

Understanding the End-Users and Technical Requirements for Real-Time Streaming Data Analytics and Visualisation

Fateme Dinmohammadi ¹, Duncan Wilson ¹

¹The Bartlett Centre for Advanced Spatial Analysis (CASA)
University College London (UCL)
Gower Street, London WC1E 6BTL, United Kingdom
Email: f.dinmohammadi@ucl.ac.uk

Abstract—Today’s digital world deals with tremendous amounts of data with high complexity and velocity. The data being generated continuously from a variety of sources at very high speeds is considered as streaming data. The number of technologies to ingest, explore, analyze, and visualize the streaming data is growing rapidly across many industries. However, the use of real-time technologies in the Architecture, Engineering & Construction (AEC) sector is still at its early stages of development and further research is needed to make the tools truly commercial. The AEC is a data-intensive, project-based industry that relies on real-time information to help projects stay on schedule and enable companies to cut costs. The technologies supporting real-time data capability in the AEC sector must satisfy several key functional and nonfunctional requirements. This paper aims to identify and prioritize both end-users and technical requirements for AEC businesses when adopting a real-time streaming data analytics and visualization technology in their production control rooms. To this aim, an expert questionnaire survey is conducted to investigate the requirements in four distinct areas, namely: (i) data collection and ingestion, (ii) stream processing, (iii) real-time analytics, and (iv) presentation and visualization. A five-level Likert scale is used in the questionnaire to allow respondents to indicate their perceptions on the importance of each of the identified requirements. The survey is distributed using Qualtrics online tool to over twenty organizations all over the world and the responses are analyzed using SPSS statistical software. The findings of the study reveal that ‘interactivity’ and ‘user assistance’ are the most important requirements from the end-users’ perspective, whereas ‘fast data processing’, ‘real-time reporting’, ‘scalable data processing’, and ‘dashboard connectivity’ are the most important requirements from the technology developers’ perspective. The findings of this research can help stakeholders and researchers get a better picture of the current challenges facing the AEC sector in terms of streaming data analytics and provide insight into the possible directions of future technological changes.

Keywords—Streaming data; Real-time analytics; Data visualization; User requirement; Architecture, Engineering & Construction (AEC).

I. INTRODUCTION

Nowadays, a tremendous amount of data is being generated at very high speeds in a variety of formats such as documents, spreadsheets, images, audios, or video files. This data is collected, often on a continuous basis, from a

wide range of sources including online transactions, mobile applications, sensors, Internet of things (IoT) devices, video-capturing systems, etc. The data being generated continuously from multiple sources is often considered as *streaming data* [1]. The amount of streaming data in the world is growing so rapidly, doubling approximately every two to three years. For instance, the amounts of audio and video streaming data transferred to mobile devices are growing at 34% and 50% Compounded Annual Growth Rate (CAGR), respectively [2]. Extracting meaningful and useful insights out of such data is highly motivated by many application domains. To make timely and efficient decisions, the data streams must be processed, analyzed, and visualized as and when they arrive (i.e., in real-time) rather than being stored and processed offline. The capture, processing, analysis, and visualization of streaming data in real-time is referred to as *stream processing* or *real-time analytics* [3].

The need for real-time streaming data management technologies is more evident than ever before. Stream processing is more complex and often more expensive than traditional batch processing, since the data arrives continuously and is delivered in a wide variety of formats. Therefore, more efficient tools are needed to deal with problems inherent to stream processing. In recent years, the number of technologies to ingest, explore, analyze, and visualize the streaming data has grown considerably across many industries. Several stream processing technologies such as Apache Storm [4], Apache Spark [5] and Apache Flink [6] have been deployed in many businesses to process their data faster and more effectively. Nevertheless, some organizations are still seeking investment or tackling the challenges of communicating the benefits of real-time data analytics and visualization to their stakeholders.

The architecture, engineering, and construction (AEC) is one of the greatest industries in the world, accounting for about 11 percent of global GNP (gross national product) [7]. The AEC is the second largest sector in Norway in terms of revenue and number of employees and the third largest sector of the American economy after government and healthcare [8]. The AEC sector is a data-intensive, project-based industry, involving multiple parties such as designers, contractors, suppliers, constructors, and owners who work together to complete the projects in a timely manner. The industry has undergone a dramatic digital transformation through the adoption of advanced technologies such as Building Information Modelling

(BIM), Internet of Things (IoT), big data analytics and machine learning (ML). These technologies have led to the generation of large amounts of datasets from many construction sites on a real-time or near real-time basis.

The definition of real time in the AEC sector varies from system to system and project to project, depending on how long elapses between data recording and their automated processing. This time ‘latency’ often ranges from a few minutes in applications such as project scheduling and monitoring, to milliseconds in crane loading/unloading operations [9]. Processing and analyzing the data collected from construction projects in real time can help the planners estimate actual construction time, assess productivity and cost of activities, and identify resource bottlenecks. Despite all the potential benefits, the use of real-time technologies in the AEC sector is yet at its early stages of development and further research is needed to make the tools truly user-friendly and technically robust.

The technologies supporting real-time data capability must satisfy several key requirements in order to be of use in generating insights and visualizing data. These requirements are broadly categorized into three groups: generic requirements, functional requirements, and nonfunctional requirements [10]. The generic requirements are those that are common across all users and in all projects. Functional requirements define what a system/product should do and what features and capabilities are expected from it. Non-functional requirements define the quality attribute of a system/product and specify how well it should perform a task. Understanding these requirements and their implications is the key first step toward successful implementation of a real-time data analytics and visualization technology.

A review of the literature shows that there is no study conducted to ‘systematically’ explore the requirements for adopting real-time stream processing in the AEC sector. To overcome this research gap, this study conducts an expert questionnaire survey with the aim of identifying and prioritizing both end-users and technical requirements for a real-time data analytics and visualization technology in production control rooms. Sixteen requirements across four distinct areas of data collection and ingestion, stream processing, real-time analytics, and presentation and visualization are identified. The experts are then asked to indicate their perceptions on the importance of each requirement using a five-level Likert scale. The survey is distributed to over twenty AEC businesses worldwide, and the responses are collected via an online tool called Qualtrics XM [11]. The results of the survey are analyzed using IBM SPSS Statistics software [12]. ‘Interactivity’ and ‘user assistance’ are identified as the most important requirements from the end-users’ perspective, whereas ‘fast data processing’, ‘real-time reporting’, ‘massively scalable data processing’, and ‘dashboard connectivity’ are found to be the most important requirements from the technology developers’ perspective.

The organization of the remainder of the paper is as follows. In the next section, the methodology of the research is described. Section 3 presents both end-users and technical requirements. Section 4 reports the results and discusses the findings. Finally, section 5 concludes the paper and gives suggestions for further work.

II. THE METHODOLOGY OF RESEARCH

The methodology of this research is based on a comprehensive literature review of user requirements in general, and a questionnaire survey in the AEC sector in particular. According to the latest scientific research [13], a real-time stream processing technology contains four key layers. These layers from ‘top to bottom’ are shown in Figure I and explained below:

Layer 1 — Presentation and visualization: This layer displays the insights in real time on screens in production control room to ‘end users’ or ‘clients’ who may be given access to such information.

Layer 2 — Data analytics: This layer lies underneath the presentation and visualization layer. In this layer, we apply logic and mathematics to data to provide insights into the construction project team.

Layer 3 — Stream processing: This layer lies underneath the data analytics layer. In this layer, we run continuous queries to filter, transform, mask, aggregate and enrich the data, to make it ready for analytics.

Layer 4 — Data collection and ingestion: This layer lies underneath the stream processing layer. In this layer, we collect data from various sources such as transactional databases, application logs, message queues, sensors, etc. and ingest them into a data lake.

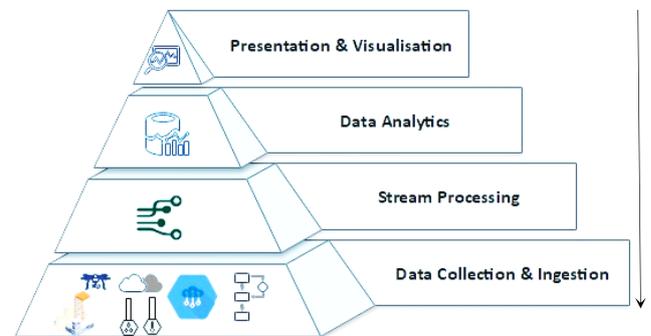


Figure I. Four layers of a real-time stream processing technology.

Following the four-layer model shown in Figure 1, we developed a questionnaire to elicit experts’ opinions about both types of end-user and technical requirements (including those that are functional, nonfunctional or generic) for adopting a real-time data analytics and visualization in the AEC production control rooms [14]. Our questionnaire included five separate parts, which are explained briefly below:

(i) General information: In this part, the respondents are asked to specify if they work with ‘data’ or not, and if yes, in what ‘capacity’ they do work with data. The answer options included: data collection, data analysis, data interpretation, data reporting, and data visualization.

In another question, the respondents are asked if their businesses have already deployed any real-time data technology or not; and, if not, whether they are planning to acquire a real-time data technology in the future. The answer options included: ‘we are yet studying about it’, ‘we are yet preparing for it’ and ‘we are implementing it’.

We also asked the experts about their experience with the speed at which data is created and processed in the AEC sector. Five ranges of frequency for data processing were

included in the response option, including: (i) every microsecond to millisecond (10^3 to 10^6 Hz), every millisecond to second (1 to 10^3 Hz), every second to minute (0.016 to 1 Hz), every minute to hour (0.00027 to 0.016 Hz), and longer than every hour (less than 0.00027 Hz).

(ii) Requirements for data collection and ingestion: In this part, the businesses are asked to specify what ‘sources’ of real-time data they have available (e.g. building sensors data, environmental sensors, smart meters, IOT sensors). We then asked the businesses about the ways they do collect and ingest information from construction sites, whether this is done by manual coding or automated tools (e.g. Actian, Diyotta, Fivetran, Informatica, Matillion, Striim, Syncsort, Talend and Tamr) [15].

In the last question of this part, the respondents are asked to identify and prioritize the requirements of data collection and ingestion on a 5-level Likert scale [16]. The answer options included: 1= not important, 2= slightly important, 3= moderately important, 4= high important, and 5= very high important.

(iii) Requirements for stream processing: In this part, the businesses are asked if they store their data and process it later offline or process the data as and when it arrives. Also, we asked them to specify any commercial tools which they use in their businesses to process the data in real-time, e.g., Apache Kafka, S4, Apache Flink and Samza [17].

In the last question of this part, the respondents are asked to identify and prioritize the requirements of stream data processing on a 5-level Likert scale.

(iv) Requirements for real-time analytics: In this part, the respondents are asked to identify and prioritize the requirements of real-time data analytics on a 5-level Likert scale.

(v) Requirements for real-time visualization: In the last part of the questionnaire, the respondents are asked to identify and prioritize the requirements of real-time data presentation and visualization on a 5-level Likert scale.

The questionnaire was then converted to an online version, enabling the survey to be completed online. A Qualtrics link was created and distributed to AEC businesses within different categories, e.g., house building and renovation, heavy industrial construction, institutional and commercial construction, infrastructure construction, etc. Figure II shows a screenshot of our questionnaire designed to identify and prioritize requirements for real-time data analytics and visualization in the AEC sector.

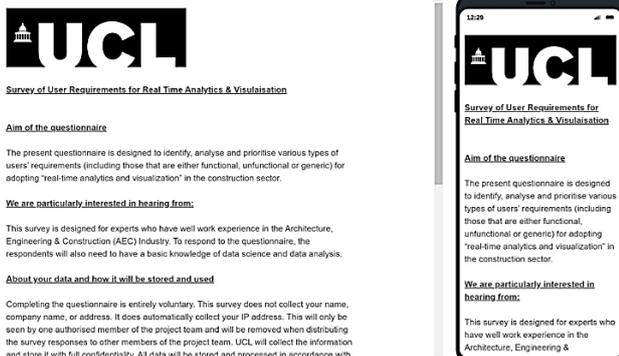


Figure II. A screenshot of our questionnaire.

III. THE END-USER AND TECHNICAL REQUIREMENTS

After reviewing the literature and consultation with the AEC businesses, we identified sixteen requirements across four layers of data presentation and visualization, data analytics, stream processing, and data collection and ingestion. These requirements are explained in detail in the following subsections.

A. Presentation and visualisation

Interactivity with end-users: This requirement represents the ability for an end-user to intuitively comprehend the data, perceive the underlying patterns, and interact with information displayed on control room screens [18].

Scalability / readability: This requirement represents the scalability and readability of real-time data across all screen sizes and in all contexts [18].

Fast retrieval: This requirement represents how fast the production control room retrieves data so that the end-users do not lose their focus and momentum when visually analyzing the information on screens.

User assistance: This requirement represents how user friendly and convenient the production control system is.

B. Data analytics

Unlimited data storage: This requirement represents the capability of the real-time technology in storing millions of transactions, ideally in a single repository.

Decision updating: This requirement refers to capability of the real-time data technology in incorporating new data samples and updating predictions.

Real-time reporting: This requirement represents the capability of the real-time data technology in recording and reporting events as and when they occur.

Dashboard connectivity: This requirement represents the support provided for open connectivity standards, including Java Database Connectivity (JDBC) and Open Database Connectivity (ODBC), to real-time dashboards.

C. Stream processing

Low message queuing time: This requirement refers to the time from when a data sample is published to the server until the processing of the data begins.

Fast data processing: This requirement represents how fast a data sample is processed.

Guaranteed only once processing: This requirement refers to how effective is the technology in processing every data sample ‘only once’.

Massively scalable: This requirement refers to the capacity of the processing unit.

D. Data collection and ingestion

High-velocity data capture: This requirement represents the ability of the AEC businesses to capture high-velocity data streams from multiple sources embedded in construction sites, such as sensors and IoT devices.

Highly reliable data capture: This requirement represents the probability that the data capture systems (e.g. IoT devices) will be operational at any time during the project

lifecycle. This requirement is measured in terms of ‘reliability’, i.e., the probability that the data collection devices performs correctly for a given period of time.

High data transfer rate: This requirement represents how fast the data can be transferred from the devices embedded in a construction site to the server. The data transfer rate may range from a few kilobytes per second to hundreds of megabytes per second in the AEC sector.

Automated data integration: This requirement represents how AEC businesses combine real-time data streams with existing reference data from databases or other data sources.

In addition to the above sixteen requirements, there are a number of recommended and optional elements in a real-time production control room which may not be required by every business in the AEC sector to acquire. These desirable elements are excluded from our analysis in the following.

IV. RESULTS AND DISCUSSION

In this section the results of the survey are presented in the form of descriptive statistics. In total, 29 responses were received from different AEC businesses in seventeen countries around the world. These responses were analyzed and are reported below.

In terms of the capacity in which respondents do work with data, 31% (i.e. 9 respondents) have responded that they are involved in data analysis, 20.7% (i.e. 6 respondents) in data collection, 17.2% (i.e. 5 respondents) in data reporting, 13.8% (i.e. 4 respondents) in data visualization, and 10.3% (i.e. 3 respondents) in data interpretation. The remaining respondents have responded that they work with data in other capacities such as data cleansing, synchronization, integration, etc. These results are shown in Figure III. This gave us confidence that the participants had good experience of working with different aspects of data science in the AEC sector.

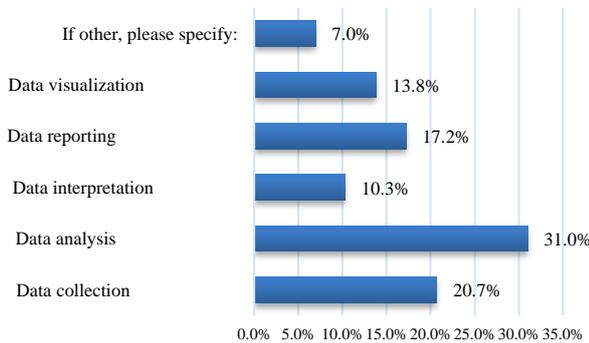


Figure III. The capacity in which respondents do work with data.

55.2% of the respondents (i.e. 16 respondents) have responded that their businesses already deployed a real-time data technology, 13.8% have responded that their businesses are studying about it, 20.7% are preparing for it, and 10.3% are implementing it.

In terms of the speed at which data is created and processed in the AEC sector, 20.7% businesses have responded that the interval with which data samples are recorded is longer than an hour, 55.2% businesses have responded from a minute to an hour, 13.8% businesses have

responded from a second to a minute, and 10.3% businesses have responded from a microsecond to a millisecond. These results are shown in Figure IV.

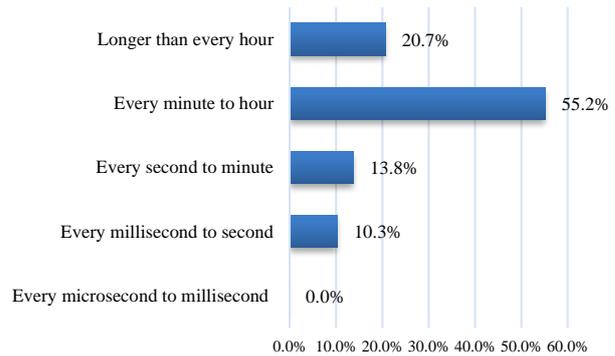


Figure IV. The frequency of real-time in the AEC sector.

In terms of sources of real-time data in the AEC sector, 10.3% have responded that their data is collected from building sensors, 37.9% from environmental sensors, 27.6% from IOT devices and 24.2% responded others. These percentages are shown in Figure V.

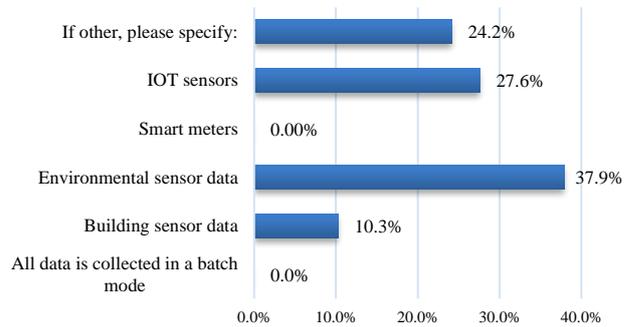


Figure V. The sources of real-time data in the AEC sector.

The percentages of responses on the relative importance of each of the identified requirements based on a 5-level Likert scale are given in Figure VI.

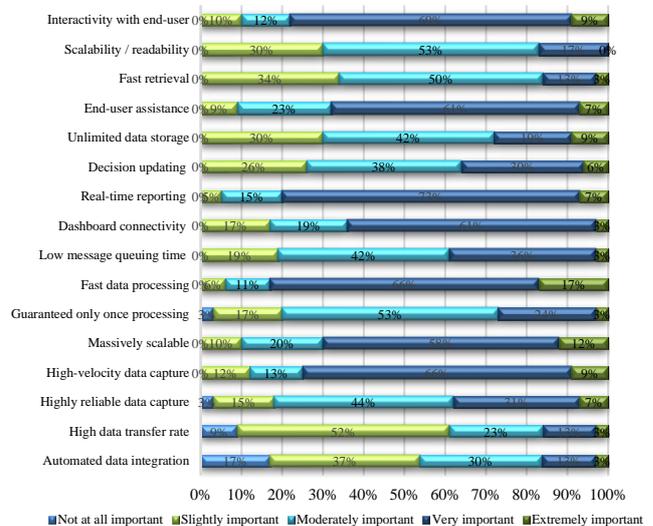


Figure VI. The responses to the 5-level Likert scales.

By aggregating the percentages of the responses from Figure VI, the relative importance of each of the identified requirements is obtained. The priorities of the sixteen requirements based on a 5-level Likert scale are presented in Table I.

Table I. The priorities of the sixteen requirements.

No.	Requirement	Priority				
		1	2	3	4	5
1	Interactivity with end-user	3.77				
2	Scalability / readability	2.87				
3	Fast retrieval	2.85				
4	End-user assistance	3.66				
5	Unlimited data storage	3.07				
6	Decision updating	3.16				
7	Real-time reporting	3.82				
8	Dashboard connectivity	3.50				
9	Low message queuing time	3.23				
10	Fast data processing	3.94				
11	Guaranteed only once processing	3.07				
12	Massively scalable	3.72				
13	High-velocity data capture	3.48				
14	Highly reliable data capture	3.24				
15	High data transfer rate	2.49				
16	Automated data integration	2.49				

As can be seen, ‘interactivity’ and ‘user assistance’ are the most important requirements of the end-users in the AEC production control rooms, whereas ‘fast data processing’, ‘real-time reporting’, ‘scalable data processing’, and ‘dashboard connectivity’ are the most important requirements from the technology developers’ perspective. In order to improve the data processing speed in production control rooms, some new technologies can be deployed. A few of these solutions are as follows:

Processing In Memory (PIM) — This solution reduces time ‘latency’ by integrating the processor in a memory chip.

In-Database Analytics — In this solution, data processing happens within the database and the analytic logic is also built into the database.

In-Memory Analytics — This solution queries data in random access memory (RAM) instead of physical disks.

Massively Parallel Programming (MPP) — In this solution, multiple processors tackle different parts of a program and each processor has its own operating system and memory.

In order to process and analyze the data in real-time, it is crucial for the AEC sector to shift towards using automated data collection and aggregation systems. A real-time streaming analytics solution must be able to ingest data from a variety of sources and offer the performance and scalability needed to analyze that data in real-time. The data formatting should also be reviewed to ensure that data from multiple sources can be aggregated seamlessly.

The volume of unstructured data (e.g. BIM images, CCTV videos) is growing many times faster than the volume of structured data. In order to process and analyze data in real-time, the AEC sector must be able to handle and analyze unstructured data in real time. This effectively means that they must use artificial intelligence (AI) – and natural language processing (NLP) – based technology.

The AEC businesses must also have systems or processes enabling them to take immediate action (or update quickly their taken action plans) based on new information arriving from various construction sites. A streaming analytics solution should be a general-purpose solution that can be integrated with, or embedded into, other applications such as crane loading / unloading, construction progress monitoring, project cost monitoring, etc.

One additional factor to consider when adopting a real-time data technology in the AEC production control rooms is ‘ease of use’. The real-time analytics is not used only by data scientists and business analysts. All construction project stakeholders such as clients, consultants, the main contractor, engineers, subcontractors and end-users may use the insights obtained and the information visualized on screens in production control rooms. Key capabilities needed to make streaming analytics available to such decision-makers include built-in data visualization and real-time dashboards.

V. CONCLUSIONS

The need for a real-time streaming data analytics and visualization technology is growing continuously in the AEC sector. Many businesses are making investments to find new ways of extracting actionable insights from their real-time data streams. These data streams are collected on a continuous basis from a wide range of sources, e.g. sensors, IoT devices, video-capturing systems, etc. across different construction projects. Processing and analyzing such data in real time can help construction managers identify whether a project is ahead or behind schedule, over or under budget, and whether it meets quality criteria or not.

The technologies supporting real-time data capability in the AEC sector must satisfy a number of functional, nonfunctional, and generic requirements. Understanding these requirements and their implications is the first step in implementing a real-time data analytics and visualization technology. To the best of our knowledge, there was no study in the literature reporting the requirements for adopting a real-time stream processing in the AEC production control rooms.

To identify and prioritize these requirements from both end-users and technical perspectives, we conducted a questionnaire survey of AEC businesses operating in different countries around the world. A top-down model with four layers of data collection and ingestion, stream processing, real-time analytics, and presentation and

visualization was proposed to identify the requirements. A five-level Likert scale was defined to allow respondents to indicate their perceptions on the relative importance of each of the identified requirements. The survey was distributed using Qualtrics online tool and the responses were analyzed using statistical software packages.

The findings of the study revealed that the six requirements of ‘fast data processing’, ‘real-time reporting’, ‘interactivity with end-user’, ‘massively scalable stream processing’, ‘end-user assistance’, and ‘dashboard connectivity’ are the most important requirements and will need more attention from the AEC businesses. The results of this study can help the AEC sector develop a strategy roadmap for adoption and implementation of real-time streaming data analytics and visualization technologies, tools and processes in a project control room system built on live data feeds.

ACKNOWLEDGEMENT

This work was supported by an Innovate UK grant (Ref. 106169) awarded to Mace Limited, University College London, Evifile Limited, 3D Repo Ltd, Imperial College London, and Mission Room Limited.

REFERENCES

- [1] B. Ellis, “Real-time analytics: Techniques to analyze and visualize streaming data”, ISBN: 978-1-118-83791-7, 432 Pages, 2014.
- [2] J. Poderys, J. Farooq, and J. Soler, “A novel multimedia streaming system for urban rail environments using Wi-Fi peer-to-peer technology”, In: Proceedings of IEEE 87th Vehicular Technology Conference, 3–6 June 2018, Porto, Portugal.
- [3] Z. Milosevic, W. Chen, A. Berry, F.A. Rabhi, “Real-Time Analytics”, In: Big data, principles and Paradigms. Chapter 2, pp. 39-61, 2016.
- [4] Apache Storm, <https://storm.apache.org/>
- [5] Apache Spark, <https://spark.apache.org/>
- [6] Apache Flink, <https://flink.apache.org/>
- [7] PricewaterhouseCoopers (PwC). Growth markets insights. March 2018. Available online: <https://www.pwc.com/gx/en/government-public-services/assets/pwc-gmc-growth-markets-insights.pdf>
- [8] E. Hjelseth, “Public BIM-based model checking solutions: Lessons learned from Singapore and Norway”, WIT Transactions on The Built Environment, vol. 149, 421–436, 2015.
- [9] T. Cheng, J. Teizer, “Real-time resource location data collection and visualization technology for construction safety and activity monitoring applications”, Automation in Construction, vol. 34, pp. 3–15, September 2013.
- [10] I. Blanquer, N. Puchalski, N. Andrade, “D7.1: End-User Requirements Elicitation”, In: Europe - Brazil Collaboration of BIG Data Scientific Research through Cloud-Centric Applications, 2016.
- [11] Qualtrics XM, <https://www.qualtrics.com/uk/>
- [12] IBM SPSS Statistics <https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-27>
- [13] Microsoft Azure. <https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-introduction>
- [14] J. Zhao, H. Olivieri, O. Seppänen, A. Peltokorpi, B. Badihi, P. Lundström, “Data analysis on applying real time tracking in production control of construction”, In: IEEE International Conference on Industrial Engineering and Engineering Management, 10-13 Dec. 2017, Singapore.
- [15] Informatica, “Change data capture for real-time data ingestion and streaming analytics”, White Paper https://www.informatica.com/content/dam/informatica-com/en/collateral/white-paper/change-data-capture-for-real-time-data-ingestion-and-streaming-analytics_white-paper_3914en.pdf
- [16] P. Bhandari, “Designing and analyzing Likert scales”, July 3, 2020. <https://www.scribbr.com/methodology/likert-scale/>
- [17] Xenonstack, “Real Time Data Streaming: Introduction, Tools, Platform and Technologies”, March 19, 2019. Available at: <https://www.xenonstack.com/insights/real-time-data-streaming/>
- [18] J. Lowe, M. Mathee, “Requirements of data visualisation tools to analyse big data: A structured literature review”, In: Conference on e-Business, e-Services and e-Society (I3E 2020): Responsible Design, Implementation and Use of Information and Communication Technology, 6-8 April 2020, Skukuza, South Africa, pp 469-480.