DEVELOPING AMARITIME BIG DATA MANAGEMENT ARCHITECTURE

USING INFORMATION ENTROPY

Khaled EL-Sakty

Arab Academy for Science and Technology and Maritime Transport

P.Box 2033-Horia, Heliopolis, Cairo, Egypt.

E-mail: Khaled.sakty@aast.edu

Keywords: Maritime supply chain, big data analytics, multi-dimensional analytics

ABSTRACT

PURPOSE

Optimisation within maritime logistics is complicated due to the uncertainty, complexity and difficult accessibility of data. Maritime logistics companies operate in an environment which requires them to become more analytical. How big data is shaping the maritime supply chains in the future became questionable. The purpose of this paper is to develop an optimisation model for a maritime seaport using Multi-Dimensional Analytical (MDA). This purpose requires exploring the big data concept in the maritime supply chains, identifying big data sources and the applications of big data concept in maritime industry, and identifying the skills to gain analytical competence.

RESEARCH APPROACH

An explanatory approach is applied. The approach of this paper starts with identifying the main tools, platforms and methods currently used to analyse large portions of data in the maritime supply chain. MDA has been applied for a maritime data analysis including two data categories; data dimensions and data measures. The data of ports' operations have been selected for data dimensions, while big data five attributes were used as data measures including volume, velocity, variety, verification and value.

FINDINGS AND ORIGINALITY

A maritime supply chain big data optimisation model is developed, displaying the core transactional data, internal maritime supply chain data, and the external sources of maritime supply chain data.

RESEARCH IMPACT

This paper aims to discuss how is big data being applied in the maritime supply chain operations? In turns, it aims to identify the value-creating big data sources and the capabilities of the future big data practitioners at the maritime supply chain context.

PRACTICAL IMPACT

The maritime supply chain is required to be more responsive, demand-driven and customer centric by leveraging the advantages of the advanced data analytics. Managing big data sources, supporting integrated ports' planning, and identifying the applications of big data at a global level are the practical impacts for decision makers at a maritime industry.

INTRODUCTION

Seaborne trade constitutes almost 90% of world trade in term of volume and about 80% in term of value. It is linked almost in every international supply chain. Hence, a maritime supply chain is facing a challenge to cope with the vast size of the network that global carriers operate.Regarding the trends in port container traffic in 2016, 701 million Twenty-foot Equivalent Units (TEU) of containers were handled on ports worldwide. Also, 10.3 billion metric tonnes of goods shipped, 1.9 billion dead-weight tonnes constitute the world commercial fleet, and the growth commercial fleet shipping capacity is about +3.1% (UNCTAD, 2017). Leading companies and shipping lines may operate a fleet of more than 500 vessels annually. There is an enormous pressure to fill this capacity of seaborne trade and utilise the efficiency benefits of the world fleet. Operating a liner shipping network is truly a big-data problem. Hence, informationare a key success performance indicator of decision makers and the quality of decisions. Industries such as a maritime industry are rich with amounts of data available to be analysed. Different data analytics, tools, models and applications need to be developed and applied in order to handle the value and knowledge of datasets(Elgendy and Elragal, 2016). At ports, managers and authorities need to reach valuable insights from high volume, velocity, variety and value of data using such reliable data analytical approach. Accordingly, data applications play an important role in present and future research applications. Overall, these applications can be divided into two different categories including Database Management (DM) and Data Analytics (DA). This paper aims to discuss the DA concept in a maritime industry. The remaining sections of the paper are as follows: section two displays the available literature of big data and its applications in supply chain and ports, sections three reveals the research work in term of big data measures and dimensions, section four represents the multidimensional data integration with analytics over big data, followed by conclusion.

LITERATURE REVIEW

Big data *The* concept of big data refers to complex sets of data, which exceeds the space of technical ability of storage system, processing, managing, interpreting and visualising of a traditional system. Also, big data can be defined as a study of skills, technologies and practices used to evaluate businesses' strategies (Wang et al., 2016). This has led many companies to develop their big data analytics capabilities. Big data analytics refers to an application of advanced analytic techniques including data mining, statistical analysis and predictive analytics. However, Tiwari et al. (2018) claimed that there is a limited performance of big data that support supply chain. Addo-Tenkorang and Helo(2016) explained five main attributes of big data"5Vs" concept; including volume, velocity, variety, verification/veracity, and value.On the other hand, Wang et al. (2016)classified big data business analytics (BDBA) into two dimensions: big data (BD) and business analytics (BA). BD refers to the ability to process data with the following qualities: velocity, variety, and volume. While, BA refers to the ability to gain insight from data by applying statistics, mathematics, econometrics, simulations, optimisations, or other techniques to help business organisations make better decisions.

Gunther et al. (2017) clarified that big data is based on large *volumes* of extensively *varied* data that are generated, captured and processed at high *velocity*. Gandomi and Haider (2015) identified five main attributes "5Vs" of big data, including big data development sources (variety), big data acquisition (velocity), big data storage (volume), big data analysis (veracity), and big data value-adding benefits to industry (value). These attributed can be displayed in supply chain as follows:

- Variety; it refers to big data development sources, including upstream sources, which are the consolidation points, and downstream sources, which are the logistics and distribution sides.
- Velocity; it refers to an efficient application of analysing big date for value-adding.
- Volume; it refers to an effective management of voluminous data-clusters in order to provide a real-time accessible way.
- Veracity; it refers to big data analytical processing tools that provides high capability of achieving veracious data.
- Value; it refers to a strategic application of big data for enhancing the flow of data (velocity) and veracity.

Gandomi and Haider (2015) discussed two additional dimensions of big dataincluding Variability and Complexity. Variability refers to the variation in the data flow rates, while complexity refers to the fact that big data are generated through a myriad of sources. This imposes a critical challenge to connect, match, and transform data received from different sources. On the other hand, Oussous et al. (2017) underlined moreattributesfor improved define Big Data, includingVision(a purpose), Verification (processed data conformed to some specifications), Validation (the purpose is fulfilled), and Immutability (collected and stored Big data can be permanent if well managed).This paper is considering the 5Vs according to Gunther et al. (2017)and Addo-Tenkorang and Helo(2016) as discussed above as they are common attributes in different articles and industries.

Supply Chain Analytics Supply chain analytics are categorised into three main categories(W*ang* et al., 2016) including descriptive, predictive and prescriptive analytics. The descriptive analytics aim at identifying problems and opportunities within existing processes and functions using different tools such as online-analytical process. The predictive analytics aim to project what will happen in the future and why it is happening using such as mathematical algorithms. The prescriptive analytics aim to improve business performance by assessing alternative decisions. Sivarajah et al. (2017) grouped the data analytics into (1) descriptive analytics that scrutinises data and information to define the current state of a business situation, (2) inquisitive analytics that is about probing data to certify/reject business propositions, (3) predictive analytics is concerned with forecasting and statistical modelling to determine the future possibilities, (4) prescriptive analytics that is about optimisation and randomised testing to assess how businesses enhance their service levels, and(5) pre-emptive analytics that is about having the capacity to take precautionary actions on events that may undesirably influence the organisational performance.On the other hand, Wang et al. (2016) clarified that statistical analysis, simulation and optimisation are common techniques in supply chain analytics. However, Tiwari et al. (2018) argued that traditional statistical methods become invalid for mining big data in supply chain analytics, considering volume, velocity, variety and veracity.

Maritime Big data*Different* maritime frameworks of big data have been recently developed.Yang et al. (2018) developed an innovative framework labelled as cooperative cognitive maritime big data systems (CCMBDSs) on the sea. They aimed to provide a more secured communication networks between ports' users. However, their model is limited to small fish boats. In addition, the China Automatic Identification System (AIS) Asia-Pacific Data Centre was established and its network with IALA-Net. The AIS's purpose is to make the worldwide ship data increasingly available.Zhang et al. (2018) proposed an automatic maritime route generationalgorithm that takes ship AIS trajectories as input, designed to consume low memory, mitigate AIS noise and disparity, and scale to a large database.Vouros et al. (2018)discussed data sources addressed in the European datAcronproject in order to support specific maritime situation. They discussed both data-at-rest (archival data) and data-in-motion (streaming data) sources. The purpose was to improve the capacities of surveillance systems to promote safety and effectiveness of critical operations and forecasting for large numbers of moving entities in large geographical areas. They integrated the heterogeneousdata coming from various data sources in order to provide a unified and combined view in maritime.

MULTI-DIMENSIONAL ANALYTICS

Semenov (2008) discussed the Polish maritime transport situation that is facing some problems in term of insufficient competitiveness and lowoperational flexibility. He started to discuss the obstacles facing the Polish maritime industry and the associated causes. A multidimensional approach was applied where a structuration of collected data about administrative barriers to innovative development of the Polish marine industry was developed, and a multilevel hierarchy of barriers to activity was used.

Research Problem

In a maritime supply chain where operations are characterised by complexity and uncertainty, ports' managers are facing several challenges when exploring Big Data sets and when extracting value and knowledge from such mines of information. The difficulties can be displayed at different levels including data capture, storage, searching, sharing, analysis, management and visualisation(Oussous et al., 2017; Gandomi and Haider, 2015). Hence, this paper has addressed the following problem: *how can maritime big data are turning into maritime big value?*

A Maritime Big Data Model Development

Multidimensional Analyticalapproach has been applied for a maritime data analysis including two data categories; data dimensions and data measures. Data of ports' operations have been selected for data dimensions. While, big data five attributes were used as data measures including volume, velocity, variety, verification and value. The following part aims to discuss the 5Vs attributes and their measures at ports as shown in Table 1.

Table 1. Measures of Big Data 5Vs Attributes

Big Data Attributes	Measures	Given Symbols
Variety (V_1)	- Technological advances - Type of data	- t_1 - t_2
Velocity (V_2)	- Data generated rate - Real time analytics	- d_1 - r_1
Volume (V_3)	- Time for gathering data - Data management technology	- t_3 - d_2
Veracity (V_4)	- Analysis tools - Valuable information generation - Management of uncertain data	- a_1 - i_1 - m_1
Value (V_5)	- Analysing large volume of data	- a_2

Source: developed by the researcher

Data and information in maritime supply chain play indispensable role in quantifying performance and cost of port operations and as a support to decision-making. The virtual and digital flows of information put a huge challenge along with global shipping and carriers (Ducruet, 2017). Regarding the measures of big data 5Vs attributes in ports in Table 1, different dimensions can be founded as follows:

- Technological advances
  - ability to provide cargo tracing (a_3)
  - Type of data dimension
  - high frequency of sailing (h_1)
  - berthing the vessels in ports (b_1)
  - stowage of containers on vessels (s_1).

- Data generated rate
  - prompt response to claim (p_1)
- Real time analytics
- on-time pick-up (o_1)
- routing containers through the physical transportation networks (p_2)

- Time for gathering data
  - courtesy of inquiry (c_1)
- Data management technology
  - ability to provide door-to-door service (s_2)
  - schedule for individual service in ports (s_3)
- Analysis tools
  - good condition of containers (c_2)
- Valuable information

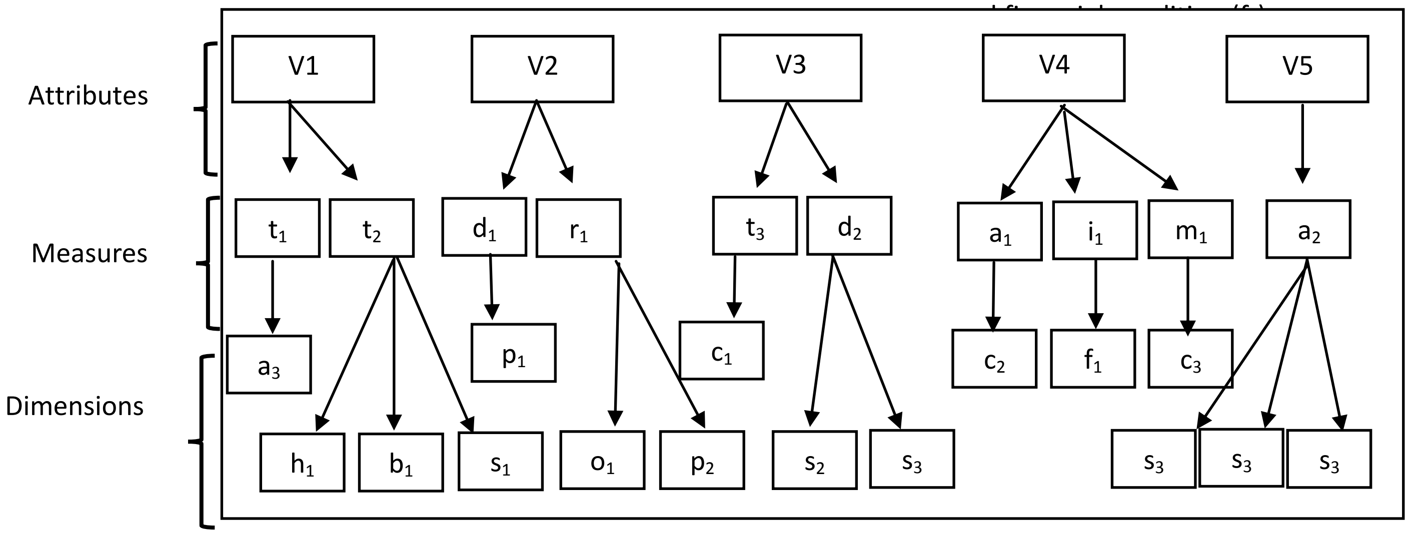

Figure 1: BIG Data Attributes and Associated Measures and Dimensions in Ports Operations

Source: developed by the researcher

Most data in maritime industry are associated with some degree of uncertainty. Accordingly, maritime supply chain is facing planning problems and challenges. Focus has been onaddressing strategic, tactical and operational problems by modern large-scale optimisationmodels. Therefore, Figure 1 displaysan optimisation big data analytics frameworkin ports wherethe 5Vs attributes constitute data required for the strategic level and decision makers at ports, the measures display the tactical level data, while the dimensions refer to the operations data in ports. In ports, the related dimensions and measures of the 5Vs attributes applied in maritime supply chain received from terminalscan be used for example to predict delays and help ships adjust sailing speed to save fuel(Lu, 2000; Brouer et al., 2016).Data at terminals can also lead to better predictions of what will happen in the future. However, optimisation within maritime logistics is complicatedby the uncertainty and difficult accessibility of data.

Research Results

This paper has addressed three results. First, the optimisation big data analytics frameworkin portsis developed as big data analytics in order help decision makers identify the value-creating big data sources and the capabilities of the future big data practitioners(see Figure 1). Second, there are many tools of big data such as storage and management, data cleaning, datamining and data analysis. Also, there are various trends of big data including displacement to the cloud, integration with Internet of Things (IoT), and security improvement. Third, big data analytics in maritime is generally facing many challenges in which can be presented in four groups (Nitaand Mihailescu, 2017) as follows:

- Security challenges where data are collected and shared by many entities in ports' community.
- Technological challenges in term of difficulties of data integration, analytical tools and flexibility.
- Human resources where there is a lack of specialised persons who handle and analyse data.
- Data governance where data are highly secured.

Multidimensional Data and Analytics

Regarding the challenges discussed above, the developed optimisation big data analytics framework needs to be adopted using big multidimensional data approach.Indeed, multidimensional data represent an add-on valuefor analytics models and methodologies. Different multidimensional data models can integrate with analytics including multidimensional abstractions, hierarchy-baseddimensional tables, multi-resolution fact tables, multi-wayaggregations (Cuzzocrea et al., 2011). This integration attains more powerful analytics capable of enhancing the developed framework in this paper.Optimisation, evaluation aspects, advanced decision support tools, building multidimensional data structures, integrating multidimensional sources, developing languages of multidimensional data, and visualisation issues, represent the second-generation big data revolution in the future.Huang et al. (2015) developed a maritime big data management architecture which includes data provision, data preprocessing, data storage, data analysis, and data application aswell as quality control and data security. Their architecture aims to enhance integration between multidimensional data with analytics as represented in the following equation. The question raised is *how to achieve an effective integration of multidimensional data models with analytics over big data?*From Figure 1,datasetsare divided into several layers according toattributes, measures and dimensions whereare stored in clouds. In order to improve the query efficiency on cloud storage, it is essential to assure the optimal utilisation of storage resource and data migration. Things like data sensitivity,data access frequency, data time length, and data size shouldbe fully considered when performing data migration. The migration function can be presented as follows:

$$\mathrm{D}\ (\mathrm{V}_1, \mathrm{V}_2\ldots, \mathrm{V}_n) = \sum_{i=1}^{n} \frac{1}{Ti}\ \mathrm{X}\ \sum_{K=1}^{n} fk\ \mathrm{X}\ \frac{1}{S}$$

Where Tirepresents time-length of the ith access of the marinedataset D per attribute V_n, fkrepresents access frequency of marine dataset Dover the period of Tk, and $S$is the size of marine dataset D. Along with a maritime big data analytics model, the previous migration function supports dynamic data migration in a maritime supply chain, where integration between multidimensional data with analytics can be attained over big data. On the other hand, operations at ports are characterised by complexity and high uncertainty. So, it is essential for turning maritime big data into maritime big value. Information entropy can be applied for this purpose as it is a logarithmic measure of the rate of transfer of information in a particular message or language.Entropy refers to disorder or uncertainty associated with each possible data value. The entropy function is shown as follows:

$$H (X) = E (I(X))$$

where *H* represents entropy, *E* is the expected big value of data, *I* is the information contents of *X*, and *X* is itself refers to the predefined dimensions of 5Vs attributes in a maritime supply chain. According to the huge data collected from different sources in a maritime supply chain, the maximum rate at which information can be transmitted over a communication channel can be formulated as follows:

$$C (V_1, V_2 \ldots\ldots V_n) = B\log_2 (1+\frac{S}{N})$$

where *C* is the channel capacity of 5Vs dimensions, *B* is the bandwidth, *S* is the average received signal power over the bandwidth, and *N* is the average power of interface over the bandwidth. It is evident that big value of big data in a maritime era requires enablers in order to qualify fact-based decision making, provide local performance monitoring and optimisation, and provide opportunity to explore bigger picture. Various enablers are highly required in maritime supply chain such as increased connectivity, increased cloud storage and computing, and new technologies for capturing, storing and analysing vast amount of data.

CONCLUSION

In recent years, new concepts and technologies received a wide attention such as big data, data mining, and data migration and data query efficiency. Existing challenges in different fields and industries are presented in term of data storage, data analysis and data security. Big data is not a single technology, technique or initiative rather it is a trend across many areas of businesses.

In a maritime supply chain, there is no doubt that study on maritime big data management is still in the initial stage of development. Hence, optimisation a maritime big data analytics model was the main purpose of this paper.The developed model aims to discover accessible maritime big data, its measures and dimensions, and the possible values that could be obtained from it. It helps to improve ports' strategic, tactical and operational planning. A migration function can be used by ports' authorities and managers to assure the optimal utilisation of data storage resources.

For further and future researches, practical and theoretical challenges in the existence of marine big data management need to be investigated.Future challenges include data dynamic scalability, controllable maritime data size, maritime database consistency, maritime database usability, maritime data analysis algorithms, and quality of maritime big data.

REFERENCES

1. Addo-Tenkorang, R. &Helo, P.T. (2016), 'Big data applications in operations/supply-chain management: A literature review', Computers & Industrial Engineering, vol. 101, pp. 528-543.
2. Braganza, A., Brooks, L., Nepelski, D., Ali, M. and Moro, R., (2017), 'Resource management in big data initiatives: Processes and dynamic capabilities'. Journal of Business Research, 70, pp.328-337.
3. Brouer, B.D., Karsten, C.V. and Pisinger, D., (2016), 'Big data optimization in maritime logistics'. In Big data optimization: Recent developments and challenges (pp. 319-344). Springer, Cham.
4. Cuzzocrea, A., Song, I.Y. and Davis, K.C., (2011), 'October. Analytics over large-scale multidimensional data: the big data revolution'. In Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP (pp. 101-104). ACM.
5. Ducruet, C. ed., (2017), 'Advances in Shipping Data Analysis and Modeling: Tracking and Mapping Maritime Flows in the Age of Big Data'. Routledge.
6. Elgendy, N. and Elragal, A., (2016), 'Big data analytics in support of the decision-making process'. Procedia Computer Science, 100, pp.1071-1084.
7. Gandomi, A. and Haider, M., (2015), 'Beyond the hype: Big data concepts, methods, and analytics'. International Journal of Information Management, 35(2), pp.137-144.
8. Günther, W.A., RezazadeMehrizi, M.H., Huysman, M. & Feldberg, F. (2017), 'Debating big data: A literature review on realizing value from big data', Journal of Strategic Information Systems, vol. 26, no. 3, pp. 191-209.
9. Huang, D., Zhao, D., Wei, L., Wang, Z. & Du, Y. (2015), 'Modeling and Analysis in Marine Big Data: Advances and Challenges', Mathematical Problems in Engineering, vol. 2015, pp. 1-13.
10. Lu, C.S., (2000), 'Logistics services in Taiwanese maritime firms'. Transportation Research Part E: Logistics and Transportation Review, 36(2), pp.79-96.
11. Nita, S.L. and Mihailescu, M.I., (2017),'Importance of Big Data in Maritime Transport'. Scientific Bulletin" MirceacelBatran" Naval Academy, 20(1), p.556.
12. Oussous, A., Benjelloun, F.Z., Lahcen, A.A. and Belfkih, S., (2017), 'Big Data technologies: A survey'. Journal of King Saud University-Computer and Information Sciences.
13. Semenov, I. (2008),'The multidimensional approach to development strategy of marine industry: Part II. Multifaceted analysis of the development outlook for the polish marine industry'. *Polish Maritime Research* 15, 4, 85-95, Available From: De Gruyter - Sciendo<https://doi.org/10.2478/v10012-008-0001-0> [Accessed 03 June 2018]
14. Sivarajah, U., Kamal, M.M., Irani, Z. and Weerakkody, V., (2017),'Critical analysis of Big Data challenges and analytical methods'. Journal of Business Research, 70, pp.263-286.
15. Tiwari, S., Wee, H.M. and Daryanto, Y., (2018),'Big data analytics in supply chain management between 2010 and 2016: Insights to industries'. Computers & Industrial Engineering, 115, pp.319-330.
16. UNCTAD (2017), 'Handbook of Statistics 2017', United Nations.
17. Vouros, G.A., Doulkeridis, C., Santipantakis, G. and Vlachou, A., (2018),'Taming Big Maritime Data to Support Analytics'. In *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)* (pp. 15-27). Springer, Cham.

18. Wang, G., Gunasekaran, A., Ngai, E.W.T. & Papadopoulos, T. (2016), 'Big data analytics in logistics and supply chain management: Certain investigations for research and applications', International Journal of Production Economics, vol. 176, pp. 98-110.
19. Yang, T., Feng, H., Yang, C., Deng, R., Guo, G. & Li, T. (2018),'Resource allocation in cooperative cognitive radio networks towards secure communications for maritime big data systems', Peer-to-Peer Networking and Applications, vol. 11, no. 2, pp. 265-276.
20. Zhang, S., Shi, G., Liu, Z., Zhao, Z. & Wu, Z. (2018), 'Data-driven based automatic maritime routing from massive AIS trajectories in the face of disparity', Ocean Engineering, vol. 155, pp. 240-250.