Showcasing Data Management Challenges for Future IoT Applications with NebulaStream

Aljoscha Lepping
TU Berlin
aljoscha.p.lepping@tu-berlin.de

Hoang Mi Pham
Laura Mons
Balint Rueb
TU Berlin
lastname@campus.tu-berlin.de

Ankit Chaudhary
TU Berlin
ankit.chaudhary@tu-berlin.de

Philipp M. Grulich
TU Berlin
grulich@tu-berlin.de

Steffen Zeuch
TU Berlin
steffen.zeuch@tu-berlin.de

Volker Markl
TU Berlin, DFKI GmbH
volker.markl@tu-berlin.de/dfki.de

ABSTRACT

Data management systems will face several new challenges in supporting IoT applications during the coming years. These challenges arise from managing large numbers of heterogeneous IoT devices and require combining elastic cloud and fog resources in unified fog-cloud environments.

In this demonstration, we introduce a smart city simulation called IoTropolis and use it to create interactive eHealth and Smart Grid application scenarios. We use these scenarios to showcase three key challenges of unified fog-cloud environments. Furthermore, we demonstrate how our recently proposed data management system for the IoT NebulaStream addresses these challenges. Visitors to our demonstration can configure and interact with the scenarios to manage electricity usage in IoTropolis or to distribute patients across different hospitals. Thereby, visitors can actively engage with the challenges showcased by IoTropolis and utilize NebulaStream to address them. As a result, our demonstration enables visitors to experience data management for future IoT applications.

PVLDB Reference Format:
Aljoscha Lepping, Hoang Mi Pham, Laura Mons, Balint Rueb, Ankit Chaudhary, Philipp M. Grulich, Steffen Zeuch, and Volker Markl. Showcasing Data Management Challenges for Future IoT Applications with NebulaStream. PVLDB, 16(12): 3930 - 3933, 2023.
doi:10.14778/3611540.3611588

1 INTRODUCTION

Over the last decade, the Internet of Things (IoT) enabled new applications for many domains, such as health care [10], public transport [4], and the energy industry [8]. These applications collect and analyze data from physical devices. For instance, monitoring systems for intensive care units (ICU) collect data from dozens of sensors to prevent emergencies [10]. To support such IoT applications, several data management systems have been proposed, e.g., IoTDB [12], Frontier [9], and NebulaStream [14]. These systems utilize compute resources in unified fog-cloud environments, i.e., resources close to IoT devices in the fog layer [3] combined with elastic cloud resources. This enables in-network processing in the fog layer, reducing network load, decreasing data processing latency, and minimizing cloud costs. To perform in-network processing, these systems must support diverse data sources (C1), handle complex network topologies (C2), and trigger the actuation of physical environments in real-time (C3). Although current IoT data management systems address these challenges (C1-C3), showcasing them in real-world scenarios remains difficult.

In this demonstration, we introduce an interactive smart city simulation called IoTropolis. We use IoTropolis to realize IoT scenarios from the eHealth and Smart Grid domains, which showcase the following three key challenges of unified-fog environments.

C1. Heterogeneous and Disaggregated Data Sources: Fog-cloud environments consist of geo-distributed and heterogeneous data sources that utilize many different messaging protocols and data formats. To showcase this challenge, IoTropolis enables visitors of our demo to integrate heterogeneous data sources, such as wind turbines and EVs. In particular, these sources can generate data streams using different messaging protocols and data formats, i.e., Parquet, CSVs, or JSON. Furthermore, visitors can place sources at different locations in IoTropolis to simulate geo-distribution, e.g., the wind park or the hospital. Thus, visitors can combine data from wind turbines using MQTT with a Kafka stream from solar panels.

C2. Complex Hierarchies of Heterogeneous Processing Nodes: Fog-cloud environments can consist of complex topologies of heterogeneous processing nodes. These nodes range from densely-connected and powerful cloud servers to sparsely connected and resource-constrained devices in the fog. To this end, IoTropolis enables visitors to configure hierarchical topologies of heterogeneous computing nodes. Based on this topology, our application scenarios leverage in-network processing to preprocess data in the fog layer. For example, visitors to our demonstration can configure the processing capabilities of hospitals in the eHealth scenario. Thus, hospitals can process queries in-network locally without exposing data to the cloud.

C3. Real-Time Actuation of Physical Environments: Effective actuation of processes in physical environments often requires real-time analysis of large amounts of raw sensor data.To showcase this challenge, our application scenarios provide different actuation endpoints that allow data management systems to influence

Proceedings of the VLDB Endowment, Vol. 16, No. 12 ISSN 2150-8097.
doi:10.14778/3611540.3611588

the simulation in real-time. First, the energy scenario provides an actuation endpoint to distribute cars to charging stations based on locally available electricity. Second, the energy scenario provides an endpoint that enables fine-grained control over the number of active street lights at night. Third, the eHealth scenario provides an endpoint to distribute patients across hospitals based on available intensive care unit beds. The smaller the actuation latency, the more effective the actuation process becomes.

To show how IoT data management systems address these challenges, we integrate IoTropolis with NebulaStream [14]. NebulaStream is designed to operate in fog-cloud environments and addresses the above challenges as follows. First, NebulaStream supports many messaging protocols and data formats and handles geo-distributed data sources (C1) [4]. Second, NebulaStream generates hardware-tailored code to efficiently utilize powerful cloud machines and resource-constrained edge devices (C2) [5–7]. Finally, NebulaStream addresses (C2) and (C3) by utilizing network topology information for query operator placement, minimizing network latency and overall network traffic [1, 2]. In particular, NebulaStream utilizes geospatial information from data sources (C1) to process data locally and thereby more efficiently (C2 & C3).

In general, we envision IoTropolis to become a testbed that illustrates the capabilities of data management systems in fog-cloud environments. In summary, our contributions are as follows:

(1) We introduce IoTropolis, an interactive smart city simulation that showcases three key challenges of unified fog-cloud environments.
(2) We leverage IoTropolis to realize two real-world IoT application scenarios.
(3) We demonstrate how NebulaStream addresses the challenges showcased by IoTropolis.
(4) We provide a UI to enable visitors to explore IoTropolis, NebulaStream, and the application scenarios interactively.

The rest of this paper is structured as follows. In Section 2, we discuss the system overview of our demonstration. Section 3 highlights two real-world IoT scenarios with which visitors to our demonstration can interact. Finally, we present our conclusion in Section 4.

2 SYSTEM OVERVIEW

In this section, we discuss the implementation details of IoTropolis and NebulaStream. We will refer to Figure 1 throughout this section, as it provides an overview of our demonstration.

2.1 IoTropolis

IoTropolis provides an interactive smart city simulation to represent challenges of unified fog-cloud environments and consists of two components, e.g., the visualization and the data generation.

Visualization: IoTropolis represents the entire simulation visually via an interactive application running in a web browser ①. For example, IoTropolis visually represents electronic vehicle (EV) sources via cars that drive across the city, deplete battery power and regularly recharge at charging stations. Furthermore, the visualization component of IoTropolis provides two interfaces for user interaction. First, IoTropolis contains a visual control menu ② allowing visitors to modify parameters, such as wind speed, interactively. Second, IoTropolis provides several actuation endpoints

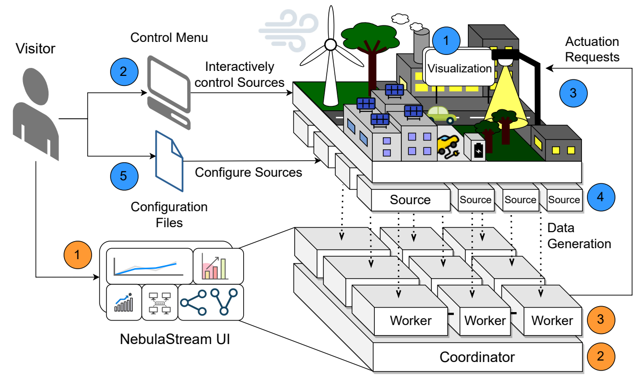

Figure 1: System Overview from the perspective of a visitor.

③ that visitors can employ by submitting queries to NebulaStream. Thereby, visitors can establish feedback loops between IoTropolis and NebulaStream, such as controlling the number of active street lights at night, based on available electricity. Furthermore, IoTropolis immediately reflects actions that impact data generation during runtime, i.e., actuation requests (C3) or control menu changes visually. Overall, the visualization component provides two ways of interaction and immediate visual feedback of generated data. The actual physical sources exist in a separate data generation layer.

Data Generation: IoTropolis simulates eight different sources ④ ranging from electronic vehicles (EVs) to hospitals that visitors can scale. Sources continuously generate and publish realistic data based on real-world properties. For example, EVs use their driven distance in combination with individual battery capacity and efficiency values to determine their current battery level. Furthermore, IoTropolis creates individual sources in Docker containers that visitors can configure freely ⑤. Visitors can configure the data format, messaging protocol, geo-distribution, and the scale factor of individual sources via configuration files (C1). Additionally, visitors can set up the network topology in which NebulaStream operates (C2).

2.2 NebulaStream

In the following, we highlight aspects of NebulaStream that are relevant to our demonstration. For a detailed description of NebulaStream, we refer the reader to the following papers [13, 14].

Visitors of our demonstration interact with NebulaStream via *NebulaStream UI* ①. The UI is a convenient web interface for NebulaStream that allows for submitting queries, managing sources, monitoring NebulaStream, and visualizing query results. The UI continuously communicates with the NebulaStream coordinator to apply user changes and update information.

The NebulaStream *coordinator* ② orchestrates query processing. It organizes all sources in the *source catalog*. Physical sources like wind turbines and solar panels can be combined into logical sources via the UI to manage large numbers of sensors (C1) conveniently. Furthermore, the coordinator organizes all computing nodes in the *topology catalog*, which benefits query optimization. The coordinator utilizes topology information to split incoming queries into local query plans that distribute the query workload across workers (C2).

NebulaStream *workers* ③ are heterogeneous and sparsely connected machines that form arbitrary hierarchies in unified fog-cloud

Figure 2: IoTropolis Overview. Red: Hospitals of the eHealth Scenario. Green: Producers in the Smart Grid Scenario. Purple: Consumers in the Smart Grid Scenario. Grey: Control Elements. All highlighted elements provide interaction with visitors.

environments (C2 & C3). Primarily, workers process local query plans and transfer the query results to the next downstream worker, thus enabling in-network processing. Additionally, workers locally manage the lifecycle of deployed query plans and inform the coordinator about any local failures.

3 IOT APPLICATION SCENARIOS

In this section, we discuss how visitors can interactively experience the challenges C1-C3 via a smart grid scenario (3.2) and an eHealth scenario (3.3). For each scenario, we first show how IoTropolis and NebulaStream realize the scenario and highlight how visitors can interact with it. In particular, we show how visitors can combine heterogeneous file formats and messaging protocols in single queries (C1). Second, we demonstrate how visitors can use the UI to monitor NebulaStream's in-network processing on fog-cloud processing node hierarchies (C2). Finally, we show how visitors can utilize IoTropolis' actuation endpoints and NebulaStream to create real-time actuation of IoTropolis' simulated environments (C3). This section will refer to Figure 2, which shows an overview of IoTropolis' visualization, and Figure 3, which shows the relevant elements of NebulaStream's UI.

3.1 Demo Setup

For our demonstration, we deploy IoTropolis on a laptop. IoTropolis runs in a web browser and uses Docker containers for the data generation layer. Furthermore, we deploy the NebulaStream coordinator on a different laptop, and NebulaStream workers no a cluster of six Raspberry Pis, which allows the simulation of a hierarchical fog-cloud topology. Each Raspberry Pi represents a fog node connected to the central NebulaStream coordinator node.

3.2 Smart Grid Scenario

IoTropolis simulates an increased need to use varying amounts of available electricity produced by renewable sources [11]. To that end, IoTropolis enables visitors to manage a smart grid by actively controlling electricity distribution and sources.

IoTropolis represents the smart grid scenario by enabling interactions between so-called electricity *producers*, e.g., wind turbines and solar panels, and *consumers*, e.g., households, factories, and street lights (see Figure 2). Visitors can configure these sources to use different file formats and messaging protocols. For example, producers could transmit JSON data via MQTT, and consumers CSV files via ZMQ. Visitors can group such physical sources by creating logical sources in NebulaStream UI (see Section 2). This enables visitors to conveniently query potentially large numbers of physical sources using logical sources. Additionally, queries can combine sources with different file formats and messaging protocols. Thereby, visitors can experience challenge C1.

When visitors submit queries via the UI, NebulaStream optimizes the queries and efficiently places operators across fog-cloud processing nodes by utilizing information from the topology catalog. Figure 3 shows the network topology for the smart grid scenario in the middle. Additionally, Figure 3 shows the operator placement of a query that calculates the difference between produced and consumed energy over a one-second window on the left. It shows that NebulaStream pushes down operators in the hierarchy to aggregate data as early as possible, thereby decreasing processing latency and network load. In this way, visitors can experience challenge C2.

The smart grid scenario provides two actuation endpoints that visitors can use in their queries. The first actuation endpoint lets visitors control the number of street lights at night. If electricity production is low because wind speeds are low, IoTropolis allows to turn off street lights via actuation queries to reduce the amount of consumed electricity. Additionally, the endpoint allows visitors to target groups of street lights in specific areas of IoTropolis, for example, to turn off street lights in the industrial area first. The second actuation endpoint allows visitors to distribute EVs across charging stations in the city. Charging stations receive electricity from their solar panels and nearby producers. Different charging stations can supply different amounts of electricity to recharge EVs over time. Visitors can use the information on the locally available electricity

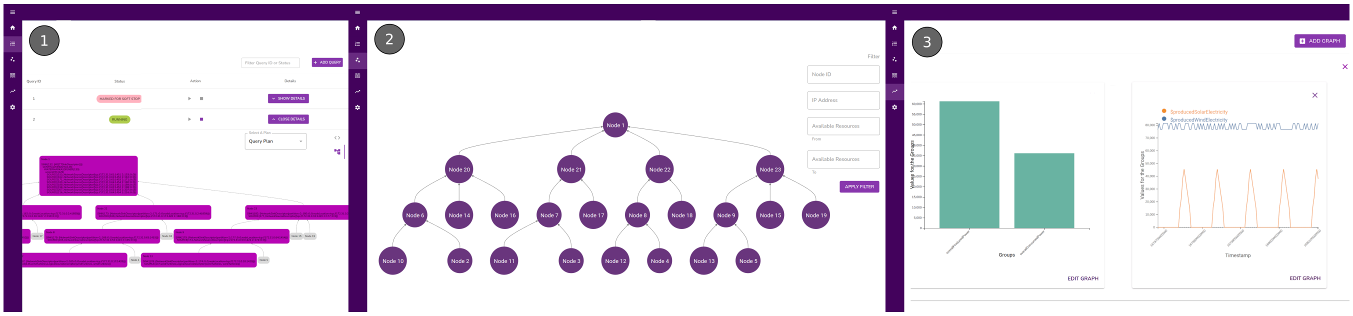

Figure 3: NES-UI: 1: Deployed Query Plan. 2: NebulaStream Topology. 3: Result Visualization.

at charging stations and the battery statuses of EVs to smartly distribute EVs across charging stations. By additionally utilizing the control menu, visitors can experience challenge C3.

3.3 eHealth Scenario

Real-time analysis of medical IoT data could be life-saving. In particular, monitoring patients' health statuses and predicting load in ICU beds can be crucial [10]. In this scenario, visitors manage the two hospitals in IoTropolis by using health status information to distribute patients across available ICU beds efficiently.

IoTropolis simulates two hospitals with limited ICU beds (see Figure 2). Additionally, IoTropolis simulates IoT sensors for patients that continuously enter the hospitals where they are treated and eventually leave again. Each hospital is a physical data source that visitors can configure concerning scale, data format, and messaging protocol. Thus, visitors can configure the hospital data sources differently, enabling queries that combine data from heterogeneous data sources. Thereby, the eHealth scenario represents C1.

Hospitals also function as NebulaStream Workers. Consequently, each hospital can (pre-)process its data. If visitors submit a query that calculates the number of available ICU beds across hospitals, NebulaStream pushes down aggregation operators to the individual hospital Workers. Hospitals only transmit the pre-aggregated data to the parent fog node afterward. In this way, the eHealth scenario enables visitors to experience challenge C2.

The eHealth scenario provides one actuation endpoint. This actuation endpoint enables visitors to control the distribution of new patients across the two hospitals. Visitors can submit queries that monitor the number of available ICU beds across hospitals and potentially predict future changes. Based on the results of the monitoring query, visitors can use the actuation endpoint of the eHealth scenario to smartly distribute new patients across the hospitals. The in-network data processing of NebulaStream enables actuation queries with very low latency. Thereby, visitors can experience C3.

Finally, visitors may also use the control menu to change the capacity of ICU beds for the individual hospitals and to adapt the overall patient load. This allows visitors to significantly manipulate the data generation and the assumptions used for patient distribution. On the one hand, IoTropolis directly reflects the effects of parameter changes via the control menu. On the other hand, visitors can also submit and visualize queries via the UI. The bar chart on the right side of Figure 3 shows an example of the difference between overall produced- and consumed electricity in IoTropolis.

Overall, the scenarios provide many different interaction points for visitors to experience the challenges C1-C3. Additionally, they enable visitors to experience potential future IoT applications in the energy industry and healthcare.

4 CONCLUSION

This paper introduces a smart city simulation called IoTropolis that we use to realize IoT scenarios from the eHealth and Smart Grid domains. Furthermore, we use these scenarios to showcase three key challenges of unified fog-cloud environments and show how our recently proposed IoT data management system NebulaStream addresses them. We show that visitors to our demonstration can submit queries that target heterogeneous data sources, are processed in-network in fog-cloud processing node hierarchies, and enable low-latency actuation. We plan to extend IoTropolis towards benchmarking capabilities and additional scenarios in the future.

ACKNOWLEDGMENTS

This work was funded by the European Union as ELEGANT (957286), the BMBF as BIFOLD (01IS18025A and 01IS18037A), moreEVS (410830482), and the Software Campus (01IS17052).

REFERENCES

[1] Chaudhary, A., et al. Governor: Operator placement for a unified fog-cloud environment. In *EDBT* (2020).

[2] Chaudhary, A., et al. Incremental stream query merging. In *EDBT* (2023).

[3] Dastjerdi, A. V., et al. Fog computing: Principles, architectures, and applications. In *Internet of things*. Elsevier, 2016.

[4] Gavriilidis, H., et al. Scaling a public transport monitoring system to internet of things infrastructures. In *EDBT* (2020).

[5] Grulich, P. M., et al. Grizzly: Efficient stream processing through adaptive query compilation. In *SIGMOD* (2020).

[6] Grulich, P. M., et al. Towards unifying query interpretation and compilation. *CIDR* (2023).

[7] Grulich, P. M., Zeuch, S., and Markl, V. Babelfish: Efficient execution of polyglot queries. *PVLDB* (2021).

[8] Lau, B. P. L., et al. A survey of data fusion in smart city applications. *Information Fusion* (2019).

[9] O'Keeffe, D., et al. Frontier: Resilient edge processing for the internet of things. *PVLDB* (2018).

[10] Pathinarupothi, R. K., et al. Rewoc: Remote early warning of out-of-icu crashes in covid care areas using iot device. In *2021 IEEE BIBM* (2021).

[11] Tuballa, M. L., et al. A review of the development of smart grid technologies. *Renewable and Sustainable Energy Reviews* (2016).

[12] Wang, C., et al. Apache iotdb: time-series database for internet of things. *PVLDB* (2020).

[13] Zeuch, S., et al. Nebulastream: Complex analytics beyond the cloud. *OJIOT* (2020).

[14] Zeuch, S., et al. The nebulastream platform: Data and application management for the internet of things. In *CIDR* (2020).