Heliyon 10 (2024) e24741

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Improving digital twin experience through big data, IoT and social analysis: An architecture and a case study

Rubén Alonso [a,c], Riccardo Locci [b], Diego Reforgiato Recupero [a,b,*]

[a] ICT Division, R2M Solution s.r.l., Via Fratelli Cuzio, 42, Pavia, 27100, Italy
[b] Department of Mathematics and Computer Science, University of Cagliari, Via Ospedale 72, Cagliari, 09121, Italy
[c] Programa de Doctorado. Centro de Automática y Robótica, Universidad Politécnica de Madrid-CSIC, Madrid, Spain

ARTICLE INFO

Keywords:
Digital twin
Big data
Internet of things
Social analysis

ABSTRACT

Industries such as construction and business companies are becoming increasingly digitized. The amount of data to be monitored and processed has increased signi cantly since the advent of the Internet of Things and the massive use of sensors. In addition to the data from these sensors, large amounts of data that require speci c handling and processing are received. Much of this data is eventually represented in digital twins as a monitoring or decision-support tool. In this paper, we present an architecture to improve digital twin-based experiences that need to represent information from multiple sources. This architecture is demonstrated using the speci c use case of a digital twin for an office of an Italian company. The implementation leverages the Matterport 3D media platform and integrates different technologies and sensors. An evaluation of the solution has also been carried out. The results show high user acceptance and the opening of multiple possibilities to enrich the virtual model with further data from different sources.

1. Introduction

The Internet of Things (IoT) and powerful data analytics have enabled the Industry 4.0 transformation. Digital twins [1] are one of the main innovations of such a transformation. The amount of data that can be used in manufacturing, healthcare, and smart city environments has expanded thanks to IoT. The establishment of a connected physical and virtual twin (Digital Twin) helps address the issue of seamless integration between IoT and data analytics. Accurate analytics can be used for efficient monitoring and to make real-time decisions in a digital twin scenario.

Digital twin instances must be linked to their corresponding real-world twins, often in real-time, in order to gather and arrange data from the associated real-world items and dynamically represent objects in the actual world. The digital twin should make it possible for computational and analytical models to examine these data and analyze the states and behaviors of real-world systems and objects in order to describe, diagnose, forecast, and mimic them. The conclusions drawn from such analysis can be used in conjunction with corporate goals and reasoning to suggest measures for improving the production processes. For intelligent industrial applications to access the data and analytical results, service interfaces must be incorporated into the design of the digital twin in order to accomplish this.

* Corresponding author.
E-mail addresses: ruben.alonso@r2msolution.com (R. Alonso), r.locci11@studenti.unica.it (R. Locci), diego.reforgiato@unica.it (D. Reforgiato Recupero).

https://doi.org/10.1016/j.heliyon.2024.e24741
Received 18 August 2023; Received in revised form 8 November 2023; Accepted 12 January 2024

Available online 19 January 2024

Digitalization in manufacturing is considered a chance to reach higher levels of productivity in today's fiercely competitive marketplaces, where mass customization of products and the growing relevance of software components are posing new problems [2].

A digital twin could be anything basic like a representation of a piece of furniture or something as complex as a model of a car or a manufacturing production line. To generate a virtual proxy, the digital twin simulates every component of the object (or group of related objects). In a digital twin of a car, the shape, tires, seats, engine, and transmission would all be modeled. Companies create a 3D replica of the original object using a digital twin, allowing teams to assess how the underlying object performs in various scenarios. Sometimes, it is possible to create digital 3D objects from 2D images [3].

There are five levels of sophistication for digital twins [4], as detailed by Autodesk.[1] The most complex models are capable of functioning autonomously, while the simplest models just incorporate input from numerous sources. The first level, the Descriptive Twin, is a visual representation of a built item, a live, editable form of design and construction data. Users define the type of data they want to extract as well as the type of information they wish to include. The second level, Informative Twin, has an additional layer of operational and sensory data. The twin collects defined data, aggregates it, and verifies it to ensure that systems work together. The third level, Predictive Twin, can learn from operational data. The fourth level, Comprehensive Twin, considers "what-if" questions and simulates future scenarios. The fifth level, Autonomous Twin, possesses the capacity to understand and act on behalf of users. Note that Levels 1 and 2 are now in use within the Architecture, Engineering, and Construction services sectors. On the horizon are levels 3, 4, and 5, which are enhanced with real-time data from embedded sensors and IoT technology.

When used effectively, digital twins can significantly save costs, enhance product designs, and increase productivity. Heavy sectors like oil and gas extraction, aerospace, and automotive have been at the forefront of the adoption of digital twins so far since they deal with enormous assets and have already embraced product lifecycle management systems. That has been changing recently as the components that enable and make use of digital twins became much more accessible, less expensive, and usable.

The price of sensors keeps going down, data collection is rapidly increasing, and cloud storage is becoming more and more accessible. Manufacturing businesses have more opportunities than ever to manage quality information. The goal is to collect and analyze all of this data from the factory floor in order to better understand not just what is happening now but also to predict what plant operators may do in the future to increase quality and productivity.

The Dutch government, an industrial consortium, and six Dutch institutions have joined forces to create the Digital Twin Program,[2] which intends to harness digital twins to address companies' biggest challenges. One of the objectives of the program is for a digital twin to be utilized as a decision-supporting tool at both an operational and machine level. For instance, personnel costs and downtime make asset maintenance expensive; predictive maintenance, using data from a digital twin, is more effective. One industrial partner of the program, Tata Steel,[3] developed a brand-new hot iron steelmaking method that is supposed to cut down energy usage and carbon emissions. They had to improve the process to make it reliable enough for commercialization. The prototype reactor experienced many outages because the procedure necessitates highly complicated technology. Since using the actual physical plant would be extremely expensive and time-consuming, experts developed a digital twin to simulate the process and used data and models to identify where the weak points were and how to fix them.

The program also collaborates with the multinational corporation Philips.[4] The company, which has been producing electric razors for decades, used data from a line of smart shavers that users could configure based on their skin sensitivity to accelerate time-to-market and streamline the design process. Sensors included in the shavers gathered data on how they were being used, and that data was put into a digital twin to suggest a certain shaver to the user.

Construction projects use digital twins to build perfect reproductions of real-world spaces. Construction teams can visually engage with the real property while in the design and planning phases thanks to these 3D models. Project teams can benefit in a number of important ways thanks to digital twins, which provide them with immersive access to crucial building intelligence in real time. Benefits that can be realized are: i) increased productivity; ii) improved contractor and trades workflows; iii) reduction of problems and Requests for Information (RFIs); iv) quicker and simpler collaboration between project parties. Construction companies may address several of the most pressing issues in the industry, such as low productivity, profitability, performance, and high error and accident rates, by integrating digital twins into the BIM (Building Information Modelling) process. Construction companies may also be able to win more bids and lower their virtual design and construction expenses with the aid of digital twins.

With the goal of providing more tools and examples for the deployment of perfect reproductions of real-world spaces, this paper presents an architecture for digital twins that is integrated with Big Data technologies, IoT, and social analysis. It exploits cutting-edge technology for transforming real-life spaces into immersive digital twin models and data sources for augmenting and improving 3D exploration. A use case has been instantiated out of the proposed architecture and represents the digital twin of an Italian company augmented with data collected from different sources.

This use case is motivated by the company's interest in making efficient use of its spaces and monitoring temperature, humidity, and air quality parameters to move towards a sustainable office where the well-being of users is taken care of, and energy consumption is efficiently managed. In turn, the use case pursues the idea of representing the information in an immersive way, beyond the traditional dashboards, in a way understandable to any user, and with high visual impact.

More in detail, the contributions we bring in this paper are the following:

1 https://redshift.autodesk.com/articles/what-is-a-digital-twin.

2 https://www.digital-twin-research.nl/.

3 https://www.tatasteel.com/.

4 https://www.philips.com/global.

- we propose an architecture for defining a digital twin integrated with Big Data technologies, IoT, and data extracted from different sources;
- we leverage the Matterport 3D media platform to create an immersive and interactive virtual model, Twitter API to retrieve tweets to be further processed by Big Data technologies, Spaceti workplace management software to monitor desk and meeting room occupation and fetch data from sensors, an Italian train service to fetch data about train schedules, and Apache Spark and Kafka as Big Data frameworks to compute analytics on the collected data and to perform real-time data extraction;
- we have set up the overarching framework and implemented it in a practical scenario involving an Italian company's office. Additionally, we have developed digital components that complement the data we have gathered.

The remainder of this paper is organized as follows. Section 2 discusses related work about digital twins and how they have been integrated with other technologies. Section 3 illustrates all the tools and technologies that we have used in the presented architecture. Section 4 depicts the proposed architecture and details each module. The use case we have defined is illustrated in Section 5. The evaluation we have carried out is described in Section 6 where the reader can find the results of the SUS questionnaire we have performed on a set of employees of the company. Future challenges are mentioned in Section 8 whereas Section 9 ends the paper with conclusions and directions on where we are headed.

2. Related works

In this section, we will examine the most recent works that have appeared in literature. In Section 2.1 we will list the surveys that have revolved around digital twin technology whereas Section 2.2 includes applications of digital twins in various domains. Finally, Section 2.3 lists the differences between our paper and the state-of-the-art works.

2.1. Survey works of digital twins

In [1], a categorical review of recent studies on digital twins has been performed by authors. More in detail, reviewed papers have been categorized by different research areas: manufacturing, healthcare, and smart cities. For each of them, an evaluation of the enabling technologies, challenges, and open research for digital twins have been assessed.

Other authors in [5] have provided an overview of the digital twin technology and its application areas, along with a full analysis of the networking needs and suggested enabler technologies to meet those requirements.

In a two-part series of papers, authors looked at the essential roles played by various modeling approaches, twinning enablers, and commonly employed uncertainty quantification and optimization methodologies in digital twins [6,7]. Moreover, with an emphasis on uncertainty quantification, optimization techniques, open-source datasets and tools, significant findings, difficulties, and future perspectives, they give a literature analysis of key enabling technologies of digital twins.

Similarly, other authors provided a thorough analysis of the digital twin technology, as well as its implementation difficulties and restrictions in the most important engineering-related areas and applications [8].

Authors in [9] presented the outcome of a study that examined the most recent definition of digital twins, their main characteristics, and the exploration of the domains where digital twins applications are developed. The study's design implications were then discussed, concentrating on socio-technical design elements and digital twins' lifespan.

At a different level, a proper methodology for visualizing the digital-twin science landscape using contemporary bibliometric tools, text mining, and topic modeling, based on machine learning models such as Latent Dirichlet Allocation (LDA) and BERTopic (Bidirectional Encoder Representations from Transformers), has been presented in [10]. The outcome of the study has been that i) on the one hand, studies on digital twins are still in the early stage of development, and ii) on the other hand, the core of the topic is growing especially related to digital twins of industrial robots, production lines, and objects.

Other authors in a recent study conducted a thorough and comprehensive review of 240 academic publications on digital twins [11]. They examined these papers from the perspectives of concepts, technologies, and industrial applications. Ultimately, the authors provided their recommendations for digital twin research, organizing them into various stages of the lifecycle.

The work conducted by the authors in [12] presents an overview of the digital twin as a genuine representation of a cyber-physical system within the context of Industry 4.0, highlighting the significance of AI in this concept. It discusses the critical enabling technologies of the digital twin, including edge, fog, and 5G, which facilitate the integration of physical processes with computing and network domains. The study identifies the role of AI within each technology domain by analyzing various AI agents at both the application and infrastructure levels. Finally, it focuses on movement prediction, which is selected and experimentally validated using real data generated by a digital twin for robotic arms, with results demonstrating its potential.

2.2. Applications of digital twins

Applications of digital twins in various domains such as smart factories in industrial production, digital models of life in the medical field, construction of smart cities, security guarantees in the aerospace field, immersive shopping in the commercial field, and so on have been illustrated in [13].

In particular, the application of digital twin technology in social media, according to [14], has the potential to transform businesses' social media strategy and influence the direction of marketing. Furthermore, the authors claimed that businesses may remain ahead of the curve and succeed in a continuously changing digital environment by adopting digital twin technology and incorporating it into their social media operations.

Within the domain of construction, other authors have proposed a definition of a digital twin and identified forty applications grouped into seven capabilities [15]. Moreover, a case study on how building authority can be integrated into digital twins has been presented as well.

Another research offers a comprehensive perspective and justification for integrating digital twin technology into model-based system engineering, as described in [16]. This study explores the advantages of combining digital twins with system simulation and the Internet of Things to support model-based system engineering. It also provides concrete examples illustrating the utilization and advantages of digital twin technology across various industries.

Other papers focused on the applications of digital twins for spacecraft protection [17], for occupational safety and health in workplaces [18], for smart aging [19], and for port systems [20].

2.3. Differences with respect to the SoA

Differently from the state-of-the-art works, in this paper, we propose an architecture for defining a digital twin integrated with Big Data technologies, IoT, and data retrieved from multiple sources. The goal is to show the 3D model of the office of an Italian company where we have designed different digital layers showing the information we have extracted from the collected data.

3. Material

This section aims to provide the reader with a thorough understanding of the tools used in this article, describing their attributes, characteristics, and uses. In detail, the following tools will be described:

- Matterport: to create immersive, 3D digital twins;
- Apache Spark: to implement the algorithms used at scale;
- Apache Kafka: to distribute different types of data through the defined architecture;
- Twitter API: to fetch raw data;
- PostgreSQL: to store the processed information;
- Spaceti: to fetch sensor data;
- HuggingFace: to perform sentiment analysis with opensource models;
- OpenWeatherData: to fetch climatic data.

3.1. Matterport

Matterport[5] is a 3D camera provider and the developer of a platform for creating immersive and interactive 3D models of physical spaces. It consists of a specialized camera that can capture every detail of a room or building, creating an accurate 3D model that can be explored from any angle. This technology is widely used by real estate agents, architects, and interior designers to showcase their properties, as it allows potential buyers to virtually “walk” through space without being physically present. In addition, Matterport offers several features such as measurement tools, virtual reality compatibility, and highlights that allow users to label and annotate specific areas within a model. Overall, Matterport offers an innovative and exciting way to digitally experience physical spaces.

3.2. Apache spark

Apache Spark[6] is an open-source distributed computing system designed to process large amounts of data quickly and efficiently. It is a unified analysis engine that provides rapid processing of data from multiple sources, including structured, semi-structured, and unstructured data. Spark can work on top of the Hadoop Distributed File System (HDFS[7]) and can process data in real-time and batch mode.

The main feature of Spark is its ability to handle in-memory processing, which means it can store data in RAM and process it faster than traditional disk-based systems. This makes it an excellent choice for applications that require real-time processing, such as financial fraud detection, sensor data analysis, and recommendation systems.

Spark offers a wide range of APIs and libraries, including Spark SQL[8] for structured data processing, Spark Streaming[9] for real-time data stream processing, and MLlib[10] for machine learning. It also supports several programming languages, including Java, Python, Scala, and R, making it accessible to a wider range of developers.

[5] https://matterport.com/.
[6] https://spark.apache.org/.
[7] https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html.
[8] https://spark.apache.org/sql/.
[9] https://spark.apache.org/streaming/.
[10] https://spark.apache.org/mllib/.

Overall, Apache Spark is a powerful and flexible tool for processing big data that has become a popular choice for data scientists, engineers, and developers. Its ability to process data quickly and efficiently, along with its flexibility and ease of use, has made it an ideal solution for many organizations that wish to manage and analyze large amounts of data.

3.3. Apache kafka

Apache Kafka[11] is an open-source distributed event streaming platform, initially developed by LinkedIn[12] and subsequently donated to the Apache Software Foundation. It is designed to handle high-performance real-time data streams, allowing users to publish and subscribe to log streams, process them in real-time, or store them persistently.

Kafka's architecture is based on a distributed, partitioned, replicated registry, allowing multiple producers and consumers to write and read data to and from the same topic in a scalable, fault-tolerant manner. Data are stored in Kafka topics, which can be partitioned for scalability and performance.

Kafka also provides several Application Programming Interfaces (APIs) and tools, including the producer API, the consumer API, and the Kafka Streams API, which can be used to create real-time data processing applications, messaging systems, or event-driven architectures.

Overall, Apache Kafka offers a scalable, reliable, and fast way to manage real-time streaming data, making it popular among various industries and use cases, such as financial services, social media, e-commerce, and IoT.

3.4. Twitter API

Twitter[13] is a popular social media platform where millions of active users exchange tweets every day. The Twitter API is a powerful tool that allows developers to interact with Twitter services programmatically. The API enables application developers to access Twitter's vast amount of data, including tweets, hashtags, user data, and more.

One of the main benefits of using the Twitter API is the ability to create custom applications that can interact with Twitter services in real time. This opens up a wide range of possibilities, including automated tracking of specific hashtags or tweets, advanced analytics based on Twitter data, and real-time interaction with customers and followers.

It is available in two versions: the REST API and the Streaming API. REST API allows developers to access Twitter data via HTTP requests and supports read and write operations, while the Streaming API provides low latency access to real-time Twitter data and is suitable for applications that require real-time data processing.

The Twitter API provides developers with a wide range of resources and tools to help them build applications quickly and efficiently. It provides access to comprehensive documentation, forums, and support services, making it easy for beginners to get started.

In conclusion, the Twitter API stands as a formidable resource for developers seeking to craft tailored applications capable of interfacing with Twitter's array of services. Through its robust data repositories and real-time engagement functionalities, the Twitter API affords a singular prospect for the inception of inventive and imaginative applications.

3.5. PostgreSQL

PostgreSQL[14] is an open-source relational database management system. It is one of the most advanced and feature-rich database systems available. PostgreSQL is often referred to as Postgres and is highly reliable, scalable, and extensible.

Many organizations around the world use it to manage their critical data. PostgreSQL supports a variety of programming languages and runs on multiple platforms, including Windows, Linux, macOS, and Unix. It follows the SQL standard and offers many extensions that allow users to customize and enhance the functionality of their database.

In addition, PostgreSQL offers advanced features such as concurrency management, full-text search, triggers, views, and JSON support. Overall, PostgreSQL is a powerful database system that provides robust functionality and performance for a variety of applications.

3.6. Spaceti

Spaceti[15] is an innovative proptech[16] company that is changing the way people interact with built environments. Spaceti provides property owners and managers with the tools they need to optimize their facilities, improve tenant satisfaction, and increase building efficiency. One of Spaceti's key offerings is its smart building platform, which uses sensors and IoT technology to collect real-time data on building occupancy, energy consumption, and air quality. This data is then analyzed using machine learning algorithms

[11] https://kafka.apache.org/.
[12] https://www.linkedin.com/.
[13] https://twitter.com/.
[14] https://www.postgresql.org/.
[15] https://www.spaceti.com/.
[16] Short for "property technology", it refers to the use of technology in the real estate industry.

to provide building managers with actionable insights and recommendations to help them make informed decisions that can save energy, reduce costs, and improve the overall occupant experience.

Spaceti additionally provides an assortment of mobile applications, affording residents novel means to engage with their surrounding environment. Ranging from individualized lighting and climate regulation to informational displays and concierge functionalities, Spaceti's applications have been meticulously designed to augment the resident experience. They facilitate individuals in customizing their surroundings in alignment with their speci c requirements and preferences.

Overall, Spaceti is driving the digital transformation of the real estate sector, helping property owners and managers create spaces that are more efficient, sustainable, and user-friendly.

3.7. HuggingFace

Hugging Face[17] is an innovative natural language processing (NLP) company founded in 2016. Based in New York City, it has quickly become one of the leading NLP companies in the industry. Their main focus is creating solutions that enable computers to understand human language and communicate more intuitively with people.

Their most popular product is the Hugging Face platform, an open-source library for natural language processing. This tool makes it easier for developers to build chatbots, automated assistants, and other AI-powered communication tools. Hugging Face uses machine learning algorithms to train chatbots and automate communication tasks.

Hugging Face's unique approach focuses on sharing model checkpoints, a way for developers to share their pre-trained NLP models with other developers. This approach has led to a collaborative environment where developers can build on each other's work to create more advanced NLP solutions.

In addition to the Hugging Face platform, the company has developed several other tools and applications, such as Hugging Face Enterprise, which is a platform for building and deploying AI applications.

Notable projects include several models based on the Hugging Face Transformers library,[18] a powerful deep learning model for NLP based on the Transformer architecture, over which are built tools like DALL-E,[19] a neural network that generates images from text descriptions, and GPT-3,[20] an AI language model for natural language text generation.

3.8. OpenWeatherData

OpenWeatherData[21] is an open and easy-to-use platform that provides weather-related data and services to developers, researchers, and businesses. It provides a wide range of weather data, including current conditions, forecasts, historical data, and much more. OpenWeatherData aims to democratize weather data and make it available for a variety of applications and use cases.

Developers can access OpenWeatherData through an API, which allows them to programmatically retrieve weather data and integrate it into their applications or services. The API provides endpoints for retrieving current weather conditions, 5- and 16-day forecasts, historical weather data, and even special data such as air pollution indices and UV indices. Developers can use this data to build weather apps, display weather information on websites or mobile apps, incorporate weather data into analyses or studies, and more.

OpenWeatherData also offers a variety of subscription plans that provide additional features and data access, such as higher request thresholds, longer forecast periods, and access to specialized weather data. These subscription plans cater to different needs and levels of use, allowing users to choose the option that best suits their needs.

The availability of OpenWeatherData enables developers and businesses to use weather data in their applications and services, improving the user experience, enabling informed decisions, and providing valuable weather-related information. The open and accessible nature of OpenWeatherData has already fostered innovation and collaboration in the use of weather data across industries and sectors.

4. Architecture

The architecture of the system we propose in this paper is depicted in Fig. 1. Basically, the user accesses through the browser (the Client) the DigitalTwin FrontEnd. The FrontEnd calls the Matterport 3DVM which provides the 3D model with all the metadata. The Backend includes the REST API Layer which allows the connection with the data sources. The support for the REST API Layer is given by the Database and Kafka layers. The rst one is employed to store information whereas the second one fetches information from the data sources in a more organized and modular way.

[17] https://huggingface.co/.

[18] https://huggingface.co/docs/transformers/model_summary.

[19] https://openai.com/research/dall-e.

[20] https://openai.com/blog/gpt-3-apps.

[21] https://openweathermap.org/.

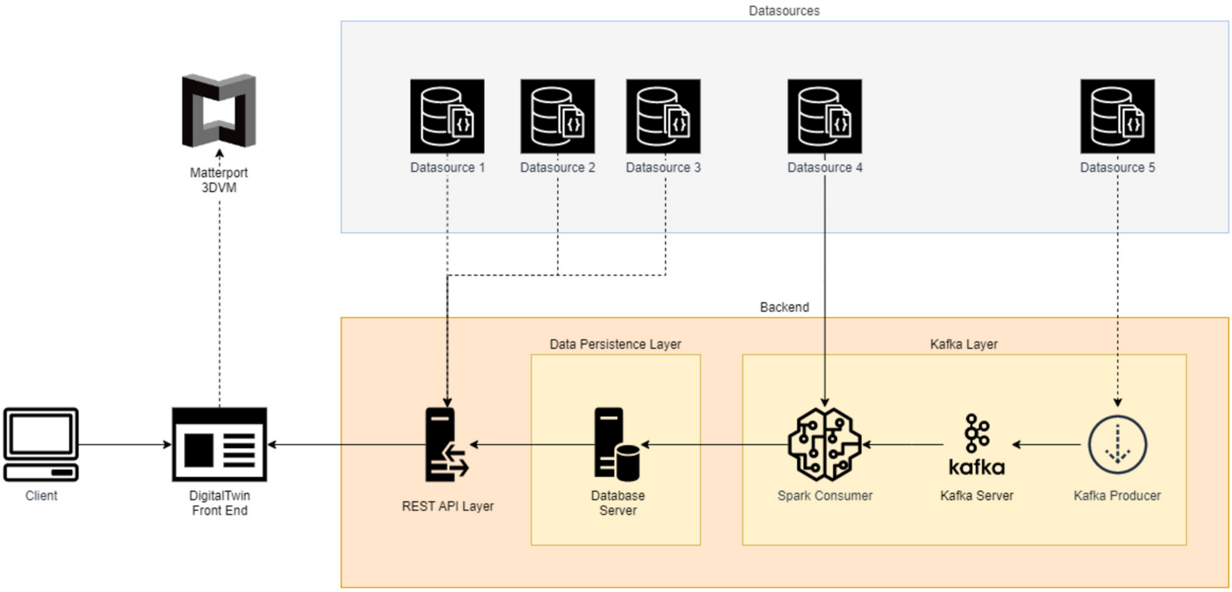

Fig. 1. Architecture diagram.

4.1. REST API layer

The REST API layer facilitates seamless communication between different systems and services by providing a standardized way of interaction: this enables the integration of the Digital Twin Front End with third-party applications. Incorporating the REST API layer into the architecture offers numerous advantages, such as its simplicity and ease of use.

REST APIs follow a straightforward design using widely accepted HTTP methods and standard data formats such as JSON or XML, making it easier for developers to understand and work with the API, speeding up development and integration processes. In addition, REST APIs offer platform and language independence: they are based on the HTTP protocol, which is universally supported by different platforms, programming languages, and devices, so that different clients, such as web browsers, mobile apps, or other services, can use it seamlessly.

REST interfaces also promote flexibility and modularity in software development. By exposing speci c resources and functions through well-de ned endpoints, granular and targeted APIs can be created. This modularity makes it easier to upgrade or add new features without affecting the overall system.

Indeed, this approach facilitates the integration of additional sensors through a sequence of relatively straightforward steps, primarily leveraging the compatibility of the API-generated data format with the front-end interface. In scenarios characterized by heightened complexity, particularly when confronted with substantial data volumes, more specialized strategies closely related to Big Data principles, may be invoked, as detailed in Section 4.3.

The procedure of adding new sources is based on 3 general steps: i) the initial step entails the development of intermediary code designed to interact with the sensor API for the purpose of acquiring the data intended for representation. In cases where the original data source fails to deliver data in the requisite format or necessitates the fusion of multiple data sources, the code is tasked with performing appropriate data processing and calculations. Although it is technically feasible to implement this API access and processing layer at the front-end level, there is a prevailing preference for its implementation within the backend infrastructure. This approach affords several advantages, including enhanced security in credential management for web services and direct access to the persistence layer.

ii) The second step is related to the de nition of how this information is going to be represented in the navigable model. This involves de ning whether it is going to be a kind of tag with the information, or if the need is to load a representative 3D model to display the data. Furthermore, the decision to incorporate information panels and the implementation of user interaction components becomes essential, especially in cases where the element under consideration supports interactivity. iii) The last step pertains to the spatial arrangement within the 3D navigable map, along with the essential process of element validation. This facet holds particular signi cance in the context of occupancy, illumination, or gas sensors, where the alignment between the real-world and virtual-world domains assumes a critical role.

4.2. Data persistence layer

The database layer provides data storage capabilities, allowing to persistently store structured or unstructured data such as user information, product details, transaction records, and more. By leveraging a database, it ensures data is securely stored and accessible even when the application or the server restarts.

The database layer serves as the foundation for data analysis and reporting. It enables complex queries, aggregations, and transformations to derive insights from the stored data. By integrating this layer with analytics and reporting tools, it is possible to generate meaningful reports, visualize data, and make informed decisions based on the extracted information.

Furthermore, the database layer facilitates efficient data retrieval: querying and indexing mechanisms allow the retrieval of specific records or perform complex searches to extract the required information. This functionality is vital for applications that need to access and present data to users or other systems.

4.3. Kafka layer

There are significant benefits to incorporating the Kafka layer into the infrastructure: one key benefit is the introduction of an event-based architecture, where events act as a key communication mechanism between different parts of the system. Kafka's distributed event streaming platform facilitates seamless event flow and enables loose coupling, scalability, and flexibility in handling different data streams and processing scenarios. This event-driven approach provides the system with real-time and reactive capabilities that enable efficient event processing, event sourcing, and event-driven workflows.

Another important advantage of Kafka is its reliable message queuing system, which ensures that messages are stored and consumed sustainably by multiple consumers without the risk of data loss. Kafka's fault-tolerant architecture ensures high availability and resilience, making it well-suited for critical data flows such as financial transactions, system logs, or IoT sensor data. By leveraging Kafka's robust messaging capabilities, organizations can build robust, scalable, and fault-tolerant applications that can handle large volumes of data with ease. In addition, Kafka's scalability and high throughput capabilities make it an ideal choice for real-time data collection, processing, and analysis. Its distributed nature enables horizontal scalability by adding more brokers to the cluster to meet growing requirements as the application grows.

4.3.1. Spark consumer

The Spark Layer works alongside the Kafka Layer to extend the distribution capabilities provided by its architecture and their combination result in a reliable and scalable framework for real-time data processing and analysis.

Both Spark Streaming and Kafka are designed to reliably handle failures and provide fault tolerance. Spark Streaming divides data into micro-batches, enabling disaster recovery, while Kafka replicates data across multiple nodes, ensuring data reliability. This combination ensures data is processed even in the face of failures and enables streaming applications to scale horizontally.

Spark Streaming provides high-level abstractions for real-time data processing, such as window-based operations and stateful computations. It enables complex analysis of the data stream, making it suitable for applications that require low-latency processing. Integration with the broader Spark ecosystem further enhances capabilities by leveraging Spark's libraries and APIs for advanced analytics, machine learning, and graph processing.

5. Use case

As a use case, we have considered one office of the Italian company R2M Solution. This company wanted to research digital twins with novel visualization and user interaction methods, in which IoT, Big Data, and 3D virtual models come together.

The digital twin of the use case implements the proposed architecture and makes use of multiple sensors and data sources to meet the requirements that were initially proposed. Among them was the ability to represent sensors, show information from social networks, display information obtained through third-party APIs, include processed information from the company's data, and finally integrate with the space and room management system so that any user of the office can be informed of what is ongoing with a glance at the digital twin. For more realism, and instead of representing all this information on a panel or on a web page, the idea was to provide an immersive interface that faithfully represents the real world.

The system we have developed can be freely visited[22] and a public video of it is publicly available online.[23] In Fig. 2 the schema of the use case we have instantiated from the architecture previously discussed is shown. In the following, we will detail the elements we have developed.

5.1. Backend

The backend provides the elements that are fundamental for the architecture to provide a scalable way of publishing its content.

The first element is the REST API Layer, which allows a unified interface to interact with the data. It also exposes the endpoints called by the DigitalTwin FrontEnd as it requests the information it will then present to the final user. In our implementation, the REST API Layer has been developed using Express.js,[24] which provides a myriad of HTTP utility methods and middleware, allowing us to create a robust API. The REST API Layer empowered us with the ability to retrieve data from third-party services (see Section 5.2), but more importantly, to query our data persistence layer.

[22] https://dt.r2m.cloud/.

[23] https://www.youtube.com/watch?v=gxyvE-Df1Zg.

[24] https://expressjs.com/.

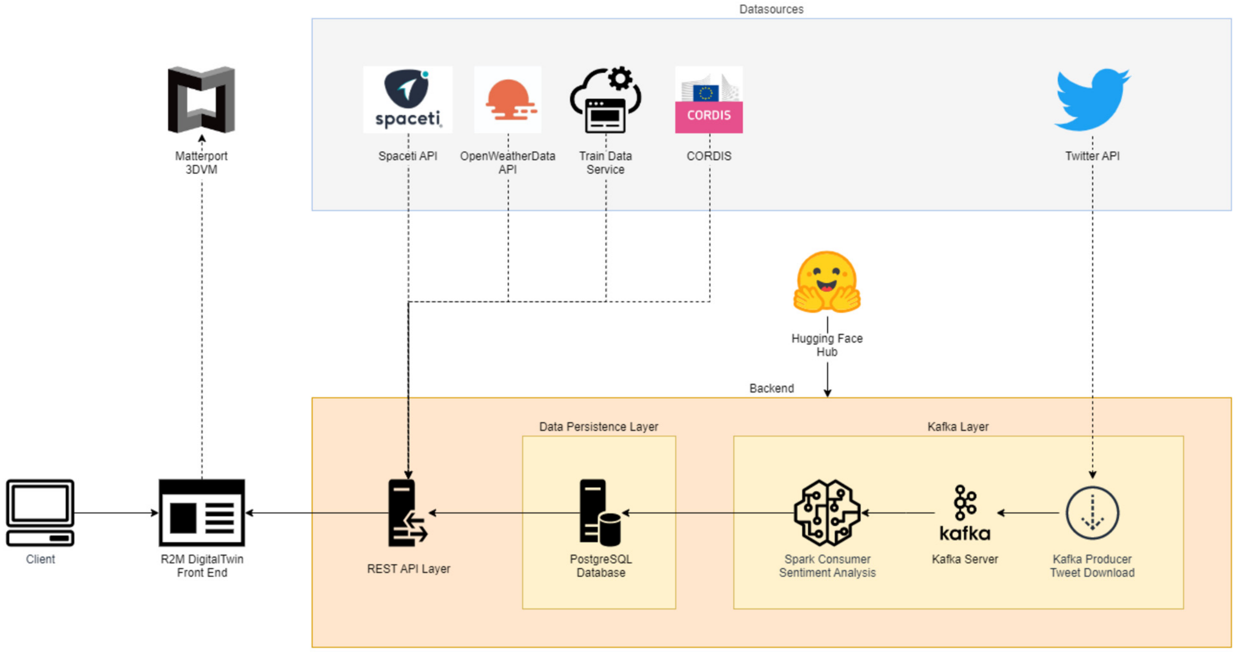

Fig. 2. Architecture of the use case.

The persistent layer has been implemented as a PostgreSQL database server. Its content is associated with the data that is manipulated in Apache Kafka. It stores data that contains a raw representation of tweets with their metadata, but also computed results that come from the application of ML algorithm from the consumer in the Kafka Layer.

Last but not least, the Kafka Layer is constituted of three major components: the Kafka Producer, the Kafka Server, and the Kafka Consumer. The Kafka Server is a simple implementation of a Kafka Broker: a Kafka topic is con gured as a dedicated topic for the tweet stream (which stores tweet data for at most a day to allow daily computation performed by the consumer).

The Kafka Producer is a Python application that daily queries the Twitter API based on lters related to the company subject of the use case to obtain data related to tweets where the company itself is being mentioned, with the usage of both hashtags and anchors. Once it has retrieved the latest tweets, they are then encoded in a format understandable by Kafka (mostly because of the emojis contained in the tweets) and sent to the Kafka Server for further processing.

The nal component of the Kafka Layer, the Kafka Consumer, runs daily and listens to the tweet's topic from the Kafka Server in order to retrieve the raw data it needs for its work. We implemented the Kafka Consumer through the development of a Spark Streaming application, such that the data could be easily preprocessed and cleaned before any further computation. By exploiting open-source ML models from HuggingFace, it runs a series of sentiment analysis algorithms to label and categorize the tweets the company has been involved with, so that the nal users understand how positive or negative the sentiment is towards the company.

Each tweet is associated with a tag that identi es the company as "positive", "neutral" or "negative", with a score indicating the index of reliability to that label (a value between 0.6 and 1.0), and then it is stored into the persistence layer to make it available to the nal application. To integrate the PostgreSQL database with the Spark Streaming application, we needed to implement also a custom sink so that the streaming process could store information directly into the database table destined for the re ned data.

5.2. Data sources

While the backend section describes the elements which are being used to retrieve, handle, process, and provide data, data sources are the very origin of the data being used in the use case. As indicated in the Datasources module of Fig. 2, we used ve different services where to extract data that have been represented in the digital twin.

The rst component is the Spaceti API, which is used by R2M to store data from its private sensors and to have an easy and reliable interface to be able to fetch and query them. Potentially, it is possible to leverage any kind of sensor. Currently, the company handles occupancy sensors to know which workstations are available, how many seats are occupied in a room, and the temperature and CO2 levels.

To better represent weather information, the use case also included some open-source data provided by the OpenWeatherData API. Therefore, the digital twin is able to show for the same room two kinds of environmental information, the temperature of the office and the external temperature of the city of Pavia.

Since the digital twin has the objective to represent information that can be useful to those attending R2M offices, one of the widgets focuses on public transportation information to reach the office. Therefore, we employed a train data service that has been queried by the backend and shows in a dedicated panel the times of the train departing and arriving at the central station of Pavia.

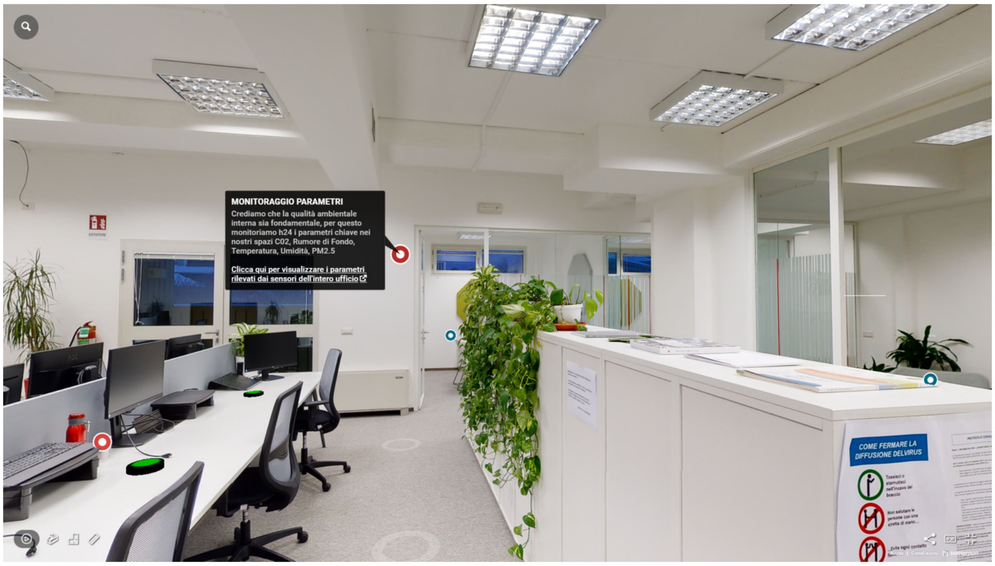

Fig. 3. Mattertag inside of the Digital Twin.

One more data source has been obtained by the data published by the Community Research and Development Information Service (CORDIS[25]) which provides a lot of information about funded research projects for European organizations. Thus, we designed a panel where we show the ranking of the company related to the obtained funding in Italy and Europe with the possibility to leverage different filters (year, type of organization - if private organizations only or both public and private).

Finally, we have employed the Twitter API to retrieve the most recent tweets about R2M and have performed a sentiment analysis classification on the tweets that have been preprocessed. All data sources except Twitter are managed in the REST API Layer. Twitter data, on the other hand, is processed through the data feed management system, and as a result, there is a component within the Kafka layer responsible for fetching data from the Twitter API.

5.3. R2M digital twin frontend

The R2M Digital Twin Frontend represents the main interface the final users interact with to attain the data provided by the proposed use case. It is a React Single Page Application built on top of the Matterport API and a 3D model of R2M offices which was digitalized (periodically updated) and stored in the Matterport Cloud.

The information is displayed using both simple Matterport SDK components (i.e., Mattertags) and advanced customized entities to better represent the capabilities and the added value provided by the Digital Twin. While Mattertags are used to display mostly descriptive information (see Fig. 3), the custom objects developed for the use case focus on more complex data and thus try to provide a better user experience when rendered in the Matterport model. These elements periodically query (based on the needs of the specific model being rendered) the REST API Layer to retrieve the data they need to display useful information.

For each available workstation, the office is equipped with a dedicated occupancy sensor that detects whether the seat is taken or not: these sensors send their value to the Spaceti service, which in turn distributes them through its API. Every sensor is represented in the R2M Digital Twin Frontend by rendering a 3D model (in this case in DAE format) that changes its color based on the latest data provided by the API (simply green to show the seat is free, red otherwise) as can be observed in Fig. 4.

Mattertags allow to show texts and even allow to present some simple HTML elements to enrich the information they provide. However, they could be really strict with some kind of representation. In the R2M Digital Twin Frontend, this is the case for the rendering of details related to train schedules: the best way to present them is to organize them in a timetable, like for the real-life timetables consulted by travelers. This kind of display has been achieved in the use case by developing a custom element that consists of an HTML canvas over which data is simply drawn to be organized as a table and showing details about possible destinations, planned time of arrival, and notes about that particular ride (Fig. 5).

While the focus of the previous elements was to just provide the user with useful information, the real potential of a digital twin is really exploited when the entities also work as actuators. The most complex feature of R2M Digital Twin Frontend is a compound of custom entities developed to work together as a wall-painted form that is used to interact with the REST API layer, which in turn communicates with the CORDIS service. In the info box shown in Fig. 6, it is possible to distinguish, from the top to the bottom:

- a text box, which shows a descriptive text;

[25] https://cordis.europa.eu/about.

Fig. 4. Occupation sensor widgets inside of the Digital Twin. At the time of the screenshot, the three seats in front of the respective workstations were free.

Fig. 5. Timetable widget inside of the Digital Twin with the train departing from Pavia.

- three different input boxes, each of them representing a parameter used by CORDIS (year, geographic area of comparison, type of organization);
- a submit button, used to send the parameters as a REST API call.

The text boxes basically act as combo boxes. However, as the interface does not offer advanced ways for interacting with those boxes, we have used just the mouse click event to interact with each of them: this is due to the fact that we are creating a custom web form rendered through a canvas. We cannot use standard HTML elements and every single interaction with the element must be developed again, increasing the complexity of every new widget. Thus, each input is con gured with a set of values and each mouse click updates the box with the next value in the list. The inputs can also interact with each other: for example, the country element allows two possible values, 'COUNTRY' or 'EUROPE', which are used to establish the geographic area of interest. If 'EUROPE' is selected, the country selector (right below the country input) is completely disabled from rendering otherwise (if 'COUNTRY' is selected), then it is also possible to interact with the eld below it to choose the country where R2M operates.[26]

[26] R2M has different branches in UK, Spain, France and, therefore, we can focus our searches on the country of each branch.

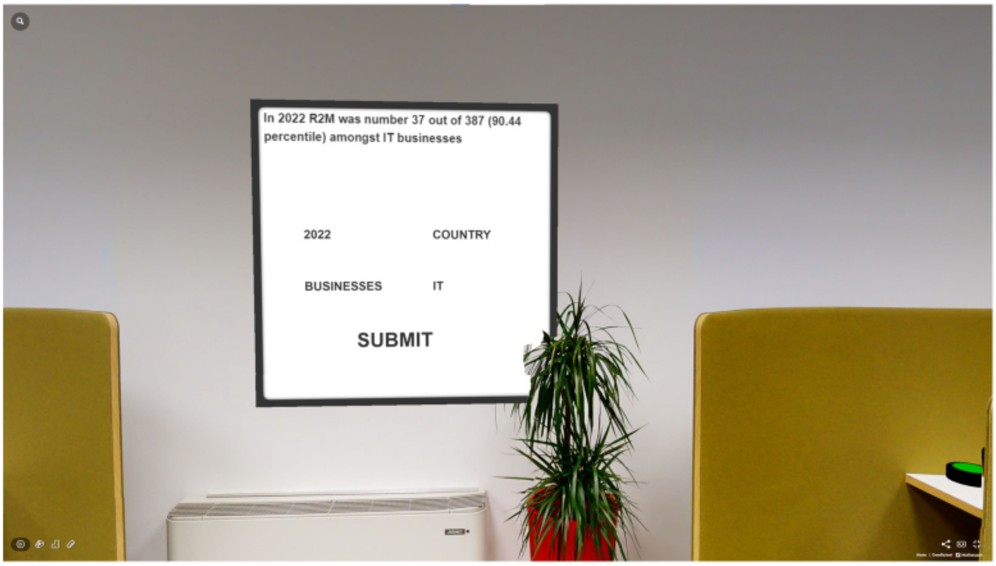

Fig. 6. InfoBox widget inside of the Digital Twin showing information about R2M through the CORDIS service.

6. User experience evaluation

We have carried out a user experience evaluation for the use case we have designed. The developed digital twin has been demoed for 20 minutes to 20 participants chosen among the employees of the company (9 project managers, 5 researchers, 1 managing partner, and 5 business developers). Six of them were females and fourteen were males. On average, they have been working on digital systems for 15 years. Moreover, they were heterogeneous as far as their areas of expertise were concerned: AI, embedded systems, robotics, energy efficiency, data science, sustainable energy, electronics, and BIM models.

They were required to complete a two-part survey to share their overall impressions. The usability of the system was evaluated in the first section using the standard *System Usability Scale* (SUS) questionnaire.[27] Seven open-ended questions about our system's advantages, disadvantages, and overall feedback were included in the second section. The first three open-ended questions included five levels of reply from 1, the lowest, to 5, the highest, and a text where to leave open answers. The last four were open questions only. The results of the user study are described in the paragraphs that follow.

6.1. SUS questionnaire

The SUS questionnaire resulted in a score of 92.5/100, or an A+ grade, putting the proposed system in the 98th percentile rank.[28]

The score distribution of the users is displayed in Figs. 7 and 8. In particular, Fig. 7 concentrates on the odd questions (the positive ones), which should score highly if the system is usable, whereas Fig. 8 concentrates on the even questions (the negative ones), which should score poorly if the system is usable.

The annotators claimed that our system is very simple to use (with an average rating of 4.05 ± 1.23[29]). Users displayed a high level of confidence while utilizing the system (4.15 ± 0.67) and concurred that they did not need any more details (1.355 ± 0.82). They also discovered very few inconsistencies (1.40 ± 0.50). The features of the proposed systems were considered well-integrated by users (3.80 ± 0.89) and believed that it was not complex to use (1.70 ± 0.80). As a result, they stated that no support will be needed in the future (1.60 ± 0.99).

6.2. Open questions

The seven open questions and their responses are covered in this section. As mentioned earlier, the first three open questions contained a field where to insert a value in [1-5] and a text where to write any related comments. Fig. 9 summarizes the numeric results of the first three questions whereas in the following we will comment on the received feedback of the whole set of seven questions.

Q1. How do you assess the quality of the interaction with the R2M Digital Twin?

As it can be seen from Fig. 9, the scores given by the annotators were very high. In general, their opinion was highly positive as well as the comments they left (e.g., *intuitive*, *the interaction is great*). Some remarks were about the panels we developed that should

[27] https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html (accessed on 7 July 2023).

[28] Interpreting a SUS score - https://measuringu.com/interpret-sus-score/.

[29] With the notation $X \pm Y$ we specify that X is the average score and Y the standard deviation.

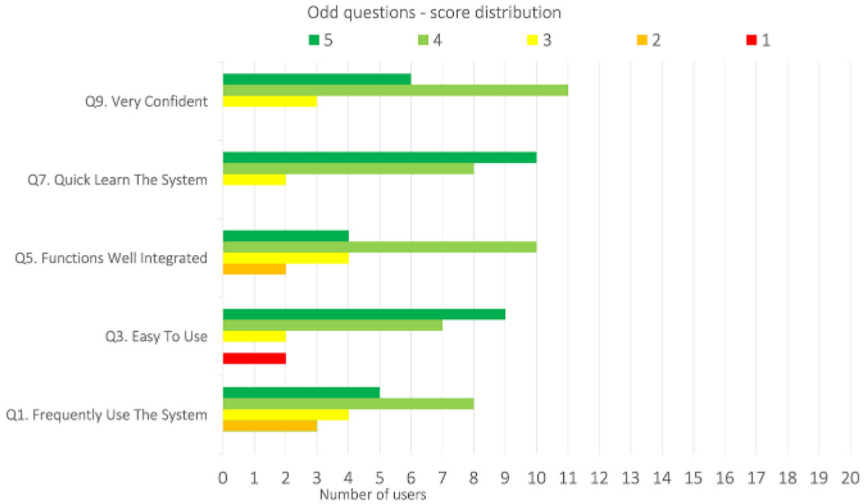

Fig. 7. Odd questions. The higher the value, the better the system.

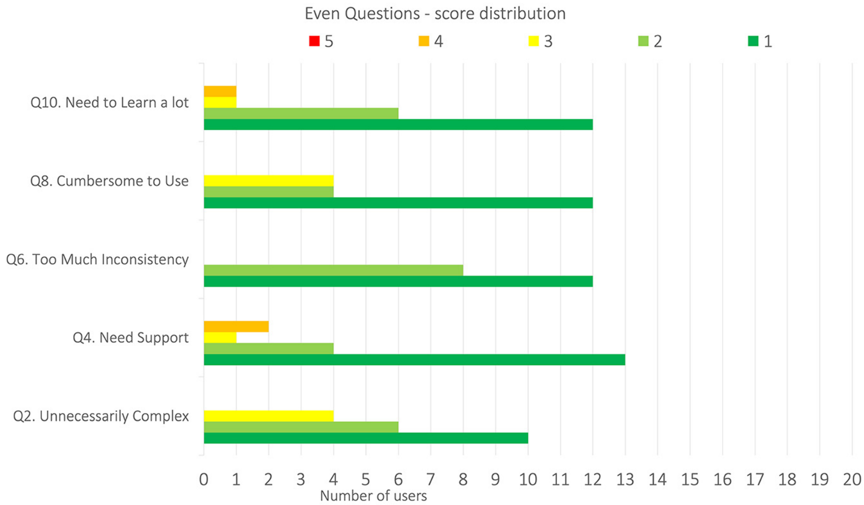

Fig. 8. Even questions. The lower the value, the better the system.

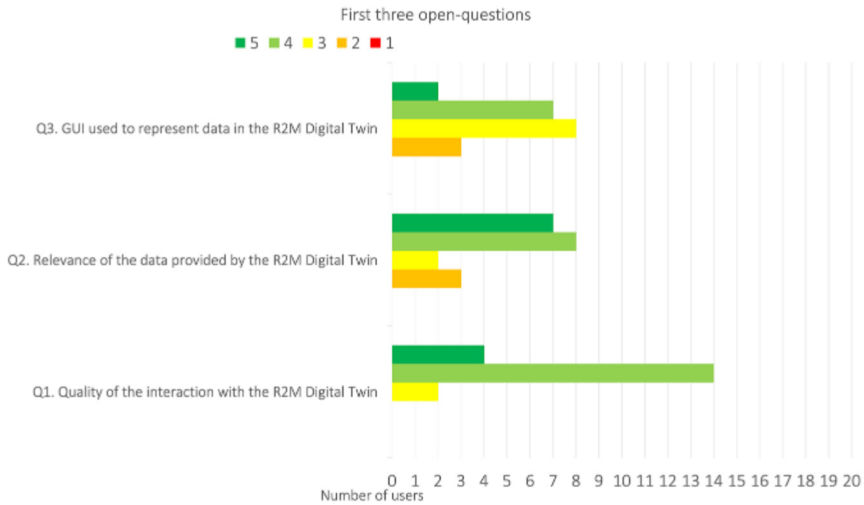

Fig. 9. First three open questions. The higher the value, the better the system.

show an animated icon to let the user understand that some computation was happening when they clicked on a button. Also, an information box explaining what each sensor was and how they work was demanded to be included.

Q2. How do you rate the relevance of the data provided by the R2M Digital Twin?

Three users gave a score of 2 and two users gave a score of 3. Nobody of these 5 users posted any comment likely because they were just not interested in the provided data. All the others gave 4 and 5. These last were satisfied with the provided data and no particular comments were left.

Q3. How do you rate the GUI (e.g., tables for CORDIS data, for train hours) used to represent data in the R2M Digital Twin?

As indicated in Fig. 9, three users gave a score of 2, 8 users scored 3 and the others 4 and 5. In general, some remarks were about the visualization of the data, and the possibility to make the panels graphically more appealing by using better colors and graphically enhanced buttons.

Q4. What are the main strengths of the R2M Digital Twin?

Five users commented that the 3D experience is great and gives the user a perfect sense of the office. Four more users appreciated the presence of the panel with the different information that could be obtained. Three users were happy to be able to daily monitor the office. Three more were impressed by the quality of the rendering and the fluidity of the demo and the virtual navigation. The integration with data from multiple sources was appreciated by five users.

Q5. What are the main weaknesses of the R2M Digital Twin?

A couple of users pointed out the lacking of animations and more appealing graphics for the developed panels. One user experienced some slow times in particular browsers and the fluidity of the navigation. A couple more users would have liked a menu with a list of all the added sensors and interactive objects. One user would have liked to see the external of the building too.

Q6. Can you think of any additional features to be included in the R2M Digital Twin?

Several suggestions have been sent by users: the inclusion of a robot that might assist the navigation; actuators to turn on/off the lights or change the temperature in the room; the possibility to book workstations and rooms from the digital twin; the inclusion of some warnings if the CO2 levels reach a certain threshold; showing the CO2 levels and other air quality data; the panels that display the historical values of temperature and other sensors.

Q7. Would you like to include other types of data in the R2M Digital Twin?

The following data were suggested to be included by different users: weather data, events in Pavia; a calendar of the working days, and the events of the company; other air quality data; a graph showing the historical occupancy of the office; the consume of electricity; if and which light of the offices are on or off; if and which window is open or close; data about current noise and history about that; some alert for bad air quality or high temperature.

In conclusion, the user study showed that the proposed system is very user-friendly and regarded as an important instrument for obtaining information about the company from its digital twin.

7. Strengths and limitations

In this section, we will list some strengths and limitations of the proposed solution.

- **Strengths:**
  - **3D navigation**: The solution offers a navigable 3D environment based on real-world information obtained through 360º cameras, distinguishing it from other digital twins that rely solely on synthetic renderings.
  - **Realism**: What the user sees in the digital twin precisely reflects what the cameras captured when creating the point cloud to generate the 3D world.
  - **Ease of updates**: If the real world undergoes changes, a new camera scan can be conducted, and the digital twin will continue to function with minimal modifications.
  - **Seamless integration with external data sources**: This feature enables the representation of various types of information within the 3D Digital Twin.
- **Limitations:**
  - **Dependency on specific SDK and manufacturer cameras**: The Digital Twin heavily relies on the SDK and, consequently, on cameras from a specific manufacturer.
  - **Challenges in rendering certain information**: Rendering certain information can be challenging, especially when trying to render interactive elements.
  - **Limited support for 3D models**: The support for 3D models is limited to a specific set of formats.

8. Future challenges

The instantiation and validation of the architecture have opened up multiple possibilities both in research and in use cases on which we can continue to extend the ideas pursued in this article. For instance, there are multiple challenges concerning the integration of new sensors and data. The interest to include more sources and information is a clear result of the validation and opens new problems in topics such as the representation of AI-based Big Data Analytics or the integration with the organization's data lakes.

Another challenge and future focus of our research is related to the actuators and that actions in the virtual environment are reflected in the real one. In the evaluation, it was clear that this functionality is among those expected by the users, for example, to control illumination in the real world or to reserve spaces.

As mentioned in the evaluation, some users pointed out that certain elements, especially the panels, lack visual appeal. This can have an impact on the sense of immersion we are aiming for. As a result, one of the priorities for future work is to enhance the panels with more representative elements, such as 3D panels or representations of displays. To achieve this, the authors are conducting various tests involving the loading and rendering of 3D models in the digital twin. In addition to this, certain performance issues have been observed. Currently, we confront a dual set of interrelated challenges. First and foremost, the front-end component has a considerable computational burden, primarily arising from the extensive 3D processing operations executed in JavaScript by our SDK. This computational load is further compounded by the heightened processing requirements necessitated by the incorporation of supplementary 3D elements within the DT framework. Secondly, the utilization of a demo server, not designed to meet the exacting standards of a production-grade server, hampers the efficacy of the backend system. In response to these challenges, the following strategy is being pursued. Firstly, we are in the process of adapting to the latest innovations introduced within the SDK, contemplating the potential migration of select processing tasks to the backend infrastructure. Secondly, within the domain of backend operations, active efforts are underway to optimize efficiency through the transition to a more streamlined server architecture.

On a separate note, allowing the replica model to perform actions in the real world is becoming increasingly common in digital twins. Not only through user interaction but also automatically, digital twins become almost autonomous, able to learn and use actuators according to this learning. This also opens up several future research and application topics, such as integration with AI models and the adoption of predictive analytics, which can be used to anticipate situations or support making informed decisions. This predictive analysis will also benefit from other future functionality of interest such as notification and alerts, either to the user or to other components integrated with the digital twin.

As for other specific challenges related to research and commercial interests, there is the integration of human avatars into digital twins. Including the human element as part of the digital twin [21] is a trend that presents challenges in different fields. On the one hand, there are the security and privacy aspects, since the human-aware digital twin may be representing sensitive information related to the human element. On the other hand, there are issues of integration, movement, and even interaction with these avatars, both human and virtual. In particular, the authors are working on the representation of avatars in the digital twin, with which the users can interact using multimodal interaction and move around the virtual model using pathfinding algorithms [22,23]. In this way, the model can be provided with conversational elements that will guide users to access different locations or specific points of the virtual model. All this integration of avatars in the digital twin is a challenge and follows similar directions of research on themes related to the metaverse and virtual worlds, one of the current challenges and emerging trends, cited, among others, in the Emerging Tech Impact Radar 2023 [24].

Finally, and as a specific challenge related to potential future work, there are also interests and research needs in terms of information visualization. In particular, for the representation of certain noxious gases it could be of interest the integration of particle systems on top of the 3D virtual models so that in different environments, it is possible to see graphically how the volume of these gases is evaluated in interiors. This functionality is of interest not only for offices like in the use case presented in this paper, but also for industries, for example in the manufacture of perfumes or even in welding facilities.

9. Conclusions

Sectors such as construction and industry are rapidly going digital. The amount of data being monitored and processed has increased significantly since the advent of the IoT. All this data, along with performance information, is presented in a variety of dashboards and serves as the basis for decision-making.

Although dashboards and panels are a useful tool, in some cases it is of interest to have this information presented on a virtual model of the real element or component we want to monitor. This facilitates understanding and helps identify potential problems or bottlenecks. Digital twins, virtual replicas that resemble a real entity, are the trend in data representation and visualization, as well as in simulation and monitoring tools.

In this paper, we presented our approach to improving digital twin-based experiences that need to present information from multiple sources. By sources, we mean not only IoT sensors but also processing and analysis of large amounts of information coming from BigData or including real-time information from social networks.

To this end, we presented an architecture that combines Big Data frameworks such as Apache Spark and Kafka with proprietary and third-party APIs, as well as visualization tools for 3D virtual environments such as Matterport, which can simplify the creation of digital twins of buildings or factories and go beyond traditional dashboards to provide immersive experiences.

From the presented architecture we implemented a use case in the construction industry. In particular, the virtual twin of the R2M office in Pavia was realized. In it, sustainability information, environmental sensors, data from transportation services, data from the company's Data Lakes, and information from social networks were represented. In this way, the different elements of the architecture were covered.

The use case is complemented by an evaluation of the implemented idea and solution. In particular, the System Usability Scale methodology (SUS) was performed with 20 users to obtain their overall impression and identify gaps to define future challenges. It should be noted that the results of the evaluation were very positive, especially in terms of usability and the ability to display relevant information at specific points and elements of the digital twin that are connected to the real world.

We also identified some challenges and future work and dedicated a separate section to them. Supporting actions in the virtual model that affect the real world, or even supporting the virtual model to represent human activities are some of the next steps that will be taken.

Overall, the architecture presented in this article, along with the particular demonstration and integration capabilities presented, opens up numerous possibilities. Sectors such as the aforementioned construction or industry can benefit greatly from these technologies and research.

CRediT authorship contribution statement

Rubén Alonso: Conceptualization, Data curation, Validation, Methodology, Writing – original draft, Writing – review & editing. **Riccardo Locci:** Data curation, Software, Visualization. **Diego Reforgiato Recupero:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: the author Diego Reforgiato Recupero is Associate Editor of the Information Science section of the Heliyon Journal.

Data availability

The data required to reproduce the above findings cannot be shared at this time due to legal reasons.

Acknowledgements

We acknowledge financial support under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.5 - Call for tender No. 3277 published on December 30, 2021 by the Italian Ministry of University and Research (MUR) funded by the European Union – NextGenerationEU. Project Code ECS0000038 – Project Title eINS Ecosystem of Innovation for Next Generation Sardinia – CUP F53C22000430001- Grant Assignment Decree No. 1056 adopted on June 23, 2022 by the Italian Ministry of University and Research (MUR).

This work has been partly supported by Horizon Europe research and innovation programme project INPERSO - INdustrialised and PErsonalised Renovation for Sustainable sOcieties (Grant agreement ID: 101069820).

References

[1] A. Fuller, Z. Fan, C. Day, C. Barlow, Digital twin: enabling technologies, challenges and open research, IEEE Access 8 (2020) 108952–108971, https://doi.org/10.1109/ACCESS.2020.2998358.

[2] T.H.-J. Uhlemann, C. Schock, C. Lehmann, S. Freiberger, R. Steinhilper, The digital twin: demonstrating the potential of real time data acquisition in production systems, Proc. Manuf. 9 (2017) 113–120, https://doi.org/10.1016/j.promfg.2017.04.043, 7th Conference on Learning Factories, CLF 2017.

[3] N. Faruqui, M.A. Kabir, M.A. Yousuf, M. Whaiduzzaman, A. Barros, I. Mahmud, Trackez: an iot-based 3d-object tracking from 2d pixel matrix using mez and fsl algorithm, IEEE Access 11 (2023) 61453–61467, https://doi.org/10.1109/ACCESS.2023.3287496.

[4] A. Mazumder, P.S. Banerjee, A. Karmakar, P. Ghosh, D. De, H. Song, Digital twin for industry 5.0: a vision, taxonomy, and future directions, in: S. Bhattacharyya, J.S. Banerjee, D. De, M. Mahmud (Eds.), Intelligent Human Centered Computing, Springer, Singapore, 2023, pp. 246–259.

[5] M. Mashaly, Connecting the twins: a review on digital twin technology & its networking requirements, Proc. Comput. Sci. 184 (2021) 299–305, https://doi.org/10.1016/j.procs.2021.03.039, The 12th International Conference on Ambient Systems, Networks and Technologies (ANT) / The 4th International Conference on Emerging Data and Industry 4.0 (EDI40) / Affiliated Workshops.

[6] A. Thelen, X. Zhang, O. Fink, Y. Lu, S. Ghosh, B.D. Youn, M.D. Todd, S. Mahadevan, C. Hu, Z. Hu, A comprehensive review of digital twin —part 1: modeling and twinning enabling technologies, Struct. Multidiscip. Optim. 65 (12) (2022) 354, https://doi.org/10.1007/s00158-022-03425-4.

[7] A. Thelen, X. Zhang, O. Fink, Y. Lu, S. Ghosh, B.D. Youn, M.D. Todd, S. Mahadevan, C. Hu, Z. Hu, A comprehensive review of digital twin—part 2: roles of uncertainty quantification and optimization, a battery digital twin, and perspectives, Struct. Multidiscip. Optim. 66 (1) (2022) 1, https://doi.org/10.1007/s00158-022-03410-x.

[8] D.M. Botín-Sanabria, A.-S. Mihaita, R.E. Peimbert-García, M.A. Ramírez-Moreno, R.A. Ramírez-Mendoza, J. Jesús Lozoya-Santos, Digital twin technology challenges and applications: a comprehensive review, Remote Sens. 14 (2022) 1335.

[9] B.R. Barricelli, E. Casiraghi, D. Fogli, A survey on digital twin: definitions, characteristics, applications, and design implications, IEEE Access 7 (2019) 167653–167671, https://doi.org/10.1109/ACCESS.2019.2953499.

[10] K. Kukushkin, Y. Ryabov, A. Borovkov, Digital twins: a systematic literature review based on data analysis and topic modeling, Data 7 (12) (2022), https://doi.org/10.3390/data7120173.

[11] M. Liu, S. Fang, H. Dong, C. Xu, Review of digital twin about concepts, technologies, and industrial applications, J. Manuf. Syst. 58 (2021) 346–361, https://doi.org/10.1016/j.jmsy.2020.06.017, Digital Twin towards Smart Manufacturing and Industry 4.0.

[12] M. Groshev, C. Guimarães, J. Martín-Pérez, A. Oliva, Toward intelligent cyber-physical systems: digital twin meets artificial intelligence, IEEE Commun. Mag. 59 (8) (2021) 14–20, https://doi.org/10.1109/MCOM.001.2001237.

[13] J. Guo, Z. Lv, Application of digital twins in multiple fields, Multimed. Tools Appl. 81 (19) (2022) 26941–26967, https://doi.org/10.1007/s11042-022-12536-5.

[14] S. Das, Digital twins: the key to unlocking industry 4.0 and beyond, https://doi.org/10.2139/ssrn.4426592, 2023.

[15] A. Ammar, H. Nassereddine, N. AbdulBaky, A. AbouKansour, J. Tannoury, H. Urban, C. Schranz, Digital twins in the construction industry: a perspective of practitioners and building authority, Front. Built Environ. 8 (2022), https://doi.org/10.3389/fbuil.2022.834671.

[16] A.M. Madni, C.C. Madni, S.D. Lucero, Leveraging digital twin technology in model-based systems engineering, Syst. 7 (1) (2019), https://doi.org/10.3390/systems7010007.

[17] H. Yavuz, E. Konacaklı, in: E. Karaarslan, Ö. Aydin, Ü. Cali, M. Challenger (Eds.), Digital Twin Applications in Spacecraft Protection, Springer, Singapore, 2023, pp. 283–294.
[18] J.-S. Park, D.-G. Lee, J. Jimenez, S.-J. Lee, J.W. Kim, Human-focused digital twin applications for occupational safety and health in workplaces: a brief survey and research directions, Appl. Sci. 13 (2023) 4598, https://doi.org/10.3390/app13074598.
[19] G. Sun, Y. Deng, Digital twin-based modeling of complex systems for smart aging, Discrete Dyn. Nat. Soc. 2022 (2022) 7365223, https://doi.org/10.1155/2022/7365223.
[20] H. Yao, D. Wang, M. Su, Y. Qi, Application of digital twins in port system, J. Phys. Conf. Ser. 1846 (1) (2021) 012008, https://doi.org/10.1088/1742-6596/1846/1/012008.
[21] E. Montini, V. Cutrona, N. Bonomi, G. Landol , A. Bettoni, P. Rocco, E. Carpanzano, An iiot platform for human-aware factory digital twins, Proc. CIRP 107 (2022) 661–667.
[22] R.A. Saeed, D.R. Recupero, P. Remagnino, The boundary node method for multi-robot multi-goal path planning problems, Expert Syst. J. Knowl. Eng. 38 (6) (2021), https://doi.org/10.1111/exsy.12691.
[23] R.A. Saeed, D.R. Recupero, P. Remagnino, A boundary node method for path planning of mobile robots, Robot. Auton. Syst. 123 (2020), https://doi.org/10.1016/j.robot.2019.103320.
[24] Tuong Nguyen, D.C. Annette Jump, Emerging Tech Impact Radar, Gartner Research, 2023, 2022.