Visualizing an Exascale Data Center Digital Twin: Considerations, Challenges and Opportunities

Matthias Maiterth*
Oak Ridge National Laboratory

Wes Brewer
Oak Ridge National Laboratory

Dane De Wet
Oak Ridge National Laboratory

Scott Greenwood
Oak Ridge National Laboratory

Vineet Kumar
Oak Ridge National Laboratory

Jesse Hines
Oak Ridge National Laboratory

Sedrick Bouknight
Oak Ridge National Laboratory

Zhe Wang
Oak Ridge National Laboratory

Tim Dykes
Hewlett Packard Enterprise

Feiyi Wang
Oak Ridge National Laboratory

2024 IEEE Visualization and Visual Analytics (VIS) | USGov | DOI: 10.1109/VIS55277.2024.00012

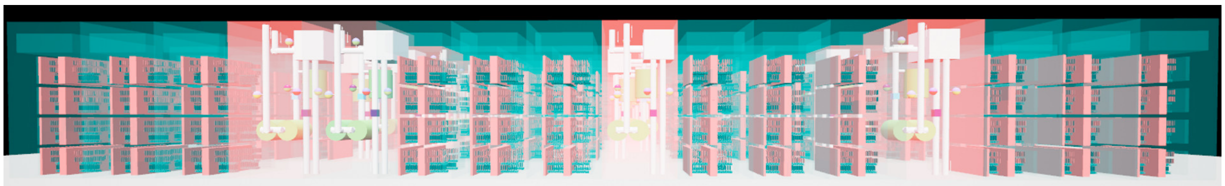

Figure 1: Prototype of the Frontier digital twin using Unreal Engine 5. All child components are visible: The first row, shows 12 compute racks and four cooling distribution units[1]. The racks show, among others: cable trays, rectifiers[1], and compute nodes, with all CPUs, GPUs[1] and DIMMs visible. ([1]: Colored according to power consumption.) Total actors: 152,906.

Abstract

Digital twins are an excellent tool to model, visualize, and simulate complex systems, to understand and optimize their operation. In this work, we present the technical challenges of real-time visualization of a digital twin of the Frontier supercomputer.

We show the initial prototype and current state of the twin and highlight technical design challenges of visualizing such a large High Performance Computing (HPC) system. The goal is to understand the use of augmented reality as a primary way to extract information and collaborate on digital twins of complex systems. This leverages the spatio-temporal aspect of a 3D representation of a digital twin, with the ability to view historical and real-time telemetry, triggering simulations of a system state and viewing the results, which can be augmented via dashboards for details. Finally, we discuss considerations and opportunities for augmented reality of digital twins of large-scale, parallel computers.

Index Terms: Digital Twin, Data Center, Information Representation, Massively Parallel Systems, Operational Data Analytics, Simulation, Augmented Reality

1 Introduction

A digital twin can be defined as an (1) evolving digital representation of the (2) historical, current and future behavior (3) of a physical object or process (4) that helps to optimize its performance [17, 11]. Simulations for digital twins in scientific domains, from applied mathematics to the life-sciences, are often run on High Performance Computing (HPC) systems [12] [21] [24]. Yet, using digital twins to model HPC systems themselves is relatively new.

HPC centers have been collecting metrics of their operations for decades, which generally served to trigger system alerts and do post-mortem data-analysis. Digital twins promise a holistic understanding of HPC systems combining telemetry with simulation, enabling insights to improve overall system efficiency.

For a digital twin, its visualization is the main point for human interaction. It ties all disaggregate information together and allows us to correlate the temporospatial information generated.

Related works have shown the impact of the visual representation of compute clusters; however, these are limited to smaller systems [3] or only focus on one aspect of digital twins [23], or do not incorporate telemetry of infrastructure [22] or simulation. At the same time, to the knowledge of the authors, no other work discusses the major challenges of visualizing large-scale data centers at high fidelity with real-time interaction, with the ability to trigger simulations visualized in the same environment.

In this work, we present: A brief overview of Frontier, a state-of-the-art supercomputer with more than 60 million parts, generating one million data-points each second; the current state of the visualization prototype for the exascale data center digital twin; and a discussion on encountered challenges and opportunities for the visualization of digital twins of large-scale parallel compute systems. The overarching ExaDigiT digital twin framework is presented in [4].

2 Background

The Oak Ridge Leadership Computing Facility (OLCF) started installation of the Frontier Supercomputer by the end of 2021, and achieved 1.1 ExaFlops of performance by the submission to the TOP500 list of June 2022 [18]. To create a digital twin and its visualization, a description of the system and an understanding of the telemetry data is needed, which is shown in the following.

2.1 System Description

The Frontier HPC system consists of 9,472 AMD compute nodes, with a total of 9,472 Central Processing Units (CPUs) plus 37,888 Graphics Processing Units (GPUs). In June 2023, 37,632 of these GPUs were used for the second TOP500 submission, achieving 1.194 ExaFlops at 22.7 MW [19].

The system is comprised of seven rows of HPE Cray EX cabinets. Three Cray EX cabinets (384 compute nodes) are supported by one Cooling Distribution Unit (CDU) for cooling. Each compute cabinet consists of eight compute chassis, plus the networking switches, for the Slingshot Dragonfly network. Each chassis itself has eight Bard Peak blades, each housing two compute nodes.

*e-mail: maiterthm@ornl.gov
Project home: https://exadigit.github.io.

DOI 10.1109/VIS55277.2024.00012

The nodes themselves consist of one AMD EPYC™7A53 "Trento" CPU with four AMD Instinct™ MI250X GPUs. Each individual MI250X GPU is composed of two chiplets, called Graphics Compute Dies (GCDs). Outside of the compute room, the central energy plant is responsible for pumping coolant through the primary cooling loop and distributing it to the CDUs as part of the secondary cooling loop. The CDUs contains the heat exchangers to supply each cabinet with cooling. Switchboards, bus-bars, rectifiers, and Super Intermediate VOltage Converters (SIVOCs) within the compute room are responsible for distributing and stepping down voltages for consumption at the nodes.

This high level overview alone shows that having a good understanding of such complex system, and how its parts interact, is not an easy task. Overall there are $\sim 60,000,000$ individual parts in the system that could be modeled. The question is which to include, model, and show; Which depends on their state, data of interest, and potential simulations to augment insights gained.

2.2 Telemetry

The Oak Ridge Leadership Facility houses two flagship HPC systems, while the installed infrastructure supports additional facilities. The fact that there is not a single central system collecting data is due to the sensible split of responsibilities according to facility, compute, scheduling, etc. A majority of data sources of relevance for the digital twin are aggregated in the Integrated Telemetry Database (ITDB), but the consolidation is an ongoing process. At the same time, several systems of relevance, e.g., to the central energy plant, are explicitly encapsulated. The collection of telemetry is a process not only required by a digital twin, but already undertaken by Operational Data Analytics (ODA) [1].

In general, there are three data retention strategies for the telemetry generated, which influences how they can be used within a digital twin: Data is either a) at full resolution and stored long term, b) available at full resolution but reduced in resolution for long-term storage, or c) only available in real-time or for a short time window and over-written in a sliding window fashion. Retention and long-term storage depend on the impact and usage of each sensor value.

Overall Frontier alone generates one million data points each second. This is a stark increase from the previous system, Summit, which generated 'only' 400,000 data points each second.

3 Considerations for the UE5 Frontier Digital Twin

The goal of the digital twin is for users and system engineers to interact with the virtual representation of the system and to understand the implications of the system setup, including telemetry replay and triggering simulations of its sub-components.

Initial tests using web frameworks served as an example to evaluate streaming of data into an Augmented Reality (AR)/Virtual Reality (VR)-Scene. This was done using A-Frame.io [14, 15], which is great for prototyping but not suitable for large scenes. To resolve this, the prototype was re-implemented in Unreal Engine 5 (UE5) [8] for AR, as it has been successfully adopted for other projects in the organization [20, 9], using HoloLens2, as well as desktop. A snapshot of the current state is shown in Fig. 1.

3.1 Current Prototype

System hierarchy: The general setup of the system is built using UE5-blueprints and follows the hierarchical setup of the system as outlined above. The Frontier system blueprint consists of rows of cabinets, as seen in Fig. 2. These are themselves either compute racks or CDU blueprints. The compute rack blueprint contains eight chassis, where each chassis is further subdivided into its network equipment, cooling manifolds, PDUs and rectifiers. Additionally, each chassis contains eight blades, each with two nodes, where

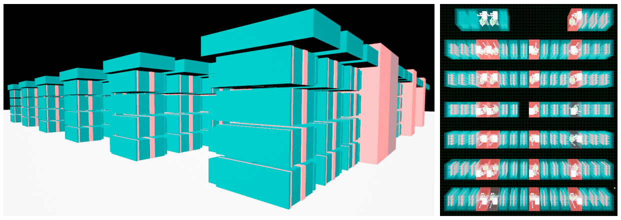

Figure 2: Desktop View: Side View and Translucent top down.

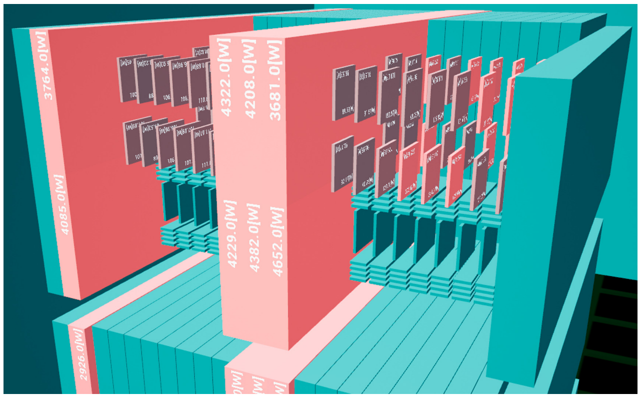

Figure 3: Left to right: Cooling manifold, two rectifiers, eight blades (containing two nodes each), six additional rectifiers, eight blades and the PDUs on the right. Each node consists of one CPU, four GPUs, eight DIMMs each. Telemetry in red for rectifiers and GPUs.

nodes contain blueprints for CPU, GPUs and Dual In-line Memory Modules (DIMMs). In case a component of the system is contained within another, the modeled UE5 actor is attached to a parent actor forming a hierarchy of actors. The detailed view of a single chassis, showing rectifiers, GPUs, CPUs, and DIMMs is shown in Fig. 3.

Component Representation: Even with the more advanced engine, the number of individual components is very large. With the current state of the implementation, the fully populated scene has over 150,000 UE5 actors. (The setup follows the blueprint hierarchy as described above, plus a few additional parts such as rectifiers, etc.) The actors can be spawned on demand, either in the editor or while running the scene, to only show the subsystems of interest. Since each component can show different information and can be spawned on demand, for visual clarity, typical techniques such as merging of meshes or instanced components are not feasible. This would prohibit AR interaction and disallow streaming of data to only subsets of a system and the change of the system in a non-static way. This is a double-edged sword: the resulting large actor counts are not good for performance, but allow to reduce visual clutter and interactive focus on areas of interest. In the default level, the user only sees the 104 cabinets (of which 74 are compute cabinets of frontier) and can interact and show internals and additional detail on demand, alleviating the large actor count.

Fig. 1 and Fig. 2 show all components spawned in different views. Displaying all components with all details and dynamic telemetry is not feasible, due to performance limitations. In an interactive view only the subsystems of interest are spawned, as seen in Fig. 4, showing the system projected onto a meeting table, or in Fig. 5, where the two twins are shown in place.

Data Representation: Each component has a blueprint with a material and variables for the metrics of interest. The metrics are used to programmatically change the material color based on the requested metric (from low to high, teal to red). Additionally,

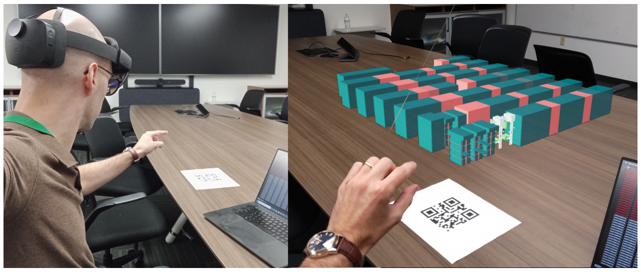

Figure 4: AR Table-top: First (right) and third (left) person view.

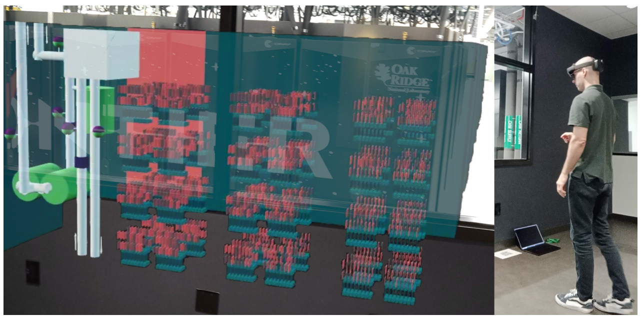

Figure 5: Digital and physical twin in-place: First (left) and third (right) person view. The user is displaying telemetry and about to trigger the simulation in AR. The QR-Code serves as spatial anchor.

the system is programmed in a way where the Frontier system has a representation of the hierarchical data and each spawned child-actor has functionality to distribute and display its data as a color via the material of the actor. The telemetry data is either stored locally or can be requested from the telemetry service which is then distributed to the correct location. This can be used to continuously stream data or project a recorded dataset into the scene. For this, we use our telemetry service and can request data of interest. In the case shown in the figures, we query power for the CDUs, the rectifiers, and the GPUs. Data structures, data ingestion and distribution is implemented in parts in blueprints but mostly C++.

3.2 Data Ingestion and Visualization

Fig. 6 shows the data flow from source to visualization. After querying, selecting, and preprocessing, the remote data is forwarded to the local workstation. This allows to query data from the beginning of data collection within under five seconds via REST-API of the telemetry service. 'Live' data via the telemetry service is available within two minutes [1] of collection. The final preprocessing step is performed locally such that the visualization can be displayed in the 3D-Scene.

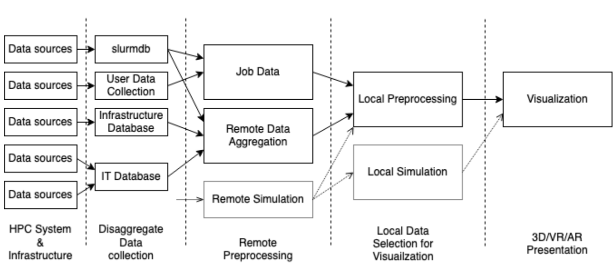

Figure 6: Data Flow from source to visualization.

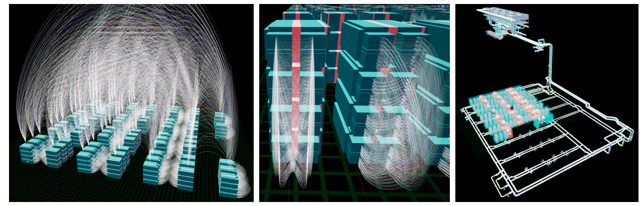

Figure 7: Future Work: Dragonfly Network, with inter-group (left) and intra-group (middle) network. And central energy plant (right)

Depending on the current actor hierarchy in the scene, the telemetry is used to populate a hierarchical shadow structure, which is then used to distribute the data to the respective actor. This is achieved by the implementation of a scatter or broadcast mechanism available to each blueprint, by dynamically distributing and routing of the information in at most $n \log n$ steps. This allows us to handle missing data, while only requiring distribution to components present in the current scene. The design allows us to incrementally add subsystems based on the state and focus of the scene. Additionally, this enables us to easily port the design to other HPC twins. A static distribution using offsets in a data table has been avoided due to reduced flexibility and portability reasons.

The data-set used in the figures of this work is the power data recorded for the TOP500 submission of June 2023 [19, 2]. In general, the data can be queried on demand by the client.

3.3 Integration of simulation for what-if scenarios

As mentioned in the introduction, "[a digital twin is a digital] representation of the (2) historical, current and future behavior [...]" of a system [17]. Therefore, not only is the display of telemetry data of interest, but also the ability to simulate subsystems behavior based on a system state. The level of fidelity (both spatially and temporally) for each of such interdependent simulation tools largely depend on how they are coupled with each other. Depending on the the simulation, this can be run in in-situ, alongside the digital twin, or be fed into the 3D-scene after the simulation completion. This depends largely on the model complexity, real-time capability and how the individual simulations are coupled.

Our current model is able to run a thermo-fluidic model to simulate the cooling system, triggered via a hand menu. This is achieved using the Functional Mock-up Interface (FMI) standard interface used to integrate packaged ODE-based system models (Functional Mock-up Units (FMUs)). The FMU model accepts real-time inputs and generates outputs which are key to coupling to other models within the digital twin modeling framework. The system-level model offers a good balance between advanced predictive capability and simulation time in comparison to Computational Fluid Dynamics (CFD) models which offer greater spatial and temporal resolution but are currently not suitable for real-time dynamic simulations. In the work presented, a Modelica [5] model was developed and the simulation of the liquid cooling loop from cooling tower to CDU is integrated and can be triggered and results visualized via FMU plugin [10], developed for the TRANSFORM library [9]. The results of the simulation are displayed in the mocked internals of the CDUs (see Fig. 5, left). The thermo-fluid simulation includes the cooling loop from cooling tower to CDU [4] and is in the process of being fully integrated with the visualization. Currently, we map simulated flow rates, pressures and temperatures to the CDUs when triggering the simulation from telemetry within the visualization.

3.4 Next steps: Addition of the network and integration of central energy plant

HPC systems rely on fast interconnects to achieve high performance at system scale, therefore, modeling and understanding the network is essential. We have started prototyping the network and can show a full Dragonfly topology, with $74*73/2$ Inter-Group plus $74*(64*63)/2$ Intra-Group links, resulting in a total of $151,885$ links. This large amount of links is implemented via the Niagara particle system in UE5 [7]. (Note: The real system has fewer links, as only 32 switches per rack are installed, where 64 is the racks maximum.) The next step is to replay and display telemetry for each link. Fig. 7 shows that all static links can be displayed, and even be streamed to the HoloLens2 without performance issues. Selectively displaying based on link-level telemetry is work in progress.

Additionally, we have added the CAD models for the central energy plant in the visualization, where integration of simulation and telemetry are ongoing work (see Fig. 7 right).

4 Challenges

In the following, we discuss the major challenges encountered:

Large Scale Systems: The system discussed is at extreme scale. This goes both for individual components, but is true for each: compute, infrastructure, and data. At the same time, the interplay of the different systems spans multiple domains (computation, electric, cooling). Most of the compute systems can be grouped hierarchically, such as the layout described in the setup above. Everything else has a complex interaction with the compute system and forms an overlay structure: the networking of the system; the electrical system; the cooling system. Mapping active workloads and jobs to the compute system and understanding its impact on the aforementioned subsystems poses a significant challenge, where a visual system can help understand theses interactions.

Big Data: As seen in the telemetry section, there are several teams required to maintain and operate the telemetry services to use the data coming from the system, in which ODA [16] is of principal importance. This is truly big data, as becomes clear by putting information of our telemetry system to the four Vs associated with big data [1]: Volume (expected 20PB); Velocity (currently 300 MB/s of telemetry in transit); Variety (Types of sensors, datatypes, sampling frequency retention strategy etc.); Veracity (missing data and faulty sensors). By projecting the collected data onto a digital twin, a better understanding of the system is formed. Yet, the real value-add is achieved by using this data to run simulations and gain additional insights, by doing what-if analysis.

Visualization: The current visualization only captures a part of the system. However, the total number of actors with only this number of components present in the system is over 150,000 already. Since we use simple cuboids the rendering and updating of data in the scene is still feasible and performance on a modern desktop system is acceptable. When streaming the scene to a HoloLens 2 device from the same desktop, not all actors can be displayed, experiencing limitations at $\sim 100,000$ actors. Even if we could stream all this data, there is more data of interest, thus we always have to make trade-offs and select areas of interest and usability, while providing overview and detail.

5 Opportunities

The combination of telemetry and simulation displayed in the same 3D scene allows for unique capabilities for better understanding complex system behavior.

Spatial and Temporal Correlation of Information: Each component within the system has a physical location. Data visualization of telemetry usually does not present the data according to its location. Digital twins present a unique opportunity to bring telemetry and sensors into a spatial context and help to understand the impact on connected systems. After the spatial correlation, the temporal correlation of the data is the next step. The distribution of telemetry observed in different subsystems can be replayed in a synchronized fashion. Depending on the fidelity of the simulations, initial conditions can be chosen based on the telemetry of the complementary subsystems and replayed based on the temporal context.

Adaptive Layout of Information: Lessons learned from information visualization should be vigorously applied to avoid information overload and visual clutter. This can be especially detrimental in AR and VR environments. At the same time, by operating inside of a game engine, the layout is not tied to the physical layout of the system. In Computer Aided Design (CAD) the exploded view of a part is an example for this. For an HPC digital twin, presenting the information in a planar view as seen in a dashboard and then transforming the individual components back to their physical layout can allow bridging the chasm between data and information visualization and 3D visualization of physical systems [13, 6].

Understanding Complex System Interaction: Within the digital twin modeling framework, one of the keys to understanding complex system interaction is to ensure that each of the constituent models can exchange dynamic data and generate system simulation which is useful to the end-user or customer. The flavor of the digital twin depends on the end-user – for planning and the design phase, for operations, or after decommissioning for retrospective use. Our use-cases are: In the facility design/planning phase, optimization of the cooling system can be used to drive the datacenter Power Usage Effectiveness (PUE) closer to 1.0. For facility operations, which is the typical use-case for a digital twin, the digital twin is useful for the facility operator to visualize the performance of the digital twin alongside the real system, allowing to localize potential deviation. The digital twin can also include reliability models informing the operator of potential component failures, such as mechanical components (cooling system), electrical components (power supply), and GPUs/CPUs. The what-if-scenario use case relies heavily on the simulation aspect of the digital twin and is descibed in [4].

Interactive Level of Detail (LoD): We identified interaction as an LoD mechanism to be an excellent way of coping with limited performance and at the same time visual clutter. Fully utilizing interactive LoD is an opportunity, yet to find good ways of general interaction is not always trivial in complex systems. The value of the digital twins is also of collaborative nature, therefore understanding how interactive LoD can be done with network replication in mind is a large opportunity that we want to leverage.

Combining AR with web-based Dashboards: Data analysis in three dimensions is not trivial, where two dimensions can have significant advantages when comparing individual data points, and timelines. As we operate in an AR environment we can bring web-based dashboards already developed for ODA into the scene, and supplement the digital twin with simulation-specific dashboards. This also improves the value of the dashboards, as it is generally not trivial to correlate their information to the spatial domain. This bridges a huge gap that only digital twins with real-time analysis capability can achieve.

6 Conclusion & Future Work

We presented the current state of the visualization for the Frontier digital twin. We showed the hierarchically spawn-able structure of our twin and the capability to replay data from the ODA telemetry servers and to trigger thermo-fluid simulations, displaying its results in the live AR environment. With the focus on interactive AR systems, resource limitations for representing such a large-scale system had to be carefully considered, resulting in a solution with scalability and interactive LoD in mind. We are in the process of capturing additional use-cases and are working with the community to make it modular and relevant for other compute systems.

The visualization of a digital twin greatly augments the understanding of a complex system, such as the Frontier supercomputer, and allows us to interact with them for greater understanding.

Acknowledgments

This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725.

The author did not use any content generated by artificial intelligence (AI) in this article.

References

[1] R. Adamson, T. Osborne, C. Lester, and R. Palumbo. STREAM: A Scalable Federated HPC Telemetry Platform, 5 2023. 2, 3, 4

[2] S. Atchley and M. Maiterth. OLCF Frontier Supercomputer 2023-04-29 HPL Power Data used for Top500/Green500 Submission, 5 2023. doi: 10.13139/OLCF/1975494 3

[3] B. Bergeron, M. Hubbell, D. Sequeira, W. Williams, W. Arcand, D. Bestor, C. Byun, V. Gadepally, M. Houle, M. Jones, A. Klien, P. Michaleas, L. Milechin, J. Mullen, A. Prout, A. Reuther, A. Rosa, S. Samsi, C. Yee, and J. Kepner. 3d real-time supercomputer monitoring. In *2021 IEEE High Performance Extreme Computing Conference, HPEC 2021*, pp. 1–7. IEEE, Waltham, MA, USA, 9 2021. doi: 10.1109/HPEC49654.2021.9622787 1

[4] W. Brewer, M. Maiterth, V. Kumar, R. Wojda, S. Bouknight, J. Hines, W. Shin, J. Webb, S. Greenwood, W. Williams, D. Grant, and F. Wang. A digital twin framework for liquid-cooled supercomputers as demonstrated at exascale. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '24. Association for Computing Machinery, New York, NY, USA, 11 2024. (accepted). 1, 3, 4

[5] F. Casella, M. Otter, K. Proelss, C. Richter, and H. Tummescheit. The Modelica fluid and media library for modeling of incompressible and compressible thermo-fluid pipe networks. In *Proceedings of the 5th International Modelica Conference*, pp. 631–640. The Modelica Association, Vienna, Austria, 2006. 3

[6] A. Cockburn, A. Karlson, and B. B. Bederson. A review of overview+detail, zooming, and focus+context interfaces. *ACM Comput. Surv.*, 41(1), jan 2009. doi: 10.1145/1456650.1456652 4

[7] Epic Games. Niagara Reference, 11 2022. 4

[8] Epic Games. Unreal Engine 5.1, 11 2022. 2

[9] M. S. Greenwood, B. R. Betzler, A. L. Qualls, J. Yoo, and C. Rabiti. Demonstration of the advanced dynamic system modeling tool transform in a molten salt reactor application via a model of the molten salt demonstration reactor. *Nuclear Technology*, 206(3):478–504, 2020. doi: 10.1080/00295450.2019.1627124 2, 3

[10] S. Greenwood. Unreal Engine - FMI Plugin (UEFMI), 2019. 3

[11] W. Hu, T. Zhang, X. Deng, Z. Liu, and J. Tan. Digital twin: A state-of-the-art review of its enabling technologies, applications and challenges. *Journal of Intelligent Manufacturing and Special Equipment*, 2(1):1–34, 2021. doi: 10.1108/JIMSE-12-2020-010 1

[12] M. Kasztelnik, P. Nowakowski, J. Meizner, M. Malawski, A. Nowak, K. Gadek, K. Zajac, A. A. L. Mattina, and M. Bubak. Digital Twin Simulation Development and Execution on HPC Infrastructures. In J. Mikyška, C. de Mulatier, M. Paszynski, V. V. Krzhizhanovskaya, J. J. Dongarra, and P. M. Sloot, eds., *Computational Science – ICCS 2023*, pp. 18–32. Springer Nature Switzerland, Cham, 2023. doi: 10.1007/978-3-031-36021-3_2 1

[13] J. Kehrer and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(3):495–513, 2013. doi: 10.1109/TVCG.2012.110 4

[14] D. Marcos, D. McCurdy, K. Ngo, and A-frame contributors. A-frame – a web framework for building 3D/AR/VR experiences, 01 2023. 2

[15] D. Marcos, D. McCurdy, K. Ngo, and A-frame contributors. aframe git, 01 2023. 2

[16] A. Netti, W. Shin, M. Ott, T. Wilde, and N. Bates. A conceptual framework for HPC operational data analytics. In *2021 IEEE International Conference on Cluster Computing (CLUSTER)*, pp. 596–603. IEEE, Portland, OR, USA, 2021. doi: 10.1109/Cluster48925.2021.00086 4

[17] A. Parrott and L. Warshaw. Industry 4.0 and the DT, 2017. 1, 3

[18] PROMETEUS Professor Meuer Technologieberatung und Services GmbH. Top500 supercomputer sites — june 2022, 05 2022. 1

[19] PROMETEUS Professor Meuer Technologieberatung und Services GmbH. Top500 supercomputer sites — june 2023, 06 2023. 1, 3

[20] M. B. Smith, D. E. Peplow, N. Nelson, N. Thompson, M. S. Greenwood, and USDOE. Viper - method for visualizing and interacting with ionizing radiation data using augmented reality, 4 2023. doi: 10.11578/dc.20230406.2 2

[21] B. G. Sumpter and V. Meunier. Digital twins in materials and chemical sciences. *Carbon Trends*, 13(C), 12 2023. doi: 10.1016/j.cartre.2023.100297 1

[22] R. Wang, Z. Cao, X. Zhou, Y. Wen, and R. Tan. Phyllis: Physics-informed lifelong reinforcement learning for data center cooling control. In *Proceedings of the 14th ACM International Conference on Future Energy Systems*, e-Energy '23, p. 114–126. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3575813.3595189 1

[23] R. Wang, X. Zhou, L. Dong, Y. Wen, R. Tan, L. Chen, G. Wang, and F. Zeng. Kalibre: Knowledge-based neural surrogate model calibration for data center digital twins. In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, BuildSys '20, p. 200–209. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3408308.3427982 1

[24] I. Zacharoudiou, J. McCullough, and P. Coveney. Development and performance of a HemeLB GPU code for human-scale blood flow simulation. *Computer Physics Communications*, 282:108548, 2023. doi: 10.1016/j.cpc.2022.108548 1