algorithm, the ships, which have been the port or are supposed to get the port are extracted using AIS data and ship classification data. Then, the location range is determined based on an estimated time of arrival. Finally, the potential customer is decided from the ships on the range in the present time. The order for parts and repair service will be prepared in advance based on the analysis result. It makes sure the MRO service is supplied on time.

· *Optimal MRO Package Analysis*

One company cannot supply all the demand for MRO service and purchase of components because millions of the components are used in the shipbuilding and the type is very diverse. Therefore, many companies supplying each parts and equipment needs to cooperate. Necessary equipment and component will be supplied at the minimum cost on time through the optimal MRO package analysis (DSME, 2016; S. Kim et al., 2016). The data for the analysis are MRO service and supplier data, ship (customer) data, and equipment data. The suppliers which can satisfy the requirement of the customer will be automatically searched and matched using this data if the customer orders the equipment or the component. The selected suppliers propose the estimated cost for the order. Then, the package for the order is composed at the minimum cost. As the result, the customer can take the service at the optimized cost and time.

- *Condition Diagnosis Analysis of Ship Equipment*

The anticipative precautionary maintenance and repair by diagnosing the state of facilities is necessary to reduce the cost that repairs the facility after the failure. The worker or system that conducts the diagnosis for the facilities does not know the property of all the types of faults in reality. Thus, data driven approach is needed to find the characteristic of the fault types. Moreover, automatic classification will be conducted based on the result through condition diagnosis analysis of the ship equipment. The used data is equipment operation data such as vibration, noise, temperature, and pressure. Regression and pattern recognition are applied to the data for the analysis (Sedo Oh, Kim, Seo, Choo, & Monica, 2012). As the example of applying regression, RMS (Root Mean Square) of vibration acceleration is a statistics value for diagnosing the faulty type (S. Park, Sim, Lee, & Lee, 2003). It is a crucial factor for determining whether the engine has a fault. In addition, the environmental factor affecting the vibration of engine is temperature. The average of the RMS is predicted using regression that analyze the relation between average temperature per day and average RMS per day. The error range of the prediction is used to manage the fault. Pattern recognition is used to classify the type of fault when the fault is detected. The signal from the vibration is vectorized to increase the clarity of pattern. The fault is classified to the type according to the pattern that is categorized by nonparametric probability density function. As the result of the analysis, the anticipative precautionary maintenance and repair will be conducted to reduce the downtime (DSME, 2016; S. Kim et al., 2016).

- *Optimal Sailing for Energy Efficiency*

Economic sailing describes the sailing ship to minimized the fuel consumption using the weather prediction information. It is needed to change geographically the sea route and control properly the engine output for successful economic sailing. Economic sailing will be implemented through optimal sailing for energy efficiency using the data. The data includes climate data (e.g., wave direction, wave height, wind direction, and wind speed), geographical information data, and ship data (e.g., type of ship, specification, and speed). In terms of the detailed algorithm, optimization and simulation is applied to the data. For the example of the optimization, estimation equation for fuel consumption is derived using the geographical information data and the climate data. The optimization algorithm, A* algorithm, is used to find the geographically shortest route (Joo, Cho, Cha, Yang, & Kwon, 2012). Then, evolution strategy algorithm calculates the detailed sea route which minimized the fuel consumption based on the initial route (Bang & Kwon, 2014). Furthermore, these results are validated by simulation using the climate scenario (Y. Yoo, Choi, & Lee, 2015). As the result of the analysis, the sea route and the way to control the engine output for optimal energy efficiency is founded considering fuel consumption and distance.

· *Optimal Sailing for Safety*

The number of ship collision accidents has not been reduced despite the development of advanced navigation devices. The 30 percent of marine accidents are collision accidents, and over the 90 percent of the cause of the collision accidents are human errors. There are many difficulties for sailors to see various facilities and take actions for one by one. The essential cause of marine accident is lack of sailors` rest and burdens of managing facilities to sailors. From types of big data analysis and foundation technology by analyzing optimal sailing for safety, it is possible to support integrated vessel navigation and realize sailing for safety. There are three types of technology related to optimal sailing for safety. The first one is software development kit (SDK) technology which develops an application supporting integrated communication for vessels. This is the tool to support developing software for operating communication facilities for communication and sailing of the ship. As a software development package, this technology supports smooth data-exchanges between facilities such as automatic navigation system, automatic identification system for vessels, radar, and satellite communication system. The technology analyzes and defines the network and protocols for exchanging information among module, and classifies contents of information and data for exchanging information. By supporting exchanges of communication data of the ship between the ships and between the ship and the land in the wireless communication environment, it is possible to monitor the state of the ship and support the ship to sail safely.

The second one is SDK technology which develops an application supporting intelligent navigation technology. This technology supports developing the system that makes sailors set plans of navigation, monitor and control the vessel`s navigation, by collecting and integrating the various functions and information of equipment for navigation with safety. The technology supports developing software adjusting to complex management of the vessel, and the environment of navigation smoothly. In addition, this technology designs the algorithm for supporting the intelligent navigation technology. By integrating and managing the information of the entire facilities and systems in the ship, the technology supports the communication of information, and sets the plans of navigation by managing comprehensive information. In addition, by detecting the location and information of other ships from exchanging information, it is possible to prevent the collision between the ships, and improve the reliability on the automatic navigation system.

The last technology is the technology for safe navigation of ship based on the big data analysis. This technology includes three types: I) a technology for evading the collision by recognizing the state of the ships; II) a technology for setting the plans for avoiding based on the big data analysis considering various external factors and performance of the ship comprehensively; III) a technology for deriving equations of control and motion of the ship, and controlling the course of the ship. From these types of technology, the following can be developed: I) a system supporting the navigation for avoiding the collision based on automatic identification system (AIS) and big data analysis; II) a navigation technology for avoiding collision considering the ships. By applying the technology of navigation for avoiding collision based on AIS data and the information of sailing to the ship, it is possible to support to avoid collisions between the ships. In addition, because of the supports for avoiding collision of the ship with the data, it is possible to prevent the marine collision accidents, and improve convenience and safety for sailors.

- *Accident Forecasting*

For efficient utilization of limited space of the ship and offshore platform, various facilities are installed densely. Due to this reason, the possible small accident during the operation can be a significant accident causing marine pollution and loss of lives and properties. It is necessary to prevent and monitor these possible accidents. By utilizing big data analysis and the types of foundation technology, as well as analyzing not only the ship and offshore platform but also weather and marine state as the cause of the accidents, it is possible to accomplish safety supervision with integrated prediction of the accidents. Accident forecasting technology includes three types of technology. The first technology is safety-supervising technology for offshore platform based on the sensors of internet of things technology. This technology is for detecting the hazard factors and accidents in the ship, and comprehensively managing the safety of workers who are exposed to extreme working environment. It also is possible to prevent the accidents from enlarging, and detecting the possible disasters in the environment of construction/navigation of the ship. From the technology, it is possible to develop four types of technology: I) a real time location system (RTLS) technology with high efficiency and reliability for offshore platform; II) a gateway node with sensors for detecting possible disasters in the environment of construction/navigation of offshore platform; III) a wearable device with high efficiency and reliability for monitoring workers` health; IV) a navigator and simulator for evacuation and rescue based on the augmented reality technology. With the data from the sensors based on internet of things technology, it is possible to detect the possible disasters in the environment of construction/navigation of the ship. In addition, by monitoring workers` health, it is possible to minimize the loss of lives from the accidents, and assure the safety of the workers.

The second one is monitoring system for offshore platform based on big data. This system is for monitoring marine environments around the offshore platform, such as waves, ocean current, and winds. It also collects the data of facilities in the offshore platform, and controls them comprehensively. By monitoring marine environments, collecting data of the facilities in offshore platform and managing them comprehensively, it is possible to monitor the possible accidents during the operation. In addition, the loss of lives and properties, as well as the marine pollution problems because of the marine accidents can be reduced by monitoring accidents.

The last one is a technology for accident prevention upgradability by generating and analyzing maritime cloud data between ocean and land in control environment. It is for exchanging the maritime information between the ship and the controller, and between the ship owner and the controller, by collecting and integrating the real-time maritime information. It also becomes the control platform, which generates the database of maritime risk information including abnormal routes of navigation, and analyzes interconnectivities among the information. For realizing this technology, it is needed to develop two types of technology: I) a technology for upgradability of safety information exchange between the ship and the control center, considering the digital maritime communication system; II) a simulator for preventing accidents with high-speed automatic searching and playing/pausing on each section, for managing the after-accident. With this technology, it is possible to generate and analyze marine risk information considering maritime information, properties of the ship, depth of the sea, and maneuvering. The marine accidents can be prevented by exchanging the safety information between the ship and the control center. In addition, it is possible to manage the navigation of the ship more efficiently by exchanging the real-time information of the ship.

Table 6. Description of Analysis Method for Service Phase

MRO Service				
Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Potential Customer Estimation	· AIS data (e.g., location of ship, speed, course) · Ship classification data (e.g., schedule of classification survey) · Ship equipment data (e.g., type of equipment, usage period) · Arrival and departure of port data	· Inference based location	· List of potential customer to enter the port	(DSME, 2016; S. Kim et al., 2016)
Optimal MRO Package Analysis	· MRO service and supplier data · Ship (Customer) data · Equipment data	· Automatic search & Matching	· Supply of right time	(DSME, 2016; S. Kim et al., 2016)
Condition Diagnosis Analysis of Ship Equipment	· Equipment operation data (e.g., vibration, noise, temperature, pressure)	· Regression · Pattern recognition	· Diagnosis of faulty	(DSME, 2016; S. Kim et al., 2016; KISTEP, 2015; Y. Lee, Lee, Bae, Jang, & Lee, 2010; Sedo Oh et al., 2012; S. Park et al., 2003)
Optima Sailing				
Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Optimal Sailing for Energy Efficiency	· Climate data (e.g., wave direction, wave height, wind direction, wind speed) · Geographical information data · Ship data (e.g., type of ship, specification, speed)	· Optimization · Simulation	· Optimal route for energy efficiency · Optimal sailing way for energy efficiency	(Bang & Kwon, 2014; Joo et al., 2012; Y. Yoo et al., 2015)
Optima Sailing for Safety	· Related technology: SDK technology to develop an application supporting integrated communication for ship, SDK technology to develop an application supporting intelligent navigation, technology for safe navigation of ship based on big data			(KISTEP, 2015)
Safety Management				
Analysis Method	**Data**	**Detailed Algorithm**	**Result**	**Reference**
Accident Forecasting	· Related technology: safety-supervising technology for offshore platform based on the sensors of IoT, monitoring system for offshore platform based on big data, accident prevention upgradability by generating and analyzing maritime cloud data between ocean and land in control environment			(KISTEP, 2015)

V. Evaluation

5.1. Interview

In order to gather further empirical evidence for the reference model to evaluate the validity of the reference model, two interviews were conducted with the experts in shipbuilding industry. They are currently working for shipbuilding companies, which Korean major heavy industries. Their tasks are involved in applying big data analysis to the field of shipbuilding. Their departments are offshore system research, process support, and quality process. This formative evaluation was executed, consisting of two parts. In the first part, the domain experts were confronted with the reference model with the explanation and asked to provide feedback on the model. In the second part, they are surveyed to evaluate the importance of each analysis methods. Thy survey was conducted in the way that each analysis method was scored of scale from 0 (Not Important) to 7 (Highly Important). As the higher the value is, the analysis method indicates more important.

5.2. Results

The general opinion as the result of interview is that the reference model provides the proper guidance for big data analysis in the shipbuilding industry. The model considers the property of shipbuilding industry since it is developed based on the value chain. Furthermore, it is evaluated that it helps to generate the idea applying big data analysis in the field. The model is expected to play a role as the blueprint for applying big data analysis to the fields of the shipbuilding industry in the future. In terms of analysis method, visualization is emphasized. It is difficult to understand the current progress on the field since the scale of shipbuilding process is quite extensive. The hidden insight such as problems that are not recognized for the worker is expected to be discovered through the visualization. Furthermore, the survey result for the importance of the analysis methods shows the reference model is generally valid considering the total average (4.90) is higher than the midpoint (3.5) of the scale (0~7) in Table 7. The average and standard deviation of all the analysis method is 4.90 and 1.54. The nine analysis methods without score were added after the survey. This part will be explained in the discussion section. As for the importance of the phase, the production phase appears the most important phase for big data analysis. The six analysis methods among top 10 analysis methods are on the production phase. Specifically, forecasting of supply chain lead time is the most important analysis method in the result.

Table 7. Importance Assessment Result

Phase	Category of Analysis	Analysis Method	Avg.	Stdev.
Production	Supply Chain Management	Forecasting of Supply Chain Lead Time	6.33	0.78
Contract	Sales Support	Estimation of Initial Production Cost	6.08	1.08
Production	Production Process Management	Production Workload Analysis	6.08	0.90
Production	Supply Chain Management	Diagnosis of Abnormal Workflow of Supply Chain	5.92	1.31
Production	Production Process Management	Production Delay Section Analysis	5.83	1.27
Contract	Sales Support	Estimation of Initial Production Period	5.75	1.22
Production	Quality Management	Optimization of Welding Quality	5.75	1.29
Design	Detailed Design Support	Forecasting of Material Requirement	5.58	1.56
Design	Production Design Support	Estimation of Duration in Production Process	5.50	1.31
Production	Production Process Management	Visualization of Production Delay	5.50	1.17
Production	Production Process Management	Performance Analysis of Production Process	5.50	1.17
Production	Production Process Management	Visualization of Production Workflow	5.33	1.23
Production	Production Process Management	Diagnosis of Abnormal Workflow in Production Process	5.33	1.44
Production	Production Planning Management	Inspection of Changes to Production Process	5.25	1.36
Production	Safety Management	Safety Analysis of Worker	5.25	1.36
Production	Production Planning Management	Feedback Analysis of Worker	5.17	1.19
Production	Maintenance	Abnormality Detection of Manufacturing Facility	5.17	1.80
Production	Maintenance	Failure Type Analysis	5.08	1.62
Production	Safety Management	Environment Analysis of Shipyard Accident	5.08	1.24
Contract	Demand Forecasting	Potential Customer Estimation	5.00	1.76
Service	MRO service	Condition Diagnosis Analysis of Ship Equipment	5.00	1.60
Design	Design Process Management	Visualization of Design Process	4.92	1.68
Production	Maintenance	Planning for Maintenance	4.92	1.68

Design	Production Design Support	Optimization of Task Allocation in Production Design	4.83	1.59
Production	Maintenance	Risk Analysis of Failure	4.83	1.59
Design	Design Process Management	Knowledge and Information of Design Analysis	4.75	1.48
Service	MRO service	Potential Customer Estimation	4.67	1.61
Production	Maintenance	Schedule Planning for Maintenance	4.58	1.44
Production	Maintenance	Deficiency Analysis	4.58	1.62
Design	Basic Design Support	Forecasting of Ship Stability	4.33	1.83
Design	Basic Design Support	Forecasting of Fatigue Life Time of Ship Materials	4.33	1.37
Service	MRO service	Optimal MRO Package Analysis	4.25	1.42
Contract	Demand Forecasting	Technology Trend & Promising Technology Analysis	4.17	1.34
Contract	Sales Support	Estimation of Earnings & Success Rate	4.17	1.40
Production	Production Process Management	Connection Relation Analysis of Worker (Facility)	4.17	0.94
Service	Optimal Sailing	Optimal Sailing for Energy Efficiency	4.08	1.08
Design	Initial Design Support	Inference of Similar Case of Ship	4.00	1.04
Service	Safety Management	Accident Forecasting	3.92	1.88
Design	Design Process Management	Performance Analysis of Design Process	3.67	2.02
Service	Optimal Sailing	Optimal Sailing for Safety	3.58	1.31
Design	Design Process Management	Cooperation Network Analysis of Design	3.42	1.56
Production	Supply Chain Management	Performance Analysis of Supply Chain Process	-	-
Production	Supply Chain Management	Visualization of Delay Time of Supply Chain	-	-
Production	Supply Chain Management	Delay Section Analysis of Supply Chain	-	-
Production	Supply Chain Management	Supplier Workload Analysis	-	-
Production	Supply Chain Management	Logistics Optimization	-	-
Design	Design Process Management	Visualization of Design Delay	-	-
Design	Design Process Management	Design Delay Section Analysis	-	-

However, phases of design and services are assessed as less important phase for big data analysis in the shipbuilding industry. Most of the bottom 10 analysis methods are on these phase and their average score of importance is less than 4. Cooperation network analysis of design is the least important in that the average score of the method is the lowest. Each average score of bottom four analysis methods are less than 4. It appears that they are not expected to make a meaningful value by applying it. Such a result could occur since departments of the interviewees are related to production phase.

According to the importance, the company is able to utilize the big data analysis in order. The importance was assessed on the result of the survey. As the lower the value is, the analysis method has the higher importance. There are three analysis methods having the highest importance, which are forecasting of supply chain lead time, estimation of initial production cost, and production workload analysis. Moreover, there exists five analysis methods that importance was not determined since they are added to the reference model after the survey.

VI. Discussion

In this section, the implication of this study is discussed. The reference model developed by the literature survey was refined and validated by the interview. The value chain was used to categorize comprehensively the data analysis methods in shipbuilding industry. The data analysis methods are arranged according to each value chain to consider the property of the shipbuilding industry since the value chain provided a good overview of the organization or the industry. In addition, the value chain was effective to categorize the data analysis methods in that there was no suitable standard for classifying the data analysis methods in shipbuilding industry. The main tasks of the four phases on the value chain were categorized to the category of analysis in each phase. The categories of the other three phases except the production phase are based on the detail of the value chain. The category on the production phase seems to be not based on the detail of the value chain in Figure 2. However, the category includes all the activities of the production phase on the value chain. The category is applied to the crossed value chain. For example, quality management as the category of analysis is the common issue on unit assembly, grand assembly, forming, cutting, and others. Such detailed activities of the production phase on the value chain were actual production process of the ship. It would be more effective to categorize the categories of the analysis on the production phase depending on the division of the department in general manufacturing industry. In the view of the practitioner who want to apply the data analysis, such a categorization on the production phase is more familiar. In addition, the data analysis methods of the literature on the production phase were classified according to such a categorization.

There were principal feedbacks for the refinement of the reference model from the interview. First, category of analysis was subdivided. The initial reference model includes “Production Operation & Production Process Management” as the one of categories of the analysis in the production phase. However, it was divided into production operation management and production process management. Production operation management deals with the operation and planning for the production. It has relevance to the planning before actual production. On the other hand, production process management handles actual production process in progress. Second, analysis methods on supply chain management was complemented in that a few issues about supply chain management were mentioned. There were two analysis methods in supply chain management of the initial reference model, which are forecasting of supply chain lead time and logistics optimization. Supply chain management is crucial in the shipbuilding since the millions of components are used in shipbuilding and the delay problem may occur seriously on the whole process of shipbuilding due to the delay of component supply. Moreover, the needs for data analysis in supply chain management exist in the field. However, they actually have tried to apply data analysis in supply chain management. They commented that process mining technique would be applicable and proper to the data analysis for the supply chain management. As the result, the five analysis methods (performance analysis of supply chain process, visualization of delay time of supply chain, delay section analysis of supply chain, supplier workload analysis, and logistics optimization) were added in the supply chain management category.

The importance of analysis methods was evaluated by the survey. We don’t know whether some factors, such as cost, data availability, and others, were considered or not when the importance of the analysis method was considered. The interviewees were surveyed to evaluate how much the analysis method is important. The importance was regarded as the importance of the area that the data analysis is applied or the impact of the data analysis. As for the feasibility, the reference model does not consider some factors for the data analysis such as cost, data availability, and others. However, the reference model provides such a feasible data analysis method that it was already implemented in the literature. The literature involves in the data analysis in shipbuilding industry or similar areas which is applicable to shipbuilding industry. Furthermore, this study offers the insight for the data analysis to the practitioner in the shipbuilding industry. Generally, the practitioner in the shipbuilding industry has a difficulty in applying big data analysis since they do not know what value can be derived from which area by implementing the data analysis. This thesis provides the guide that which data analysis can derive on finding results for a certain area in the shipbuilding industry. This guidance will increase the understanding of the practitioner for applying the data analysis in the shipbuilding industry.

VII. Conclusion

This thesis developed the reference model for the big data analysis in the shipbuilding industry. The reference model provides the big data analysis guideline according to four phases such as contract, design, production, and service. These phases were categorized according to the value chain of shipbuilding industry to offer the practical guideline considering the property of the industry. Each phase is composed of three levels of data analysis, i.e., category of analysis, analysis method, and detailed algorithm. Moreover, each analysis method is described with data, detailed algorithm, analysis result, and related technology. The proposed reference model was refined and validated through the interviews with the experts in the shipbuilding industry. The importance of the analysis methods was determined based on the survey results in the interviews. The final reference model consists of four phases, fifteen categories of analysis, and forty-eight analysis methods.

There are three contributions for this paper. First, the reference model providing the guidance for big data analysis in shipbuilding industry is important in academia. The existing studies were weighed to the technology for big data analysis such as algorithm, architecture, analysis method, etc. Furthermore, a few studies for the way of applying the big data analysis in practice were at the level of the whole industry or the particular problem. This study has a value in that it provides the guidance in detailed about various areas in a certain industry. Secondly, this study contributes to increase the applicability of big data analysis in the shipbuilding industry. It helps the company to generate ideas about analyzing the data to derive a value. The guidance for data analysis is highly concrete in that it offers the data, analysis method, and the expected result. Additionally, the importance of analysis methods is provided. When the company introduces the big data analysis, the data analysis, which is expected to make a great ripple effect, will be preferentially selected. Lastly, the reference model will be utilized for the others as well as personnel in the shipbuilding industry. Even though the reference model was developed for the shipbuilding industry, the model may be used for others such as the companies related to big data including big data platform, data analysis service, and others. It will be helpful for the companies to get through to the area of big data in the shipbuilding industry. In addition, similar method can be considered as an introduction in this study to develop such a reference model in other industry.

There exists some limitation in the research despite the contributions of the paper. The model was developed by literature surveys and the interviews with the experts in related fields. Although the interview was conducted to validate the reference model considering both qualitative and quantitative aspects, the number of interviewees was only twelve. More interviews are needed to guarantee the validity of the reference model. Besides, four phases were considered in this study according to the value chain. However, the value chain of shipbuilding industry can be subdivided in detail. The reference model could be extended according to the subdivided value chain. Even though most of the analysis methods are from the literature, there exists some analysis methods which has not been verified. These analysis methods in the reference models are required to be implemented in practice. As for future work, more researches about the way to apply the big data analysis in shipbuilding industry should be conducted. For the big data analysis, there are some prerequisites such as data acquisition, processing, storage, and management. These activities appear in the value chain of big data. Although this study only focused on the data analysis, it could be applied to developing a reference model on the value chain of big data, from data generation to data analysis, in the shipbuilding industry. Similar studies in other industries are necessary. With the development of analysis technique, the reference model is required to be continuously updated. The range of the analysis will be extended as currently impossible analysis will be applicable in the future.

REFERENCES

Ahlemann, F. (2009). Towards a conceptual reference model for project management information systems. *International Journal of Project Management*, *27*(1), 19-30.

Aramja, A., Kamach, O., & Chafik, S. (2015). Creating a baseline model of Manufacturing Execution Systems using Petri nets : Literature overview. *Xème Conférence Internationale : Conception et Production Intégrées*.

Bang, S., & Kwon, Y. (2014). Economic Ship Routing System by a Path Search Algorithm Based on an Evolutionary Strategy. *The Journal of The Korean Institute of Communication Sciences, 39*(9), 767-773.

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.

Chen, M., Mao, S., Zhang, Y., & Leung, V. C. M. (2014). Big Data Analysis. *Big Data: Related Technologies, Challenges and Future Prospects* (pp. 51-58). Cham: Springer International Publishing.

Cho, I., & Kim, N. (2011). Recommending Core and Connecting Keywords of Research Area Using Social Network and Data Mining Techniques. *Journal of Intelligence and Information Systems, 17*(1), 127-138.

Davenport, T. H. (2006). Competing on analytics. *harvard business review, 84*(1), 98-107.

Delatte, B., & Butler, A. (2003). An object-oriented model for conceptual ship design supporting case-based design. *Marine Technology, 40*(3), 158-167.

DSME. (2016). Developmet of Ship Product Based on Big Data and MRO Service for Ship. *Plant conference*.

ECMiner. (2015). Prediction of Material Requirement and Man Hour for Shipbuilding

Retrieved from http://ecminer.com/sample/%ec%84%b1%ea%b3%b5%ec%82%ac%eb%a1%80/%ec%a0%9c%ec%a1%b0%eb%b6%84%ec%95%bc-%ed%85%8c%ec%8a%a4%ed%8a%b8-%ed%8e%98%ec%9d%b4%ec%a7%80/

Fan, J., Han, F., & Liu, H. (2014). Challenges of Big Data Analysis. *National Science Review, 1*(2), 293-314.

Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70*(5), 849-911.

Fettke, P., & Loos, P. (2004). Reference Modeling Research. *WIRTSCHAFTSINFORMATIK, 46*(5), 331-340.

Fisher, D., DeLine, R., Czerwinski, M., & Drucker, S. (2012). Interactions with big data analytics. *interactions, 19*(3), 50-59.

Gandomi, A., & Haider, M. (2015). Beyond The Hype: Big Data Concepts, Methods, and Analytics. *International Journal of Information Management, 35*(2), 137-144.

Ham, D., Lee, Y., & Woo, J. (2016). Big Data Based Simulation for Fittings Supply Management in Shipyard. *Korean Institute of Industrial Engineers Joint Spring Conference*, 3142-3149.

Hao, W., Osen, O. L., Guoyuan, L., Wei, L., Hong-Ning, D., & Wei, Z. (2015). Big Data and Industrial Internet of Things for The Maritime Industry in Northwestern Norway. *TENCON 2015 - 2015 IEEE Region 10 Conference*, 1-5.

Hwang, H., Kim, B., Shin, I., Song, S., & Nam, G. (2016). A Development of Analysis System for Vessel Traffic Display and Statistics based on Maritime-BigData. *Journal of the Korea Institute of Information and Communication Engineering, 20*(6), 1195-1202.

Hwang, W., Kim, J., Jang, W., Hong, J., & Han, D. (2008). Fault Pattern Analysis and Restoration Prediction Model Construction of Pole Transformer Using Data Mining Techniques. *The transactions of The Korean Institute of Electrical Engineers, 57*(9), 1507-1515.

Jiang, L., Bastiansen, E., & Strandenes, S. P. (2013). The International Competitiveness of China's Shipbuilding Industry. *Transportation Research Part E: Logistics and Transportation Review, 60,* 39-48.

Joo, S., Cho, T., Cha, J., Yang, J., & Kwon, Y. (2012). An economic ship routing system based on a minimal dynamic-cost path search algorithm. *KIPS Transactions on Computer and Communication Systems, 1*(2), 79-86.

Kalagnanam, J., Lee, Y. M., Ide, T., & Han, H. (2015). Manufacturing Data Analytics System Library Supporting Optimal Resource and Operation Managment. *KEIT (Korea Evaluation Institute of Industrial Technology), 15*(9), 33-51.

KBIZ. (2012). Shipbuilding Process. Retrieved from http://www.kosic.or.kr/2012/info/mateala_sc5.asp

Kim, P., Hong, J., & Koh, S. (2014). Big Data Technology R&D Trend through Patent Analysis. *Electronics and Telecommunication Trends, 29*(2), 33-41.

Kim, S., Nam, M., & Sun, M. (2016). Global Big Data Fusion Casebook. *K-ICT Big Data Center*.

Kim, Y., Park, C., & Oh, K. (2005). Cost-Estimation Support of Make-To-Order Production System using Data Mining. *Korean Institute Of Industrial Engineers*, 526-529.

KISTEP. (2015). ICT Fusion Industry 4.0 (Marine Shipbuilding Industry). *Korea Institute of S&T Evaluation and Planning*.

Kwak, J., & Kim, C. (2013). Adaptive Clustering Based k-nearest Neighbor Algorithm for Process Fault Detection. *Korean Institute Of Industrial Engineers*, 1169-1175.

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. *META Group, 6*.

LaValle, S., Hopkins, M. S., Lesser, E., Shockley, R., & Kruschwitz, N. (2010). Analytics: The New Path to Value. *MIT Sloan Management Review, 52*(1), 1-25.

Lee, D., Choi, J., Park, B.-J., Kang, H.-J., & Lim, S.-N. (2009). A Study on the Damage Safety Assessment in Ship Design Stage. *Journal of the Society of Naval Architects of Korea, 46*(3), 343-350.

Lee, S., Kim, B., Huh, M., Park, J., Kang, S., Cho, S., . . . Lee, D. (2014). Knowledge discovery in inspection reports of marine structures. *Expert Systems with Applications, 41*(4), 1153-1167.

Lee, Y., Lee, K., Bae, S., Jang, H., & Lee, J. (2010). Defect Detection and Defect Classification System for Ship Engine using Multi-Channel Vibration Sensor. *The KIPS transactions, a17*(2), 81-92.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. *McKinsey Global Institute*.

Min, K. Y., Kim, H. T., & Ji, Y. G. (2014). A Pilot Study on Applying Text Mining Tools to Analyzing Steel Industry Trends: A Case Study of the Steel Industry for the Company “P”. *Journal of Society for e-Business Studies, 19*(3), 51-64.

Oh, J., Park, J., Kim, K., & Jung, M. (2014). Distributed Network Environment Construction for Ship Design Data Management System. *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, 148-149.

Oh, K., & Park, C. (2005). Process Planning Method under Make-to-Order Production System using Data Mining. *IE interfaces, 18*(2), 148-157.

Oh, S., Kim, Y., Seo, H., Choo, G., & Monica, C. (2012). Diagnose Technique Study Applying Data Mining and Pattern Recognition Classification. *Korean Institute Of Industrial Engineers*, 1531-1550.

Oh, S., & Lee, B. (2015). The Accident Prediction Mechanism Using Maritime Big Data. *Proceedings of the 2015 Winter Conference on KIISE*, 960-961.

Park, M., Song, M., Baek, T. H., Son, S., Ha, S. J., & Cho, S. W. (2015). Workload and Delay

Analysis in Manufacturing Process Using Process Mining. In J. Bae, S. Suriadi, & L. Wen (Eds.), *Asia Pacific Business Process Management: Third Asia Pacific Conference, AP-BPM 2015, Busan, South Korea, June 24-26, 2015, Proceedings* (pp. 138-151). Cham: Springer International Publishing.

Park, S., Sim, M., Lee, H., & Lee, M. (2003). Design of Condition Based Maintenance Expert System using FFT Algorithm. *Korea Information Science Society, 30*, 514-516.

Park, W., & Hwang, S. (2016). Big Picture of Trend in 2016. *KT Economic Research Institute*.

Patil, A., & Giffi, C. (2015). Big data and Analytics in The Automotive Industry. *Deloitte*.

Sagiroglu, S., & Sinanc, D. (2013). Big Data: A Review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 42-47.

Sandryhaila, A., & Moura, J. M. F. (2014). Big Data Analysis with Signal Processing on Graphs: Representation and processing of massive data sets with irregular structure. *IEEE Signal Processing Magazine, 31*(5), 80-90.

Schütte, R. (1998). *Grundsätze ordnungsmäßiger Referenzmodellierung*: Gabler Verlag.

Sohn, T. (2011). A Study on Associations among Number of Bidders, Contract Award Rate and Profitability on International Construction. *JOURNAL OF THE KOREAN SOCIETY OF CIVIL ENGINEERS D, 31*(2D), 247-253.

Son, M., & Kim, T. (2012). *Optimized Task Allocation for Production Desgin Manager of Body of Ship Using BPM*. Paper presented at the Korean Institute Of Industrial Engineers.

Song, H. J., Park, K. S., Jung, H. E., & Song, M. (2013). Trend Analysis of Korean Economy in the Economic Literature by text mining techniques. *Korea Society for Information Management*, 47-50.

Sowar, N., & Gromley, K. (2011). Sharper View: Analytics in The Global Steel Industry. *Deloitte*.

Thomas, O. (2006). Understanding the Term Reference Model in Information Systems Research: History, Literature Analysis and Explanation. In C. J. Bussler & A. Haller (Eds.), *Business Process Management Workshops: BPM 2005 International Workshops, BPI, BPD, ENEI, BPRM, WSCOBPM, BPS, Nancy, France, September 5, 2005. Revised Selected Papers* (pp. 484-496). Berlin, Heidelberg: Springer Berlin Heidelberg.

Tsai, C.-W., Lai, C.-F., Chao, H.-C., & Vasilakos, A. V. (2015). Big Data Analytics: A Survey. *Journal of Big Data, 2*(1), 21.

Watson, I. (2001). Knowledge Management and Case-Based Reasoning: A Perfect Match? *FLAIRS Conference*, 118-122.

Yoo, C., Kim, K., Suh, Y., Shim, Y., Ha, Y., You, W., & Choi, H. (2012). An Experimental Study on Fatigue Life Evaluation of Welded Joints under Storm Loading. *Journal of the Society of Naval Architects of Korea, 49*(1), 99-108.

Yoo, Y., Choi, H., & Lee, J. (2015). Comparative Results of Weather Routing Simulation. *Journal of the Society of Naval Architects of Korea, 52*(2), 110-118.

Yoon, H., & Zhang, J. (2008). Evaluation for Probabilistic Distributions of Fatigue Life of Marine Propeller Materials by using a Monte Carlo Simulation. *Transactions of the Korean Society of Mechanical Engineers - A, 32*(12), 1055-1062.

Younus, M., Hu, L., Yong, Y., & Yuqing, F. (2009). Realization of Manufacturing Execution System for a Batched Process Manufacturing Industry. *Proceedings of the International Multi Conference of Engineers and Computer Scientists 2009, 2*, 1337-1341.

Zhang, H., Liu, L., Wang, X., & Gruen, J. R. (2007). Guideline for Data Analysis of Genomewide Association Studies. *Cancer Genomics-Proteomics, 4*(1), 27-34.

Appendix

Website (http://www.kosbis.org) is available for the contents of the reference model.

Figure 9. Website for A Reference Model