VR-EDStream+EDA:
Immersively Visualizing and Animating Event and Data Streams and Event-Driven Architectures in Virtual Reality

Roy Oberhauser
Aalen University
Germany

Based on the conference paper in eKNOW 2023:
"VR-EDStream+EDA: Immersively Visualizing and Animating Event and Data Streams and Event-Driven Architectures in Virtual Reality"

Presenter: Roy Oberhauser

- Since 2004 Professor of Computer Science at Aalen University in Germany, teaching in the areas of software engineering.
- Prior to that, worked for 14 years in the software industry (Silicon Valley and Germany).
- Research interest is to leverage technologies, methods, techniques, and tools to innovate, automate, support, and improve the production and quality of software for society.

Contents

- Current challenge & problem
- Solution
- Implementation
- Evaluation
- Conclusion

Introduction

- “Data is the new oil”: Data plays a fundamental role in the digitalization and automation across industry and organizations.
- Events (a.k.a. record or message) are specific data consisting of a record of an occurrence.
- Modern software architectures are often event-driven
  - E.g., microservices or other distributed, decoupled reactive apps
- Microservice adoption in enterprises is growing
  - IDC reporting 77% and GitLab reporting 71% of organizations (partially) using microservices [1][2].
- The size of software applications has grown in size and complexity, as has the number of different apps or services and their interdependence or coupling within enterprises.
  - 57% of enterprises report using between 1000-5000 business applications [3].
  - Enterprise Service Bus architecture, based on events and messaging, is one example of integrating and coupling diverse apps and services without these needing to be aware or adapted.
- Consequently, more coupling and software reuse of (micro)services results in additional coupling and additional data and event traffic

Challenges & Motivation

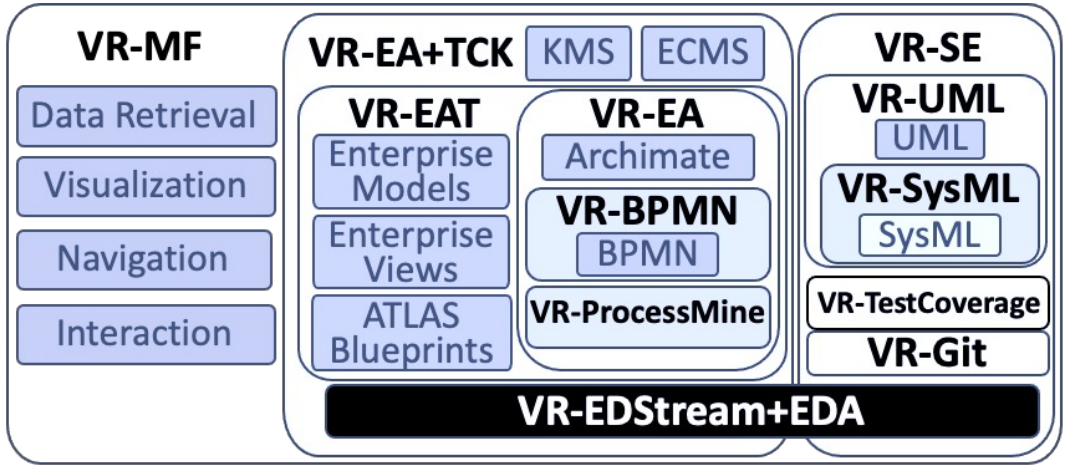
- A visually scalable application- and tool-independent capability to support data or event stream analysis is needed to address large scale IT enterprise scenarios
  - Analogous to the increase in and monitoring of network traffic, where network analysis supports analysis down to the packet level without awareness or adaptation of the involved software
- Tool and app independent visualization and analysis of large heterogeneous dynamic data and event sets and their relations remains opaque
- Grassroot or enterprise citizen accessibility for such analysis becomes imperative, as data and event streams become more relevant to the digital enterprise and stakeholders become more digitally savvy
  - Relevant and challenging to include non-expert stakeholders in such data and event analysis

Motivation for a VR Solution Approach

- VR is a mediated simulated visual environment in which the perceiver experiences telepresence.
- VR provides an unlimited space for visualizing a growing and complex set of enterprise models and processes and their interrelationships simultaneously in a spatial structure.
- As the importance, scale, inter-dependence, and coupling of data and events for IT infrastructure grows, an immersive environment can provide an additional visualization capability to comprehend and analyze both the structurally complex and interconnected static relations and the dynamic interactions between digital elements.
- Support for possible benefits of an immersive VR experience vs. 2D for analysis tasks:
  - Müller et al. [21] investigated a software analysis task that used a Famix metamodel of Apache Tomcat source code dependencies in a force-directed graph.
  - Found that VR does not significantly decrease comprehension and analysis time nor significantly improve correctness (although fewer errors were made).
  - While interaction time was less efficient, VR improved the UX (user experience), being more motivating, less demanding, more inventive/innovative, and more clearly structured.

Map of Prior Solution Concepts vs. VR-EDStream+EDA

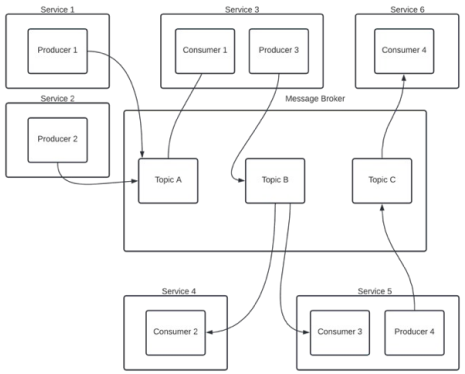
- We developed various prior solutions for the broad and complex enterprise arena
- VR-EDStream+EDA utilizes our generalized VR Modeling Framework (VR-MF) [6], which provides a VR-based domain-independent hypermodeling framework, which addresses four primary aspects that require special attention when modeling in VR: visualization, navigation, interaction, and data retrieval.

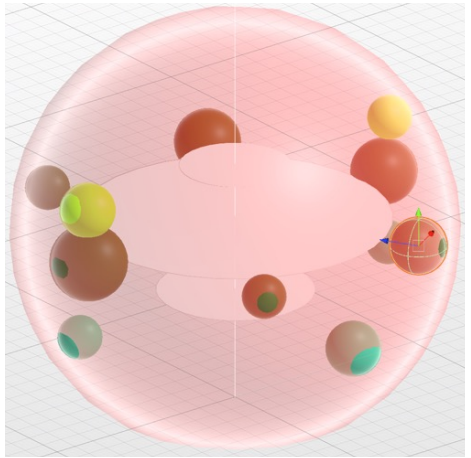


- VR-EA [6] provides specialized direct support and mapping for EA models in VR, including both ArchiMate as well as BPMN via VR-BPMN [5].
- VR-ProcessMine [6] provides support for (business or software) process mining in VR.
- VR-EAT [8] extends this further with integration of EA tools for accessing dynamically generated diagrams and models from an EA tool in VR.
- VR-EA+TCK [9] extends these capabilities by integrating further enterprise knowledge, information, and content repositories such as a Knowledge Management Systems (KMS) and Enterprise Content Management Systems (ECMS).
- VR-EDStream+EDA (shown in black) spans both the enterprise and software engineering (VR-SE) areas, applicable to relevant stakeholders depending on their focus and intention.

Solution Concept and VR Aspects

- Visualization in VR
  - Generically portraying an EDA (or dynamic stream of data/events or packets) is as a set of nodes and a Directed Acyclic Graph (DAG) to indicate the producer (source) and consumer (sink) (see top figure).
    - Events (messages) can be grouped and stored in topics, accessible to multiple producers or consumers.
  - In the immersive space of VR, analysis is affected by the distance to objects, thus ideally an initial automatic placement should place them in relative proximity to avoid delays due to traveling to objects to interact with them.
    - While a force-directed graph rebalances the distance of object automatically, it takes time to reach a steady state, while any manual element replacements by a user can cause side-effects.
  - Concept: a 3D sphere can be used to place nodes, with a 3D empty pipe (straw) used for the graph, and an event depicted as a 3D capsule that is dynamically animated within the pipe.
  - Inspired by 2D chord diagrams used in visual data analytics, we considered how to use the third dimension to reduce clutter, reduce connector collisions, and retain order and legibility.
    - To support scalability while minimizing the collision of connectors, nodes are initially placed on the outer edge of an imaginary sphere, while node groups follow along a planar circle on the sphere's edge (see bottom figure).
    - The largest sized group (based on number of nodes) is placed near the equator and serves as the basis for the sphere circumference, while smaller groups are placed accordingly closer to the poles. This grouping creates an implicit layering effect.
    - Nodes in the same group have the same color, and the size of a node (sphere) is dependent on the number of connectors (streams), with the smallest having none.
- Navigation in VR
  - Besides flythrough, also teleporting offered to reduce likelihood of potential VR sickness symptoms
- Interaction in VR
  - Since interaction with VR elements has not yet become standardized or intuitive, in our VR concept, user-element interaction is handled primarily via the VR controllers and a virtual tablet.
  - VR-Tablet provides detailed context-specific element information, and can provide a virtual keyboard for text entry fields (via laser pointer key selection) when needed.





Capabilities of the VR-EDStream+EDA Solution Concept

- A network-based mechanism for monitoring and collecting data or events
- A common data and event storage mechanism
- EDA definition and configuration
- Node grouping, placement, naming, coloring
- Store and load VR model changes
- Define event flow time period
- Dynamic event flow step and speed control

Realization

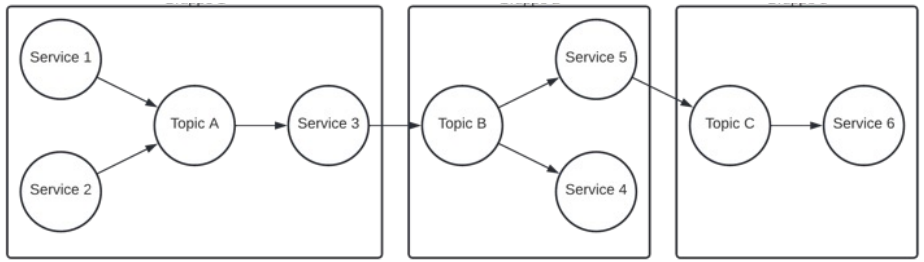
- Our prototype is partitioned into a data hub, a backend for data processing, and a front end responsible for VR visualization.
- The data hub was implemented as a microservice. For storage, the InfluxDB was used as a database due to: 1) its time series support and 2) since its storage requirements were deemed significantly smaller for large time series datasets than any alternatives, a benefit when scaling the solution.
- For receiving events generically, a microservice RESTful interface for receiving JSON event or data records was realized in Python using the FastAPI web framework.
- In addition to the REST interface, Telegraf (part of InfluxData platform) offers an open source server-based agent written in Go for collecting and sending metrics and events from databases, systems, and sensors to the InfluxDB.
  - Either interface can be flexibly used to extract or collect events, applying an interceptor, proxy, or decorator pattern as appropriate.
- VR was implemented with Unity and SteamVR

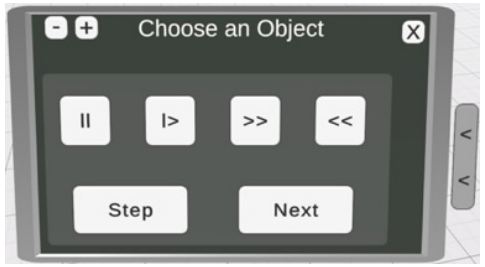
Realization

- Integration with two different event systems was performed.
- Apache Kafka is an open-source distributed event streaming platform implementing the message broker pattern.
- Kafka Connect supports data integration between databases, key-value stores, search indexes, and file systems.
- The connectors receive and transmit data to and from topics as a source or sink, and various extensible implementations are available.
  - Examples include a Source Connector that streams database updates to a topic, collect server metrics to a topic, forward topic records to Elasticsearch, etc.
- Confluent ksqlDB was used in the test applications as a database supporting queries in SQL syntax for stream processing applications based on Kafka Streams.
- To ensure a generic solution, a second popular publish/subscribe message broker event system, RabbitMQ, was also utilized in the evaluation.
- Metainformation collected via REST or Telegraf and retained in the database with each record are as follows: source, target, timestamp, payload.
  - Thus, the payload can be data, an event, a message, etc. If no target exists, then any null or fake named node can be used (equivalent to a null device in Unix).

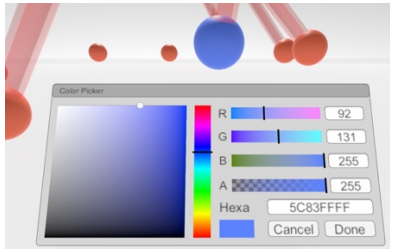
Realization

- For interaction support with the data, a VR-Tablet menu offers these display modes:
  - Animated Timeline for controlling dynamic stored or real-time playback
  - Additional settings for storing and fetching configurations are not shown here
- Configurations stored in JSON format permit stakeholders to flexibly group nodes and streams, in essence defining the EDA (e.g., based on microservices) the way they wish based on their interest. An example cross-service EDA is shown below



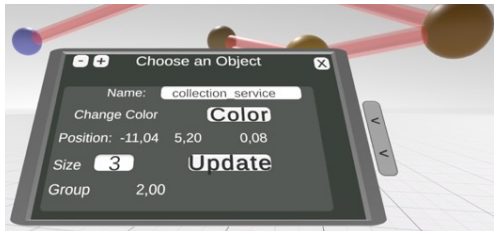


Dynamic animation interface



Color customization

Querying the event or data store



Example object details for a selected node

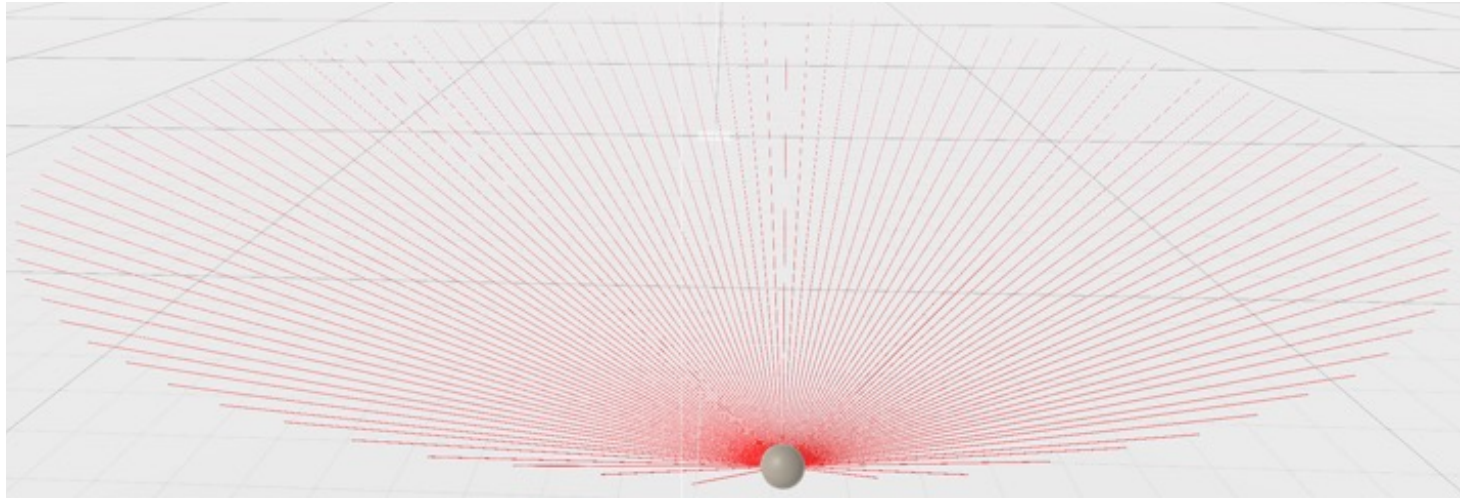
Example event details after selecting a capsule

Evaluation: Event System Integration Tests

- First configuration: Confluent Quickstart Demo using ksqlDB in combination with Kafka Connect
  - Two connectors to the topics pageviews und users
- Second configuration: Confluent Kafka
  - One producer and three consumers in Python
- Third configuration: demonstrating Kafka independence
  - RabbitMQ with our microservice

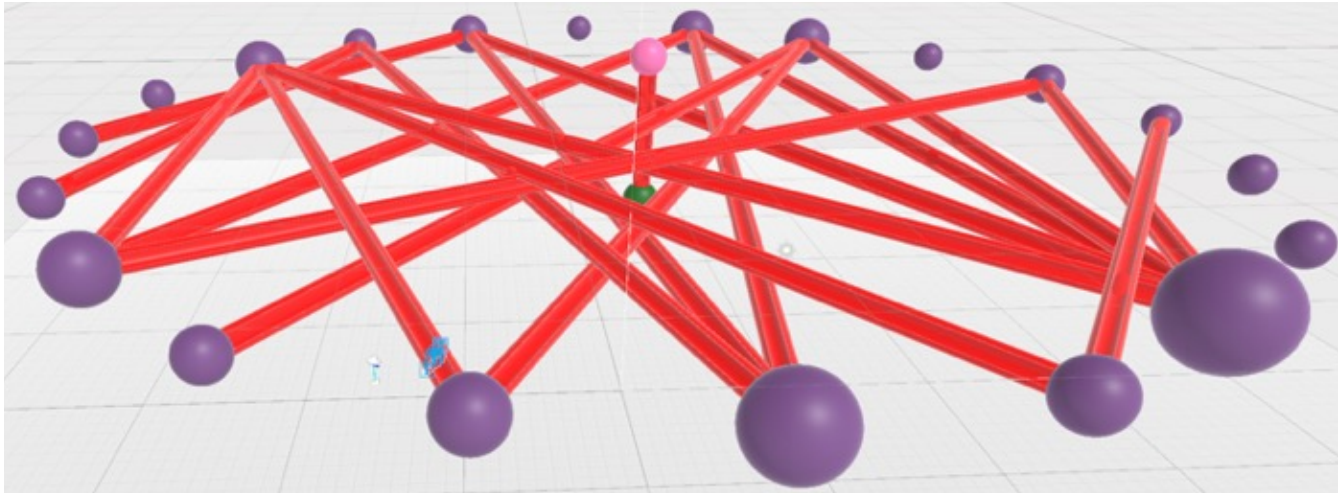
Evaluation: Single Large Group Connected to One Node

- As a scalability scenario, a single group of 100 nodes all connected is shown.
- Note that in VR, due to its unlimited space, there are no actual limitations in navigating to nodes and comprehending large models



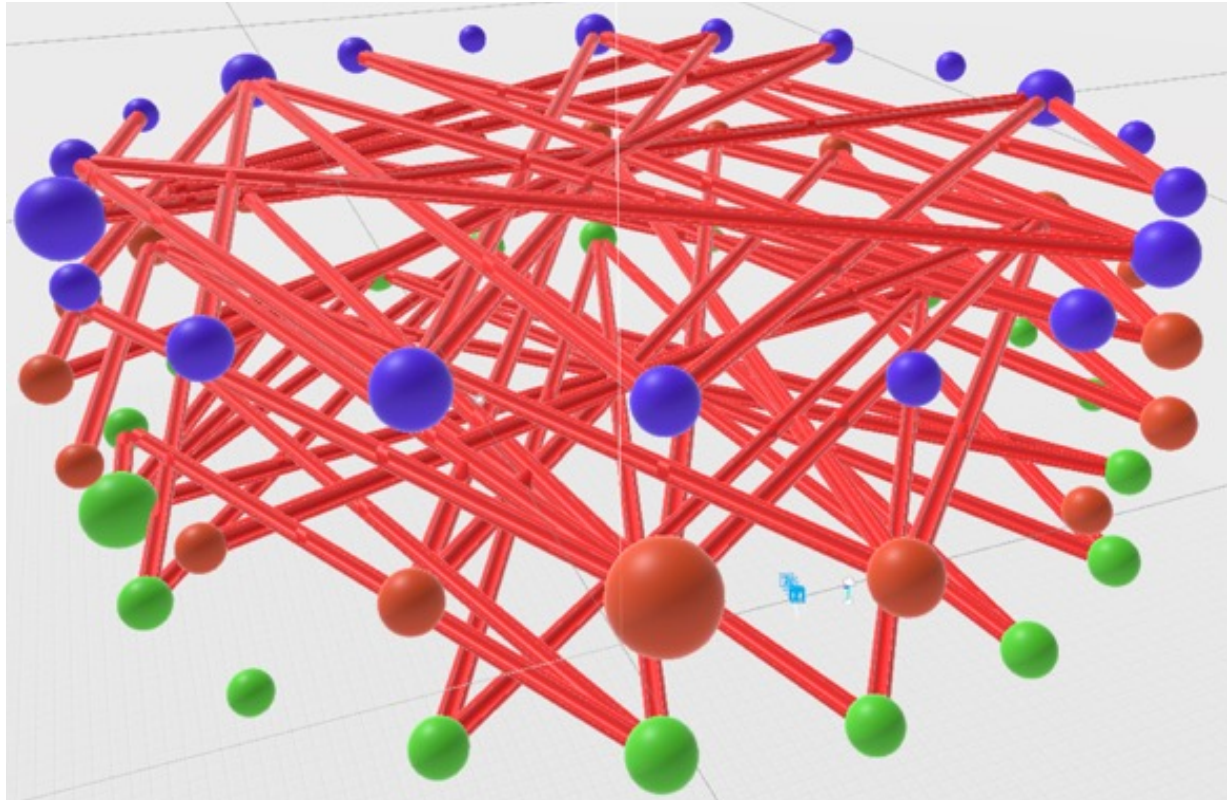
Evaluation: Unbalanced Groups Randomly Interconnected

- This scenario consisted of three unbalanced groups:
  - one group with 20 randomly intra-connected nodes, and
  - two inter-connected groups consisting of a single node each
- Note each group has a different node color, and more connected nodes are larger, and smaller groups are near the poles of the sphere, with the largest group at the equator



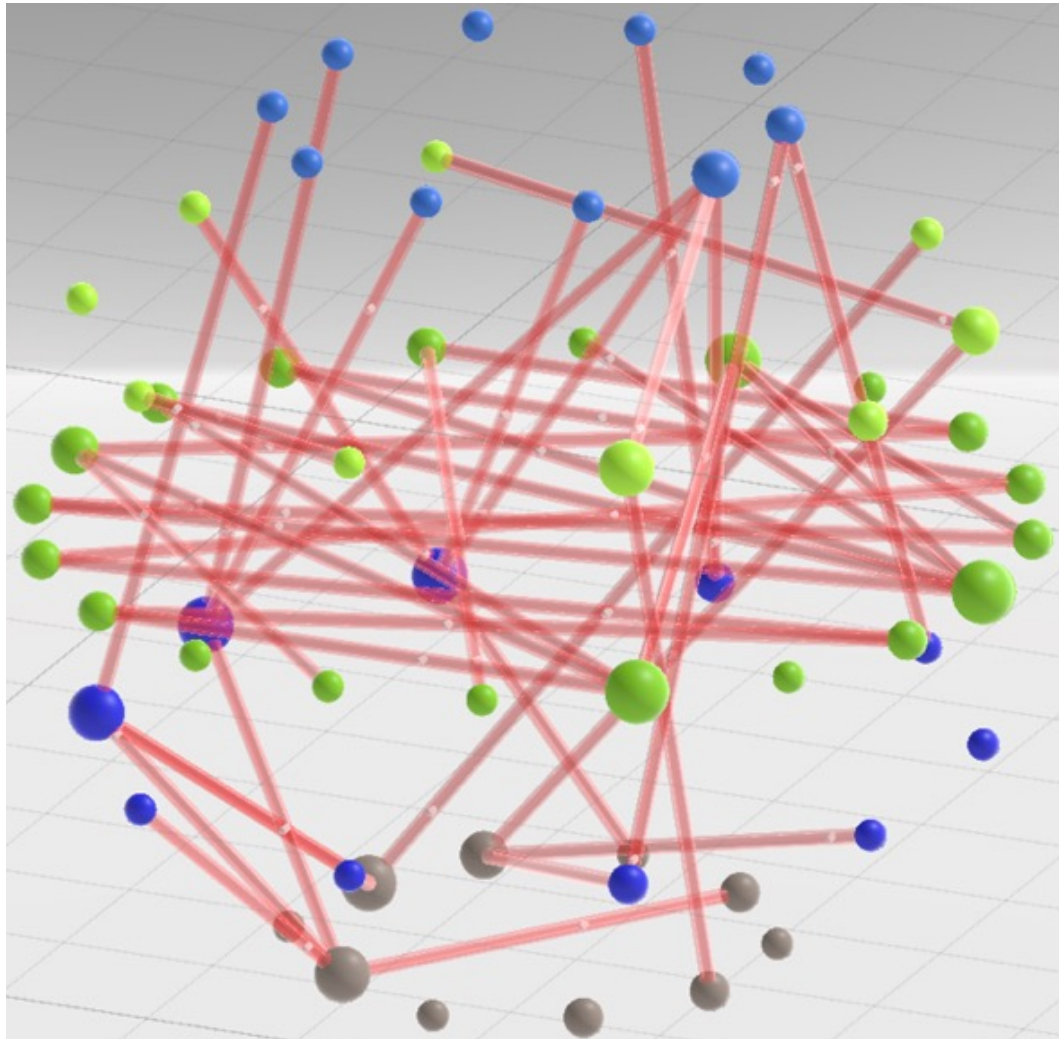
Evaluation: Multiple Balanced Highly Interconnected Groups

- In this scenario, three balanced groups of 20 nodes each are randomly inter- and intra- connected with other nodes



Evaluation: Multiple Unbalanced Groups Irregularly Interconnected

- To test many unbalanced groups with different degrees of connectedness, this scenario had five groups, one group with 20 nodes and the rest consisting of 5-10 nodes with random unbalanced coupling



Evaluation Discussion

- The scenarios with their various node grouping configurations show various possibilities and capabilities of the generic solution concept.
- Our concept generically simplifies the understanding of complex software systems for those stakeholders only concerned with event and data flow.
  - By immersively depicting sources and sinks as nodes in a spatially compact (3D spherical) layout, and animating any interaction between them.
- In focusing only on the essential flows and communication streams for data and events, while hiding all else, it is more readily scalable and immersive without cognitive overload.
- By immersively visualizing these key aspects, various (grassroot) stakeholders can now access and comprehend the flow of event data in an animated fashion.
- The default placement of an EDA configuration provides a starting point for analysis, and users can move and recolor nodes as desired, and query applicable datasets

Conclusion

- Access to and comprehension and analysis of the underlying events and data, their stream processing, and correlation with enterprise events and activities will become increasingly vital for a larger set of stakeholders.
- VR-EDStream+EDA contributes a VR solution for immersively visualizing and interacting with event and data streams or pipelines and EDA.
- Our realization showed its feasibility.
- A case-based evaluation provided insights into its capabilities, showing its ability to deal with balanced and unbalanced node group configurations and various coupling scenarios.
- Its customizable node placement in addition to the event animation provides an immersive visualization alternative for various stakeholders to comprehend the dynamic event and data flow data that form the basis of the IT system interactions in enterprises.

References

1. M. Loukides and S. Swoyer, “Microservices Adoption in 2020,” O’Reilly Media, Inc., 2020. [Online]. Available from: https://www.oreilly.com/radar/microservices-adoption-in-2020/ 2023.03.25
2. GitLab, “A Maturing DevSecOps Landscape,” 2021. [Online]. Available from: https://about.gitlab.com/images/developer-survey/gitlab-devsecops-2021-survey-results.pdf 2023.03.25
3. “New Ponemon Study Reveals Application Security Risk At All Time High: 1 in 2 Enterprises Need Better Protection,” 2015. [Online]. Available from: https://www.businesswire.com/news/home/20151210006098/en/New-Ponemon-Study-Reveals-Application-Security-Risk-At-All-Time-High-1-in-2-Enterprises-Need-Better-Protection 2023.03.25
4. M. Shirer, “IDC Survey Illustrates the Growing Importance of Developers to the Modern Enterprise,” IDC, 2021. [Online]. Available from: https://www.idc.com/getdoc.jsp?containerId=prUS48058021 2023.03.25
5. R. Oberhauser, C. Pogolski, and A. Matic, "VR-BPMN: Visualizing BPMN models in Virtual Reality," In: Shishkov, B. (ed.) Business Modeling and Software Design (BMSD 2018), LNBIP, vol. 319. Springer, Cham, 2018, pp. 83–97, doi.org/10.1007/978-3-319-94214-8_6.
6. R. Oberhauser, "VR-ProcessMine: Immersive Process Mining Visualization and Analysis in Virtual Reality," the Fourteenth International Conf. on Information, Process, and Knowledge Management (eKNOW 2022), IARIA, 2022, pp. 29-36.
7. R. Oberhauser and C. Pogolski, "VR-EA: Virtual Reality Visualization of Enterprise Architecture Models with ArchiMate and BPMN," In: Shishkov, B. (ed.) Business Modeling and Software Design (BMSD 2019), LNBIP, vol. 356, Springer, Cham, 2019, pp. 170–187, doi.org/10.1007/978-3-030-24854-3_11.
8. R. Oberhauser, P. Sousa, and F. Michel, "VR-EAT: Visualization of Enterprise Architecture Tool Diagrams in Virtual Reality," In: Shishkov B. (eds) Business Modeling and Software Design (BMSD 2020), LNBIP, vol 391, Springer, Cham, 2020, pp. 221-239, doi.org/10.1007/978-3-030-52306-0_14.
9. R. Oberhauser, M. Baehre, and P. Sousa: VR-EA+TCK: Visualizing Enterprise Architecture, Content, and Knowledge in Virtual Reality. In: Shishkov, B. (eds) Business Modeling and Software Design (BMSD 2022), LNBIP, vol 453, Springer, Cham, 2022, pp. 122-140, doi.org/10.1007/978-3-031-11510-3_8.
10. R. Oberhauser, “VR-UML: The unified modeling language in virtual reality – an immersive modeling experience,” International Symposium on Business Modeling and Software Design (BMSD 2021), Springer, Cham, 2021, pp. 40-58, doi.org/10.1007/978-3-030-79976-2_3
11. R. Oberhauser, “VR-SysML: SysML Model Visualization and Immersion in Virtual Reality,” International Conference of Modern Systems Engineering Solutions (MODERN SYSTEMS 2022), IARIA, 2022, pp. 59-64.
12. R. Oberhauser, “VR-Git: Git Repository Visualization and Immersion in Virtual Reality,” The Seventeenth International Conference on Software Engineering Advances (ICSEA 2022), IARIA, 2022, pp. 9-14.
13. X. Qin, Y. Luo, N. Tang, and G. Li, “Making data visualization more efficient and effective: a survey,” The VLDB Journal, 29, pp.93-117, 2020.
14. A. Fonnet and Y. Prié, "Survey of Immersive Analytics," in IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 3, pp. 2101-2122, 2021.
15. A. Fonnet, F. Melki, Y. Prié, F. Picarougne, and G. Cliquet, “Immersive Data Exploration and Analysis,” Student Interaction Design Research Conference, Helsinki, Finland, hal-01798681, 2018, https://hal.science/hal-01798681.
16. P. Reipschläger et al., “DebugAR: Mixed dimensional displays for immersive debugging of distributed systems,” In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, pp. 1-6.
17. R. Sicat et al., “DXR: A toolkit for building immersive data visualizations,” IEEE transactions on visualization and computer graphics, 25(1), 2018, pp.715-725.
18. M. Cordeil et al., "IATK: An Immersive Analytics Toolkit," 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 2019, pp. 200-209, doi: 10.1109/VR.2019.8797978.
19. S. Hubenschmid, J.Zagermann, S. Butscher, and H. Reiterer, “Stream: Exploring the combination of spatially-aware tablets with augmented reality head-mounted displays for immersive analytics,” Proc. 2021 CHI Conference on Human Factors in Computing Systems, 2021, pp. 1-14.
20. A. Satyanarayan, R. Russell, J. Hoffswell, and J. Heer, "Reactive Vega: A Streaming Dataflow Architecture for Declarative Interactive Visualization," In: IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 1, pp. 659-668, 2016, doi: 10.1109/TVCG.2015.2467091.
21. R. Müller, P. Kovacs, J. Schilbach, and D. Zeckzer, "How to master challenges in experimental evaluation of 2D versus 3D software visualizations," In: 2014 IEEE VIS International Workshop on 3Dvis (3Dvis), IEEE, 2014, pp. 33-36.
22. P. Dobbelaere and K.S. Esmaili, “Kafka versus RabbitMQ: A comparative study of two industry reference publish/subscribe implementations,” In: Proc. 11th ACM int’l conference on distributed and event-based systems, 2017, pp. 227-238.
23. A.R. Hevner, S.T. March, J. Park, and S. Ram, “Design science in information systems research,” MIS Quarterly, 28(1), 2004, pp. 75-105.