

Enabling Interactive Visualizations in Industrial Big Data ^{*}

Aguinaldo Bezerra ^{*} Vitor Greati ^{*} Vinícius Campos ^{*}
Ivanovitch Silva ^{*} Luiz A. Guedes ^{*} Gustavo Leitão ^{*}
Diego Silva ^{*}

^{*} *Federal University of Rio Grande do Norte, Natal, Brazil*
{aguinaldo, greati, vinihampos}@ufrn.edu.br, {ivan, gustavo.leitao}@imd.ufrn.br,
affonso@dca.ufrn.br, diego@ect.ufrn.br

Abstract: Industries are considered data rich but information poor environments. Mainly due to systems design restrictions, to the lack of adequate processing power and to a sector culture notably focused on collecting, selecting, storing and preserving historical series in on-demand access repositories, massive data generated in industrial operations is traditionally neglected (or simply took aside). This huge amount of unprocessed data resting in these repositories is a latent and rich source of information that could be used to improve industrial processes. This work then proposes an approach in which an elastic processing engine is designed to be plugged-in to currently installed industrial information infrastructure to provide it with the ability of performing visual analytics on massive industrial data. A case study where an interactive visualization application is made possible in real-world industrial data scenario of over 100 million records is presented to attest the effectiveness and potential of the proposed approach in enabling interactive visualizations to Industrial Big Data.

Keywords: Knowledge discover (data mining), intelligent decision support systems in manufacturing, industrial Big Data, visual analytics, distributed and parallel processing.

1. INTRODUCTION

Current trends in modern manufacturing point towards the undeniable reality of data orientation, in which processing data into timely and valuable information is a requirement to support decision making and continuous improvement of industrial processes. As the possibilities to collect and store data increase at a faster rate than the ability to use it in decision making, data acquisition is no longer the driving problem. There remains, therefore, the challenges related to the acquisition of the ability to identify methods and models that can turn the data into reliable and provable knowledge (Keim et al., 2008).

In this context, Industry 4.0 emerges as a watershed in manufacturing, from the synergy of available innovative information technology with the consumer demand for high quality and customized products. Referring to the so-called 4th industrial revolution, Industry 4.0 is an umbrella term which outlines a series of paradigm shifts by which industries have been going through to ensure survival in a high competitiveness global scenario (Bartodziej, 2017). Although the term is defined from various and diverse perspectives, a convergence point is that it represents a revolution towards digitization and computerization of manufacturing which is transforming production and management (Bartodziej, 2017).

As a natural and direct consequence of these paradigm shifts, volume, velocity and variety of data industries

^{*} The authors would like to thank Federal University of Rio Grande do Norte and Petróleo Brasileiro S.A. for supporting this work.

have to manage is exploding at really high rates (Obitko et al., 2013). This information overload and the inability of dealing properly with enormous data volumes make data exploration and analysis a laborious task for humans, obfuscating improvement opportunities in the sector.

To cope with that, some enabling Big Data technologies which have been typically and solidly employed in the Information Technology (IT) domain, have been also applied to establish an infrastructure capable of embracing this mass of data while attaining industrial reality and needs (Gokalp et al., 2016). However, direct and large-scale application of mainstream big data tools and methods is not a trivial task, due to domain-specific challenges such as diverse communication standards, proprietary information and automation systems, heterogeneous data structures and interfaces, inflexible governance policies, allied with the lack of inherent support from those tools and methods to industrial applications (Gokalp et al., 2016).

Recent literature have come on Industrial Big Data subject with a particular regard to problems and challenges to be overcome in face of the infeasible absorption of this new paradigm in automation. In this field, a significant challenge is to provide the ways to simply and effectively unveil hidden knowledge and value resting in those large, unexplored and potentially precious industrial data repositories. Upon this premise, the development of data visualization and exploration strategies for information seeking in large amounts of data constitutes a fundamental objective to be pursued in industry. Analytical reasoning and effective understanding of large industrial datasets can

be facilitated by combining automated analysis techniques with interactive visualizations (Keim et al., 2008). Visual analysis of industrial data can assist production operators with intuitive production monitoring and on-site troubleshooting and also provide production managers with deep insights into non-real-time historical data for process improvement and innovation (Zhou et al., 2019).

Providing on-the-fly scalable visualizations, navigation and analysis over large, great variety and dynamic data constitutes a challenge in this data-driven era (Bikakis et al., 2016) as most existing data visualization approaches do not scale well from small datasets to enormous ones (Godfrey et al., 2016). Apart from the common visual scalability issues incurred when proceeding visualization of huge amounts of data such as data occlusion and display overloading (Shneiderman, 1996), also emerges data processing difficulties regarding processing power, storage space, bandwidth and display resolution necessary to computing the visualization of large size and dimensionality data (Godfrey et al., 2016). The endeavor gets even more challenging considering interactive visualizations, which impose to the underlying infrastructure a severe processing demand to enable properly data navigation, once an action result should be quickly visible (Limet et al., 2014).

Thence, to reach timely, smooth and flexible data visualizations, the underlying computational infrastructure must have the ability of processing large-scale data within an interval congruent with human reasoning. Still, visualization on large datasets requires coupling efficient exploration techniques with mechanisms for information selection, abstraction, aggregation and summarization (Bikakis et al., 2016), as well as the appropriate usage of automatic data analysis techniques such as clustering and classification as preprocessing stage (Keim et al., 2008).

Assessing the aforementioned challenges, this work provides a brief review regarding the subject of Industrial Big Data, with particular focus on its related infrastructure and data visualization. Then, it is introduced an approach in which mainstream open-source tools are teamed to setup an engine meant to be plugged-in to conventionally installed industrial infrastructures, providing them with the ability to perform graphically-rich Big Data analysis over large industrial data repositories. Then, a case study focused on demonstrating engine processing capabilities in a visual analytics application scenario is carried out.

The remainder of the paper is structured as follows. In Section 2, paper background subject of industrial Big Data infrastructure is discussed. Section 3 introduces our approach to cope with massive industrial data. In Section 4, a case study with some test scenarios and their respective results are presented. Finally, Section 5 concludes the paper and envisions future directions.

2. INDUSTRIAL BIG DATA INFRASTRUCTURE

Data-driven approaches have been supporting decision-making over the last years in a myriad of sectors and this was enabled by technologies pioneered in the IT realm (Obitko et al., 2013). Not unlike, data-drivenness has also made its path through industry and rapidly gained prominence (Diez-Olivan et al., 2019). It happened in a

gradual and natural way, once manufacturing generates and stores more data than any other sector (Brian Hartmann and Narayanan, 2015; Obitko et al., 2013) and given the already customary and well-accepted industry convergence with IT-technologies.

Current industrial computing infrastructure, in terms of management and processing of plant-related data, is mainly focused on collecting, selecting and storing data at appropriate rates, preserving historical series in on-demand access repositories (Obitko et al., 2013). Mostly due to a design restriction, any additional processing such as deeper queries or analysis are beyond the capacity of typically installed computing infrastructure. Thus, to put industrial systems on track of this data inundation scenario, a current trend is to use Big Data precepts as a mean to enable processing of a huge and dormant amount of data generated in industrial plants, generally not suitable for processing in conventionally installed infrastructure.

Industrial Big Data is an already well-known concept that refers essentially to the absorption of Big Data in Industry. It inherits the defining characteristics of general purpose Big Data concept such as volume, variety, velocity, variability and veracity (5 Vs), as well as extends this concept by adding two new Vs: visibility, which regards to the discovery of unexpected insights of existing processed data; and value, which puts emphasis on creating new value from massive data (Basanta-Val, 2018).

Perform collection, aggregation, handling and processing of such a data mass is indeed a challenge to be overcome. To make most data-driven approaches feasible and viable in such a voluminous data and time-restricted scenario, a hardware and software infrastructure capable of meeting the great computational demands of those approaches is needed. Literature confirms this matter as a high relevance interest topic and some progress has been accomplished in expanding industrial processing abilities to the new demands, mainly in terms of high performance, distributed and parallel computing, online processing, cloud computing, distributed file systems, fast and robust communications infrastructure, as well as an analysis systematization based on the specific plant knowledge (Wan et al., 2017; Basanta-Val, 2018; Geng et al., 2019).

3. INDUSTRIAL VISUAL ANALYTICS

In most industries, since research and production costs are extremely high, operation efficiency and safety are major concerns. The advances of Industry 4.0, with the exponential growth and increasing complexity of manufacturing data that hinders analytical tasks, opens up unprecedented opportunities for manufacturers to engage in data-driven science (Wu et al., 2018) and make the need for extracting useful information from data even more urgent than before (Zhou et al., 2019). The ability of processing large-scale data within a short period of time (online analysis) is therefore a current requirement in the manufacturing industry that, if satisfied, can broaden its horizons for a wide range of prognosis, diagnosis and prediction scenarios.

In this context, visualization has become an important aspect in complex data analysis, which can effectively combine machine intelligence with human intelligence to gain

insight from the data to support informed decision-making under (Zhou et al., 2019). Many scientific visualization researches within the industrial domain focus in providing both additional means for better understanding of process operation and novel visualization schemes for effectively communicating results from industrial data analysis (Al-Dabbagh et al., 2018; Bezerra et al., 2019; Dorgo et al., 2018). However, much less has been suggested as to how these industrial data visualization approaches can be made possible given the large size that commonly characterizes industrial datasets. The larger the dataset to be handled gets, the more difficult it gets to manage, analyze, and visualize data effectively (Keim et al., 2008).

With that regard, visual analytics, an approach for effective understanding, reasoning and decision making on the basis of large and complex data, is propitiated by interactive visual interfaces (Keim et al., 2008). It takes full advantage of advanced computational power and human cognitive abilities in a semi-automated analytical process, where humans and machines interact and collaborate using their respective distinct capabilities for the most effective results (Keim et al., 2008; Wu et al., 2018). It therefore represents a human-in-the-loop system in which extensive professional knowledge and domain experience are aggregated to ensure more granular and intelligent industrial data analysis (Zhou et al., 2019), sustaining a more efficient evaluation and improvement of processes and models.

Visual analytics tools must support smooth and flexible use of visualizations at rates resonant with the pace of human thought (Heer and Shneiderman, 2012) in a massive data environment where interacting with it is onerous and potentially of high latency. A valuable guidance in the area is the celebrated Shneiderman's visual information seeking mantra, namely *overview first, zoom and filter, then details on demand* (Shneiderman, 1996), describes a clearly iterative process, where users launch visual analysis, browse and navigate through interactions and highlight or select areas of the visualization for further study. Under this mantra and relying on an adequate underlying processing infrastructure, visual analytics in large-scale industrial data proves to be viable task.

4. INDUSTRIAL BIG DATA PROCESSING ENGINE

Industry is an inherently conservative sector with respect to its policies, methods and procedures. Thus, besides being a current top-priority topic for most industries, the pursuit for Industry 4.0 is a progressive and forward-looking continuous improvement process that has just started and has a long way ahead. A plausible move in this gradual approach to Industry 4.0 is to improve operational performance and governance by proactively and timely exploring the great amount of potential valuable information hidden in massive industrial data repositories such as loggers and historian servers.

As a step towards this scenario, this work proposes the establishment of a plugable industrial Big Data processing engine that is capable of dealing with those voluminous data repositories without interfering on or impacting current industrial processes and plant. This engine, hereafter referred to as Industrial Big Data Processing Engine (IBiDaPE), is a work-in-progress meant to be a process-

ing and visualization engine targeted to those industrial data repositories and to be attached to currently installed industrial information infrastructure.

IBiDaPE is primarily conceived to provide the management/planning layer (layer 3 of the classical automation pyramid), mainly composed by Process Information Management Systems (PIMS) and Manufacturing Execution Systems (MES), with the ability of extracting value from large amounts of mostly neglected data. Data representatives in this layer are (although not limited to) the logs of plant events and alarms from alarm management systems, mostly fated to be dormant in a logger SQL-based server. Figure 1 outlines the placement and role of IBiDaPE in the functional scope of the automation pyramid.

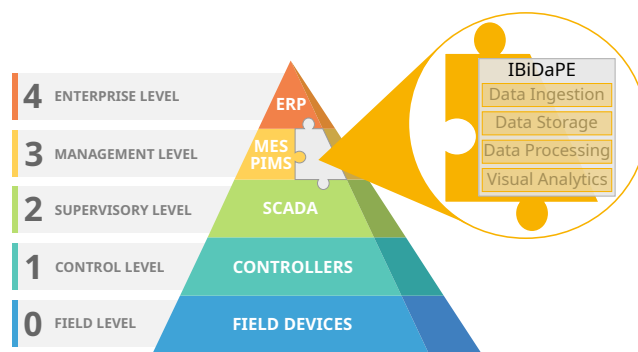


Fig. 1. IBiDaPE role in the automation pyramid.

IBiDaPE is a processing engine built on top of mainstream open-source software solutions. The gears that make up IBiDaPE rely on a locally-installed container orchestration environment based on Kubernetes (Hightower et al., 2017), a portable, extensible, open-source platform for managing containerized workloads and services in a computer cluster. Kubernetes then provides an elastic and self-healing infrastructure of containerized applications for the engine, taking care of scaling and fail-over to allow continuously and resiliently running of the whole system components as well as ensuring a more rational and flexible use of available computational resources.

IBiDaPE makes use of vertical and horizontal scaling capabilities of the supporting elastic infrastructure to establish on-demand processing capability for interactive visualization applications, enabling diverse and complex Big Data analysis and visual analytics tasks. IBiDaPE is structured in a 4-layered application stack illustrated in Figure 2 and further detailed.

4.1 Data Ingestion Layer

The Data Ingestion Layer (1) is the bridge between current installed information infrastructure and IBiDaPE. It continuously and asynchronously gathers stored data from industrial SQL-based event and alarm loggers and makes it available in the distributed file system of the engine. It uses *Sqoop* (Oussous et al., 2017) as data ingestion tool to automate bi-directional data transfer between the Hadoop Distributed File System (HDFS) (Shvachko et al., 2010), a fast and reliable distributed file system, further on explained, and conventional relational databases.

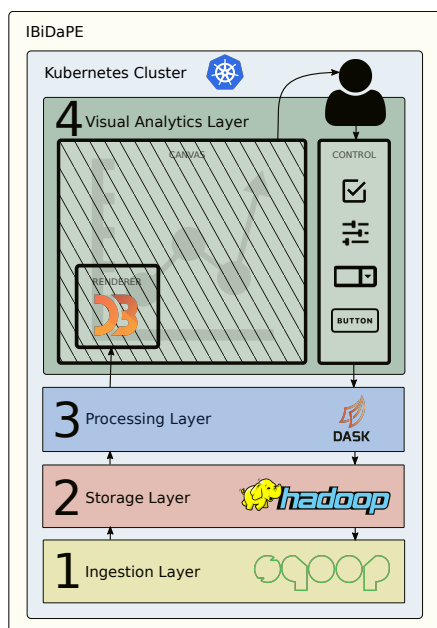


Fig. 2. IBiDaPE architecture.

4.2 Data Storage Layer

The Data Storage Layer (2) depends on a fast and reliable data storage system capable of handling voluminous data supporting replication and redundancy for fault tolerance. It relies on HDFS, a wide adoption distributed file system that met the requirements of the layer, to speed up file system jobs over massive volumes of both structured and unstructured data. HDFS is designed for high-latency batch processing operations and has great portability and scalability across commodity hardware and software platforms (Oussous et al., 2017). *MapReduce* programming model is employed in file system operations to ensure effective and timely access to data, reducing network congestion and increasing system performance by moving computations near to data storage (Oussous et al., 2017). HDFS then provides redundant, fault tolerant and fast data access for the Data Processing Layer of IBiDaPE.

4.3 Data Processing Layer

Data Processing Layer (3) implements a parallel processing cluster to speed up general computations over massive data volumes and counts on *Dask* (Daniel, 2018), a flexible library for distributed and parallel computing written in Python language. *Dask* is suited to solve a wide variety of large data handling and analysis problems in scientific computing and general-purpose distributed computing. It can natively scale other well-established Python libraries which comprise the Python Open Data Science Stack (PODSS) such as *NumPy*, *Pandas*, and *Scikit-Learn* (Daniel, 2018) to allow an easy switching from single-node to cluster computing power as data size scales.

In *Dask*, directed acyclic graphs (DAGs) are used to compose, express and manage the execution of parallel computations (Daniel, 2018). *Dask* then endows IBiDaPE with the ability of splitting and distributing computing tasks demanded by the applications at the top of the stack across the nodes of a parallel processing cluster.

4.4 Visual Analytics Layer

The Visual Analytics Layer (layer 4) leverages the processing infrastructure provided by the lower layers to allow building graphically-rich and large-data interactive visualization applications. As such, massive data querying, handling and processing operations demanded by each class of interactive visualization application are structured in distributed file system and parallel processing operations, triggered to deliver the desired results at a fast rate that is expected from an interactive visual application.

An interactive visualization in this layer is a 3-component application as also shown in Figure 2. The *Canvas* component uses Hyper-text Markup Language (HTML) to draw graphics and animations in a two-dimensional grid. The *Control* component is responsible for capturing user interactions with the application and turning them into processing queries to the lower layers. Note that depending on each interactive visualization specificities a different set of intermediary queries, temporary tables and data buffers may be required. Finally, the *Renderer* component takes the results from those processed user-input interactions and manipulates the *Canvas* component to produce their respective visualizations.

5. CASE STUDY: AN INTERACTIVE VISUAL ANALYTICS APPLICATION

5.1 Target Data

Target data in this case study is sourced from a production industrial alarm and event SQL-based log database of a petrochemical plant. Composed of time-stamped records, the database reflects episodes regarding the operational dynamics of industrial plant assets such as sensors, actuators, controllers, transceivers, among others. From the source database, a snapshot of about 10^8 entries was selected for this case study.

Table 1 shows some representative sample entries of the target dataset in tabular format with some columns omitted because of limited horizontal space and data sanitized to safeguard sensitive information.

5.2 Sunburst Interactive Diagram

Assets in an industrial plant are arranged according to a hierarchy that may be reflected in those alarm and event logs. An hierarchy-awareness vision over plant episodes favors a better reconnaissance of plant dynamics. Hierarchy can be more easily understood in a quantitative and qualitative manner with the Sunburst diagram, a space-filling native interactive visualization technique in which items in a hierarchy are laid out radially, with the top of the hierarchy at the center and deeper levels farther away from the center (Stasko et al., 2000).

Sunburst diagram is setup so that each section arc in a ring refers to a plant entity and has size proportional to the frequency of the hierarchy entity appearance in the logs. In this diagram, quantitative and qualitative aspects of the 4-level hierarchy of plant assets referred in target data, defined by columns *Area*, *Subarea*, *Node* and *Module* are made explicit as shown in Figure 3.

Table 1. Sample entries of dataset under analysis.

Date_Time	Type ^a	Area ^b	Subarea ^b	Node ^b	Module ^b	...	Level ^a	Desc1 ^c	registry_id ^d
19-02-21 14:16:04	EVENT	AREA_01	SUBAREA_C	NODE_06	MDL-015783HIC	...	NaN	RATE	18794
19-02-23 08:45:36	EVENT	AREA_25	SUBAREA_E	NODE_04	MDL-414572PID	...	INFO	ERROR CLEAR	12946
19-02-22 12:21:32	ALARM	AREA_12	SUBAREA_F	NODE_17	MDL-45420VEC	...	ADVISORY	COMM FAIL	5948
19-02-23 15:38:42	ALARM	AREA_07	SUBAREA_G	NODE_02	MDL-1557452DI	...	CRITICAL	HIHI	6397
19-02-22 16:41:19	EVENT	AREA_23	SUBAREA_M	NODE_21	MDL-1029ASMB	...	NaN	MODBAD	24883

^a Registry attributes; ^b Registry entities; ^c Registry description; ^d Registry unique identification.

The application permits the user to navigate through the plant hierarchy tree translated into the Sunburst diagram. Each arc section in the diagram is drawn from a contingency table (frequency distribution table) calculated on-the-fly from base data at each user interaction. The main insight gained from this kind of diagram concerns the importance (in this case, given by the frequency of appearance) of different plant entities at different hierarchy levels in the logs. Users can also interact with internal control components of the graph such as limiting the levels of the hierarchy or the percentage of the dataset under analysis.

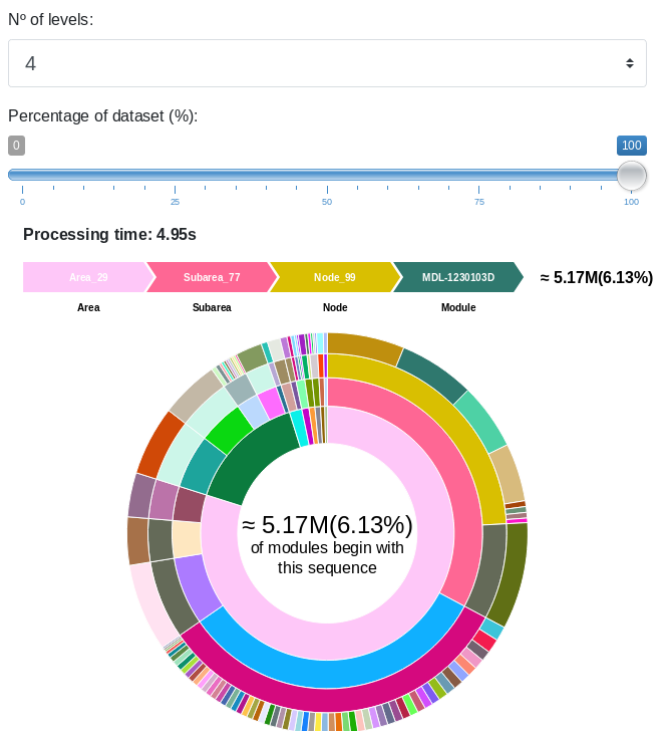


Fig. 3. Sunburst Interactive Diagram.

5.3 Experimental Setup

Containerized applications in this experiment used computational resources made available by a 6-node Kubernetes cluster composed of identical 8-core Intel Xeon machines with 16 GB of RAM memory each. From now on, architecture elements referred as nodes correspond directly to containers running on the Kubernetes cluster.

The experiments were conducted in a fixed but enough configuration for the Ingestion (1 node running *Sqoop*) and Storage (4 HDFS datanodes under the default replication setting) layers, since fine tuning for these layers proved not to significantly affect the processing performance of the visual analytics application. The Processing Layer is

then set up with identical *Dask* nodes allowed to make full usage of Kubernetes node resources, thence running one processing thread per available core and using near all available memory. The cluster is horizontally scaled, starting from a minimum working set of nodes (3 nodes) to the maximum set of nodes available for the experiment (5 nodes).

Thus, the considered experimental variables were the number of *Dask* nodes, with values in {3, 4, 5}, and the fraction of dataset size expressed in percentages, with values {25, 50, 75, 100}. The diagram application was executed 30 times under each combination of those variables and the running time of each execution was recorded.

5.4 Results

The average running times for each configuration are expressed in Figure 4, alongside with the corresponding confidence intervals represented by error bars (95% of confidence level). The main observation about these results is that the proposed architecture, under the considered configuration, enabled the interactive visualization of about 10^8 entries of industrial categorical data in approximately 5s per interaction loop, an acceptable response time for scenarios of interactive visual analysis of massive data.

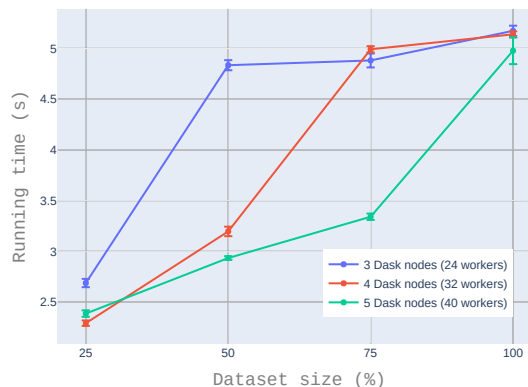


Fig. 4. Running times per dataset size and number of Dask worker nodes.

Another important fact is that the intrinsic overhead of the proposed architecture – mainly caused by network latency and memory management, common in distributed computing scenarios – hinders the performance when dealing with quantities of data entries about 25% of the dataset. This phenomena is also observable for amounts of data exceeding about 100% of the dataset, indicating that a more robust configuration of the computing cluster is necessary, as occurred for the cases of 50% and 75% of the dataset, in which increasing the cluster size in one node caused a great performance impact.

6. CONCLUSION AND FUTURE WORKS

This paper discussed the use of interactive visual analytics in industry to support systematic reasoning. The proposed solution demonstrated to be a simple and flexible approach in providing the capability of timely analyzing and visualizing large industrial data volumes to current installed industrial information infrastructure. The Sunburst diagram, implemented as a case study, showed that the proposed architecture can enable diverse visual analytics applications over large volumes of industrial data.

It is opportune to clarify that although IBiDaPE was conceived and tested as an attachable engine to the management level, the approach can be extended to be used in other levels of the automation pyramid. As well, many other visual analytics applications can be implemented according to the analysis needs for those levels, relying on the underlying processing infrastructure.

Work on fine-tuning the building blocks of the solution is a subsequent step towards improving its performance and reliability. Future work also includes endowing the architecture with the ability of dynamically and automatically scaling the elastic containerized infrastructure according to the size and complexity of the demanding visualization applications.

REFERENCES

- Al-Dabbagh, A.W., Hu, W., Lai, S., Chen, T., and Shah, S.L. (2018). Toward the Advancement of Decision Support Tools for Industrial Facilities: Addressing Operation Metrics, Visualization Plots, and Alarm Floods. *IEEE Transactions on Automation Science and Engineering*, 15(4), 1883–1896.
- Bartodziej, C.J. (2017). *The Concept Industry 4.0*. Springer Fachmedien Wiesbaden, Berlin.
- Basanta-Val, P. (2018). An Efficient Industrial Big-Data Engine. *IEEE Transactions on Industrial Informatics*, 14(4), 1361–1369.
- Bezerra, A., Silva, I., Guedes, L.A., Silva, D., Leitão, G., and Saito, K. (2019). Extracting Value from Industrial Alarms and Events: A Data-Driven Approach Based on Exploratory Data Analysis. *Sensors (Basel, Switzerland)*, 19(12), 2772.
- Bikakis, N., Papastefanatos, G., Skourla, M., and Sellis, T. (2016). A Hierarchical Aggregation Framework for Efficient Multilevel Visual Exploration and Analysis. *Semantic Web*, 8(1), 139–179.
- Brian Hartmann, W.P.K. and Narayanan, S. (2015). Digital manufacturing: The revolution will be virtualized. Available at <https://www.mckinsey.com>.
- Daniel, J.C. (2018). *Data Science at Scale with Python and Dask*. Manning Early Access Program. Manning.
- Diez-Olivan, A., Del Ser, J., Galar, D., and Sierra, B. (2019). Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. *Information Fusion*, 50, 92–111.
- Dorgo, G., Varga, K., and Abonyi, J. (2018). Hierarchical Frequent Sequence Mining Algorithm for the Analysis of Alarm Cascades in Chemical Processes. *IEEE Access*, 6, 50197–50216.
- Geng, D., Zhang, C., Xia, C., Xia, X., Liu, Q., and Fu, X. (2019). Big Data-Based Improved Data Acquisition and Storage System for Designing Industrial Data Platform. *IEEE Access*, 7, 44574–44582.
- Godfrey, P., Gryz, J., and Lasek, P. (2016). Interactive visualization of large data sets. *IEEE Transactions on Knowledge and Data Engineering*, 28(8), 2142–2157.
- Gokalp, M.O., Kayabay, K., Akyol, M.A., Eren, P.E., and Kocyigit, A. (2016). Big Data for Industry 4.0: A Conceptual Framework. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, 431–434. IEEE.
- Heer, J. and Shneiderman, B. (2012). Interactive dynamics for visual analysis. *Communications of the ACM*, 55(4), 45.
- Hightower, K., Burns, B., and Beda, J. (2017). *Kubernetes: Up and Running: Dive Into the Future of Infrastructure*. O'Reilly Media.
- Keim, D., Andrienko, G., Fekete, J.D., Görg, C., Kohlhammer, J., and Melançon, G. (2008). *Visual Analytics: Definition, Process, and Challenges*, 154–175. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Limet, S., Poquet, M., and Robert, S. (2014). Modulight: A Framework for Efficient Dynamic Interactive Scientific Visualization. *Procedia Computer Science*, 29, 692–702.
- Obitko, M., Jirkovský, V., and Bezdíček, J. (2013). Big Data Challenges in Industrial Automation. In V. Mavřík, J.L.M. Lastra, and P. Skobelev (eds.), *Industrial Applications of Holonic and Multi-Agent Systems*, 305–316. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Oussous, A., Benjelloun, F.Z., Ait Lahcen, A., and Belfkih, S. (2017). Big Data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences*, 30(4).
- Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*, volume 101, 336–343. IEEE, Maryland.
- Shvachko, K., Kuang, H., Radia, S., and Chansler, R. (2010). The hadoop distributed file system. In *Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, MSST'10. IEEE Computer Society, Washington, DC, USA.
- Stasko, J., Catrambone, R., Guzdial, M., and McDonald, K. (2000). Evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human Computer Studies*, 53(5), 663–694.
- Wan, J., Tang, S., Li, D., Wang, S., Liu, C., Abbas, H., and Vasilakos, A.V. (2017). A Manufacturing Big Data Solution for Active Preventive Maintenance. *IEEE Transactions on Industrial Informatics*, 13(4), 2039–2047.
- Wu, W., Zheng, Y., Chen, K., Wang, X., and Cao, N. (2018). A Visual Analytics Approach for Equipment Condition Monitoring in Smart Factories of Process Industry. In *2018 IEEE Pacific Visualization Symposium (PacificVis)*, volume 2018-April, 140–149. IEEE.
- Zhou, F., Lin, X., Liu, C., Zhao, Y., Xu, P., Ren, L., Xue, T., and Ren, L. (2019). A survey of visualization for smart manufacturing. *Journal of Visualization*, 22(2), 419–435.