VR-DataOps: Towards an Immersive Visualization of DataOps in Virtual Reality

Roy Oberhauser[0000-0002-7606-8226]

Computer Science Dept., Aalen University, Aalen, Germany
roy.oberhauser@hs-aalen.de

Abstract. Data is crucial in the information age and effectively leveraging it essential for businesses and organizations. DataOps supports data operations automation through practices, processes, and technologies, yet faces a number of challenges. These include a large, heterogenous data landscape and integration; cross-silo collaboration impediments for teams and stakeholders; and accessible visualization, live monitoring, and analytic insights. Towards addressing these challenges, this paper contributes VR-DataOps, a solution concept for immersively visualizing DataOps in Virtual Reality (VR). It provides holistic insights into data reality, including data pipeline transformations, pipeline evolution, and integrates heterogeneous data tools within a single immersive environment. A prototype demonstrates its feasibility, while a scenario-based empirical evaluation assesses its capabilities and suitability.

Keywords: DataOps, Data Analytics, Virtual Reality, Visualization.

1 Introduction

"Data is the new oil," as M. Palmer expounded, but like crude oil, data must be refined to provide value [1]. DataOps is a set of practices, processes, and technologies that provide an integrated and process-oriented perspective on data automation and methods, enabling continuous improvement in data quality, speed, and collaboration [2]. A simpler working definition is the application of DevOps principles to data engineering and management [3]. The increasing importance of DataOps is evident in the DataOps adoption rates and positive impacts, confirmed by cross-sector survey [4] (summarized in [5]) where 91% had (or were defining) a formal DataOps strategy and 80% reported over 100 data sources (indicating large data landscapes). Insights into data and the associated automation operations/processes have thus become increasingly relevant for businesses and stakeholders [6].

Yet DataOps also faces various Challenges (C) this paper seeks to address, consolidated from the State of DataOps survey [6] and the DataOps literature review by Chung and Molléri [7]. This includes: C1) *large heterogenous data landscape and integration*. For instance, 42% of survey respondents require over 25 vendors to execute their DataOps strategy [6], making unified access and integration a challenge for stakeholders. C2) *data collaboration impediments to teams and stakeholders.* As [7] found, Data

Silos makes it challenging to collaborate and provide analytics; Team Structure requires combining varied expertise across different departments with varying availability and proficiency in data analytics; and achieving a data-driven Organizational Culture is challenging due to communication and collaboration barriers across teams (business, ITOps, software development, and data consumers). C3) *data visualization, live monitoring, and analytic insights.* Among the data management activities consuming the most time and effort were managing data quality, data monitoring/auditing, and data visualization/reporting [6].

This paper proposes and investigates an immersive solution approach using Virtual Reality (VR). Our prior work with VR includes VR-EA+TCK [8] that supports Enterprise Architecture (EA) models, integrating enterprise repositories, Atlas EA tool, IT blueprints, and knowledge/content management systems; VR-EvoEA+BP [9] animates enterprise evolution and Business Processes (BPs); and VR-DevOps [10] immersively visualizes Continuous Delivery (CD) pipeline models for software developers. Towards addressing the three aforementioned challenges, this paper contributes VR-DataOps, a solution concept for immersively visualizing DataOps in Virtual Reality (VR). It offers holistic insights into the data reality, pipeline transformations, pipeline evolution, and heterogeneous data tools in an integrated immersive environment. Our realization demonstrates its feasibility, while an empirical scenario-based case study explores its potential and suitability.

The remainder of the paper is structured as follows: Section 2 discusses related work; Section 3 presents our solution concept; Section 4 details our realization; our evaluation is described in Section 5, followed by a conclusion.

2 Related Work

The paper's contribution is not directly related to work on data tools, data analytics, or data visualization, as it integrates existing data tools and their capabilities into VR. Rather, its visualization is DataOps-centric and focuses on data pipeline processing. The DataOps literature review by Chung and Molléri [7] does not mention VR, however, certain challenges and implications align with those ours is addressing, including cross-silo collaboration, data quality insights, interactive data visualizations, and integration across data systems. The systematic review of DataOps by Fannouch et al. [5] identified DataOps challenges, including multi-platform integration, fragmented data silos, collaboration, and ongoing monitoring of data pipeline quality/integrity; while analytics and visualization are mentioned, VR is not. Krishnaswamy's book on DataOps [11] mentions visualization in relation to data dashboards and real-time data insights, without considering VR. No other directly related work was found searching for {DataOps} with any terms {immersive, virtual reality, extended reality, VR, XR}.

Related to VR visualization, Millais et al. [12] compared 2D data visualizations vs. 3D VR-native data exploration tools regarding workload and insights (DataOps unmentioned). Task workload was equivalent, but VR users gained more accurate and insightful results. Sudár, A. and Csapó [13] compared dashboard-like Web-based 2D text, image, and video content in a traditional 2D vs. a VR desktop. They found no performance differences overall, but VR led to higher satisfaction, increased participant engagement and collaboration, fewer errors, likely due to greater immersion and presence. For detailed, in-depth insights, however, 2D displays were easier, with VR involving more physical effort. Yang et al. investigated immersive collaborative sensemaking [14], and while our paper does not evaluate the collaborative component, we believe these results can be transmuted to our concept. As DevOps serves as a model for DataOps, our own prior work in this area VR-DevOps [10] immersively visualizes a DevOps pipeline.

Consequently, immersive DataOps visualization appears underexplored; although the C1–C3 challenges are frequently articulated, we argue that our contribution is uniquely positioned towards immersively addressing these.

3 Solution Concept

Towards addressing the aforementioned DataOps challenges, our proposed VR-based solution concept provides insights into the feasibility and benefits an immersive VR environment could offer.

3.1 Grounding of our Solution Concept in VR-Related Research

Incorporating VR into our solution concept is grounded in prior research in areas related to modeling, analysis, and collaboration, some of which we highlight here. Akpan & Shanker's systematic meta-analysis [15] demonstrated significant advantages of VR and 3D in discrete event modeling, encompassing model development, analysis, and Verification and Validation (V&V) (which could carry over to DataOps data quality and analysis). Out of 23 articles analyzing analysis in 3D, 95% found it more effective than 2D, such as evaluating model behavior or conducting what-if analyses. Furthermore, the consensus was that 3D/VR can present results for decision-makers convincingly and understandably. Additionally, 74% of 19 papers concluded that 3D/VR significantly improved model development tasks, supporting teams and enhancing precision and clarity. Narasimha et al. [16] conducted a card sorting collaboration experiment to explore VR's suitability for analytical tasks within an information architecture. VR was at least as effective as in-person card sorting, and for certain variables, even superior to both conventional and video-based conditions. During collaborative interaction, qualitative data revealed that participants were aware of their task, others, and their context, mirroring an in-person setting. Also, the qualitative data revealed positive perceptions of VR. These findings suggest that VR can create a sense of presence and collaboration comparable to an in-person environment. 177 papers were analyzed in the survey on Immersive Analytics (IA) by Fonnet & Prie [17], finding concurring evidence that IA offers benefits over non-IA for graph and spatial data analysis, especially when scene complexity exceeds 2D display limitations. However, advantages for multidimensional data were more task-dependent. They note that while IA can facilitate the exploration of vast data worlds, context-aware navigation techniques were insufficiently exploited. Based on these findings, we conclude VR's immersive, visual, contextual, and collaboration properties offer the potential to also visually and comprehensively depict large and diverse DataOps landscapes and models in 3D while supporting modeling, analysis, decision support, and stakeholder collaboration, and hence this approach should be investigated.

3.2 VR-DataOps Solution Concept

On this basis, our VR-DataOps solution concept utilizes VR towards addressing the aforementioned challenges. To address C1, our Data Hub concept (i.e., hexagonal architecture pattern) simplifies and enables integration of data platforms/applications (Airflow, Kafka, etc.), including APIs and data models via specific adapters that transform differences into our common internal format. Towards addressing C3, clearly structured and interactive UI elements are offered in a unified, accessible, and immersive data landscape that supports exploration (e.g., zoom, UI element clarity, tooltips). Support for complex analyses in VR is addressed via hybrid 2D/3D visualizations (as 2D and 3D views have differing visual advantages). The visual capabilities addressing C3 also support C2 collaboration aspects, providing immersively accessible DataOps information for non-experts and cross-silo collaboration support. Further solution capabilities towards addressing C2 and C3 include:

Data pipeline visualization and evolution in VR. Classic UIs depict data pipelines differently (nested tables, text files, DAG visualizations), hindering comprehension for newer stakeholders; VR should create a holistic visualization to improve comprehension and utilize visual space to indicate its evolution.

Configurable pipeline/table depictions. Stakeholder types vary by interest and visualization needs: 3D supports spatial orientation, immersion, and collaborative interaction, while 2D enables more precise analysis of individual points.

Traceability support. Data structures or content can change or be adjusted; thus, each run's changes are stored/displayed. A 2D pipeline view can highlight changes and evolution, and data pipeline I/O are depicted as tables to readily capture data changes; individual processing steps are traced in time and logic to support understanding.

Differential I/O support. Data transformation differences between input and output are highlighted.

Live integrated pipeline monitoring. A live, unified, integrated visualization of heterogeneous infrastructure supports transparency and comprehension. Data changes can occur opaquely between processing steps, so a visual change representation (e.g., via animated data objects or comparison views) supports pipeline logic understanding.

Highlight errors and issues. Error messages are often text-centric, distributed, or difficult to interpret. Error status is visualized (e.g., via red nodes or icons) to support data quality and analysis, while error logs are directly accessible to isolate causes.

VR tablet and keyboard for interaction. A tablet menu in VR provides a familiar and accessible unified interface, lowering the barrier to entry and improving navigation. A virtual keyboard supports searching and commands.

Integrated web browser in VR tablet. Users can access remote tools or documentation via a web browser within VR, mitigating desktop switching for information.

Desktop streaming in VR. Ensures native data applications with their conventional (non-web) interfaces and capabilities are fully accessible in a unified way within VR.

Multiple simultaneous browser or desktop streaming windows. Leverage VR's virtually infinite space to offer multiple 2D windows simultaneously.

In combining these various capabilities within an immersive experience, VR-DataOps contributes a unique, single solution concept towards addressing the C1-C3 challenges.

4 Realization

While the solution concept focuses on the C1-C3 challenges, besides typical technical implementation issues such as integration, our realization faced additional VR-related challenges. Visualization challenges included: precision of interaction via freely held 3D controller (versus steady mouse), minimizing cognitive load via attention design principles (VR has more degrees of freedom, larger visual expanse, moving, grasping, etc. and can overwhelm), ergonomics, and fatigue (extended arm use, neck, and eye strain in VR). Technical challenges included: performance optimization (VR apps necessitate a minimum frame rate to mitigate VR sickness, so visual complexity must be managed), dynamic visual synchronization and consistency with live data.

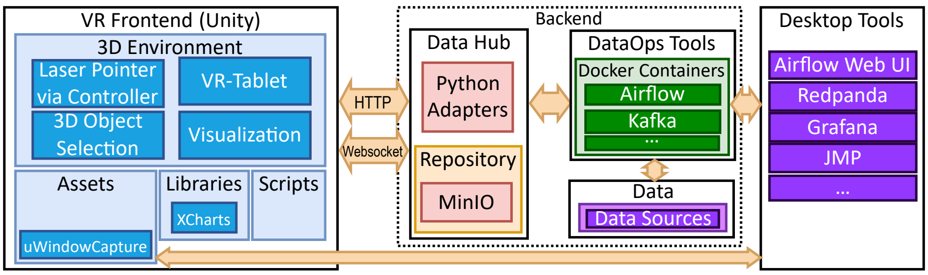

Fig. 1. Logical architecture of VR-DataOps.

The logical architecture of our VR-DataOps solution concept is shown in Fig. 1. Components used in the implementation include: Unity XR Toolkit with SteamVR and Steam Link, uWindowCapture, and XCharts, a Unity data visualization library to visualize data flowing through a Kafka topic. Graphviz is used to depict 2D diagrams of pipeline executions in VR, due to its support for multiple layout types and automatic layout placement. Communication between the Unity app and the Backend is via FastAPI using asynchronous WebSocket and REST connections. The Data Hub and its adapters are implemented in Python. It and each DataOps application (Apache Airflow, Apache Kafka, etc.) runs in a Docker container, supporting portability, dependency management, isolation, reproducibility, scalability, and rapid deployment. MinIO is used for storage inside the Docker containers to snapshot the files to support queries, history, and traceability in our internal JSON format. Each pipeline is assigned its own bucket.

Our Unity VR app includes a main menu to access functionality (Fig. 2), divided into: *Browser*, *Pipeline*, *Tools*, and *Config* sections, positioned directly in front of the user, and (de)activated via the X button on the left controller. *Browser* (see Fig. 2 left) enables interaction with a web browser from within VR (e.g., Airflow Web UI) via controllers and virtual keyboards. Creation of additional browser windows (via the *New Browser Window* button) supports parallel use and free windows placement. *Pipelines* enables 2D/3D switching via the layout button on top (Fig. 2 right); all pipelines are listed in a scroll window and all those selected (green) are depicted.

Fig. 2. VR Browser view (left) showing Airflow Web UI and keyboard; Pipeline view (right).

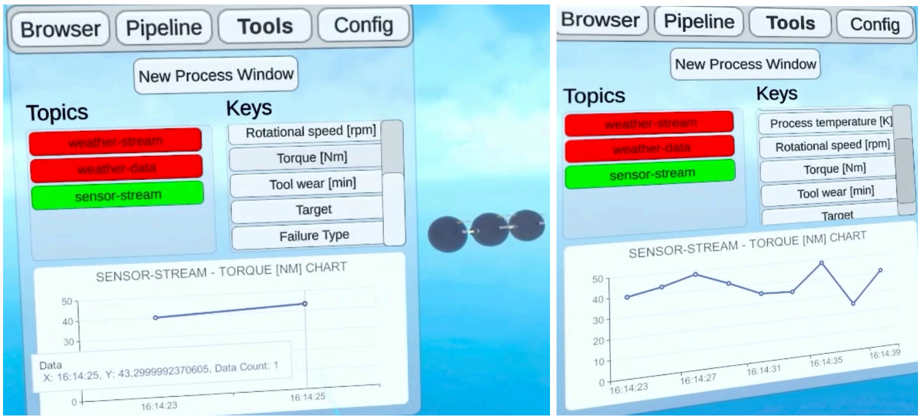

Fig. 3. VR Tools view (left) showing live pipeline data (right); spheres represent Kafka topics.

The *Tools* view (Fig. 3 left) also offers a *New Process Window* button for new windows that broadcast a desktop program live from the operating system (shown later in Fig. 11), offering interaction options via controllers and virtual keyboard. The left scroll window lists all Kafka *Topics*, any number of which can be selected (green) to display a graph (Fig. 3 left bottom). A Kafka representation consists of Topic, Consumer Groups, and Consumer as nodes (3D spheres in Fig. 3 left). When selecting a topic, all JSON keys that are sent to the topic as data are displayed in a right scroll window. When a JSON key is selected (Torque in Fig. 3), values are queried in real time and the data displayed as a line chart to support live data monitoring (Fig. 3 right). The *Config* view (not shown) offers settings, e.g., a slider limiting table text length, plane separation, cell size, transparency, reducing tables displayed to those within proximity of the user (to enhance performance and reduce visual clutter).

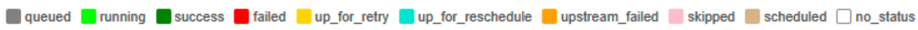

Fig. 4. Status coloring legend for pipeline execution.

A pipeline execution depicts a node's name, duration, and status colored according to the legend in Fig. 4. For 3D pipelines depictions, nodes as 3D cubes represent the executed functions (see Fig. 5). 3D visualizations offer both a 3D tree directed upwards (Fig. 5 left) or 2D plane-aligned (Fig. 5 right) directed left-to-right. 2D pipelines can show their input data table (left) and output data table (right), as seen in Fig. 6 (left). Graphviz layout options are offered for 2D pipeline images. Filtering options include modified data only and timestamp ranges, as shown in Fig. 6 (right).

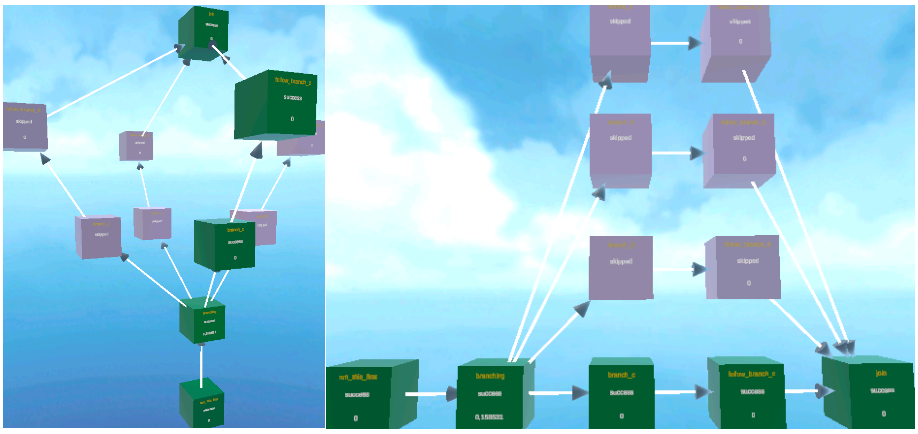

Fig. 5. VR-DataOps 3D pipeline visualization: 3D tree (left) and aligned to a 2D plane (right).

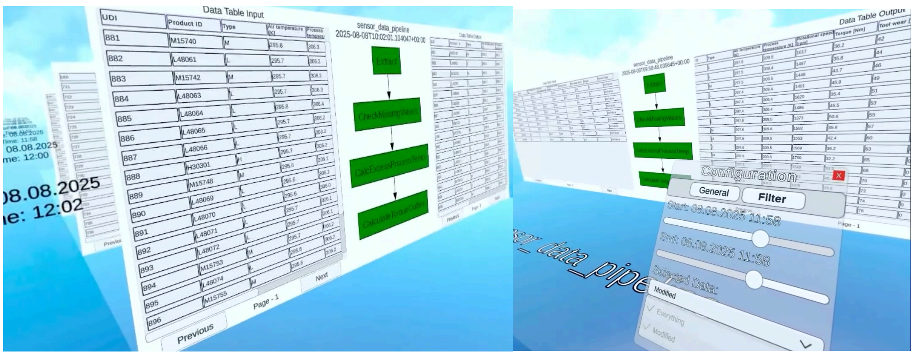

Fig. 6. On left, 2D pipeline run (center) with input data table (left) and output data table (right); on right, Filter selected Modified (so table on left shorter since only modified data displayed).

Fig. 7. Visual pipeline execution, status, and evolution.

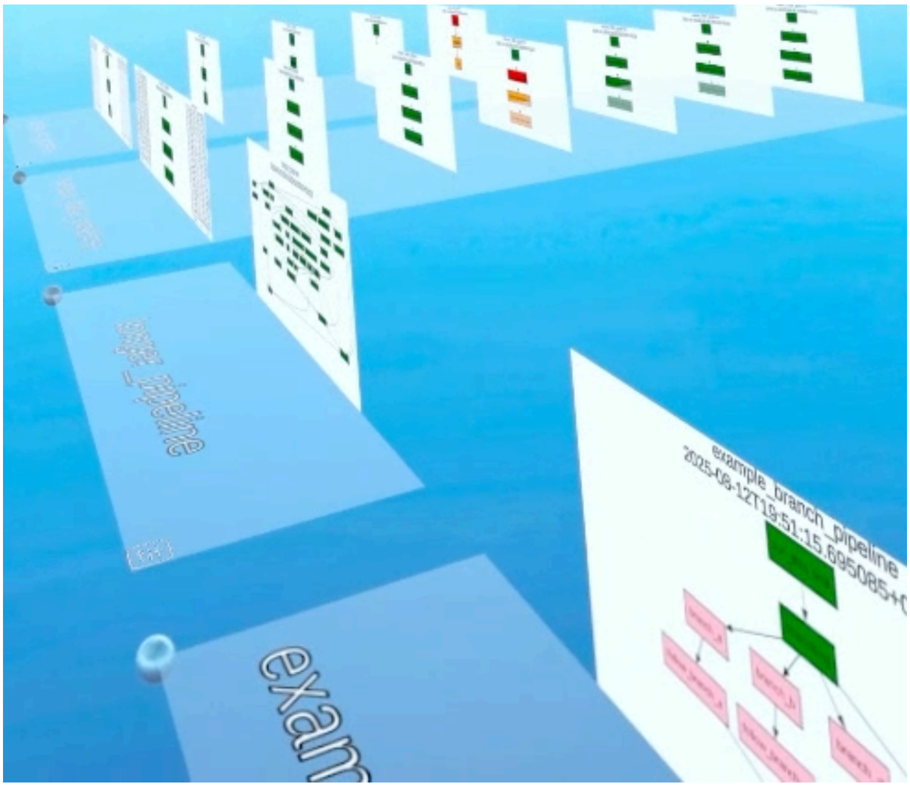

Fig. 8. Multiple pipeline models.

As shown in Fig. 7, a single run is depicted as a 2D vertical plane with a central graph image of the pipeline centered with a title (name and start date of the pipeline), and optionally input data (left) and output data (right) displayed in tabular form; hovering offers data type as tool tip. A horizontal plane, having a corner sphere move/collapse affordance, is used to depict a pipeline's run history, status, and evolution. Multiple pipeline models can be displayed and compared simultaneously (Fig. 8).

Furthermore, the *pipeline evolution* is readily accessible via highlighting, whereby a red outline indicates that that node is removed in the following run (Fig. 9 left) and green indicates it is an addition from the previous run (Fig. 9 right). The outline highlight coloring applies to the data tables as well. The data input and output tables can not only be filtered (described previously), but also sorted as shown in Fig. 10; tables offer scrolling and next/previous page buttons to address larger data tables.

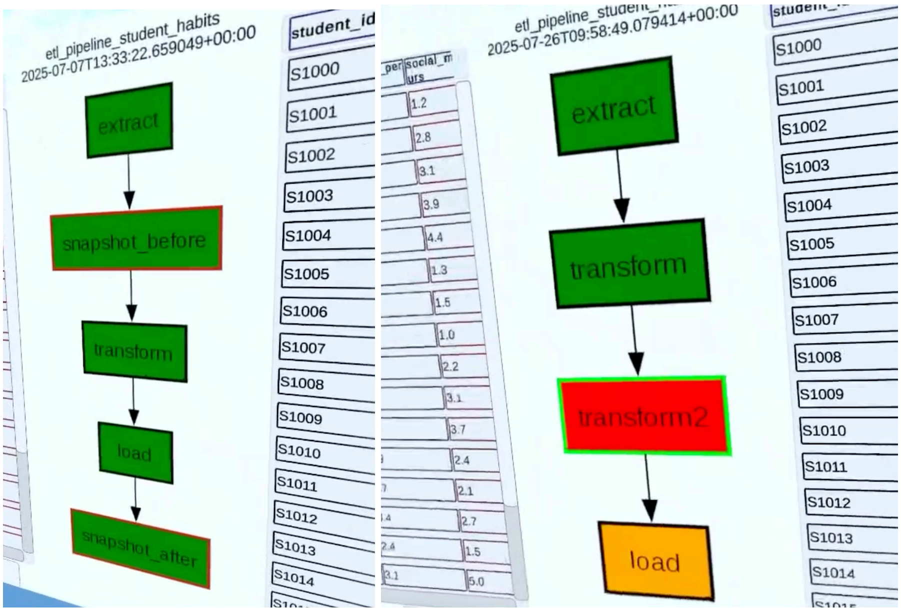

Fig. 9. Outlines (red=removed on next, green=added to previous) highlight pipeline changes.

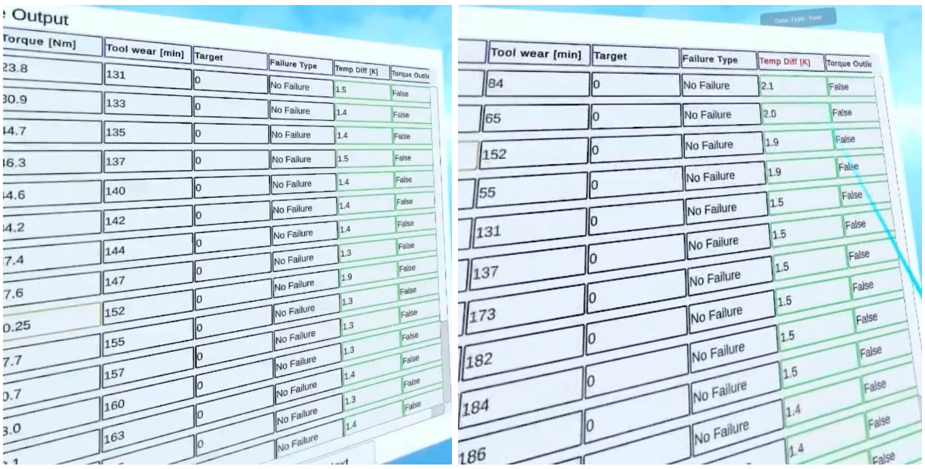

Fig. 10. Sort table by column (left=before, right=after) toggles ascending/descending.

Fig. 11. Live desktop process streaming access in VR with multiple parallel windows.

The streaming desktop windows offered in VR transmit a live display of any running desktop process/program from the operating system (Fig. 11). The title indicates the current process shown. Interaction in VR is done via the controllers and a virtual keyboard. Several windows can be created in parallel or can switch between them (without having to navigate) via the side arrow buttons. These windows, like the browser windows, are freely movable.

5 Evaluation

In evaluating our solution concept, we refer to the design science method and principles [18], in particular a viable artifact, problem relevance, and design evaluation (utility, quality, efficacy). Using our prototype realization, our scenario-based case study focuses on visualization, contextualization, analysis/configuration support, and tool integration. However, evaluating VR vs 2D desktop tasks empirically presents challenges regarding: *task equivalence* (even small differences can distort results); *learning and novelty effect* (VR usage may be infrequent versus extreme exposure to desktop environments); *controlling disruptive factors* (VR users may stand or sit, VR has headset latency while a desktop doesn't); *insight evaluation* (subjective evaluation if certain insights found are sufficiently correct or relevant); *evaluation*: different (subjective) metrics can lead to different overall assessments. Due to the elaborate VR setup and individual support during testing, a qualitative approach was chosen with the aim to obtain concrete feedback on the usability, data comprehensibility in the virtual environment, and the effectiveness of the visualizations.

Initially, a scenario in a single VR context involving individual tasks that maximize functional coverage was conceived. Thereafter, desktop applications with comparable functionality were considered. Functions of the VR application that are not directly or difficult to compare are highlighted in the evaluation. Subsequently, survey questions were created, focusing on qualitative evaluation of the individual functions, the comparison of VR to a desktop context, and the overall impression of VR-DataOps.

Sample: Due to resource/time constraints, our convenience sample consisted of ten subjects aged 24 to 25 (computer science students and software developers), two use VR regularly and two had tried VR, and seven had used DataOps apps (professionally or during their studies). The hardware consisted of**:** a Windows PC i7-10700KF CPU@3.8GHz, NVIDIA RTX 3070, 32GB RAM, 2TB SSD, 5GHz WiFi, using Steam Link to stream the VR app to the MetaQuest 3 VR glasses (Snapdragon XR2 Gen 2, 512GB RAM, 2064×2208 resolution per eye, and max 120Hz frame rate).

Scenario: To ensure VR app functionality utilization, our fictive sensor data pipeline for production quality control monitors sensor data (e.g. temperature, torque, etc.) from several simulated production machines. The data as JSON is collected via Kafka, passing through an Airflow pipeline for validation and transformation, and is monitored and analyzed live in the VR app. In Kafka, a topic, producer (sensor data of a machine) and consumers for the sensor data (e.g., Airflow) were created. The Kaggle dataset [19] consists of 10K data points and 14 characteristics as columns. Sections of this data set are used and sent to Kafka. Incorrect or missing values are integrated (e.g., unrealistic jumps). The prepared Airflow pipeline is started, processing any currently existing Kafka data. Fig. 12 shows the simple pipeline used in the scenario. In parallel, certain tasks use another, more complex pipeline (see Fig. 13 for 2D layout in VR and Fig. 14 for 3D in VR) to better evaluate the advantage of the spatial depiction capability in VR. The pipeline creates a snapshot of the run and the input/output data (raw data to cleaned values). The comparable desktop apps used in this scenario are shown in Table 1.

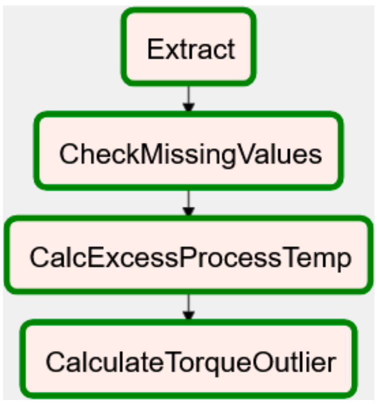

Fig. 12. Test scenario simple pipeline.

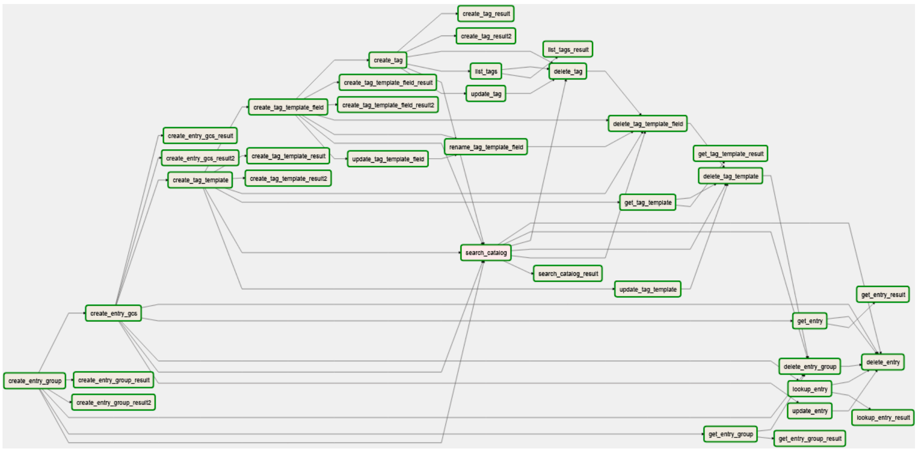

Fig. 13. Test scenario complex pipeline.

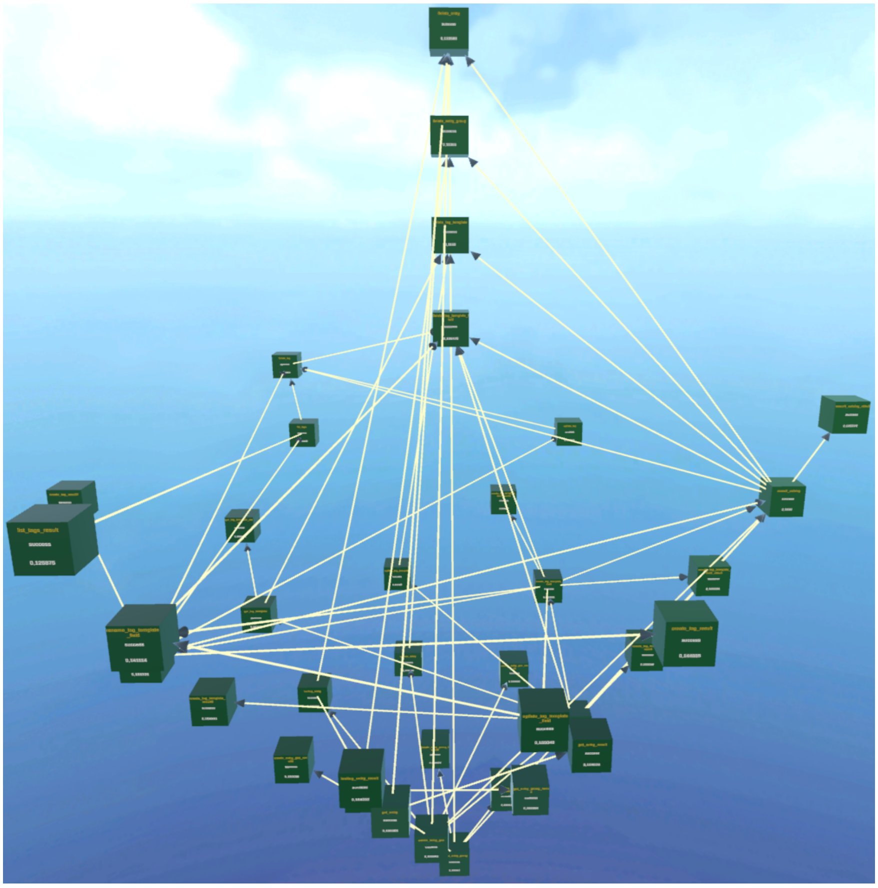

Fig. 14. Test scenario 3D rendering in VR of the complex pipeline.

Procedure: Test subjects were initially familiarized with the general control in VR, but not instructed about the functions of the tools or the VR app in order to investigate their intuitive usability. However, a brief time was allocated before measuring to permit test subjects to familiarize themselves with the app for general orientation. Afterwards, the scenario tasks were performed sequentially (following goal and task number) initially for the Desktop (D) (Table 2) and thereafter the VR (V) tasks (Table 3). The tasks in the table are enumerated by medium {D,V}, goal type {d,c,a,k,t}, and then task sequence number. For practical reasons, task ordering was media-specific based on media context, with the VR table indicating its equivalent Desktop task in parentheses. While not all VR tasks had an equivalent desktop task, they provided additional metrics and may influence the subjective evaluation.

Table 1. Data tools with comparable or similar functions.

VR function	Desktop Tools
2D graph representation of pipelines	Airflow Web UI, GraphViz
Monitoring and error reporting	Airflow Logs
Table interaction such as sorting, filtering, etc.	JMP
Analyze data	Grafana, Jupyter Notebook
Compare data (input/output)	JMP, Pandas DataFrame, Excel
Live charts for data	Grafana (with InfluxDB and Telegraf)
Kafka Topics and Data	Redpanda Console
Open and use the VR browser	Standard browser (e.g. Chrome or Firefox)

Table 2. Desktop (D) user task sequence.

Goal 1: Find and display (d) pipelines via Airflow Web UI
Dd1: Open Airflow Web UI in the browser; locate the AirflowPipeline *sensor_data_pipeline* and view the pipeline as a 2D graph *longer_pipeline*
Dd2: Choose a different graph layout for the 2D graphs
Goal 2: Analyze (a) pipeline flows via JMP (Statistical Discovery)
Da1: Via JMP open input/output data as CSV; adjust cell size and/or table width as needed
Da2: Sort the data according to the temperature difference in descending order
Da3: Display the existing data type of the different columns
Da4: Compare both input and output data tables using the appropriate function (compare tables); determine input/output differences
Goal 3: Kafka (k) use with Redpanda Console, Grafana
Dk1: In the browser, open Redpanda Console; determine the consumer groups and consumers associated with the topic *sensor-stream*
Dk2: Open Grafana in browser; view torque as a live graph; determine fluctuations / trends
Goal 4: Tools (t) combinations (Airflow Web UI, Editor)
Dt1: Edit a value of the pipeline *sensor_data_pipeline* minimally with an editor
Dt2: In the browser, open the Airflow Web UI and start a new run; view the changed data

Table 3. VR (V) user task sequence (referencing comparable Desktop (D) tasks)

Goal 1: Find and display (d) pipelines; gain familiarity with main menu and pipeline display
Vd1(Dd1): Select the Airflow pipeline *sensor_data_pipeline* and plot it as a 2D graph Vd2: Select the *longer_pipeline* for a parallel display; remove the pipeline display
Goal 2: Configure (c) pipeline representations
Vc1: In 2D timeline of *sensor_data_pipeline* runs, use menu to view only last week's runs Vc2(Da1): Adjust the visual distance between tables, cell size, and/or table width as needed
Goal 3: Analyze (a) pipeline flows; differentiate runs and compare input and output data
Va1(Da2): In 2D *sensor_data_pipeline*, sort by temperature difference in descending order Va2(Da3): Determine the data type of any column Va3(Da4): Filter to only changed data; determine input/output differences; determine how data of each run changed over the course of the week Va4(Dd2): Choose another graph layout for the 2D graphs Va5: Switch from 2D to 3D representation and select the *longer_pipeline* Va6: Choose another graph layout for the 3D graphs
Goal 4: Kafka (k): familiarization with Kafka integration and visualization of live data
Vk1(Dk2): Via the Tools menu, select the topic *sensor-stream*; view the torque as live data on the line graph; determine fluctuations or trends Vk2(Dk1): Display topic *sensor-stream* as a 3D graph; determine insights (Kafka's depiction consists of Topic, Consumer Groups and Consumer as node)
Goal 5: Tool (t) usage: use of DataOps tools such as browser and desktop programs in VR
Vt1(Dt2): Create window for live broadcast of computer processes; switch to the Chrome browser and open the Airflow Web UI; navigate through the pages with controller and virtual keyboard and start a new run. Vt2(Dt1): Switch to the Notepad window and slightly change a value of the Pipeline function.

5.1 Results

Efficiency as task duration for comparable tasks in the other medium is depicted in Fig. 15, with VR 13% slower (371 vs. 329 seconds) overall. With additional VR training and app familiarity, this difference could potentially be reduced. In VR, primarily the tool integration usage tasks (Vt) were significantly slower (3.5 and 2.8 times), due to slower VR virtual keyboard use and controller precision to use desktop apps in VR vs. the familiar physical keyboard and mouse (for which such desktop interfaces were designed). Since the Vt tasks were not designed to be VR-native (desktop app use within VR), if these tasks are excluded, VR was 19% faster (241 vs. 287 seconds). This implies that if the interfaces for these desktop tools also offered true VR-native (rather than mouse and keyboard) interfaces, then VR might be equivalently efficient.

As to interventions, considering only comparable desktop and VR tasks (excluding VR-only tasks), total interventions (assistance in the form of guidance) were 6 for desktop and 7 for VR; 19 total errors were exhibited for desktop and 13 for VR. The assistance and errors for the desktop were related to a complex data analysis task with JMP, whereas in VR it related to users being confused regarding a dropdown interface offered for 2D and 3D graph layouts. Note that for VR, this was an interface issue related to our VR prototype implementation, which can be readily remediated in the future.

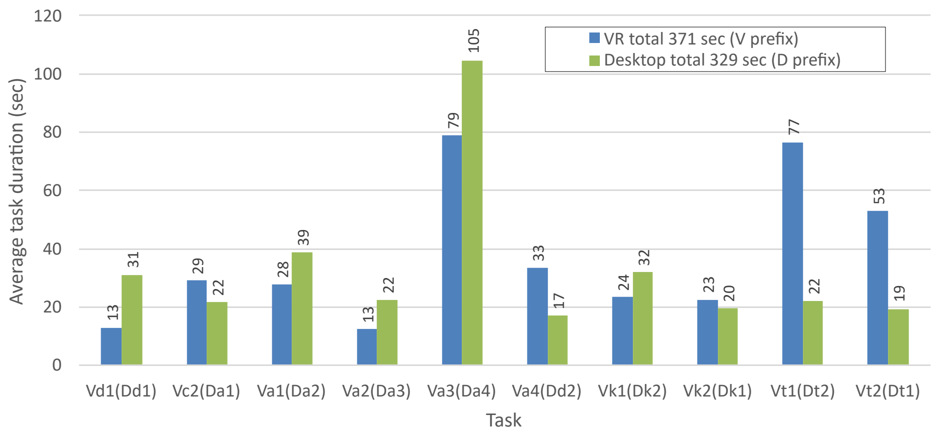

Fig. 15. Average user task duration for VR (V) vs. (equivalent Desktop (D)) tasks in seconds.

Based on the subjective survey, VR-DataOps functional features were in general rated positively (with neutral the lowest), as shown in Fig. 16. The table filtering and sorting, pipeline configuration, and Kafka live-tracking were the top-rated functions. The lowest rated were input/output data coloring, visualizing the 2D pipeline, and using streamed desktop apps in VR, yet the averages were still neutral or higher. Our explanation is that these may be more basic and not provide significant added value or a wow-factor.

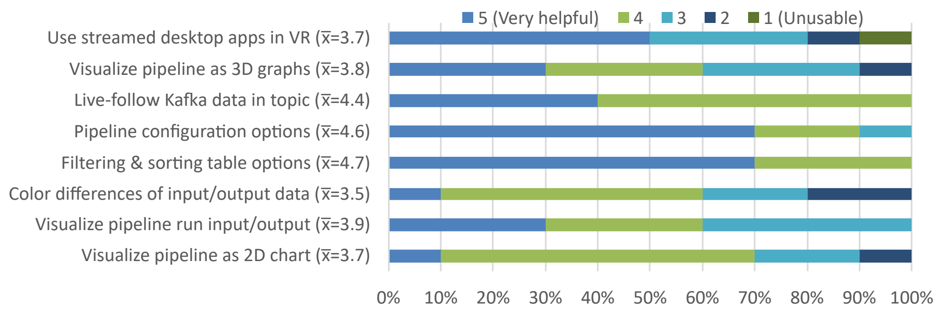

Fig. 16. VR-DataOps Functionality.

The survey results regarding VR usability and User eXperience (UX), rating enjoyment, comfort, efficiency, accessibility, clarity, and intuitiveness, is shown in Fig. 17. The VR app was well received, with the lowest average 3.9 and still above neutral. VR's strongest points appear to be emotional (enjoyment) and visual, with clarity, comfort, and accessibility receiving no disagreement. Efficiency received the lowest rating at 3.9, which is likely due to perception and the lower familiarity with VR.

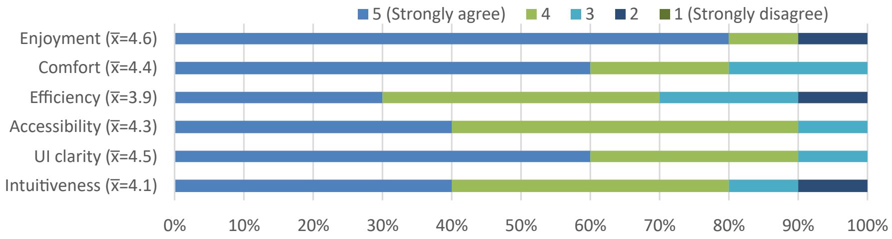

Fig. 17. VR-DataOps Usability and UX.

The survey results in Fig. 18 directly compares VR to the desktop experience, with VR scoring much better for enjoyment (4.5 average), and the spatial overview of the pipeline structure (3.9 average). Support for analysis was was 3.4 on average, yet 90% thought it better or were neutral. VR faired less well for operational complex tasks (3.1) compared to the more familiar desktop tools, perhaps due to immersive navigation and for text- and menu-heavy tasks, where mouse and keyboard remain much more familiar.

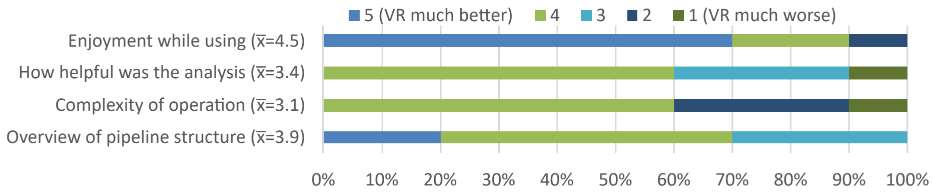

Fig. 18. VR-DataOps vs. 2D Desktop.

In summary, in terms of efficiency, VR-DataOps is comparable to desktop DataOps, averaging within +/-20% for tasks. However, accessing native desktop DataOps tools within VR was about three times less efficient due to discontinuities involving VR glasses/controller and keyboard/mouse interfaces. Creating VR-native app-specific interfaces could address this. Error rates and interventions were on par. Survey results for VR were mostly positive or neutral, except for complexity of operation, attributed to text- and menu-centric tasks. The same VR-native interface remedy could also address this (cf. Millais et al. [12]). Thus, VR-DataOps can be a viable positive addition to the DataOps environment and further stakeholder inclusion without necessitating prior VR or DataOps expertise. Threats to validity include the academic vs. business setting, the participant background and expertise, the small sample size, the sample diversity, potential bias, scenario coverage, potential subject fatigue, medium ordering (VR vs. desktop first), and differentiating VR competency and profiles. Nevertheless, we interpret the results to be indicative and promising for the potential of VR towards addressing the C1-C3 challenges, and that further and comprehensive investigation is warranted.

6 Conclusion

VR-DataOps contributes a unique, immersive solution concept offering holistic visualization and insights into the DataOps landscape. Towards addressing the aforementioned C1-C3 challenges, our contribution supports a large heterogenous data landscape by enabling integration and immersive accessibility to data, tools, and platforms. It immersively and intuitively offers access and insights to a broader set of (enterprise and non-expert) stakeholders, reducing impediments and furthering cross-silo collaboration opportunities. Its comprehensive immersive visualization supports VR-native DataOps comprehension, live monitoring, and analytic insights into data pipeline transformations and pipeline evolution. It also supports non-VR (web and desktop) data tool user interface access from within VR. The implementation demonstrated the concept's feasibility. Our empirical evaluation contributed initial insights into the concept's ability to support various DataOps scenarios related to C1-C3, as well as opportunities for further investigation and improvement.

Both native desktop and web-based DataOps apps will remain familiar, established, and useful; their replacement is not being advocated, nor can VR realistically replace the desktop entirely. Yet our evaluation has shown VR-DataOps can complement desktop apps in a meaningful way, providing an immersive holistic visualization and heterogeneous integration without significant liabilities. It thus enables DataOps accessibility and collaboration among a larger set of (enterprise or domain expert) stakeholders. With greater support by data tool vendors for VR-native interfaces, additional VR-DataOps capabilities could become readily accessible and beneficial to a wider stakeholder audience.

Future work includes improved VR-native interfaces, collaboration enhancements, integration with our prior VR EA and business process solution capabilities, and a comprehensive industry empirical study.

Acknowledgements. The author would like to thank Rafael Konecsni for his assistance with the design, implementation, evaluation, and figures.

References

1. Arthur, C.: Tech giants may be huge, but nothing matches big data. *The Guardian* (23 August 2013). https://www.theguardian.com/technology/2013/aug/23/tech-giants-data. Accessed 2026/05/14
2. Ereth, J.: DataOps - Towards a Definition. Proc. Conf. Lernen, Wissen, Daten, Analysen (LWDA 2018), Vol. 2191, pp. 104-112 (2018)
3. Berberi, L., Kozlov, V., Nguyen, G., Sáinz-Pardo Díaz, J., Calatrava, A., Moltó, G., Tran, V.. López García, Á.: Machine learning operations landscape: platforms and tools. Artificial Intelligence Review, 58(6), p.167 (2025)
4. Aslett, M.: DataOps Unlocks the Value of Data. 451 Research (2020)
5. Fannouch, A., Gharib, J., Gahi, Y.: Enhancing DataOps practices through innovative collaborative models: A systematic review. International Journal of Information Management Data Insights, 5(1), 100321 (2025)

6. Catanzano, S.: State of DataOps: Unleashing the Power of Data. Enterprise Strategy Group, TechTarget (2023) https://research.esg-global.com/reports/515201716. Accessed 2026/05/14
7. Chung, E.S., Molléri, J.S.: A Multivocal Literature Review on DataOps—Concepts, Benefits, and Challenges. In: Iglesias, A., Shin, J., Patel, B., Joshi, A. (eds) Information Systems for Intelligent Systems. ISBM 2024. Lecture Notes in Networks and Systems, vol 1255. Springer, Singapore. (2025). doi: 10.1007/978-981-96-1747-0_18
8. Oberhauser, R., Baehre, M., Sousa, P.: VR-EA+TCK: Visualizing Enterprise Architecture, Content, and Knowledge in Virtual Reality. In: Shishkov, B. (eds) Business Modeling and Software Design (BMSD 2022), pp. 122-140. LNBIP, vol 453. Springer, Cham (2022). doi: 10.1007/978-3-031-11510-3_8
9. Oberhauser, R., Baehre, M., Sousa, P.: VR-EvoEA+BP: Using Virtual Reality to Visualize Enterprise Context Dynamics Related to Enterprise Evolution and Business Processes. In: Shishkov, B. (eds) Business Modeling and Software Design. BMSD 2023. LNBIP, vol 483. Springer, Cham (2023). doi: 10.1007/978-3-031-36757-1_7
10. Oberhauser, R.: VR-DevOps: Visualizing and Interacting with DevOps Pipelines in Virtual Reality. In: Proceedings of the Nineteenth International Conference on Software Engineering Advances (ICSEA 2024), IARIA, pp. 43-48 (2024)
11. Krishnaswamy, P.: Winning with DataOps: Harnessing Efficiency in the Enterprise. Libertatem Media Private Limited (2023)
12. Millais, P., Jones, S.L., Kelly, R.: Exploring data in virtual reality: Comparisons with 2d data visualizations. In *Extended abstracts of the 2018 CHI conference on human factors in computing systems* (CHI EA '18), ACM, Paper LBW007, pp. 1–6 (2018) doi: 10.1145/3170427.3188537
13. Sudár, A., Csapó, Á.B.: Comparing desktop 3D virtual reality with web 2.0 interfaces: Identifying key factors behind enhanced user capabilities. Heliyon, 10(11), e31717 (2024) ISSN 2405-8440, doi: 10.1016/j.heliyon.2024.e31717
14. Yang, Y., Dwyer, T., Wybrow, M., Lee, B., Cordeil, M., Billinghurst, M., Thomas, B.H.: Towards immersive collaborative sensemaking. Proceedings of the ACM on Human-Computer Interaction, 6(ISS), pp.722-746 (2022) doi: 10.1145/3567741
15. Akpan, I. J., Shanker, M.: The confirmed realities and myths about the benefits and costs of 3D visualization and virtual reality in discrete event modeling and simulation: A descriptive meta-analysis of evidence from research and practice. Computers & Industrial Engineering, 112, 197-211 (2017)
16. Narasimha, S., Dixon, E., Bertrand, J. W., Madathil, K. C.: An empirical study to investigate the efficacy of collaborative immersive virtual reality systems for designing information architecture of software systems. Applied ergonomics, 80, 175-186 (2019)
17. Fonnet, A., Prie, Y.: Survey of immersive analytics. IEEE transactions on visualization and computer graphics, 27(3), 2101-2122 (2019)
18. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Quarterly, 28(1), pp. 75-105 (2004)
19. https://www.kaggle.com/datasets/shivamb/machine-predictive-maintenance-classification. Accessed 2026/05/14