

Exploiting Marine BD to Develop MLDB and Its Application to Ship Basic Planning Support

Mohammad Danil Arifin¹, Fanny Octaviani²

(Received: 20 December 2021 / Revised: 21 December 2021 / Accepted: 24 December 2021)

Abstract— recently, the global marine logistics industry has changed significantly because of the global movement of goods. Where the amount of available data and attention paid to extensive data analysis in various topics exponentially grows, it is possible to obtain vast amounts of marine BD. However, the collection of BD groups is difficult to organize and frequently redundant. This is why the database can be so important. If these BD are effectively utilized, great innovation can be achieved in the marine industry. In this study, we develop a marine logistics database to ship basic planning support in the future. The database under study consists of BD sets, i.e. port, ship, route, international trade, and ship operation information from automatic identification system data. As a result, the relational database was developed. The effectiveness of the database is evaluated and extracted data from the database necessary for ship basic planning is discussed.

Keywords—marine big data, MLDB, ship basic planning, ship allocation.

I. INTRODUCTION

Big Data (BD) is more than a word, and the business benefit of utilizing BD is widely understood. It is found that data-driven organizations perform 5% to 6% better per year [1]. BD is already playing a more significant role in shaping the maritime industry's future. By analyzing the data, the opportunity to drive improved efficiency and quality could be managed by the marine logistics players [2]. Moreover, it is widely known that BD can aid in improving forecasts, and BD can be adequate for forecasting demand and planning processes [3][4].

A popular definition of BD is high volume, high velocity, and high variety information assets requiring new processing forms to enable enhanced decision making, insight discovery, and process optimization [5] [6]. The characteristics of BD are defined as the three Vs (volume, velocity, and variety), as shown in **Figure 1**. It can be further described as follows:

- Data volume defines as the amount of data, and many factors can contribute to the volume increase in data. It could be the number of hundreds of tera/petabytes of information that is generated everywhere [7].
- Data velocity is defined as the generation of data that is rapid; acquiring, processing, and analyzing it requires fast mechanisms. The velocity emphasizes the real-time processing power of BD for enterprise needs [8].
- Data variety define as data types of BD, which includes structured and unstructured data such as text, audio, video, sensor data, posts, log files, and many more [8].

Significant potential and high value are hidden in the huge volumes of data widely used in various kinds of areas, including the maritime field. Because of the global movement of goods, global marine logistics has changed significantly. Hence, it is essential to develop ships that meet the specifications of the market needs. Simultaneously, marine logistics BD, e.g. port, ship, AIS data, etc., can be acquired more efficiently. If these data are effectively harnessed, great innovation might be obtained. Since BD are frequently redundant, to organize them, the database should be developed.

By considering the data available from marine BD, the objective of this study is to develop marine logistics database (MLDB) by exploiting marine BD and its application to ship basic planning support by extracting the data from MLDB. The bulk carriers which operate between Australia to Japan, Korea, and China are taken as an example. The effectiveness of the databases is discussed.

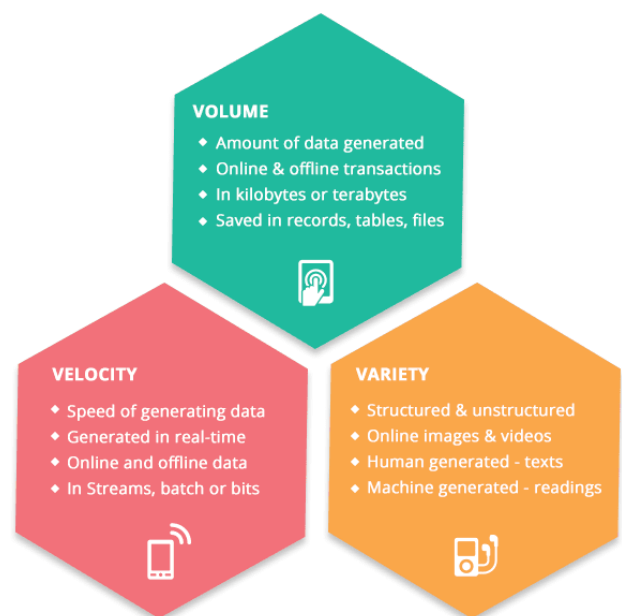


Figure 1. The 3Vs of Big Data

Mohammad Danil Arifin, Department of Marine Engineering, Darma Persada University, Jakarta, 13450, Indonesia.

E-mail: danilarifin.mohammad@gmail.com

Fanny Octaviani, Department of Naval Architecture, Darma Persada University, Jakarta, 13450, Indonesia.

E-mail: fanny_octaviani@yahoo.com

II. METHOD

A. BD in the marine field

Many studies have applied BD in the marine field. The high accuracy of block component measurement method for construction applications has been developed by using the point cloud data of the 3D scanner [9]. Aoyama et al. [10] proposed new methods of extracting and utilizing monitoring data by introducing two different monitoring technologies and considering the reliability of each for advanced shipbuilding construction management. Perera et al. [11] analyzed large ship performance datasets to propose a model for evaluating ship performance under various seagoing conditions in

the operations field.

Ando et al. [12] and Yoshida et al. [13] proposed a data collection platform called the Ship Information Management System. They utilized the data collected for many purposes (e.g., energy efficiency determinations, ship performance monitoring, and engine monitoring). MD Arifin et al. [14] [15] [16] develop a ship allocation model using marine logistics data that can forecast the demand for bulk carriers and examine the adequate principal particulars of ships for cargo transportation. Based on [12], BD application areas in the marine field are shown in **Table 1**.

TABLE 1.
BD APPLICATION AREAS IN THE MARINE FIELD

Role	Function	Example of BD application
Ship operator	Operation	<ul style="list-style-type: none"> ▪ Energy-saving operation ▪ Safe operation ▪ Schedule management
	Fleet Planning	<ul style="list-style-type: none"> ▪ Fleet allocation ▪ Service planning ▪ Chartering
	Technical management	<ul style="list-style-type: none"> ▪ Safe operation ▪ Hull & propeller cleaning ▪ Condition monitoring and maintenance ▪ Environmental regulation compliance ▪ Energy-saving retrofit
Shipowner	New building	<ul style="list-style-type: none"> ▪ Design optimization

Limited studies have employed BD to improve ship construction, operation, and performance; few have examined the use of BD for basic ship planning [14] [16] [17]. Therefore, the aim of this study is to develop ship allocation for ship basic planning support by using marine BD.

B. The basic concept of MLDB

Marine logistics database is developed by integrating marine BD, i.e. operation information from AIS data, ship, port, route, and international trading information, into a relational database. The illustration of MLDB and its application for ship basic planning is shown in **Figure 2**.

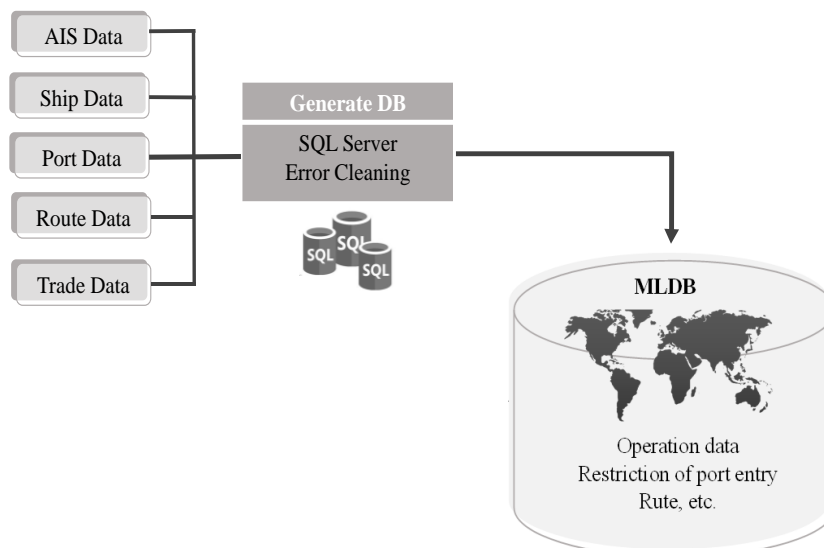


Figure 2. Basic Concept of MLDB

To extract valuable information from MLDB, the following requirements should be considered.

- The data structure of MLDB should be defined to make a relation between the data (operation, ships, port, route, and trade data) integrated into the relational database.
- An error cleaning is required to remove some errors to ensure the quality of the data.
- The cargo volume should be estimated to confirm the effectiveness of the database in this study

C. The basic concept of MLDB

The MLDB is developed by considering the following data:

- Automatic Identification System (AIS): indicated speed, indicated draft, ship position, timing arrival and departure dates, and arrival and departure port from MINT [18].
- Port data that is collected from Sea-web Port, e.g., port name, longitude, latitude, port dimension, and cargo handling [19].
- Ship data that is collected from Sea-web Ship, e.g., ship name, DWT, IMO number, ship classification, principal dimension, operator, shipbuilder, ship status, and build year collected [20].
- Route data that is collected from Sea-web Port and IHS-Fairplay, e.g., departure, arrival, route, and route distances between the port [19] [21].
- Trade data that were collected from UN Comtrade, e.g., commodity trade data, trade periods, trade value, commodity code, trade quantity, reporter, and partner [22].

D. The basic concept of MLDB

The structure of MLDB in this study is shown in **Figure 3**. As shown in the figure below, all input data are organized and connected as a relational database. The relational concept faces challenges in handling BD and providing horizontal scalability, availability, and performance required by BD applications.

In this study, the effort of integration and relation among data ensures that valuable knowledge from vast amounts of BD 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 and port data with operation data, some information related to the basic information of ships' operation state can be analyzed.

A relational database effectively organizes data in tables (or relations). The relationships created with the tables enable a relational database to efficiently store vast amounts of data and effectively retrieve selected data. The relational database design process is described in the following steps:

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

- Step 3. Create relationships between tables

E. Error cleaning

In a high-density shipping area where thousands of ships may transmit AIS messages, it is a challenge for the AIS system to collect, process, and download all the messages efficiently. It results that many messages being lost, and sometimes error data collection occurs. Below are the samples of errors from the Automatic Identification System, as follows:

- The draft value (d) is zero (0).
- Duplication information.
- Null information or blank space.

Therefore, to ensure the quality of the data, the duplicate and NULL data should be eliminated, and the draft data should be evaluated.

F. Generating cargo information

Cargo information on an operating ship is essential for forecasting demand and understanding ship usage. However, that information is unavailable in AIS data. Therefore, the cargo type and volume of each operation are estimated. The cargo type is selected from 3 types: iron ore, coal, grain & others.

- *Checking the data reliability*

Confirmation of data's reliability is required for a good cargo volume estimation. In our study, data reliability was evaluated by using Equation (1). Where d_i is defined as draft rate, $d_{sail(i)}$ (m) is defined as the sailing draft, and $d_{max(i)}$ (m) is defined as the maximum draft of the ship.

$$d_i = \frac{d_{sail(i)}}{d_{max(i)}}, \quad (1)$$

- *Checking the cargo type by using the information of the port*

The cargo of each operation was estimated by analyzing the cargo type from port data. As shown in **Table 2**, the estimation of cargo type was identified by checking the combination of arrival and departure ports cargo. For example, in operation from departure Port A to arrival Port D, the only common cargo is identified as coal. Therefore, the cargo type was estimated as coal. However, in the case of operation from departure Port B to arrival Port D, common cargos were identified as iron ore and coal.

In such a case, the cargo type was defined as multi-cargo and should be estimated by using the ship size

- *Checking the cargo type by using ship size information*

If the two or more common cargo types exist in port data, the cargo types were estimated using ship size. The distribution of ship size from Australia to Japan, Korea, and China for fixed cargo is shown in **Figure 4** to **Figure 6**. Since the ship size and cargo type are closely related, the remaining operation could be estimated.

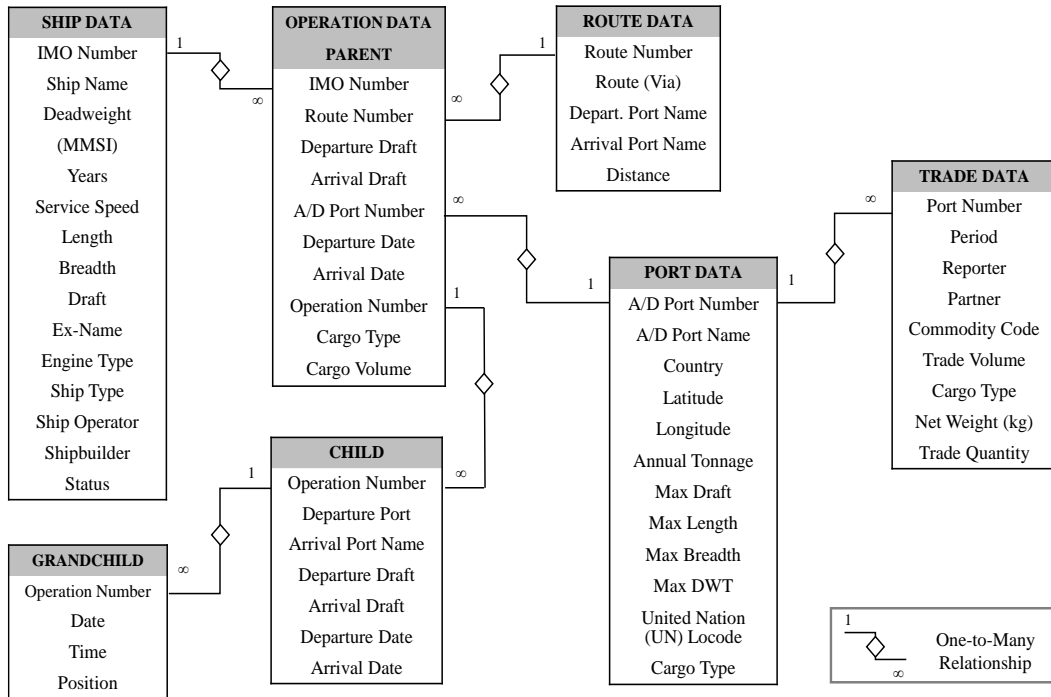


Figure 3. Structure of MLDB

TABLE 2.
 CHECKING CARGO TYPE COMBINATIONS USING A MATRIX

Port Name	Port D (Coal, Iron Ore)	Port E (Coal, Iron Ore, Grain & MB)	Port F (Coal, Iron Ore)
Port A (Coal, Grain & MB)	Coal	Coal, Grain & MB	Coal
Port B (Coal, Iron Ore)	Coal, Iron Ore	Coal, Iron Ore	Coal, Iron Ore
Port C (Iron Ore, Grain & MB)	Iron Ore	Iron Ore, Grain & MB	Iron Ore

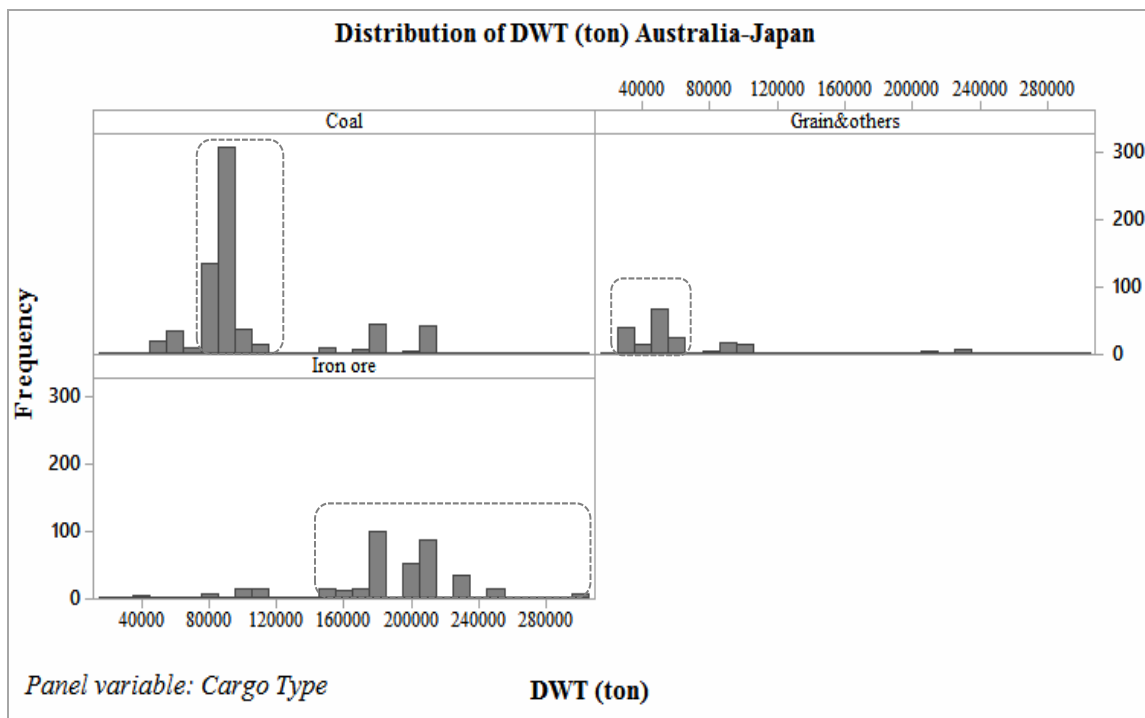


Figure 4. Distribution of ship size from Australia-Japan

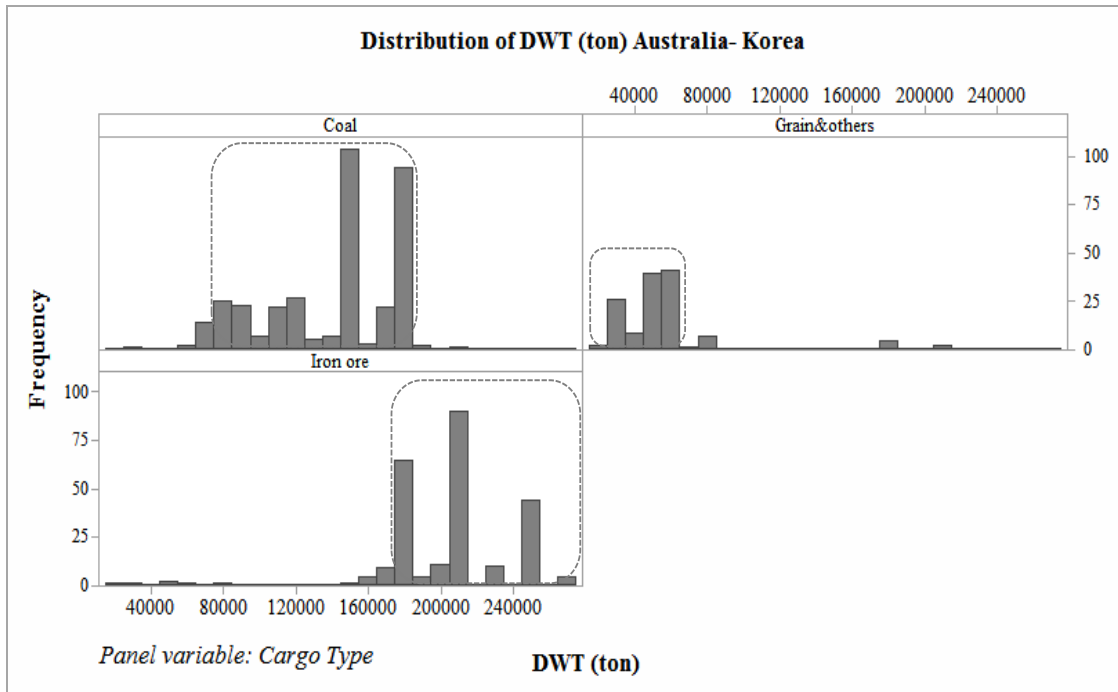


Figure 5. Distribution of ship size from Australia-Japan

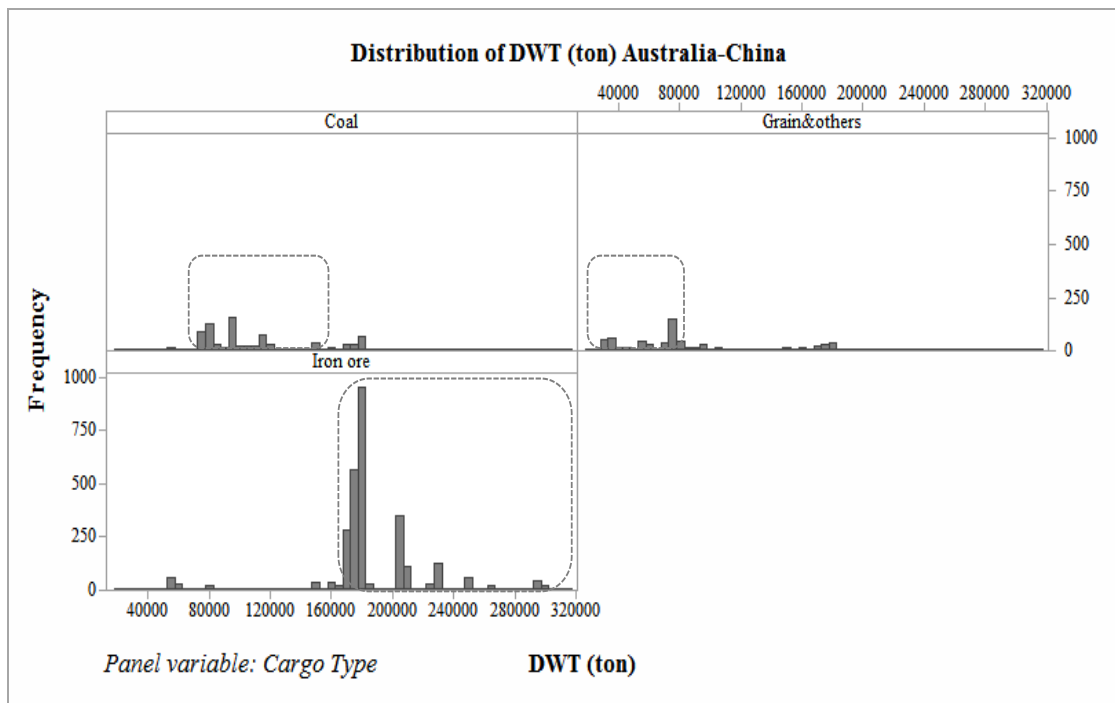


Figure 6. Distribution of ship size from Australia-Japan

• *Cargo Volume Estimation*

The cargo volume was estimated by considering the maximum draft and deadweight of the ship and sailing draft that is extracted from AIS data. The cargo volume was estimated by using Equation (2). Where V_i (ton) is defined as cargo volume, DWT_i is defined as deadweight, and d_i is the draft rate.

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

III. DISCUSSION

A. *Evaluation of the cargo estimation*

To verify the cargo estimation result in Section 2.6.4, we evaluate the estimation results by comparing the results with actual trade value from UN Comtrade data, using bulk carriers operating from Australia to Japan, Korea, and China in 2014. The estimation result covered around 93% from Australia-Japan, 93% from Australia-Korea, and 88% from Australia-China, as shown in **Figure 7-9**. Since the estimation result is compatible with the actual condition, it validates our cargo estimation method.

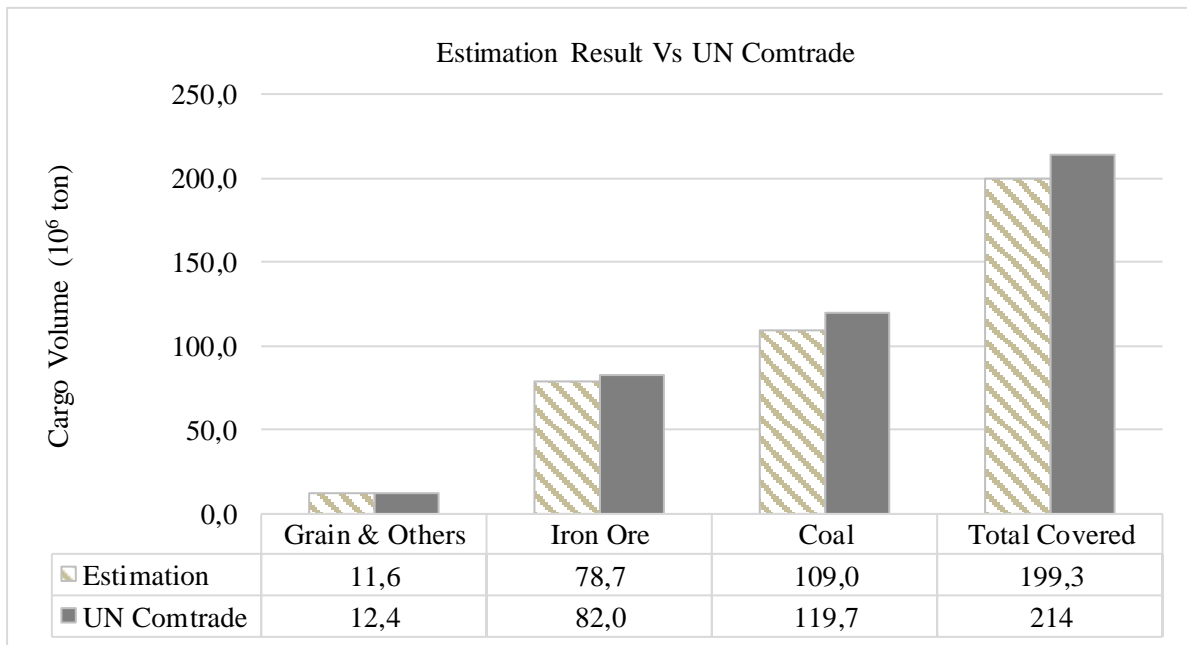


Figure. 7. Estimation result for Australia-Korea

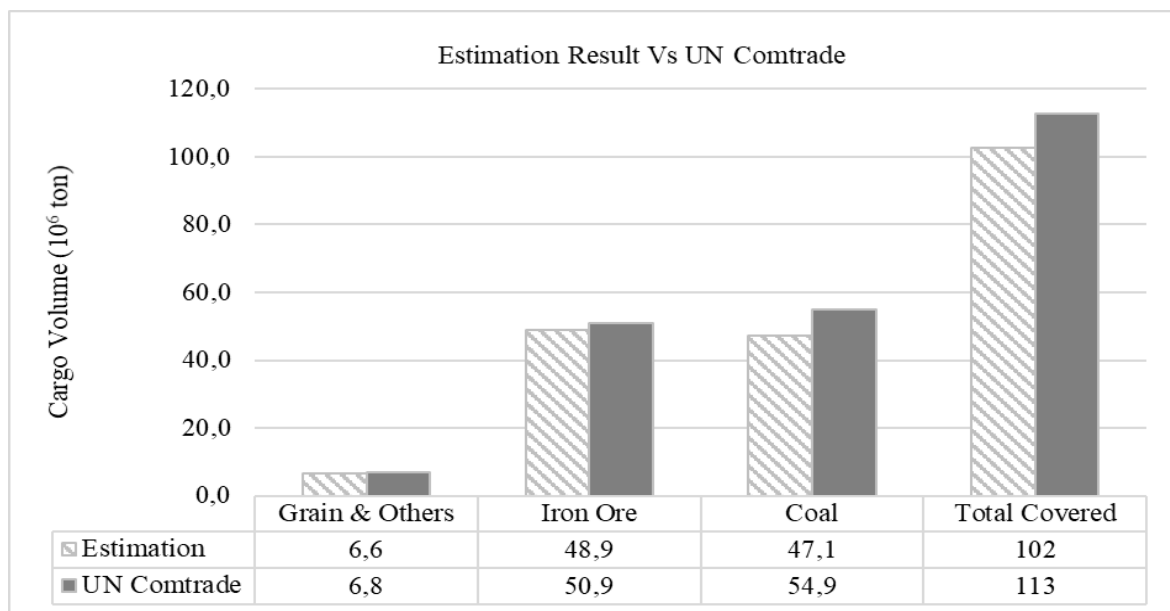


Figure. 8. Estimation result for Australia-Korea

B. Extracted data for ship basic planning

The relational database structure in MLDB allows a user to get some valuable knowledge. The operation information and other information, e.g. LOA (m), DWT (ton), design speed (knot), B (m) d (m), could be easily extracted. By harnessing the extracted data from MLDB, the characteristics of the bulk carrier (Australia-Japan) and (Brazil-Japan) from 2014/01/01 ~ 2014/12/31 could be identified as shown in **Table 3**. Moreover, the critical information for predicting the ship's demand in the future can be obtained by using the ship allocation model.

Ship basic planning support using the ship allocation model will be developed for future applications. A ship allocation model is developed by using the information from MLDB. The ship allocation model predicts the ship allocation when the user inputs the trade volume, economic situation, etc. Therefore, when the user inputs the future scenario, e.g., the state of the world economy, fuel price fluctuation, canal and port expansion, the new ship allocation will be generated so that the effective ship specification can be simulated and estimated. The future application architecture by exploiting the extracted data from MLDB is highlighted in **Figure 10**.

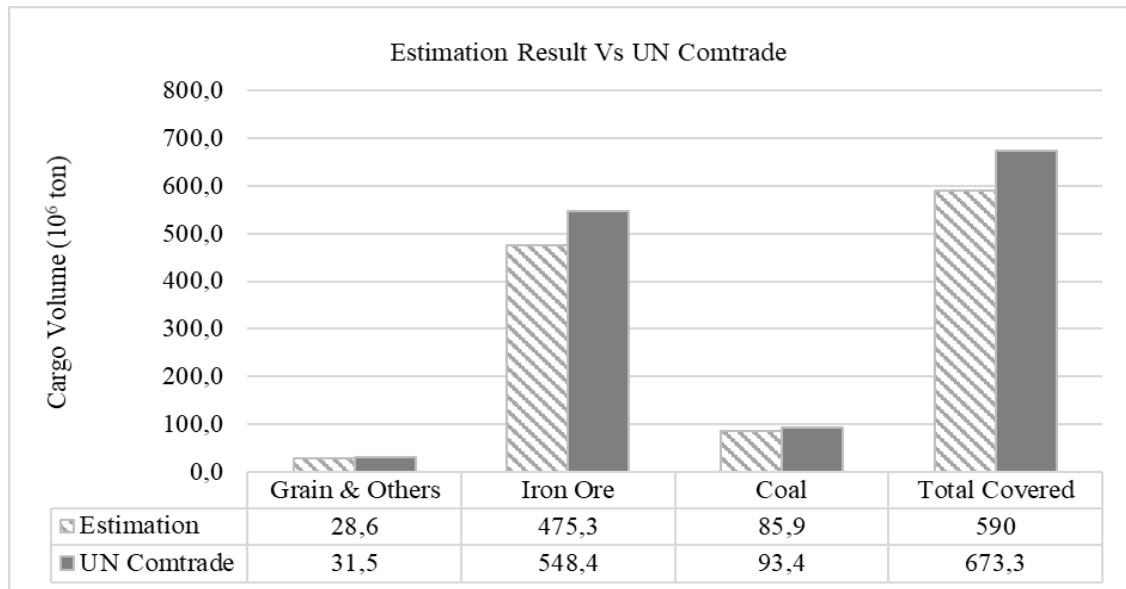


Figure. 9. Estimation result for Australia-China

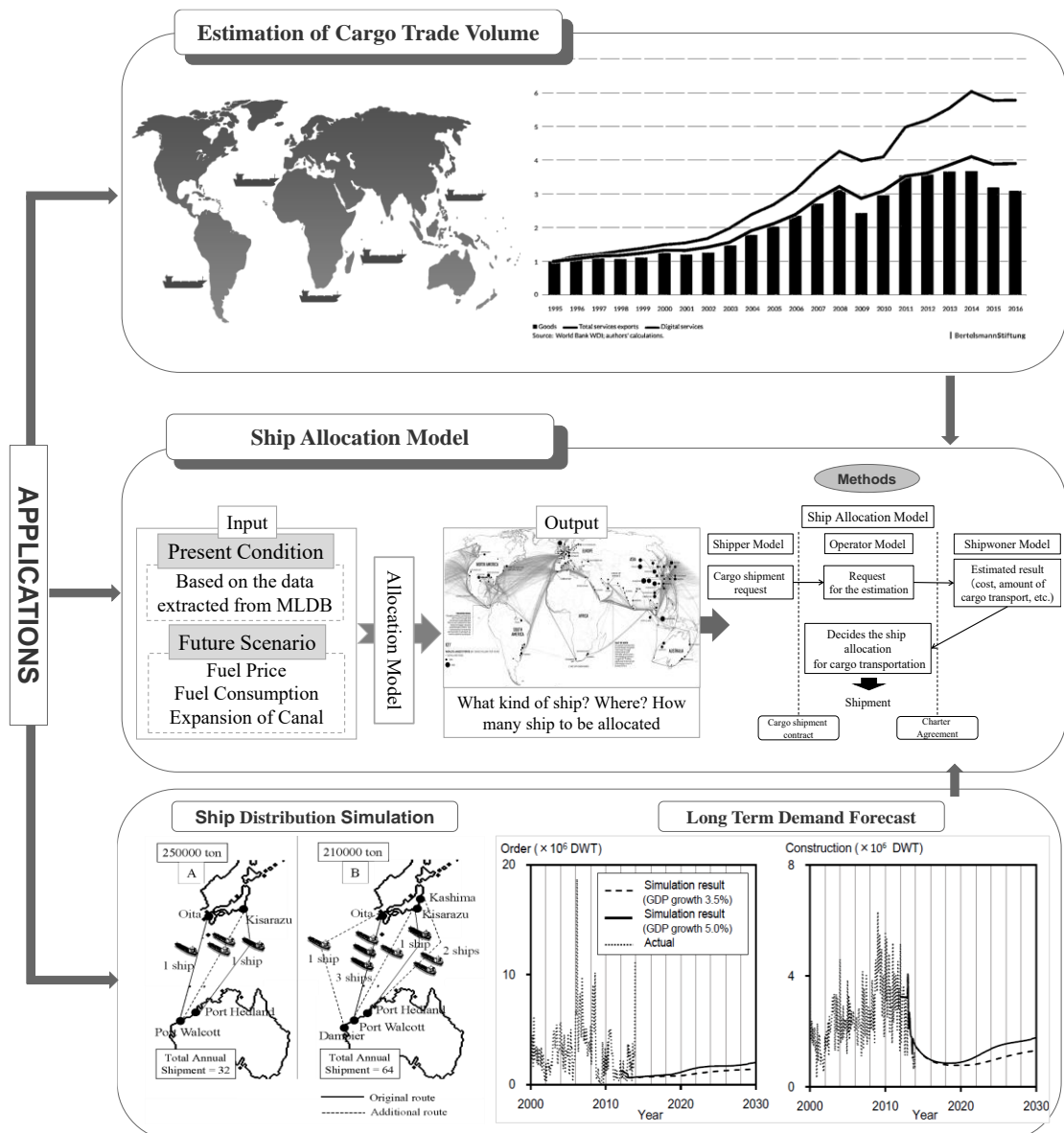


Figure. 10. The architecture of the future application by extracting data from MLDB

TABLE 3.
CHARACTERISTIC OF THE BULK CARRIER FROM AUSTRALIA-BRAZIL TO JAPAN

		DWT (10 ⁴ t)	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

IV. CONCLUSION

In this study, MLDB using marine logistics BD is developed. Integrating BD and error cleaning, which is essential for MLDB, is executed. The estimation methods of cargo type and volume are proposed, and the valuable information that is important for ship basic planning support was extracted. The evaluation of the proposed methods is confirmed by comparing the estimation result with the international trade data from UN Comtrade. The critical data for basic planning support is extracted from the MLDB and discussed. The architecture of the future application by extracting data from MLDB is illustrated.

REFERENCES

- [1] Esben P 2016 BD in the Digital Age – The Impact on Ships and Ports the International Maritime Transport & Logistics Conference (MARLOG 5) pp 1-10
- [2] Martin K 2015 Ship Insight-The Buzz of BD *the Quarterly Journal* pp. 35-41
- [3] Hossein H, Emmanuel S S 2015 'Forecasting with BD *Springer Ann.* pp.5-19
- [4] Chase Jr, Charles W 2013 Using BD to Enhance Demand-Driven Forecasting and Planning *Journal of Business Forecasting* Volume 32 Issue 2 pp. 26-32
- [5] D. Laney 2001 3D Data Management Controlling Data Volume Velocity and Variety
- [6] D. Laney 2012 The importance of BD A Definition Gartner Retrieved 21 June
- [7] Arifin M D 2016 Development of Marine Logistic Database by Using AIS and Statistical Data (TEAM 2016) 30th Asian-Pacific Technical Exchange and Advisory Meeting on Marine Structures pp. 395–402
- [8] S. Shukla, V. Kukade, A. Mujawar 2015 BD Concept, Handling and Challenges: An Overview *International Journal of Computer Applications* (0975 –8887) Volume 114 –No. 11 pp. 6-9
- [9] Hiekata. K., Yamato H 2014 Study on Improving Accuracy for Edge Measurement Using 3D Laser Scanner, IFIP Advances in Information and Communication Technology pp. 427–434
- [10] Aoyama K, Liu J, and Hiro Y 2015 Monitoring System for Advanced Shipbuilding Construction Management: Extracting and Utilizing Monitoring Data by Considering the Reliability of Monitored Data, ICCAS pp. 143–154
- [11] Perera LP, 2016 Statistical Filter-based Sensor and DAQ Fault Detection for Onboard Ship Performance and Navigation Monitoring Systems, IFAC pp. 323–328
- [12] Ando H 2015 Smart Operations: The Application of BD for Ship Operational Efficiency MTI pp. 1–30
- [13] Yoshida T 2016 Utilizing BD and the Internet of Things in Shipping Sea Japan 2016 MTI pp. 1–32
- [14] Arifin M D, Hamada K, Hirata N, Yuki K 2017 Development of Ship Allocation Models Using Marine logistics Data and Its Application to Bulk Carrier Demand Forecasting and Basic Planning, JASNAOE Volume 27 pp. 139-148
- [15] Arifin M D, Hamada K, Hirata N, Ihara K 2017 A study on the support system of ship basic planning by using marine logistics BD International Conference on Computer Applications in Shipbuilding pp. 61-68
- [16] Arifin M D, Hamada K, Hirata N, Ihara K 2017 Development of ship allocation model using marine logistics data and its application to the demand forecasting of ship, Annual meeting of the JASNAOE pp. 529-532
- [17] Wada Y, Hamada K, Hirata N, Yui I, Arifin M. D, Shohei M, and Kazutaka S, A 2015 Study on System Dynamics Model for the Demand Forecasting of Shipbuilding, 12th IMDC, 2, pp. 167–182
- [18] Market Intelligence Network, "Automatic Identification System Data," Available at <<https://www.marketintelligencenetwork.com/app/>>, Start using from 2015-present.
- [19] IHS Maritime & Trade Sea-web Ports Start using from 2015-present.
- [20] IHS Maritime & Trade Sea-web Ships Start using from 2015-present.
- [21] IHS Fairplay Ports & Terminals Guide, 2013
- [22] United Nations Commodity Trade Statistics Database UN Comtrade