A No Code Approach to Infrastructure Provisioning in Support of Science

Ciprian Popoviciu
Tech Systems Department
East Carolina University
Greenville, NC 27858, USA
popoviciuc18@ecu.edu

Jerome Sobieski
ECE Department
George Mason University
Fairfax, VA 22030, USA
jsobiesk@gmu.edu

Bijan Jabbari
ECE Department
George Mason University
Fairfax, VA 22030, USA
bjabbari@gmu.edu

Colby Sawyer
Tech Systems Department
East Carolina University
Greenville, NC 27858, USA
sawyerc021@ecu.edu

Liang Zhang
ECE Department
George Mason University
Fairfax, VA 22030, USA
Lzhang36@gmu.edu

Abstract - Communications infrastructures and compute resources are critical to enabling advanced science research projects. Science cyberinfrastructures must meet clear performance requirements, must be adjustable to changing requirements and must facilitate reproducibility. These characteristics can be met by a programmable infrastructure with guaranteed resources such as the BRIDGES infrastructure enabling cross Atlantic research projects. While programmability should be a foundational design principle for research cyberinfrastructures, by itself might not be sufficient to enabling scientists who have no or limited experience with advanced IT technologies operate their testbeds independent of IT support teams. The trend of offering "no code" platforms enabling users without IT core competency to achieve business goals should manifest itself in the context of research and educational infrastructures as well. In this paper we describe the architecture of a "no code" platform which would enable scientists to easily configure and modify a programmable infrastructure by using a large language model-based interface integrated with the composable services language of the infrastructure. The BRIDGES testbed is used as an example for such an integration where the functionality benefits projects operated by large, diverse teams.

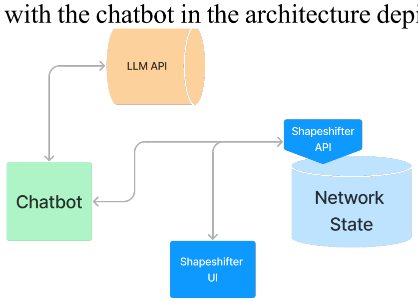
Keywords— Cyber Infrastructure, Virtualization, Cloud, Orchestration, Large Language Models, No Code, Science Testbeds

I. Introduction

With the emergence of easily accessible, powerful Large Language Models (LLM) [1], Artificial Intelligence (AI) entered the mainstream and is now accepted as a powerful tool for many human activities and endeavors. The broad popularity of AI proves to be an accelerant to trends that were already developing with respect to its use in various domains, including the management of cyberinfrastructures [2]. Machine Learning (ML) specifically has been making steady inroads in various aspects of IT operations [3]. The pattern discovery and pattern recognition capabilities of ML have been applied at first towards issue and threat detection and then towards predictive analytics in support of cybersecurity, performance, and incident management [4]. These natural initial use cases represent a first stage in the evolution towards the broader use of AI tools in cyberinfrastructure operations. In the current phase of AI adoption in IT operations, AI is used to enhance or optimize the resource managers to handle a significantly larger set of scenarios than past resource managers had in scope. Additionally, AI has been applied towards forwarding decision making [5]. These developments, however, should not focus exclusively on commercial infrastructures, they should be used in the context of research cyberinfrastructures as well [6]. Modern cyberinfrastructures must provide on-demand access to computational, data, and other resources that allow scientists to tackle increasingly complex problems and have the flexibility to adapt to changing test requirements. Moreover, the management of these resources should align with the current user expectations related to no-code or low-code provisioning and chatbot assisted operations.

II. No-code Automated provisioning

The BRIDGES (Binding Research Infrastructures for the Deployment of Global Experimental Science) deployed a cross-Atlantic cyberinfrastructure capable of delivery 100Gbps bandwidth, on demand virtual circuits and virtual compute resources on it points of presence [7]. The BRIDGES primary mission is to facilitate scientific collaboration between the United States and European countries however, the test bed is more than a simple connectivity provider, the testbed virtualized all its resources in the context of a complete virtualization model meant to support multitenancy with dedicated, flexible slices providing guaranteed performance to each individual tenant. These characteristics make BRIDGES a perfect candidate for the various applications of AI in IT operations [8]. The generalized virtualization model proposed and implemented by BRIDGES [Programmable Cyber-Infrastructure Architectures] goes beyond the traditional, technology siloed, virtualization concepts, it aims to accommodate the type of resources typically found in research infrastructures: Compute, Network, Storage, Sensors, Instruments such as radio telescopes and more. All these resources are supporting virtual objects, objects that are linked using connectors while virtual circuits can be terminated by virtual objects. The technology stack developed to support the generalized virtualization model is called Shapeshifter and it provides, by default, an easy-to-use web-based interface that fits in the category of no-code provisioning [9]. Shapeshifter uses JSONs for environment definition and management so it can easily receive inputs from an off the shelf Chatbot such as OpenAI's ChatGPT. The chatbot can complement, rather than replace, the Shapeshifter UI. It will support user interactions with the API, allowing users to manage their allocated resource. Users will be able to ask questions about their resources as well as assist in provisioning them. The Shapeshifter API will handle interactions with the chatbot in the architecture depicted below.



The chatbot functionality will be broken into a few basic components: Chat (handling user requests related to troubleshooting and operations), Knowledge (related to contextualized information), and Synchronization (related to current system information). One of the key implementation decisions was to adhere to the use of OpenAPI and Shapeshifter automated update of API documentation. This approach allows the chatbot layer to dynamically adapt to changes made in Shapeshifter. The chatbot will utilize the catalog feature of Shapeshifter where users can persist specific virtual resource structures for future use. The chatbot will understand the individual user's catalog of composite objects as part of the initial synchronization process so users can directly enact or modify their catalog.

For reliability purposes, the chatbot provides a mechanism for validating the provisioned environment against the original request of the user and offers users the ability to make ongoing modifications based on past settings. This critical functionality is implemented with the help of a combination of pre-defined generic test cases that are instantiated to the specific environment that is validated. For example, if a user is setting up a set of compute resources interconnected by links with specific requirements, the system must verify the functionality and performance of the provisioned environment. This reliability test is performed regardless of the chatbot integration. Nevertheless, the chatbot integration provides another unique opportunity in cyberinfrastructure operations, automated testing of provisioned environment. The generation of the testbed configuration JSON is paired with a testbed verification JSON which instantiates pre-defined tests. The predefined tests combine functionality verification, simple connectivity tests between the provisioned testbed nodes and performance tests where IPERF can be leveraged to verify the expected performance of the testbed links. Additionally, the Shapeshifter-Chatbot integration enables users to query the provisioned testbeds for operational information and desired modifications.

III. Conclusions

In this paper we discuss the opportunities presented by the integration of an LLM based chatbot with the orchestrator of the BRIDGES infrastructure, Shapeshifter. Two specific use-cases were discussed: chatbot assisted provisioning and chatbot assisted query and reporting. Going beyond the evident benefit of simplifying the interaction between non-technical users, we see a chatbot integration providing two major benefits: a level of abstraction that makes it easy for the orchestrator to change and evolve without having to worry about the front end, and to provide the tools to automate tests that will ensure the accuracy and reliability of a provisioned testbed. In a fully virtualized environment, users can provision, test, and reprovision testbeds in a sandbox until happy with the outcome before moving the validated JSON to production.

The authors are working on a demonstrator for the concepts discussed. The results of the demonstrator will be presented in a subsequent paper. It is the belief of the authors that chatbots will become a front end for cyberinfrastructure operations and that non-technical users will expect a chatbot interface based on the rapid adoption of chatbots such as ChatGPT.

This material is based upon work supported by the National Science Foundation under Grant No. 2029218, 2029221.

References

[1] H. Naveed, et. al, A Comprehensive Overview of Large Language Models, Accessed on line August 2023, https://arxiv.org/abs/2307.06435

[2] S. Stoykova, N. Shakev, Artificial Intelligence for Management Information Systems: Opportunities, Challenges, and Future Directions. *Algorithms* **2023**, *16*, 357. https://doi.org/10.3390/a16080357

[3] M. Jordan, T. Mitchell, Machine learning: Trends, perspectives, and prospects, Science, Vol. 349, Issue 6245, July 2015, pp 255-260, https://www.science.org/doi/10.1126/science.aaa8415

[4] M. Ahsan, K.E. Gomes, M.M. Chowdry, N. Rifat, J.F. Connolly, "Cybersecurity Threats and Their Mitigation Approaches Using Machine Learning—A Review", *J. Cybersecur. Priv.* **2022**, *2*, 527-555. https://doi.org/10.3390/jcp2030027

[5] H. -N. Quach, C. Choi and K. Kim, "Dynamic Network Provisioning with AI-enabled Path Planning," *2020 21st Asia-Pacific Network Operations and Management Symposium (APNOMS)*, Daegu, Korea (South), 2020, pp. 271-274, doi: 10.23919/APNOMS50412.2020.9236957.

[6] P. A. Buitrago, N. A. Nystrom, "Strengthening the Adoption of AI in Research and Cyberinfrastructure", NSF Smart CI, February 25-27, 2020, Arlington, VA.

[7] B. Jabbari, J. Sobieski and C. Popoviciu, "Programmable Cyber-Infrastructure Architectures for Science Applications," *2021 IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems (COMCAS)*, Tel Aviv, Israel, 2021, pp. 84-88, doi: 10.1109/COMCAS52219.2021.9629107.

[8] B. Jabbari, J. Sobieski, C. Popoviciu, "BRIDGES - Binding Research Infrastructures for the Deployment of Global Experimental Science," April 2020 See https://CNL.gmu.edu/BRIDGES.

[9] Y. Zhaohang, "The Impacts of Low/No-Code Development on Digital Transformation and Software Development", Accessed on lin August 2023 https://arxiv.org/abs/2112.14073