Development of marine logistic database using Automatic Identification System (AIS) and statistical data

Conference Paper · October 2016

3 authors, including:

Mohammad Danil Arifin
Universitas Darma Persada

Proceeding of the 30th Asian-Pacific Technical Exchange and Advisory Meeting on Marine Structures (TEAM 2016)
Mokpo, Republic of Korea, 10-13 October 2016

Development of marine logistic database using Automatic Identification System (AIS) and statistical data

Mohammad Danil Arifin[1*], Kunihiro Hamada[2], Noritaka Hirata[3], Kazetaro Ouchi[4]
1) Doctoral Student, Graduate School of Engineering, Hiroshima University, Japan*
e-mail: d145238@hiroshima-u.ac.jp
2) Professor, Graduate School of Engineering, Hiroshima University, Japan
e-mail: hamada@hiroshima-u.ac.jp
3) Assistant Professor, Graduate School of Engineering, Hiroshima University, Japan
e-mail: nhirata@naoe.hiroshima-u.ac.jp
4) Master Student, Graduate School of Engineering, Hiroshima University, Japan
e-mail: m163539@hiroshima-u.ac.jp

Abstract
Nowadays, the use of Big Data has been implemented and studied in various fields. In marine industries, Big Data such as port data, ships data, routes data, international trade data and AIS (*Automatic Identification System*) data are recently available. However, the collection of data groups is difficult to organize and frequently redundant. This is why the database can be so important. In consideration of the condition above, the authors have been developing MLDB (*Marine Logistic Database*) and it is aimed at supporting the basic planning of ship in the future. This database represents the large collective data set assembled and provides information which is related to logistics distribution. The database under study consists of sets of the latest information, i.e. operation information from AIS data, ships, port, routes and international trading information. As a result, relational database model had been developed. Overview of the MLDB structure is shown. Estimation methods of cargo type of bulk carrier and crude oil tanker are taken as an example of the effectiveness of the database. The evaluation of the proposed methods is shown. Moreover, data extracted from database application which necessary for the basic planning of ship development are shown and discussed.

Keyword: Big Data; MLDB (*Marine Logistic Database*); basic planning; bulk carrier; crude oil tanker

1. INTRODUCTION

Recently, Big Data is paying attention and getting popular in the marine industries. The future competitiveness of marine industries will be affected by how rapidly we take advantage of it. Big Data in marine industries is providing information that can make a maritime operation more efficient [1]. Moreover, it is widespread believed that Big Data can aid in improving forecasts provided that we can analyze and discover hidden patterns [2]. Furthermore, Big Data which is effective for forecasting demand and planning process is the oil of the information economy [3]. By identifying the real information extracted from Big Data, we can take the advantage of its full value to help organizations to be more efficient and profitable. The other research found that Big Data has a big potential to improve operational efficiency in shipping [4].

A popular definition of Big Data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization [5] [6]. The characteristics of Big Data are defined as the three Vs (volume, velocity, and variety) [6]. They can be further described as follows:

- Volume: refers to the amount of data and there are many factors that can contribute to the volume increase in data. It could amount up to hundreds of terabytes or even petabytes of information generated for everywhere [7].
- Velocity: refers to the rate change and about combining data sets that are coming with different speeds [8].
- Variety: from an analytic perspective, a variety of data is the biggest challenge to be used [9]. Some researchers believe that taking the data variety and volatility is the key to Big Data analysis [10].

In fact, data are frequently redundant. In order to organize and manage them, the database is important to be developed.

In consideration of available data and in order to organize and manage a huge amount of Big Data which is important for marine industries, the objectives, and target of ships of this study are described as follows:

(1) Paying attention into Big Data which is effective for basic planning and demand forecast of ships.
(2) Developing MLDB conducted by cleaning and integrating Big Data.
(3) Extracting data for supporting the basic planning of ships by using MLDB.
(4) The target of ships is focused on bulk carrier and crude oil tanker.

2. METHODOLOGY

2.1 Big Data in Marine Industries

Recently, a huge amount of Big Data potentially available within the marine industries, considering the number of ships carrying vast amounts of goods to and from the numerous ports worldwide. For example, Big Data such as voyage data, machinery data, weather data, business data, trouble data, AIS data, port data, ships data, route data and trade data are available in the marine industries. To date, only a fraction of this information is used [11].

Basically, that huge amount of Big Data is potentially important within the marine industries and become useful knowledge bank if handled carefully. Hence, big companies are largely investing in the research and harnessing on Big Data [12]. Based on [4] [13] Big Data application areas in marine industries are shown in ***Table 1***.

Table 1. *Big Data application areas in marine industries*

Role	Function	Example of Big Data application
Ship operator	Operation	▪ Energy saving operation ▪ Safe operation ▪ Schedule management
	Fleet Planning	▪ Fleet allocation ▪ Service planning ▪ Chartering
Ship owner	Technical management	▪ Safe operation ▪ Hull & propeller cleaning ▪ Condition monitoring and maintenance ▪ Environmental regulation compliance ▪ Energy saving retrofit
	New building	▪ Design optimization

The actual examples of Big Data application in marine industries are shown as follows:

(1) Operation optimization with an in-service ships performance model.
By combining the ships performance model with past weather records data, contributes to optimization ship's operation [4] [13] [14].
(2) Optimum weather routing with performance monitoring.
Ship model and weather forecast are inherently included errors. But the feedback loop by monitoring can make this system work better [4] [14].
(3) High accuracy of shape measurement method for actual shipbuilding component.
High accuracy of shape measurement method by using a 3D laser scanner and point cloud is proposed [15]. The results give suggestions for having the possibility of applying this proposed method for the actual Shipbuilding component.

From the explanation above, Big Data has a big potential in marine industries, especially in ship's operation and ship construction.

2.2 Basic Concept of Support System using MLDB

Generally, Big Data is very effective in understanding the demand of products [3]. However, in the shipbuilding, the use of Big Data in ships operation and ship construction are paid attention to [13][15], and few types of research pay attention to the use of Big Data in basic planning. Therefore, this study pays attention to basic planning. The basic concept is shown in the ***Fig. 1***. As shown in the ***Fig. 1***, Big Data on the ship, port, route, operation, and trade are integrated as MLDB by using SQL server in order to make a relation among Big Data.

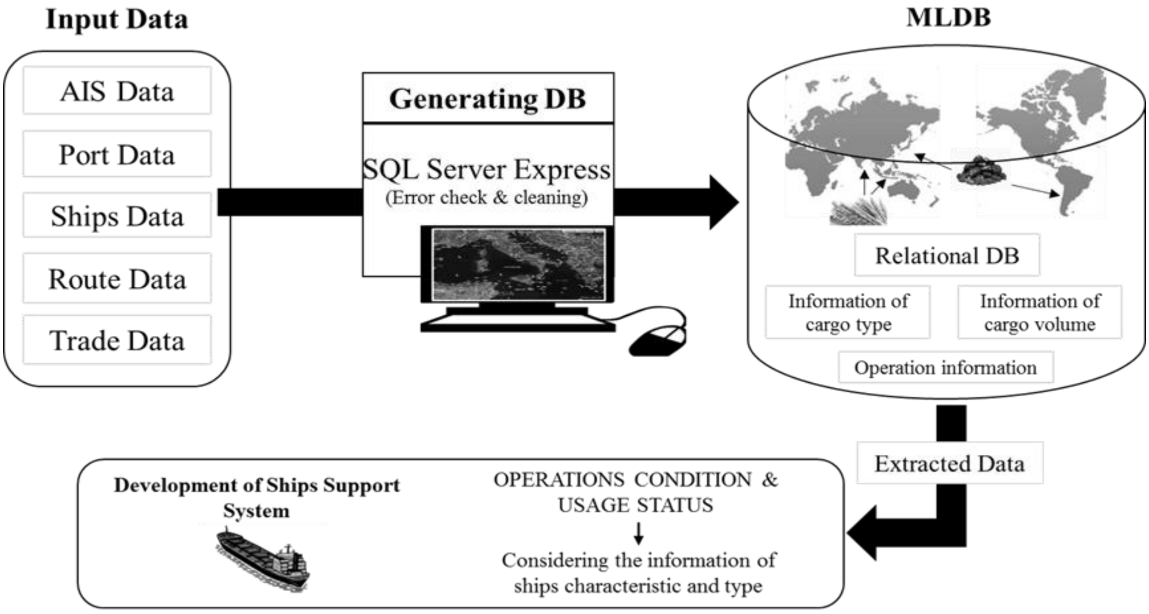

Fig. 1 *Basic concept of support system of ships*

Insightful information extracted from relational database MLDB are used for supporting the basic planning of ships. By utilizing information from MLDB, ships characteristic which is important for ship development could be identified and analyzed. In order to extract valuable information from MLDB for supporting basic planning, the following requirements should be considered.

(1) Data structure of MLDB should be defined in order to make a relation between Big Data (operation data, ships data, port data, route data and trade data) integrated into the relational database.
(2) In order to forecast the demand, generation of cargo information such as cargo type and cargo volume is required. Moreover, such information is also important for supporting the basic planning of ship.
(3) Generally, Big Data includes some errors. Therefore, an error cleaning is required in order to remove some errors to keep the quality of data.

2.3 Input Data

In this study, input data of MLDB are described as follows:

- ***AIS Data*** is an automatic tracking system used on ships to identify and locate vessels by electronically exchanging data. In this study, AIS data is collected from Market Intelligence Network (MINT) [16]. The information is provided by AIS data e.g. indicated speed, indicated draft, ships position, arrival and departure date, arrival and departure port, etc.
- ***Port Data*** refers to port information worldwide. In this study, port information is collected from Sea-web Port [17]. The information provided by port data e.g. port name, port number, longitude, latitude, port dimension, cargo handling, etc.
- ***Ships Data*** refers to ship information worldwide. In this study, ships information is collected from Sea-web Ships [18]. The information provided by ships data e.g. ships name, DWT, IMO number, classification, ships dimension, operators, shipbuilders, ships status, years of build, etc.
- ***Route Data*** refers to route information worldwide. In this study, route information is collected from Sea-web Port and IHS-Fairplay [17] [19]. The information provided from route data e.g. distances, route, etc.
- ***Trade Data*** refers to international trade statistics data detailed by commodities and partner countries. In this study, international trade data is collected from UN Comtrade [20]. The information provided by trade data e.g. commodity trade, a period of trade between country, code, trade value ($), trade quantity, reporter, and partner etc.

2.4 Data Structure

The structure of MLDB in this study is shown in ***Fig. 2***. As shown in the figure below, all of the input data are organized and connected to each other as a relational database. The relational concept is facing challenges in handling Big Data and providing horizontal scalability, availability, and performance required by Big Data applications [10].

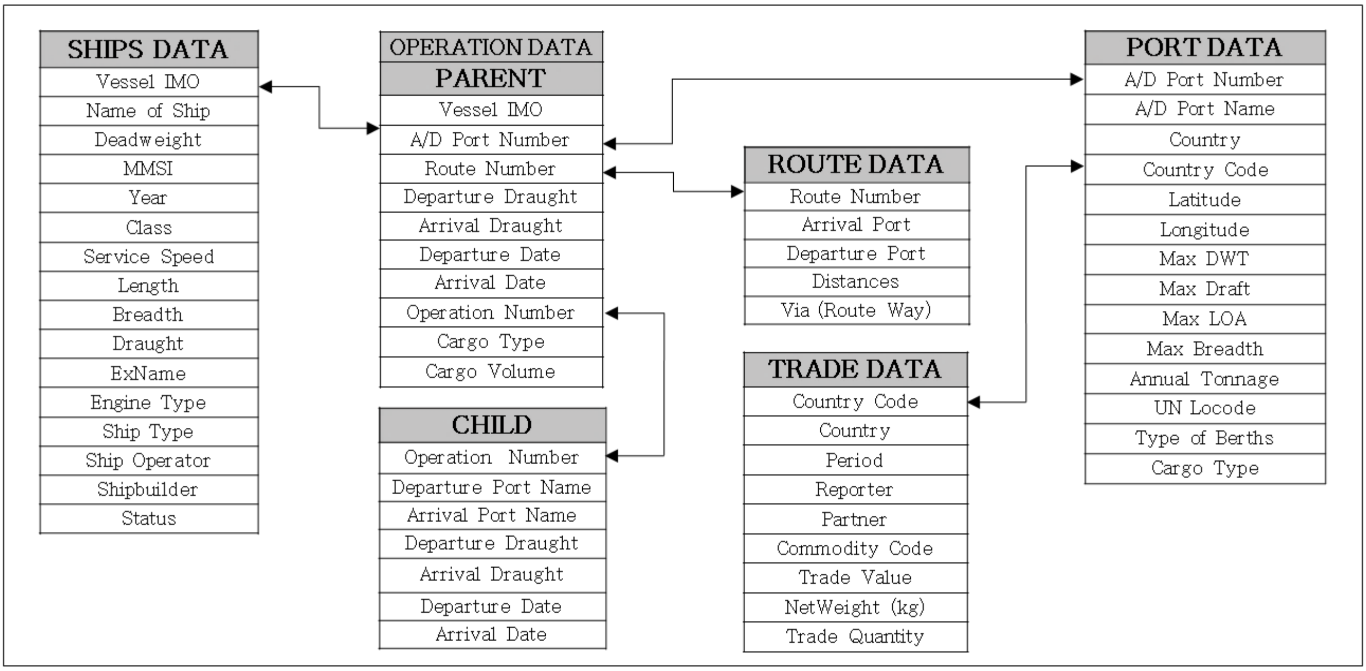

Fig. 2 *Structure of MLDB*

In this study, the effort of integration and relation among data are aimed to ensure that valuable knowledge from huge amounts of Big Data can be found and easy to extract. For example, by integrating ships data and routes data and utilizing the extracted data from it e.g. by considering speed and distances, the actual speed of ships can be identified. Moreover, by integrating ships data and port data with operation data, some information related to the actual information of ships operation state can be analyzed (*berthing*, *anchoring* or *sailing*).

A relational database is effective for organizing data in tables (or relations). The relationships that can be created with the tables enable a relational database to efficiently store huge among of data and effectively retrieve selected data. The relational database design process is described in the following steps:

(1) Step 1. Define the data structure as shown as in ***Fig. 2***
(2) Step 2. Gather data, organize in tables and specify the Primary Keys (PK)
(3) Step 3. Create relationships between tables

2.5 Data Cleaning

The AIS is a VHF (*Very High Frequency*) radio tracking system that is used by ships and vessel traffic services to identify and locate vessels by electronically exchanging data with other nearby ships, AIS base stations, and satellites. In high-density shipping area where thousands of ships may transmit AIS messages, it is a challenge for the AIS system to efficiently collect, process, and download all the messages. It results that many messages are lost and sometimes error data collection occurred. Invalid information also occurred due to human factor during input data to AIS [21]. The samples of errors from AIS data are shown as follows:

(1) Draft (d) information is not changed during operation (arrival & departure draft), as shown as in ***Table 2.***
(2) The value of draft (d) information is zero (0).
(3) Unrealistic of ships movements (too fast movement).
(4) Duplication information (duplication record of arrival port & departure port)
(5) Null information or blank space.

Table 2. *A sample of error from NS. CENTURY*

Port Name	Country	Arr. Date	Dept. Date	Arr. Draft (m)	Dept. draft (m)
Beaumont	USA	2015/7/26 0:53	2015/7/27 5:48	11.5	11.5
Europoort	Netherlands	2015/7/29 1:37	2015/7/29 1:37	11.5	11.5
.	.	.	.	.	.
.	.	.	.	.	.
Houston	USA	2015/9/4 18:55	2015/9/8 12:55	12.7	0
Houston	USA	2015/9/4 18:55	2015/9/8 0:252	12.7	0

Therefore, data cleaning of such errors is required. In this study, error cleaning is conducted by eliminating duplicate and NULL information, checking draft information between arrivals, departure and so on.

2.6 Generation of Cargo Information

Cargo information on operating ship is important for demand forecasting and understanding the use case of a ship. However, such information does not exist in AIS data and so on. Therefore, cargo type and cargo volume of each operation are estimated in this study. In the case of a bulk carrier, cargo type is selected from three types of cargos i.e. iron ore, coal, grain & others. In the case of the tanker, cargo type already exists in ship data. Therefore, only cargo volume is estimated.

The estimation methods in this study are shown as follows:

(1) ***Step 1***. Estimation of cargo type using port data.
As shown in the ***Fig. 2***, port data include the cargo type which is handled at the port. By using this data, cargo of each operation is estimated. Some examples are shown in ***Table 3***. As shown in ***Table 3***, estimation of cargo type is conducted by checking the combination of cargo from arrival port and departure port. In the case of operation from Newcastle to Fukuyama, common cargo is only "Coal". Therefore, cargo type is estimated as "Coal". On the other hand, in the case of operation from Newcastle to Chiba, there are two common cargos, i.e. Coal and Grain & Others. In such a case, cargo type is decided by executing ***Step 2***.

Table 3. *Checking combination of cargo using matrix*

	Fukuyama (Iron ore, Coal)	Chiba (Iron ore, Grain, Coal & Others)	Higashi-Harima (Iron ore and Coal)
Newcastle (Aus) (Grain, Coal, Others)	Coal	Coal, Grain & Others	Coal
Parker Point (Iron ore, Others)	Iron ore	Iron ore, Others	Iron ore
Fremantle (Iron ore, Grain, Coal & Others)	Iron ore, Coal	Iron ore, Grain, Coal & Others	Iron ore, Coal

(2) ***Step 2***. Estimation of cargo type using ships size
By executing ***Step 1***, cargo type of 75% operation is fixed. ***Fig. 3*** shows the estimation result of ***Step 1***, from Australia to Japan. From the figure, it is understood that ship size and cargo type have a close relation. By using this data, cargo types of remained operations are estimated. In the case of operations from Australia to Japan, cargo types of remained ships are estimated as follows:
- if DWT (ton) >150.0000 cargo type is defined as "Iron ore",
- if 60.000≤ DWT≤150.000 cargo type is defined as "Coal",
- if DWT (ton) <60.000 cargo type is defined as "Grain & Others".

Where the threshold is defined in consideration of distribution between import country and export country.

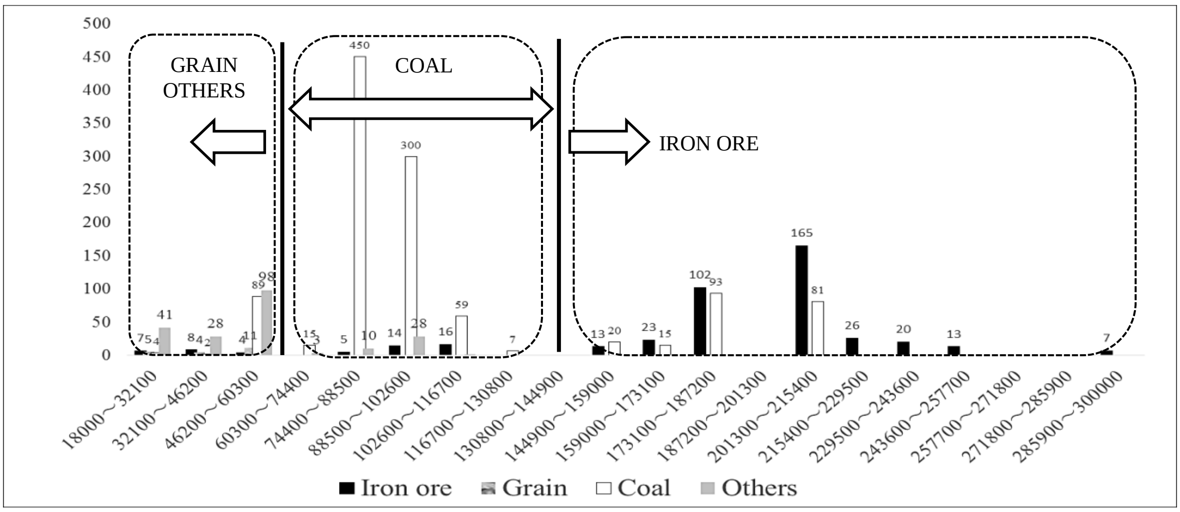

Fig. 3 *Distribution of DWT (ton) from Australia-Japan*

(3) ***Step 3***. Confirmation of reliability of data

The confirmation of the reliability of data is required in order to get a good result during estimating cargo volume. In this study, the confirmation of the reliability of data is done by checking draft rate $\boldsymbol{d_i}$ **(%)** of the ship during operation. Basically, $\boldsymbol{d_i}$ **(%)** is defined as following equation:

$$\boldsymbol{d_i}[\%] = \frac{d_{sail_i}[m]}{d_{max_i}[m]} \tag{1}$$

Where ***dsail***$_i$ ***(m)*** is sailing draft and ***dmax***$_i$ ***(m)*** is max draft of ship

As shown in ***Table 4.*** Operation condition of ships is defined as loading, ballast, and unknown condition.

Table 4. Operation condition of ship based on draft rate

Draft Rate [%]	Arrivals>Departure	Arrivals = Departure	Arrivals<Departure
100+	Unknown	Unknown	Unknown
65~100	Loading	Unknown	Unknown
45~65	Unknown	Unknown	Ballast
0~45	Unknown	Unknown	Unknown

(4) ***Step 4.*** Estimation of cargo volume

Ship data have information on deadweight and max draft of the target ship, while AIS data has information in the sailing draft. By using these, cargo volumes are basically estimated by using the following equation:

$$\boldsymbol{V_i[ton] = DWT_i[ton]} \times \frac{(d_i - 0.2)}{(1 - 0.2)} \tag{2}$$

Where $\boldsymbol{V_i}$ ***(ton)*** is cargo volume, $\boldsymbol{DWT_i}$ ***(ton)*** is deadweight and $\boldsymbol{d_i}$ **(%)** is the draft rate.

In the case of "*unknown*" condition, estimation of cargo volume is conducted by considering an average draft of fix ships which is operating on the same route and same size (DWT).

2.7 Evaluation of Cargo Information

In order to verify the result from proposed methods, the estimation result is evaluated and compared with the actual value of trade information from UN Comtrade. For the bulk carrier, estimation of cargo volume from Australia to Japan and Brazil to Japan are taken as an example. In the case of the tanker, estimation of cargo volume from The Middle East to Japan, Korea and China are taken as an example. Estimation result of the bulk carrier is shown in ***Table 5.*** The present result for coal is covered about 94% of total cargo reported into UN Comtrade. Moreover, the estimation result of the tanker is covered 99% of total cargo reported into UN Comtrade. Therefore, we can conclude that effectiveness of cargo estimation in this study is confirmed.

Table 5*. Estimation result of bulk carrier (Australia-Japan)*

CARGO TYPE	Estimation (t)	UN (t)	RATIO
Iron Ore	7.43×10^7	8.28×10^7	90%
Coal	1.12×10^8	1.19×10^8	94%
Grain & Others	1.11×10^7	1.15×10^7	97%
TOTAL	1.97×10^8	2.13×10^8	94%

Table 6. *Estimation result of tanker from Middle East to Japan, Korea & China*

Country	Estimation (t)	UN (t)	RATIO
Japan	1.62×10^6	1.68×10^6	96.5 %
Korea	1.30×10^6	1.30×10^6	100.3%
China	2.71×10^6	2.70×10^6	100.5%
TOTAL	4.63×10^6	4.68×10^6	99%

3. EXTRACTED DATA FOR SHIP PLANNING USING MLDB

The structure of the relational database in MLDB allows the user to get some valuable knowledge related to information for ship planning. In this study, the target of data extracted from MLDB is a bulk carrier (Australia-Japan) and (Brazil-Japan), crude oil tanker (The Middle East – Japan), (The Middle East – Korea) and (The Middle East – China) which the time period is 2014/01/01 ~ 2014/12/31. By utilizing valuable information from MLDB, characteristics of target ships are identified. The average value of the principal dimension of bulk carrier and crude oil tanker is shown in the ***Table 7.*** and ***Table 8***.

(1) Bulk Carrier

The characteristics of the bulk carrier from Australia-Brazil to Japan are shown in ***Table 7.*** As shown in ***Table 7***, for iron ore and coal commodity, shipments from Australia and Brazil to Japan particularly shipped by Capesize type. Where the DWT (10^4ton) LOA (m) and design speed (knot) are almost the same, but the value of B (m) and d (m) are different. However, in the case of grain commodity mostly shipped by Handymax size, which the size of ships from Australia to Japan is smaller than Brazil to Japan.

(2) Crude Oil Tanker

The characteristics of crude oil tanker briefly explained as follows. As shown in ***Table 8.*** Basically, the dimension of crude oil tanker B (m) and LOA (m) are same for all routes compared with existing crude oil tanker in the world i.e. B = 60 m and LOA = 333 m. The varied value is shown for DWT (10^4ton), D (m), design speed (knot) and horsepower (HP). Compared with others, DWT (10^4ton) and the draft (m) of a crude oil tanker from The Middle East – Japan are smaller. However, the actual speeds of crude oil tanker from The Middle East – Japan as shown in the ***Fig. 5***, are faster than others.

Table 7. *Characteristic of bulk cargo from Australia-Brazil to Japan.*

		DWT (10^4ton)	LOA (m)	B (m)	d (m)	Design Speed (knot)	Amount of Ships
Iron Ore	Australia	19.7	298.4	48.7	17.8	14.5	427
	Brazil	19.9	298.7	47.9	18.2	14.5	170
Coal	Australia	10.3	238.8	38.9	14.3	14.4	1164
	Brazil	-	-	-	-	-	-
Grain & Others	Australia	5.5	194.7	33	11.9	14.4	224
	Brazil	6.9	209.6	33	13.3	14.4	38

Table 8. *Characteristics of Crude Oil Tanker*

Route	DWT (10^4ton)	LOA (m)	B (m)	d (m)	Design Speed (knot)	Horse Power (hp)	Amount of ships
The Middle East - Japan	30.5	333	60	21	15.6	36.977	125
The Middle East - Korea	30.9	333	60	21.9	15.7	37.716	213
The Middle East - China	31.2	333	60	22.1	15.6	38.109	279
Existing Crude Oil tanker	30.8	333	60	21.9	15.6	37.336	-

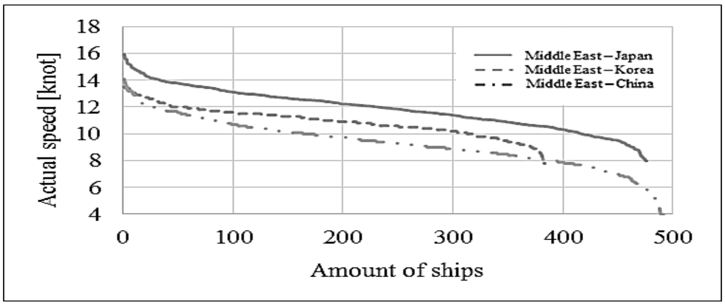

Fig. 5 *Comparison of actual speed from The Middle East to Japan, Korea & China*

From the explanation above, we can conclude that the characteristics of bulk carrier and crude oil tanker which important for the basic planning of ship development are could be easy to identify by utilizing information extracted from MLDB.

5. CONCLUSIONS

In this study, MLDB using AIS data and statistical data are developed. Integration and relation of Big Data and error cleaning, which important for MLDB are conducted. In order to extract valuable information for supporting the basic planning of ship, estimation methods of cargo type and cargo volume are proposed. The evaluation of the proposed methods is confirmed by comparing the estimated result by utilizing information from MLDB with the international trade value from UN Comtrade. Extracted data from MLDB which is important for supporting the basic planning of ships are shown and discussed.

REFERENCES

[1] Martin Kits van Heyningen., 2015. “Ship Insight-The Buzz of Big Data,” The Quarterly Journal |Summer.

[2] “Hossein H. and Emmanuel S.S., 2015. “Forecasting with Big Data: A Review,” Ann. Data. Sci. (2015) 2(1):5–19 DOI 10.1007/s40745-015-0029-9

[3] Charles W. Chase, Jr., 2013. ” Using Big Data to Enhance Demand-Driven Forecasting and Planning,” Journal of Business Forecasting, Volume 32, Issue 2, 2013.

[4] Hideyuki Ando, MTI (NYK group), 2016. “The Application of Big Data for Ship Operational Efficiency,” Japan –Norway Workshop Future Technology and Finance in the Maritime Sector, 12th February 2016.

[5] D. Laney., 2001. “3D Data Management: Controlling Data Volume, Velocity and Variety,” Gartner Inc.

[6] D. Laney., 2012. “The Importance of 'Big Data': A Definition”. Gartner. Retrieved 21 June 2012.

[7] Avita Katal, Mohammad Wazid and R H Goudar. 2013” Big Data: Issues, Challenges, Tools and Good Practices”, IEEE, 978-1-4799-0192-0/13, 2013.

[8] Sicular. 2013.”Gartner's Big Data definition consists of three parts” Forbes, 27 March 2013.

[9] Khalid. A.I H, Mohammed A.I F, Jasni M.Z and Mazlina A. M., 2015. “Big Data Analysis and Storage,” Proceedings of the 2015 International Conference on Operations Excellence and Service Engineering Orlando, Florida, USA, September 10-11, 2015

[10] Dandawate, Y., 2013. “Big Data: Challenges and Opportunities”. Infosys Labs Briefings. vol. 11, Infosys Labs.

[11] H. Segercrantz, “BIG DATA & Big Savings for Maritime Ops,” Available at <http://www.marinelink.com/news/maritime-savings-data405723.aspx>, [Accessed June 8th, 2016]

[12] S. Shukla, V. Kukade and A. Mujawar, “Big Data Concept, Handling and Challenges: An Overview,” International Journal of Computer Applications (0975 –8887) Volume 114 –No. 11, March 2015

[13] Hideyuki Ando, MTI (NYK group), “How to utilize Big Data and IoT in the shipping sector,” Japan –Norway Workshop Future Technology and Finance in the Maritime Sector, 12th February 2016.

[14] Hideyuki Ando, MTI (NYK group), 2015. “Optimizing fleet performance with smart ship management”, The IoT Show 2015 Asia22-23 September Suntec Convention& Exhibition Centre, Singapore, 22nd September 2015.

[15] Kazuo Hiekata, Hiroyuki Yamato, Jingyu Sun & Naoji Toki., 2014. “Study on Improving Accuracy for Edge Measurement Using 3D Laser Scanner”, IFIP Advances in Information and Communication Technology 442:427-434 · July 2014.

[16] Market Intelligence Network, “Automatic Identification System Data,” Available at <https://www.marketintelligencenetwork.com/app/>, Start using from 2015-present.

[17] IHS Maritime & Trade, “Sea-web Ports,” Available at <http://maritime.ihs.com/Areas/Seaweb/authenticated/authenticated_handler.aspx?control=portsearch&bp=1&type=clear>, Start using from 2015-present.

[18] IHS Maritime & Trade, “Sea-web Ships,” Available at <http://maritime.ihs.com/Areas/Seaweb/authenticated/authenticated_handler.aspx?control=shipsearch&bp=1&type=clear>, Start using from 2015-present.

[19] IHS Fairplay Ports & Terminals Guide, 2013

[20] United Nations Commodity Trade Statistics Database, “UN Comtrade,” Available at <http://comtrade.un.org/db/ce/ceSearch.aspx>, Start using from 2015-present.

[21] Rino B, Ljupko S and Zvonko K., 2012. ” Automatic Identification System in Maritime Traffic and Error Analysis,” Transactions on Maritime Service-TOMSs, Trans. marit. sci. 2012; 02:77-84.