Likhith Kanigolla[1], Gaurav Pal[1], Karthik Vaidhyanathan[1,2], Deepak Gangadharan[1], and Anuradha Vattem[1]

[1]Smart City Research Centre, International Institute of Information Technology -Hyderabad (IIIT-H)
[2]Software Engineering Research Center, International Institute of Information Technology -Hyderabad (IIIT-H)

May 09, 2024

Architecting Digital Twin for Smart City Systems: A Case Study

Likhith Kanigolla[1], Gaurav Pal[1], Karthik Vaidhyanathan[1,2], Deepak Gangadharan[1], Anuradha Vattem[1]

[1]Smart City Research Centre, International Institute of Information Technology - Hyderabad (IIIT-H), India
[2]Software Engineering Research Center, International Institute of Information Technology - Hyderabad (IIIT-H), India
likhithkanigolla@gmail.com, gaurav.pal@research.iiit.ac.in
karthik.vaidhyanathan@iiit.ac.in, deepak.g@iiit.ac.in, anuradha.vattem@research.iiit.ac.in

***Abstract*—Urbanization, driven by technological advancements, has brought about improved connectivity and efficiency, especially with the rise of Internet of Things (IoT) devices. Smart cities use these innovations to manage resources better and enhance resident's quality of life. However, implementing smart city initiatives comes with challenges like monitoring, maintaining, and testing urban infrastructure. Digital Twin (DT) entails the connection of physical facilities or devices with their digital counterparts, facilitating real-time monitoring, manipulation, and predictive analysis of their behavior. This concept offers a virtual replica of assets, processes, and systems, enabling insights into their real-time performance and predictive behaviors. By simulating real-world scenarios, DT aids in planning maintenance activities and conducting comprehensive testing, thereby enhancing the resilience and efficiency of smart city systems. Particularly in the context of managing water networks, DT technology holds significant promise. Visualization capabilities provide intuitive insights into the system's behavior, facilitating informed decision-making. This visualization, coupled with actuation capabilities, enables control actions based on predictive analytics and optimization algorithms, allowing for proactive management of water resources and infrastructure. To this end, in this paper, we present the architecture of WaterTwin, a DT developed for water quality networks in smart city systems. We demonstrate our approach through the use of a water quality network at the smart city living lab, IIIT Hyderabad campus.**

***Index Terms*—Digital twin, Smart City, Software Architecture, IoT.**

I. Introduction

The advent of the Internet of Things (IoT), coupled with advancements in computing has given rise to many Smart City initiatives across the globe [1].From an opportunity perspective, as cities grow and expand, smart and innovative solutions are crucial for improving productivity, increasing operational efficiencies, reducing management costs, etc. [3]. Smart cities leverage IoT devices to collect different types of data to efficiently and effectively manage water quality, energy consumption, air quality, waste management, etc. From a challenges standpoint, architecting IoT-enabled smart city systems requires managing challenges related to heterogeneity, interoperability, scalability, mobility, connectivity, security, etc. [2]. Interoperability remains a significant hurdle due to the diverse range of devices, platforms, and communication protocols used in smart city deployments. The absence of interoperability can be crucial, as it makes it difficult to comprehensively understand city dynamics and delays development. Maintenance of smart city infrastructure is another critical challenge. Regular maintenance is essential to ensure the reliability and performance. Rigorous testing is essential pre-deployment to verify the functionality, interoperability, and security of smart city solutions. The complexity and interconnected nature of urban systems necessitate thorough testing of solutions before full-scale implementation. However, existing testing methods may not adequately capture the intricate interactions and dynamic behavior of real-world urban environments, leading to unforeseen challenges and potential disruptions upon deployment. Real-time analysis and response capabilities are also crucial for effective smart city management. While data collection has become increasingly sophisticated, the ability to analyze this data in real time and trigger appropriate actions remains limited. This is particularly critical for identifying and responding to emerging issues and anomalies across different urban domains, such as traffic congestion, energy fluctuations, or environmental hazards. Additionally, adaptability to dynamic changes is essential for building resilient and sustainable smart cities. Urban environments are constantly evolving due to factors like population growth, economic development, and climate change. Existing smart city systems often struggle to adapt to these dynamic conditions, leading to inefficiencies and increased vulnerability to disruptions.

In spite of the various challenges, the notion of smart cities has acquired substantial support as a solution to the emerging issues of fast urban growth. At the heart of these smart city efforts is the concept of a digital twin, an innovative method that promises to transform municipal infrastructure management [4]. The term "digital twin" (DT) refers to a dynamic virtual representation of a physical product or system that uses real-time data to facilitate comprehension, learning, and reasoning throughout its existence [5]. By creating a digital twin of smart city infrastructure, planners and operators gain a powerful tool for continuous monitoring and analysis. Digital twins enable proactive maintenance by providing insights into the condition and performance of assets, allowing for predictive maintenance strategies to be implemented effectively.

Digital twins facilitate comprehensive testing checks before deployment by simulating real-world scenarios and identifying potential issues in a risk-free environment. Through the use of advanced analytics and simulation techniques, digital twins help optimize the design, operation, and management of smart city systems, ultimately enhancing their resilience, efficiency, and sustainability.

However, developing digital twins presents different challenges related to data modeling, developing simulation support, etc. Our research contributes to bridging this gap by proposing a novel digital twin architecture that addresses these challenges through its focus on interoperability, real-time processing, simulation, and actuation capabilities. To this end, in this paper, we present an approach that we used to architect a digital twin for a water network in a smart city scenario. We demonstrate our approach by applying the concept of developing a DT of the water network in the Smart City Living Lab of IT Hyderabad. The initial implementation and evaluation of WaterTwin demonstrate its potential for improving the management of water quality networks and suggest its adaptability to other smart city domains.

II. Related Works

Digital Twin (DT) technology has rapidly gained traction in recent years, captivating the interest of both academic researchers and industry practitioners. Its applications span a diverse range of domains, from manufacturing and healthcare to urban planning and infrastructure management. In the healthcare sector, researchers like LAAMARTI have demonstrated the potential of DTs for personalized medicine and proactive healthcare management [15]. By collecting and analyzing health data from individuals, these digital twins can provide insights into potential health risks and enable early interventions, ultimately leading to improved well-being. Similarly, Barat et al showcased the utility of DTs in modeling and understanding complex public health crises, such as the COVID-19 pandemic [19]. Their work demonstrated how these virtual models can help track the spread of infectious diseases and evaluate the effectiveness of various intervention strategies.

The manufacturing industry has also embraced DT technology to optimize production processes and enhance efficiency. Meueller-Zhang et al proposed a novel approach that combines digital twins with reinforcement learning and digital twins for production efficiency [21].

Smart cities are another domain where DTs hold immense promise. Mohammadi et al have conceptualized digital twins for entire urban environments, envisioning their ability to enhance decision-making, optimize resource allocation, and improve the overall quality of life for city residents [17]. However, despite the exciting potential of DTs, their development and implementation present significant challenges. Researchers like Barn et al have highlighted the complexities of modeling socio-technical aspects within DTs, emphasizing the need to account for human behavior and societal factors [18]. Additionally, Macias et al proposed a domain-driven design-based approach [20].

In this work, we present the architecture of WaterTwin, a DT for a water quality network in the smart city domain (through a case study of smart city living lab), differently from the above and aligned with the concepts presented in [20]. We further provide a concrete implementation of the DT with some initial observations. We believe this will provide a foundation for the future development of DT in the Smart City domain.

III. Case Study and Motivation

A. Case Study: Smart City Living Lab

The Smart City Living Lab of IIITH[1] is a research platform and test bed for smart city applications comprising more than 300 IoT nodes spanning various domains such as air quality, water quality and quantity, solar power monitoring, home automation, etc. The smart campus is created to enhance three value domains: social, economic, and environmental [7].

This sensor network comprises over 300 sensors of various types, including air quality sensors, water quality sensors, water quantity sensors, crowd monitoring sensors, energy monitoring, smart room monitoring, solar monitoring, and occupancy sensors. The 60-acre campus of IIIT-H situated in the city of Hyderabad, Telangana state will be converted into a smart campus as part of the project. The supreme objective of this project is to transform the campus into a platform for learning, experimentation, and showcasing new ideas and approaches. It is essentially a live setup of a building, campus, or facility that gets used for its intended purposes along with being a ‘living’ testbed for research and innovation. The deployment of these sensors enables real-time monitoring of environmental parameters, infrastructure utilization, and human activities within the campus. By collecting and analyzing data from these sensors, Smart City Living seeks to gain insights into campus dynamics, identify areas for improvement, and optimize resource allocation.

Among the various domains, one of the important ones of interest to the lab is water quality and quantity. The IIITH campus hosts a sewage treatment plant and various water quality and quantity sensors. These sensors play a crucial role in monitoring the quality of water resources and tracking water consumption patterns across the campus. By collecting data on water quality and usage, Smart City Living aims to enhance the efficiency of water management practices and promote conservation efforts. However, there are some challenges in water treatment, which further motivates the need for an advanced approach to managing the water quality/quantity networks.

B. Challenges in Water Treatment

In a hypothetical scenario, let’s envision the sewage treatment plant on the campus grappling with a malfunctioning module tasked with monitoring water quality. This malfunction results in the reporting of inaccurate values, consequently

[1] https://smartcitylivinglab.iiit.ac.in (Last accessed 20 Feb 2024)

leading to inadequate treatment of water. Such an event of contamination poses a significant risk to the safety of the treated water, potentially endangering public health and the environment alike [11].

This scenario underscores the critical importance of advanced monitoring and control mechanisms to swiftly detect anomalies and optimize system performance. It also sheds light on the primary problems within the current IoT elements of water treatment, as illustrated by [12]. One notable issue is the absence of fallback measures in cases where water is not treated correctly before being discharged back into the source.A critical concern in current water treatment systems is the lack of adequate fallback measures to prevent the discharge of inadequately treated water back into the environment.In the event of equipment malfunctions, sensor failures, or unexpected fluctuations in water quality, existing systems often lack the necessary safeguards to prevent contaminated water from entering rivers, lakes, or other water bodies.

Based on the data collected from these sensors, administrators determine the appropriate actions for the filtration process. This scenario is not specific to the Smart City Living Lab. Globally, there have been instances of fatalities resulting from waterborne illnesses caused by inadequately treated water, a scenario that could occur if any of the sensors malfunction [9]. Furthermore, cases have been documented where contaminated water has affected numerous consumers over an extended period, leading to various illnesses among people [10]. To overcome this issue, manual checks should be performed daily. However, it may not be feasible, could result in time, and can cause potential errors.

C. Digital Twin: Possibilities and Challenges

The above use case presents a classic scenario where DT can be of immense use. DT creates virtual replicas of physical assets or systems, enabling real-time monitoring, predictive analysis, and simulation of real-world scenarios. By integrating data from various sensors and analytics tools, digital twins offer insights into system behavior and facilitate informed decision-making. However, building DT for a given domain presents its challenges. These include: *CH1: Data collection:* Building a digital twin requires collecting good quality data to model the real behavior; *CH2: Interoperability* Ensuring seamless interoperability between the digital twin and the physical system presents a significant challenge due to the heterogeneous nature of smart city environments. Additionally, interoperability challenges extend to *actuation manager*, as the digital twin needs to be able to send control signals to various devices with potentially different control interfaces and protocols.; *CH3: Real-time data processing* The data needs to be processed in near-real time which calls for robust data pipelines; and *CH4: Scalability* As smart cities continue to expand and incorporate more sensors and devices, scalability becomes a crucial concern. The digital twin architecture must be able to handle the increasing volume of data generated by these devices without compromising performance or efficiency.Furthermore, the system should be designed to efficiently manage communication and coordination between a large number of devices, ensuring real-time responsiveness and preventing bottlenecks. Failure to address scalability challenges can limit the digital twin's effectiveness and hinder its ability to keep pace with the evolving needs of the smart city.

Addressing some of the challenges above, in this paper, we propose an architecture for building a digital twin catering to Smart City systems. We make use of the water treatment use case in the remainder of the paper to explain our approach. Further, we also provide an initial realization of our architecture.

IV. The Water Twin Architecture

Figure 1 provides a high-level view of the architecture of the DT System for water quality networks. Although this DT is built for water networks, the components used can be generalized to different domains. As depicted in Figure 1, the architecture consists of different components for facilitating data collection from IoT Nodes, processing, analysis, visualization engine, simulation engine, and actuation manager. In the rest of this section, we provide the details of the different functionalities of the digital twin and thereby explain the role of each of the components and their responsibility. The first step of building a digital twin is data collection. This is facilitated by the interoperability layer, thereby handling challenges *CH1 and CH2* (refer Section III, 3.3). Further, the combination of temporal and long-term data storage handles the challenge related to real-time processing and scalability (*CH3* in Section III).

Data Collection and Analysis The foundation of WaterTwin lies in its robust data collection and analysis capabilities, which provide the essential building blocks for creating a comprehensive and accurate digital representation of the physical water network. This process begins at the *IoT Nodes Layer*, consisting of a diverse array of sensors and devices strategically deployed throughout the city. These nodes continuously monitor various aspects of the water system and gather real-time data on critical water quality parameters(e.g.: TDS, pH levels, etc.).

The collected data is then transmitted from the IoT nodes to the digital twin platform through the *Communication Layer*. This layer handles diverse communication protocols, ensuring reliable and efficient data transfer. Various communication technologies may be employed depending on the specific requirements of the network and the types of sensors deployed, including cellular networks, Wi-Fi, LoRaWAN, or other low-power wide-area networks.

To overcome the challenges of heterogeneity and ensure seamless data from different types of water nodes (in general different IoT devices) is facilitated with the support of a standardized *Interoperability Layer*, ensuring consistency and compatibility across different devices and devices. To this end, we leverage the oneM2M standard [2], which is one of the globally recommended interoperability standards for

[2]https://www.onem2m.org/

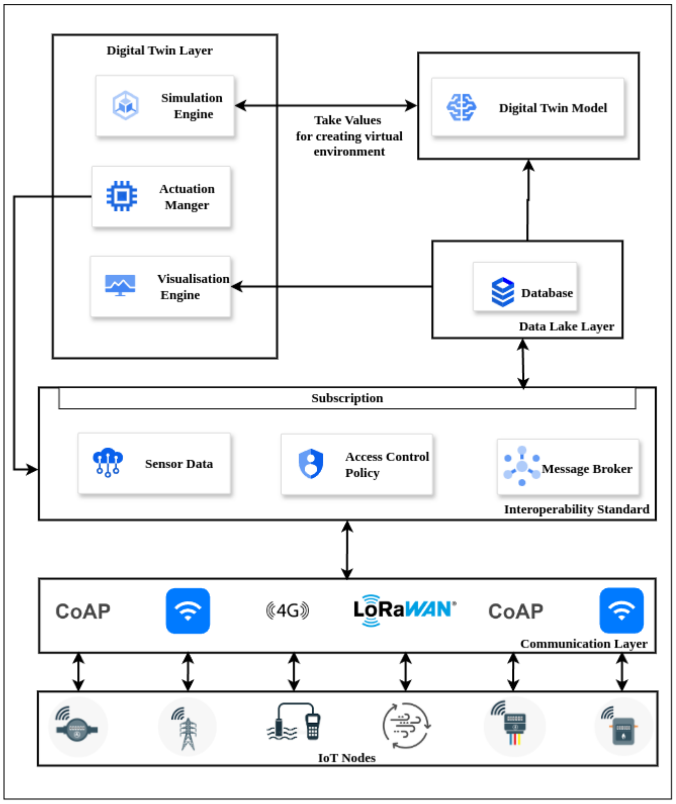

Fig. 1. The WaterTwin Architecture

IoT systems. This layer supports various functionalities, such as device registry and management, data subscriptions, data management, security, etc. By adhering to a unified protocol, the interoperability layer breaks down data and enables efficient data aggregation and analysis, providing a holistic view of the water network's status Through the subscription feature of oneM2M interoperability layer data passes through message broker which effectively distributes real-time data from the Interoperability Layer to different parts of the system, guaranteeing scalability and smooth data flow. It optimizes data distribution, decouples data producers and consumers for autonomous scalability, and speeds up decision-making. Also, this message broker is responsible for taking the requests from the digital twin layer and communicating to the interoperability layer which communicates with the devices for taking necessary actions.

Data Lake serves as the foundation for powerful visualization, the data received from the IoT devices through the interoperability layer will be stored here for further analysis. This allows the model to gain insights into the water network's performance and behavior. This data is used to build Interactive dashboards, maps, and charts to monitor key performance indicators, analyze historical trends, and identify potential issues, enabling data-driven decision-making and enhancing situational awareness. Also, this historical data helps in building the *Digital Twin Model* for predictive analytics.

Visualization Engine WaterTwin incorporates a comprehensive monitoring and alert system that continuously tracks the state of the water network and promptly notifies stakeholders of potential issues. This functionality is embedded within the *Digital Twin Layer*, where the virtual representation of the water network plays a crucial role in analyzing real-time data and identifying anomalies. The digital twin model continuously compares the calibrated sensor values received from the IoT nodes against predefined thresholds established by stakeholders. These thresholds, customizable through user-friendly web interfaces, represent acceptable ranges for various water quality parameters and operational conditions. If any sensor reading exceeds its designated threshold, indicating a potential anomaly or deviation from normal operating conditions, the digital twin model automatically triggers alerts and notifications to the relevant stakeholders. These alerts serve as early warning signals, prompting immediate attention and investigation. This leverages the power of machine learning within the Digital Twin Model Layer for predictive maintenance, aiming to prevent problems before they occur. Upon detecting threshold breaches, the system stores information about the affected devices, including timestamps, parameter values, and historical data, within the *Data Lake*. This also has the information about the timelines for the maintenance enabling responsible persons to notify them. The combination of real-time monitoring, automated alerts, and predictive maintenance capabilities ensures that potential issues are identified and addressed promptly, enhancing the overall reliability and resilience of the network.

Actuation Manager Beyond monitoring and providing insights, WaterTwin empowers stakeholders to actively control and optimize the water network through its actuation capabilities. Based on the continuous analysis of real-time data, predictive analytics, and the understanding of device behavior derived from the *digital twin model*, the system can recommend adjustments to various control parameters within the water network. These recommendations may include adjusting valve settings to optimize water flow distribution, controlling pump operations to maintain desired pressure levels, or activating or deactivating treatment processes in response to changes in water quality. Upon review and confirmation by authorized personnel for critical actions, these recommended control actions are transmitted through the *Interoperability Layer* to the corresponding devices within the physical water network. The system then meticulously monitors the impact of these actions, analyzing feedback from sensors to ensure that the desired outcomes are achieved and the system returns to a stable and optimal state. This closed-loop control system allows for continuous adaptation and optimization of the water network based on real-time conditions and predictive insights, ensuring efficient water usage, maintaining water quality, and preventing potential issues before they escalate. While the digital twin model strives to provide accurate predictions and recommendations, WaterTwin acknowledges that unforeseen problems or complex scenarios may arise where manual intervention is necessary. In such cases, the system allows operators to directly modify control parameters or override automated actions through the user interface, leveraging their expertise and understanding of the specific situation to ensure the best possible outcomes. Yet, certain unknown issues may arise during testing on hardware devices, potentially leading to device failures and significant financial losses. To mitigate such risks, a virtual mirror model [8], which supports simulation, can be employed.

Calibration of hardware sensors over time is crucial for ensuring accurate and reliable measurements. This process helps maintain the precision and consistency of sensor readings, compensating for factors such as environmental changes, component degradation, and wear and tear. The *Data Lake* plays a key role in this process by forwarding the collected data to a dedicated ML Model Component for further processing and analysis. This ML model, trained on historical data and reference sensor readings, is designed to identify and correct potential errors or inconsistencies in the sensor measurements. It achieves this by detecting outlier values or patterns that deviate significantly from expected behavior, indicating potential sensor malfunctions or environmental anomalies. Based on the identified variances and error patterns, the ML model can estimate calibration values to adjust the raw sensor data and improve its accuracy. Additionally, the model analyzes the variance of each sensor's readings compared to reference sensors or established benchmarks, providing insights into the accuracy and reliability of individual devices and these values can be updated automatically or through *Actuation Manager* which communicates with *Interoperability Layer* which make the changes to the IoT Devices. This continuous calibration process enhances the overall data quality and reliability of the digital twin model, allowing WaterTwin to accurately reflect the real-world state of the water network and enabling more informed and effective management decisions.

Simulation Engine plays a pivotal role in WaterTwin, providing a powerful tool for risk mitigation, optimization, and proactive problem-solving. In situations where issues persist despite proactive measures and real-time control actions, stakeholders can leverage the simulation engine to troubleshoot problems and explore potential solutions in a virtual environment. The simulation engine component utilizes data from the *Data Lake*, including historical sensor readings, environmental parameters, and device characteristics, to create a highly accurate digital replica of the water network. This virtual model incorporates various environmental factors that can influence water quality and system behavior, such as pipe length and diameter, water temperature, and external environmental conditions. Furthermore, the simulation engine component allows for the integration of virtual sensors, which replicate the behavior of newly installed or ideal sensors with high accuracy. This enables stakeholders to test and validate the placement and performance of additional sensors within the network, optimizing their deployment for maximum effectiveness. By conducting experiments and analyzing results within the simulation engine environment, users can gain valuable insights into the root causes of persistent issues, evaluate the potential impact of different interventions, and devise practical solutions without disrupting the real-world water system. The ability to simulate various scenarios helps to proactively identify potential problems, test different control strategies, and optimize the performance of the water network. This leads to improved decision-making and reduced risks associated with real-world interventions.

V. Implementation

In this section, we provide details on implementing the architecture presented in Section IV. We first present the hardware setup, followed by the details of the software stack.

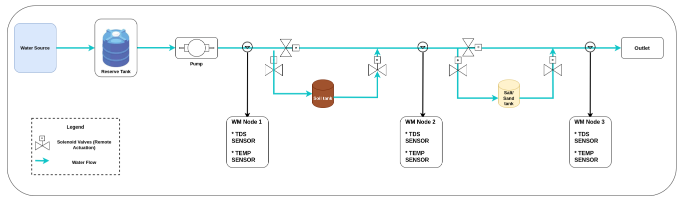

Fig. 2. The hardware setup with three water nodes, One Soil, and One Sand Container

Hardware Setup shown in Figure 2 provides the overview of the hardware setup. It consists of three water quality nodes built using ESP32 microcontrollers [22]. Each node functions as a self-contained monitoring unit, equipped with sensors such as analog TDS Sensor and Temperature Sensor specifically chosen to measure essential water quality parameters [23] [24]. Two of the containers are placed within the setup one is filled with sand and soil. Solenoids are placed in the pipes to control the flow of the water through pipes and containers. These parameters include uncompensated Total Dissolved Solids (TDS), which provides a general indication of the overall concentration of dissolved substances in the water; compensated TDS, which offers a more accurate measurement by accounting for the influence of water temperature; temperature and voltage [25]. To facilitate data collection and integration with the digital twin platform, all sensor nodes are connected to a central server via the Internet. This connection enables the continuous transmission of real-time data, ensuring that the digital twin remains synchronized with the current state of the physical water network. The collected data is then systematically stored in a database, adhering to the data flow depicted in Figure 1. This organized storage facilitates subsequent analysis, evaluation of the WaterTwin architecture, and the training of machine learning models.

Following the establishment of the hardware setup, a rigorous data collection process was undertaken to gather a comprehensive and diverse dataset. Approximately 30,000 data points were collected for each test case, encompassing variations involving sand, salt, and soil mixtures introduced into the water to simulate different water quality conditions and potential contaminants. As an initial step following the hardware setup, we collected about 30,000 data points for each test case, including variations involving sand, salt, and soil combinations. It enables the analysis of the system's adaptability to external changes, providing insights into how the digital twin model and control mechanisms respond to fluctuations in water quality parameters and varying environmental conditions. The rich dataset facilitates the exploration of potential correlations and the discovery of deeper insights into the complex relationships between different water quality parameters, helping to identify key factors influencing the overall health and behavior of the water network. This meticulous data collection approach ensures a comprehensive evaluation of WaterTwin's capabilities and provides valuable insights into its performance and adaptability within a controlled environment, paving the way for further refinement and optimization of the system.

Software Setup leverages a combination of powerful open-source tools and technologies, carefully selected for their robustness, scalability, and suitability for smart city applications. For the critical interoperability layer, we employed Eclipse OM2M[3] a leading open-source implementation of the oneM2M standard. OM2M provides a comprehensive set of functionalities for device management, data exchange, and communication between heterogeneous devices and platforms within the Internet of Things (IoT) ecosystem. By utilizing OM2M, WaterTwin ensures seamless interoperability and facilitates the integration of diverse sensors and devices within the water network, regardless of their manufacturer or communication protocol. To manage the vast amount of data generated by the sensor network, we leveraged the existing data lake infrastructure of the smart city living lab. This data lake utilizes a combination of PostgreSQL and MongoDB, two widely adopted and highly scalable database management systems. PostgreSQL, a relational database, provides efficient storage and retrieval of structured data, while MongoDB, a NoSQL database, offers flexibility in handling unstructured and semi-structured data. This hybrid approach ensures efficient data management and accommodates the diverse data types and formats encountered within the smart city environment.

For the implementation of the message broker, we utilized Apache Kafka[4], a distributed streaming platform renowned for its high throughput, low latency, and fault tolerance. Kafka efficiently handles handle real-time data feeds from the interoperability layer, ensuring reliable and scalable data distribution to various components of the WaterTwin architecture. Its ability to manage high-velocity data streams makes it ideal for processing the continuous influx of information from the sensor network, facilitating real-time analysis and timely responses to changing conditions. Utilizing the gathered data, in-depth analyses are performed to calculate variance values, providing insights into the system's dynamic behavior across different conditions. In addition, a user-friendly web interface is developed to visualize real-time data, facilitating monitoring of key parameters. Also through *Actuation* interface features remote control functionality, enabling manipulation of solenoids within the hardware setup from any location on the campus network to regulate water flow rates shown.

Furthermore, algorithms are employed to compute variance values based on factors like pipe distances, soil and salt dissolution rates, and changes in parameters such as Total Dissolved Solids (TDS) and temperature. These calculations offer crucial insights into system performance and inform decision-making processes.

The gathered data undergoes in-depth analyses to calculate variance values, offering insights into dynamic system behavior across diverse conditions. To this end, we leveraged the different data analysis packages provided by Python. Algorithms compute variance values based on pipe distances, soil and sand dissolution rates, and changes in parameters like Total Dissolved Solids (TDS) and temperature, informing decision-making. Using react.js [5], we provide a user-friendly web interface that visualizes real-time data, enabling remote control of solenoids to regulate water flow rates across the campus network. Also using the Actuation Manager shown Fig 3 provides a seamless user experience and facilitates immediate response to detected issues, ensuring timely mitigation measures can be enacted to maintain water quality standards to control devices through the web interface.A simulation Engine is also created, presenting a virtual representation of the system architecture shown in Fig 4. The simulation engine within WaterTwin serves as a powerful tool for understanding, predicting, and managing the behavior of the water network. Virtual representation of the system, allows users to interact with it and explore various scenarios without affecting the real-world infrastructure. Real sensor nodes, positioned according to their actual GPS coordinates, are displayed on a virtual map using the Leaflet library[6], creating a visual replica of the physical network layout. Pipes connecting the nodes are then drawn, accurately depicting the water flow pathways. The key functionality of the simulation page lies in its ability to incorporate virtual sensors and virtual containers(soil/sand), enabling users to explore hypothetical situations and their potential impact on water quality. This feature allows users to simulate potential contamination scenarios by introducing these substances at various points within the virtual network. The system then calculates the impact of such contamination on water quality parameters, enabling stakeholders to assess the potential risks and develop mitigation strategies. Users can gain experience by placing virtual sensors at any point along the pipeline, and the system dynamically calculates the expected values of water quality parameters at those locations. This calculation takes into account various factors, including the distance from the previous node, solenoid status, the presence and composition of any added substances, and temperature data. By leveraging mathematical models or machine learning algorithms trained on historical data, the simulation engine predicts parameters like TDS, providing insights into how water quality evolves as it flows through the network.

An alert system triggers notifications based on collected values, automatically monitoring solenoid states to detect potential contamination and prompt further investigation. Users

[3]https://projects.eclipse.org/projects/iot.om2m
[4]https://kafka.apache.org/
[5]https://react.dev/
[6]https://leafletjs.com/

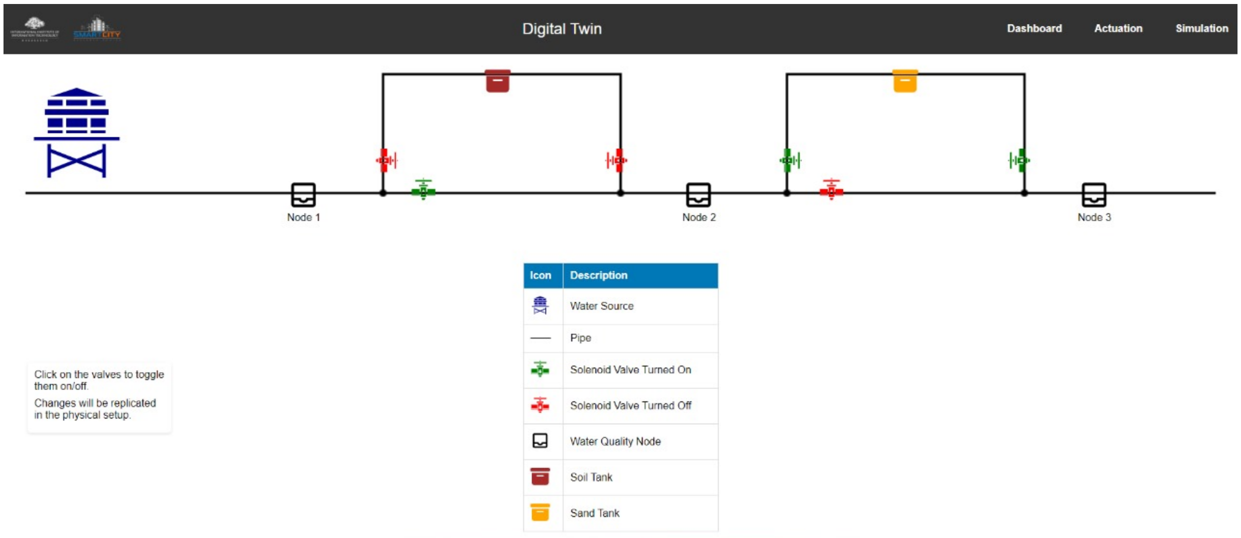

Fig. 3. Actuation interface of the realized Digital Twin for water networks

can control solenoids directly from the web interface through the actuation manager page, ensuring immediate response to detected issues and maintaining water quality standards seamlessly. Through a combination of visualization, simulation engine, and actuation capabilities, users can gain valuable insights into system behavior and directly control devices from the web interface.

Initial Results with WaterTwin yielded promising results, demonstrating the effectiveness of the architecture and its ability to adapt to dynamic changes within the water network. Fig. 3 showcases the actuation page of the user interface, which provides a clear and intuitive representation of the system's status. Symbols depicting devices are displayed with color-coded indicators (red for off, green for on), allowing users to readily monitor and control device states directly through the interface. This real-time control functionality enables immediate responses to changing conditions and facilitates proactive management of the water network.

To evaluate the accuracy and adaptability of WaterTwin, a series of experiments were conducted using a controlled setup. We prepared 1000ml water samples mixed with varying amounts, in tablespoon quantities (approximately 15g each), of salt and sand combined in one container, simulating different contamination scenarios. The system's response to these varying conditions will be monitored and analyzed, with a particular focus on the Total Dissolved Solids (TDS) levels The results, illustrated in the graph within Fig. 5, demonstrate the variance of the TDS values with respect to the quantity of soil and sand.

The simulation page in Fig. 4, further enhances the user's ability to understand and predict system behavior. A virtual marker, displayed in blue along the pipeline, represents the predicted values of water quality parameters at that specific location. The container added at the start of the pipe is a virtual sand container. This visual representation of predicted conditions aids in decision-making and supports proactive maintenance efforts by identifying potential problem areas before they escalate. The distance calculation and validations play a crucial role. Once the virtual node is placed we first calculate whether a given point lies near a line segment defined by four points. Let (x_1, y_1), (x_2, y_2), (x_3, y_3), and (x_4, y_4) represent the coordinates around the virtual node, and (x, y) denote the coordinates of the virtual node which is placed. Then Slopes m_1, m_2, m_3, and m_4 are calculated using the formula $m = \frac{y_2 - y_1}{x_2 - x_1}$, and distances val_1, val_2, val_3, and val_4 are computed about the placement of the marker from the pipe. If $val_1 > 0$, $val_2 < 0$, $val_3 > 0$, and $val_4 < 0$, the virtual node is deemed to be near the pipeline and ensuring that the node is placed on the pipeline else an error will comes up stating to place the virtual marker within the pipe. Similarly same validation applies while adding a virtual soil/sand container ensuring the virtual components are placed on the pipeline.

Upon validating data, we calculate the perpendicular distance from a given point to a line segment defined by two other points. Let (x_1, y_1) and (x_2, y_2) represent the coordinates of the endpoints of the line segment, and (x, y) denote the coordinates of the point. The distance formula employed in this function is derived from principles of vector algebra and Euclidean geometry.

$$d = \frac{|(x_2 - x_1)(y_1 - y) - (x_1 - x)(y_2 - y_1)|}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$

Using this formula, we calculate the distances between the virtual node position and each node in the physical network. Additionally, we determine the pipe section in which the virtual node is located. By identifying the position of the node and its distance from the containers and actual nodes, we can determine the dissolution rate at the location of the virtual node. This rate will be incorporated into the initial results,

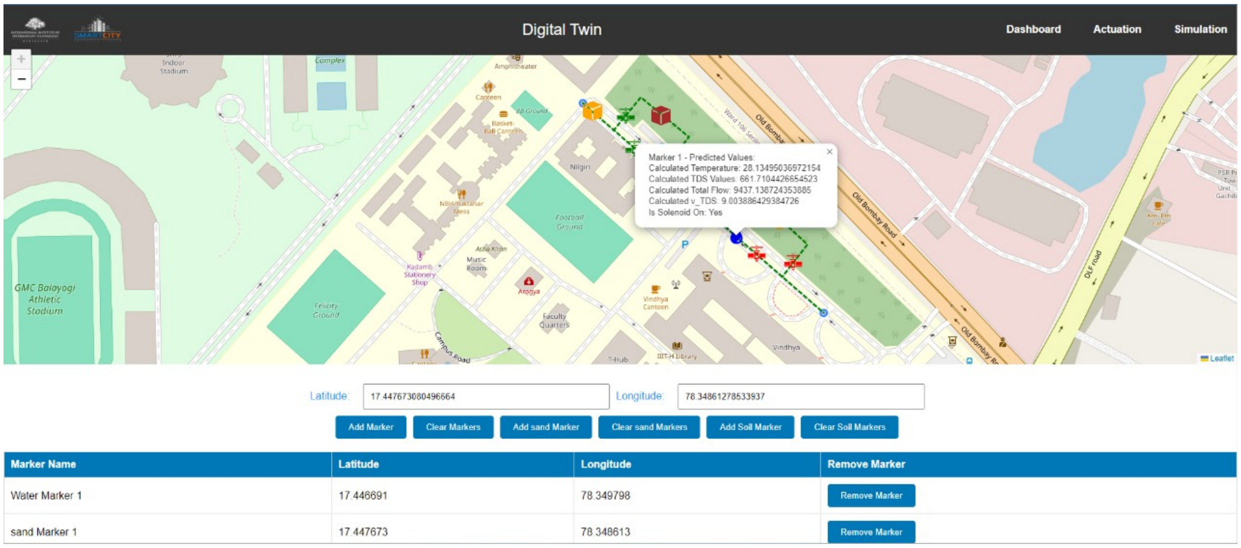

Fig. 4. Simulation page interface of the realized Digital Twin for water networks

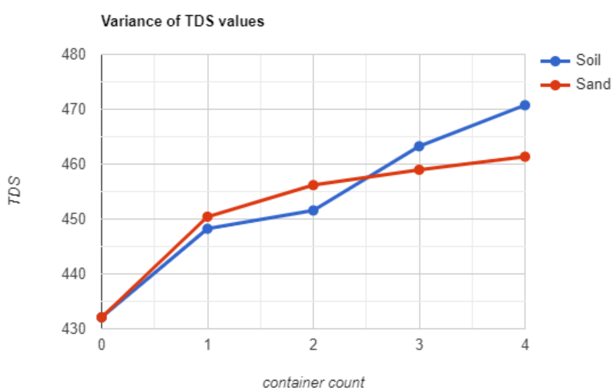

Fig. 5. Line graph showing the variance of TDS of salt and sand based on the contatiners

and the predicted Total Dissolved Solids (TDS) value at the virtual marker will be calculated based on the rate of impurity dissolution relative to the quantity of impurities.

Furthermore, we tested the system's adaptability by introducing new virtual nodes into the simulation. The predicted values generated by these new nodes were compared against expected values based on established models and expert knowledge, revealing a high degree of accuracy and consistency with over 95% confidence. This validates WaterTwin's ability to accommodate and adapt to changes in the network structure and configuration. In addition to evaluating the accuracy and adaptability of the digital twin model, we also measured the system's responsiveness to control actions. We observed that, on average, it takes approximately 800 milliseconds for the system to execute a triggered actuation command, demonstrating its ability to respond quickly and efficiently to changing conditions or emerging issues within the water network. These initial experiments provide promising evidence of WaterTwin's potential as a valuable tool for managing and optimizing smart city water systems. Future research will focus on expanding the scale of the system, incorporating additional water quality parameters, and further refining the digital twin model to enhance its predictive capabilities and adaptability."

VI. Conclusion and Future Works

This paper has presented an architecture for realizing digital twins for smart city systems, focusing on water quality networks. By leveraging advanced sensor technologies and analytics tools, digital twins offer the potential to enhance the efficiency and effectiveness of urban management practices. Although we have focused on one domain, we believe that the process followed and the conceptual architecture can be extended to other domains as well.

Future works include but are not limited to i) Extending this implementation to various domains within the smart city living lab of IIITH; ii) Implementing WaterTwin in a larger-scale, real-world water network to evaluate its performance and impact under diverse and dynamic conditions; iii) Experimenting with different ML models for improving the prediction accuracy and iv) Building a generic flow-based programming platform for realizing digital twins in the smart city domain.

Acknowledgment

The research was partly supported by Ministry of Electronics and Information Technology (MeitY), Government of India under grant no. 3070665 (2020) as part of Smart City Living lab project and being sustained by the corporate funding. We are sincerely thankful to Sasidhar Varada for his contribution to hardware development.

References

[1] Hammi, Badis Khatoun, Rida I& Zeadally, Sherali I& Fayad, Achraf I& Khoukhi, Lyes. (2018). IoT technologies for smart cities.

[2] G. V. Ihita, V. K. Acharya, L. Kanigolla, S. Chaudhari and T. Monteil, ”Security for oneM2M-Based Smart City Network: An OM2M Implementation,” 2023 15th International Conference on COMmunication Systems I& NETworkS (COMSNETS), Bangalore, India, 2023, pp. 808-813, doi: 10.1109/COMSNETS56262.2023.10041334.

[3] Khatoun, Rida I& Zeadally, Sherali. (2016). Smart cities: Concepts, architectures, research opportunities. Communications of the ACM. 59. 46-57. 10.1145/2858789.

[4] Julia Robles and Cristian Martín and Manuel Díaz. (2023). OpenTwins: An open-source framework for the development of next-gen compositional digital twins.

[5] Nast, Benjamin, Reiz, Achim, Sandkuhl, Kurt, Stirna, Janis. (2022). ”Integrating Organizational and Technological Aspects of Digital Twin Engineering”.

[6] SCRC Homepage, https://smartcitylivinglab.iiit.ac.in/. Last accessed 12 Feb 2024

[7] Pranavasri, Vjs & Francis, Leo & Mogadali, Ushasri & Pal, Gaurav & Vaddhiparthy, Suhas & Vattem, Anuradha & Vaidhyanathan, Karthik & Gangadharan, Deepak. (2023). Scalable and Interoperable Distributed Architecture for IoT in Smart Cities. 10.36227/techrxiv.24118458.v1.

[8] Zheng, Yu & Yang, Sen & Cheng, Huanchong. (2019). An application framework of digital twin and its case study. Journal of Ambient Intelligence and Humanized Computing. 10. 10.1007/s12652-018-0911-3.

[9] Benedict, K. M., Reses, H., Vigar, M., Roth, D. M., Roberts, V. A., Mattioli, M., … Hill, V. R. (2017). Surveillance for Waterborne Disease Outbreaks Associated with Drinking Water - United States, 2013-2014. MMWR: Morbidity & Mortality Weekly Report, 66(44), 1216–1221. https://doiorg.login.library.coastal.edu:8443/10.15585/mmwr.mm6644a3

[10] J. Shelley, M. Pillai and S. Kumar, ”Data Analysis of Lead Contamination in New York,” 2020 IEEE Symposium Series on Computational Intelligence (SSCI), Canberra, ACT, Australia, 2020, pp. 1614-1619, doi: 10.1109/SSCI47803.2020.9308492. keywords: Water pollution;Lead;Contamination;Water resources;Urban areas;Regulation;Fracking;Water;Lead;Contamination;New York Schools;R statistical analysis,

[11] Goparaju, S.U.N., Vaddhiparthy, S.S.S., Pradeep, C., Vattem, A. and Gangadharan, D., 2021, June. Design of an IoT system for machine learning calibrated TDS measurement in smart campus. In 2021 IEEE 7th World Forum on Internet of Things (WF-IoT) (pp. 877-882). IEEE.

[12] El-Shafeiy, E.; Alsabaan, M.; Ibrahem, M.I.; Elwahsh, H. Real-Time Anomaly Detection for Water Quality Sensor Monitoring Based on Multivariate Deep Learning Technique. Sensors 2023, 23, 8613.

[13] S. Mante, S. S. S. Vaddhiparthy, M. Ruthwik, D. Gangadharan, A. M. Hussain and A. Vattem, ”A Multi Layer Data Platform Architecture for Smart Cities using oneM2M and IUDX,” 2022 IEEE 8th World Forum on Internet of Things (WF-IoT), Yokohama, Japan, 2022, pp. 1-6

[14] S. S. Suhas Vaddhiparthy, G. Sreya, P. R. Turlapati, D. Gangadharan and H. Kandath, ”A Comprehensive Evaluation on the Impact of Various Spoofing Scenarios on GPS Sensors in a Low-Cost UAV,” 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE), Auckland, New Zealand, 2023, pp. 1-6,

[15] TY - JOUR TI - An ISO/IEEE 11073 Standardized Digital Twin Framework for Health and Well-Being in Smart Cities T2 - IEEE Access SP - 105950 EP - 105961 AU - F. Laamarti AU - H. F. Badawi AU - Y. Ding AU - F. Arafsha AU - B. Hafidh AU - A. E. Saddik PY - 2020 JO - IEEE Access IS - SN - 2169-3536 VO - 8 VL - 8 JA - IEEE Access Y1 - 2020 ER

[16] L. Raes et al., ”DUET: A Framework for Building Interoperable and Trusted Digital Twins of Smart Cities,” in IEEE Internet Computing, vol. 26, no. 3, pp. 43-50, 1 May-June 2022

[17] N. Mohammadi and J. E. Taylor, ”Smart city digital twins,” ¡em¿2017 IEEE Symposium Series on Computational Intelligence (SSCI)¡/em¿, Honolulu, HI, USA, 2017, pp. 1-5

[18] Balbir S. Barn, Tony Clark, Souvik Barat, and Vinay Kulkarni. 2023. Towards the Essence of Specifying Sociotechnical Digital Twins. In Proceedings of the 16th Innovations in Software Engineering Conference (ISEC ’23). Association for Computing Machinery, New York, NY, USA, Article 18, 1–5

[19] Barat, S., Paranjape, A., Parchure, R., Darak, S. and Kulkarni, V., 2022, December. Agent Based Simulatable City Digital Twin to Explore Dynamics of Covid-19 Pandemic. In 2022 Winter Simulation Conference (WSC) (pp. 557-568). IEEE.

[20] Macías, A., Navarro, E., Cuesta, C.E. and Zdun, U., 2023, March. Architecting Digital Twins Using a Domain-Driven Design-Based Approach. In 2023 IEEE 20th International Conference on Software Architecture (ICSA) (pp. 153-163). IEEE.

[21] Zai Mueller-Zhang, Pablo Oliveira Antonino, Thomas Kuhn, Integrated Planning and Scheduling for Customized Production using Digital Twins and Reinforcement Learning, IFAC-PapersOnLine, Volume 54, Issue 1, 2021, Pages 408-413, ISSN 2405-8963,

[22] ESP32 Microcontroller Datasheet, https://www.espressif.com/sites/default/files/documentation/esp32_datasheet_en.pdf. Last accessed 18 Feb 2024

[23] Analog TDS Sensor, https://wiki.dfrobot.com/Gravity__Analog_TDS_Sensor___Meter_For_Arduino_SKU__SEN0244. Last accessed 18th Feb 2024

[24] ds18b20 Temparature Sensor Data Sheet, https://www.quick-teck.co.uk/Management/EEUploadFile/1420644897.pdf. Last accessed 18 Feb 2024

[25] Seán Moran, Chapter 6 - Clean water characterization and treatment objectives, Editor(s): Seán Moran, An Applied Guide to Water and Effluent Treatment Plant Design, Butterworth-Heinemann, 2018, Pages 61-67, ISBN 9780128113097, https://doi.org/10.1016/B978-0-12-811309-7.00006-0.