Real-Time Ship Management through the Lens of Big Data

Stathis Plitsos
Dep. of Management Science & Technology
Athens University of Economics & Business
Athens, Greece
stathisp@aueb.gr

Takis Varelas
Danaos Research Center
Danaos Management Consultants
Piraeus, Greece
drc@danaos.com

***Abstract*—In this paper we describe a scenario from the Shipping industry, that employs analytics, stream processing, monitoring, alerting and vessel route optimization over big data. This includes the business process modelling, infrastructure management and monitoring along with dimensioning and deployment of focused services requiring different stakeholders roles for their parameterization and enactment. Apart from analysing the domain requirements and user roles, we show how BigDataStack, i.e., a high-performance data-centric stack for big data applications and operations, incorporates, supports and facilitates all these requirements.**

***Index Terms*—shipping, preventive maintenance, requirements analysis, big-data, analytics**

I. Introduction

Vessels are multi-million assets travelling around the world, transporting millions tons of goods, consuming millions tons of fuel [2], [4]. The every-day process of managing a vessel includes people working on-board and ashore [2]. The successful cooperation of both parties is based on two principles: mutual trust and transparency. Nowadays, the Internet-of-Things (IoT) on-board enables monitoring the state of the vessel in more detail and in real-time [10]. Thus, IoT becomes an enabling technology that enhances transparency between people on-board and ashore. On one hand, having sensors on-board measuring various metrics (e.g., speed through water, draft, rotations per minute of the main shaft, etc, usually on a per-minute basis) and storing this information enables an online way of reporting ashore [10]. On the other hand, this flow of measurements increases the volume of produced data. Converting this into actionable information requires sophisticated analytics that can handle the large and increasing volume of data. Last but not least, this should be based on a high-performance technology stack and a flexible and efficient data storage.

BigDataStack is a European research project that delivers a complete high-performance data-centric stack of technologies as a unique combined and cross-optimized offering that addresses the emerging needs of data operations and applications [8]. BigDataStack introduces the paradigm of a new front-runner data-driven architecture and system ensuring that infrastructure management will be fully efficient and optimized for data operations and data-intensive applications. The management system will be scalable and run-time adaptable to facilitate the deployment and management of computing, storage and networking resources, while also considering their inter-dependencies. As a holistic solution, BigDataStack will incorporate approaches that range from data-focused application analysis and dimensioning, process modelling, cluster resources / nodes characterisation, management and run-time optimization, to information-driven networking.

The research leading to the results presented in this paper has received funding from the European Union's funded Project BigDataStack under grant agreement no 779747.

In this paper we focus on the following scenario: A vessel has to complete its route within a specific time-frame. When a part of the main engine fails unexpectedly, the ship risks staying off-hire. This can be very damaging to a shipping company, as chartering revenues decrease, while replacing a spare part immediately increases cost [3]. Thus, identification of potential failure allows (a) timely ordering, or replacement of spare parts before failure [9] and (b) adjusting the route to the destination [13] without putting a heavy load on the main engine. Furthermore, identifying malfunctions on sensors installed on-board can reduce data-loss [10]. Last, being alerted when policies of the shipping company are violated allows corrective actions in a timely manner. For this scenario we provide user requirements and present the BigDataStack-components of major importance being used. This includes business process modelling tools, high-performance stream processing techniques and querying engines along with scalable data storage.

The remaining of this paper goes as follows; Section 2 describes the user requirements along with the involved user roles, Section 3 describes the business process model, Section 4 focuses on the BigDataStack components of major importance and Section 5 concludes the paper.

II. User Requirements and Roles

The Real-Time Ship Monitoring scenario in this paper addresses the following goals:

- Monitoring the state of vessels;
- Identification of sensor malfunctions with regard to condition-based rules;
- Identification of company policy violations;

- Identification of malfunction patterns in the data of the main engine;
- Notification of the end user accordingly;
- Route adjustment without putting a heavy burden on the main engine, taking into consideration the weather conditions ahead.

The expected business impact achieving these goals is:

- Advanced monitoring of key components on-board and ashore;
- Monitoring the compliance of crew-members with company policies;
- Minimization of machinery failures that cause the ship to go off-hire;
- Minimization of maintenance costs;
- Efficient engagement of the involved departments when unexpected malfunctions occur.

All these objectives require the involvement of several roles within the organization. First, the engineers of the Technical department have the overview of the key electrical/mechanical components on-board. Second, the people from the Operations department, require detailed reports regarding the state of vessel (voyage plan, speed, consumption, company policies). People from the Supplies department handle requests on spare parts, when needed. The design and development of analytical tasks over vessel data requires a Data Scientist. Existing applications containing data or implemented functionalities that are required for the achievement of these goals are maintained by Application Experts. Data provisioning is the responsibility of the Data Owner. Last, a Business Analyst is required to model the process and the interaction of the required business components.

The key activities these actors are involved in are the following:

- Business process modelling: A Business Analyst wishes to model the interaction of individual business components. This should be mapped to existing systems along with business rules.
- Acquire data: The Data Owner wishes to set a data source from which data are obtained;
- Select attributes: The Data Scientist wishes to select a set of attributes depending on custom criteria;
- Monitor the selected attributes: The Data Scientist wishes to monitor the data;
- Deploy the application services: The Application Engineer wishes to deploy existing applications or services;
- Use an analytics algorithm or inject a new one: The Data Scientist wishes to use a custom or general-purpose analytics algorithm;
- Produce alerts depending on monitoring: An engineer of the Technical department wishes to receive predictive maintenance alerts; An employee of the Operations department wishes to receive company policy violation alerts;
- Order spare parts when necessary: An employee of the Supplies department wishes to receive a notification to order new spare parts;
- Re-route the vessel accordingly: An employee from the Operations department wishes to be informed about the route adjustment.

Regarding the main engine malfunctions and the identification of such cases, it is obvious that this is a vast problem. Here, we focus on a malfunction that occurs on a specific component of the cylinders, namely, the cross-head bearing. To the best of our knowledge, there is no known correlation of main engine data that can pinpoint this malfunction. Up to this moment, this malfunction is identified only by human eyes with an on-sight inspection. Nevertheless, we strongly believe that this wear, as a physical phenomenon, is hidden somewhere in the data and wish to reveal it.

Regarding the violations of company policies, we focus on a specific policy, namely, the charter party agreement. That is, a contractual agreement of the shipping company with the charterer, where the first guarantees, among others, the fuel consumption of a vessel under certain circumstances (weather conditions, draft, speed). Monitoring the vessel state and checking for charter party violations is a daily operation of high importance performed by the Operations department of a shipping company.

Both reductions above make the scenario more concrete, without loss of generality of the described requirements. In the following sections we describe how BigDataStack assists these requirements taking into consideration the large volume of required data and the hard constraints these business decisions put on systems in terms of efficiency and accuracy.

III. Business Process Modelling

In this section we describe how Business Process Modelling (BPM) can be assisted and enhanced through BigDataStack. We start with this, since this is where business requirements are defined and business decisions can be automated. In a big data context, several aspects should be taken into consideration. On one hand, big data seems difficult to manage and process. On the other hand, big data provides new prospects for business process management research [7]. This is indeed a great opportunity to promote "evidence-based BPM" [12]. What is currently missing, is the connection of event data to process models and process mining which represents the missing link between analysis of Big-Data and business process management [12]. Consequently, enterprises are unable to deal with complex decisions [5], [14]. Furthermore, business process analytics is often not able to provide insights in a timely manner as data is only provided by Extract-Transform-Load (ETL) jobs batched overnight, not allowing real-time insight into processes which is necessary for some cases, such as this one in the shipping domain and various others [6].

Process analytics uses a set of measurement and analysis techniques to evaluate the past, to understand what is happening at the moment and to predict the future in the context of a business process [15]. On the initialization or execution of a process instance, multiple events occur that relate to the start and finish of process activities, such as internal process engine operations and other system and application functions. Process mining and other analytical techniques usually extract this process history data from a process execution environment and submit the data to the process analytics environment for processing [16]. However, the ability to process these logs in a sophisticated manner is absent in many Business Process Management Systems (BPMS) [1]. Also, process analytics is limited to analysing the past and monitoring the present only. Last, hard-wiring BPMS and process analytics systems is mandatory in order to achieve end-to-end insight-to-action [11].

In BigDataStack, apart from the ability to model processes and components, process mining is performed by analysing events to identify patterns and dependencies including both stored and streaming data. In this way, BigDataStack merges the functionality of a BPMS and a process analytics system. The pattern mining aims at process chains identification and optimization. A feedback loop from the process mining outcomes is realized in order to enable interactions with stakeholders through the process modelling framework. The latter is used by the Business Analyst to model, optimize and adapt business processes.

In the case of the Real-Time Ship Management scenario the process goes as shown in Fig. 1. We enrich this model by adding the business constraints with raw text on the sides, bottom or top of each component. First, data related to vessels are acquired through the Data Acquisition module. In parallel, weather-related data are also acquired through the Weather Data Acquisition module. These two components, allow the ingestion of historical data and the flow of fresh data into BigDataStack. There are numerous attributes in both data-sets, but only a subset of these are valuable to this case. This is where the Attribute Selection module takes place. Data flowing out of this components end to two different modules: (a) the Monitoring and (b) the Predictive Maintenance component. Both components serve a different purpose. The Monitoring component enables the online monitoring of the vessel state. Then, data flow to the Rule-Based Alert component where the malfunction of sensors on-board and the violation of a company policy is identified and an alert is produced that becomes available to the end-user. The Predictive Maintenance component is the analytics algorithm running over the incoming data, producing an alert when a malfunction pattern is identified. This alert becomes available to the Technical department and triggers an additional alert via the Spare Part Order component for the Supplies department to order the required spare part. Last, the Route Adjustment component provides all the functionality that allows the adjustment of the vessel route with regard to the main engine malfunction and its limitations, along with the weather conditions towards the destination.

Once the Business Analyst defines the whole process, the Data Analyst through the Data Toolkit, i.e., another BigDataStack offering [8], takes control and refines this model by adding extra information, such as, entity names in databases, storage size, configuration of external services, communication protocols, etc. This refinement includes the addition of extra components, if required. Fig. 2 shows the Data Analyst perspective, i.e., how the model of the previous step incorporates the required technical information.

IV. BigDataStack & Real-Time Ship Management

The Real-time Ship Management scenario exploits the BigDataStack environment with an emphasis on the data as a service offering for big data management, its analytics and methods for real-time monitoring, preventive maintenance, visualization of the vessel state and final results. Within this scenario, all components of BigDataStack are utilized [8], from the previously described BPM framework to the visualization of results. In this section we present and shortly describe the higher-level architecture of BigDataStack and how its components are utilized to implement this scenario. Additionally, we map the roles and required functionalities of each role with the respective BigDataStack layer and component of major importance. Fig. 3 shows the high-level BigDataStack architecture.

The Data-driven Infrastructure Management layer [8] ensures that the infrastructure management is fully efficient and optimized for data operations and data-intensive applications. The management system is scalable and run-time adaptable to facilitate the deployment and management of computing, storage and networking resources, while also considering their inter-dependencies. The Data Owner uses the components of this layer to ingest the available data and scale the system accordingly.

The "Data as a Service" component [8] is the cornerstone on which the four upper platform capabilities of BigDataStack rely on. This component offers a set of services that provide the building blocks of an efficient and modern data infrastructure covering all the major phases of data life cycle and usage, i.e., data ingestion, data storage and data analytics. Here, the Data Scientist employees the Data Quality Assessment tool to clean the data and feed his custom preventive maintenance algorithm with quality-assessed data. Furthermore, this layer offers the Complex-Event-Processing (CEP) component, where streaming data are filtered, monitored and queried over custom rules. In the Real-time Ship Management scenario, these rules that run over CEP are the identification of malfunctions on sensors on-board and the company policy-rules that are based on sensor data. Thus, alerts on these cases are produced mainly by this component. Last, the Seamless Data Analytics component provides a novel storage solution that federates two very different data stores: a transactional database, where fresh data are stored and an object store, where historical data reside. The seamless component permits to store and query a data-set that has been split between these two data stores as a single logical data sets. This component is utilized by custom services deployed on BigDataStack.

The Dimensioning Workbench [8] aims at enabling the dimensioning of applications in terms of predicting the required data services, their inter-dependencies with the application micro-services and the required underlying resources. Here,

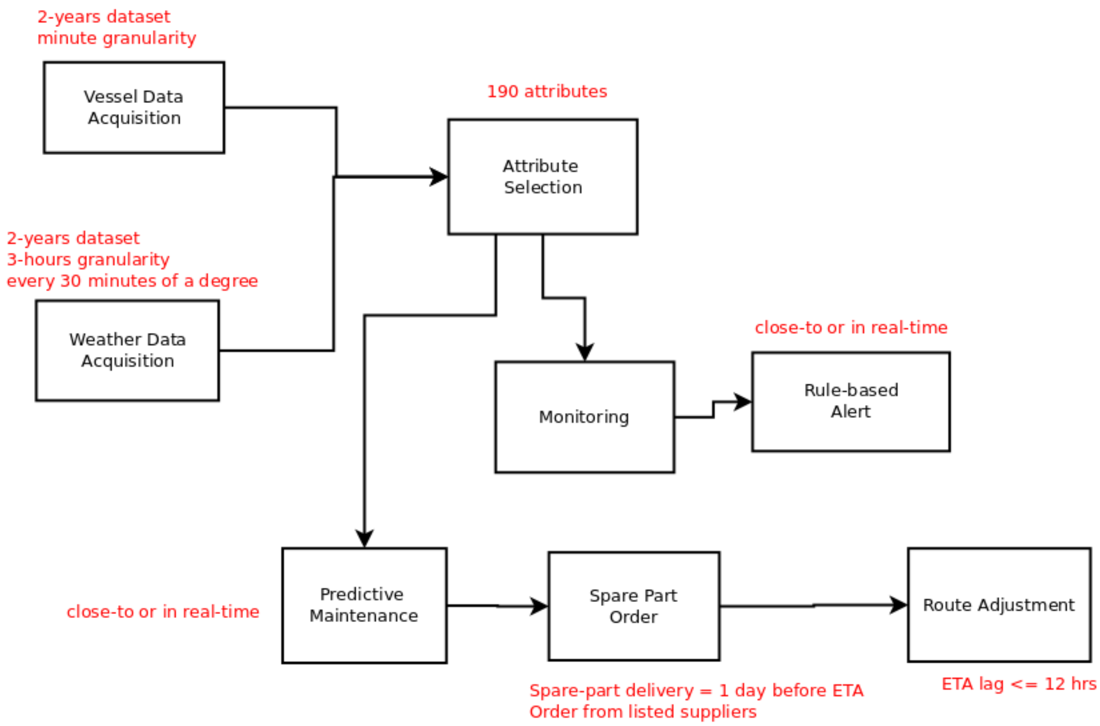

Fig. 1. Business Process Model: The Business Analyst perspective

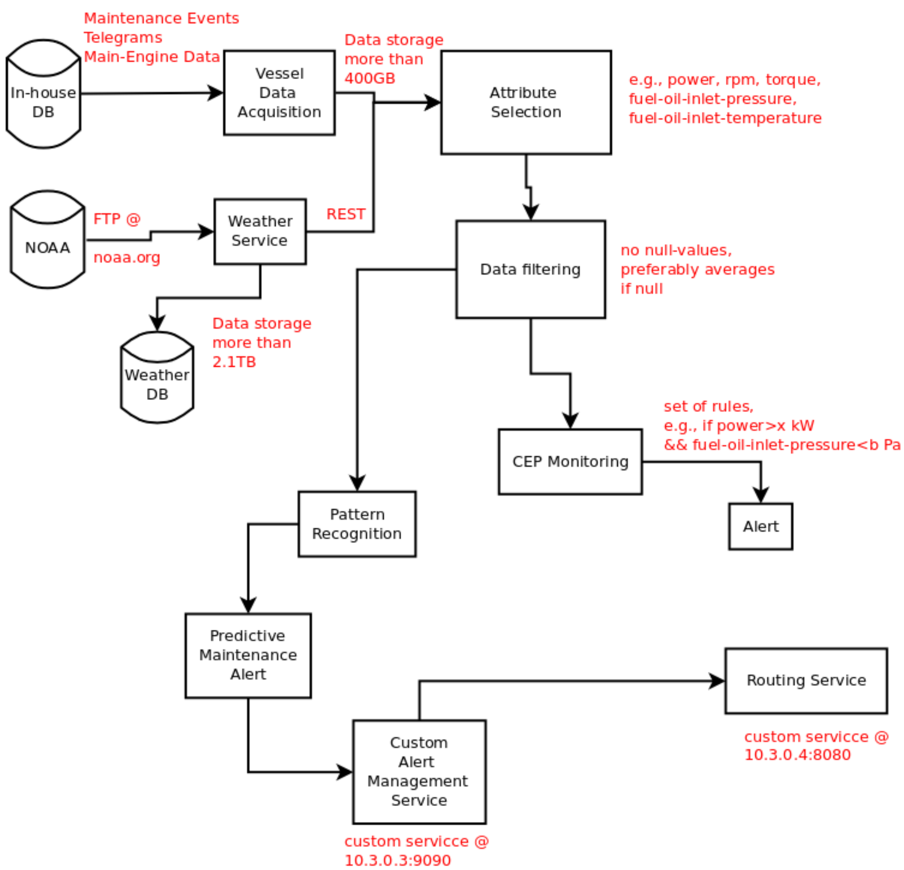

Fig. 2. Business Process Model: The Data Analyst perspective

the Application Engineers experiment with their applications and dimension it in terms of its data needs and data-related properties.

The Process Modelling component [8], as thoroughly discussed in the previous section, provides a framework allowing for declarative and flexible modelling of process analytics in order to enable their execution. Functionality-based process modelling will then be concretized to technical-level process mining analytics, while a feedback loop will be implemented towards overall process optimization and adaptation.

The Data Toolkit [8] allows the ingestion of data analytics functions and the definition of analytics in a declarative way, providing at the same time "hints" towards the infrastructure / cluster management system for the optimized management of these analytics tasks. Furthermore, the toolkit allows data scientists and administrators to specify requirements and preferences both for the infrastructure management (e.g. application requirements) and for the data management (e.g. data quality goals, incremental analytics, information aggregation "levels", etc.). The use of this component is described in the previous section as a logical sequel of the BPM step. Also, through this toolkit the Data Analyst is able to inject into BigDataStack custom analytics algorithms, in this case, the preventive maintenance algorithm.

The last core offering of BigDataStack is the Data Visualization [8] component going beyond presentation of data and analytics outcomes to adaptable visualizations in an automated way according to application analysis and data semantics. Visualizations covers a wide range of aspects (interlinked if required) besides data analytics, such as computing, storage and networking infrastructure data, data sources information, and data operations outcomes (e.g. cleaning outcomes, aggregation outcomes, etc.). Moreover, the visualizations are incremental, thus providing data analytics results as they are produced. Here, all involved roles get to visualize their results.

To conclude, Table I, shows the matching of user roles to components and the core functionalities required to assist this scenario.

V. Conclusion

In this paper we describe a scenario from the Shipping industry, that employees analytics, stream processing, monitoring, alerting and vessel route optimization over big data. This includes the infrastructure management and monitoring, deployment of custom services, process modelling and application dimensioning that different roles within this scenario require. Apart from analysing the domain requirements and user roles, we show how BigDataStack, i.e., a high-performance data-centric stack for big data applications and operations, incorporates, supports and facilitates all these requirements.

References

[1] Melissa Cheung and Jan Hidders. Round-trip iterative business process modelling between bpa and bpms tools. *Business Process Management Journal*, 2011.

[2] John Coustas, Takis Varelas, Dimitrios Theodosiou, Iraklis Lazakis, and Sofia Archontaki. *Operations Research & Information Science in e-Maritime*. 2018.

[3] Ayse Sena Eruguz, Tarkan Tan, and Geert-Jan van Houtum. A survey of maintenance and service logistics management: Classification and research agenda from a maritime sector perspective. *Computers & Operations Research*, 85:184–205, 2017.

[4] Andrew P Goodwin. The economic value of shipping and maritime activity in europe, 2016.

[5] Michael Hammer and James Champy. (1993)."reengineering the corporation: A manifesto for business revolution". *Paperback edition, HarperBusiness, New York*, 1993.

[6] Christian Janiesch, Martin Matzner, Oliver Müller, Robert Vollmer, and Jörg Becker. Slipstream: architecture options for real-time process analytics. In *Proceedings of the 2011 ACM Symposium on Applied Computing*, pages 295–300, 2011.

[7] Tim Kraska. Finding the needle in the big data systems haystack. *IEEE internet Computing*, 17(1):84–86, 2013.

[8] D. Kyriazis, C. Doulkeridis, P. Gouvas, R. Jimenez-Peris, A. J. Ferrer, L. Kallipolitis, P. Kranas, G. Kousiouris, C. Macdonald, R. McCreadie, Y. Moatti, A. Papageorgiou, M. Patino-Martinez, S. Plitsos, D. Poulopoulos, A. Paradell, A. Raouzaiou, P. Ta-Shma, and V. Vianello. Bigdatastack: A holistic data-driven stack for big data applications and operations. In *2018 IEEE International Congress on Big Data (BigData Congress)*, pages 237–241, July 2018.

[9] I Lazakis, Y Raptodimos, and T Varelas. Predicting ship machinery system condition through analytical reliability tools and artificial neural networks. *Ocean Engineering*, 152:404–415, 2018.

[10] Yiannis Raptodimos, Iraklis Lazakis, Gerasimos Theotokatos, Takis Varelas, and Leonidas Drikos. Ship sensors data collection and analysis for condition monitoring of ship structures and machinery systems. *Smart Ship Technology*, 2016.

[11] Bernd Schwegmann, Martin Matzner, and Christian Janiesch. A method and tool for predictive event-driven process analytics. In *Wirtschaftsinformatik*, page 46, 2013.

[12] Wil MP Van der Aalst. A decade of business process management conferences: personal reflections on a developing discipline. In *International Conference on Business Process Management*, pages 1–16. Springer, 2012.

[13] Takis Varelas, Sofia Archontaki, John Dimotikalis, Osman Turan, Iraklis Lazakis, and Orestis Varelas. Optimizing ship routing to maximize fleet revenue at danaos. *Interfaces*, 43(1):37–47, 2013.

[14] Alejandro Vera-Baquero, Ricardo Colomo Palacios, Vladimir Stantchev, and Owen Molloy. Leveraging big-data for business process analytics. *The Learning Organization*, 2015.

[15] Jan Vom Brocke and Michael Rosemann. *Handbook on business process management 2*. Springer, 2010.

[16] Michael zur Muehlen and Keith D Swenson. Bpaf: A standard for the interchange of process analytics data. In *International Conference on Business Process Management*, pages 170–181. Springer, 2010.

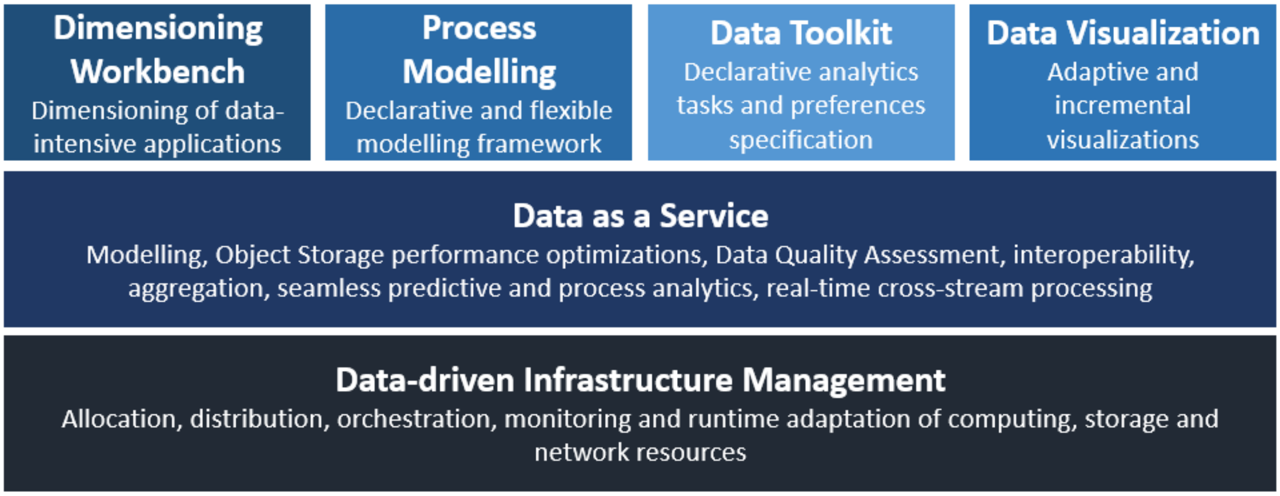

Fig. 3. BigDataStack - Components and high-level architecture

Role Name	Component	Description
Data Owner	Data-driven Infrastructure Management	BigDataStack offers a unified Gateway to obtain both streaming and stored data from data owners and record them in its underlying storage infrastructure that supports SQL and NoSQL data stores.
Business Analyst	Process Modelling	BigDataStack offers the Process Modelling Framework allowing business users to define their functionality-based business processes and optimize them based on the outcomes of process analytics that will be triggered by BigDataStack. Mapping to specific process analytics tasks will be performed in an automated way.
Data Scientist	Data Toolkit	BigDataStack offers the Data Toolkit to enable data scientists both to easily ingest their analytics tasks and to specify their preferences and constraints to be exploited during the dimensioning phase regarding the data services that will be used (for example preferences for the data cleaning service).
Application Engineer	Dimensioning Workbench	BigDataStack offers the Application Dimensioning Workbench to enable application owners and engineers to experiment with their applications and dimension it in terms of its data needs and data-related properties.
Technical Department	Data Visualization	BigDataStack offers the Data and Services Monitoring component to monitor the produced streams of data. Results of the analytics algorithm for preventive maintenance are alerts that the engineer of the Technical department wishes to further investigate through the respective data.
Operations Department	Data Visualization	BigDataStack offers the Data and Services Monitoring component to monitor the produced streams of data. In this case, the deployed services allow to an employee of the Operations department to adjust accordingly the vessel's route given a malfunction and a set of limitations on the main engine.
Supplies Department	Data Visualization	BigDataStack offers the Data and Services Monitoring component to monitor the produced streams of data. In this case, an employee of the Supplies department wishes to be alerted, when a spare part is needed.

TABLE I
ROLES, FUNCTIONALITIES & BIGDATASTACK COMPONENTS