Empowering Privacy-Preserving Research: Introducing the vantage6 Framework and GUI

Djura SMITS [a] Walter BACCINELLI [a]
[a] *Netherlands eScience Center*
ORCiD ID: Djura Smits https://orcid.org/0000-0003-4096-0260, Walter Baccinelli https://orcid.org/0000-0001-8888-4792

Abstract. Federated learning is a technique that enables researchers to combine sensitive data while respecting privacy. vantage6 is a framework for performing this type of privacy preserving analysis. We will showcase vantage6 through the new user-interface, and show available algorithms through the accompanying algorithm store.

Keywords. federated learning, privacy preserving analysis, secure multiparty computation

1. Introduction

Answering questions in healthcare often requires combining data from different sources. However, bringing together this data raises serious privacy concerns. Moreover, centralizing the data in one place is often not legally possible due to privacy regulations.

Techniques for federated analysis aim to enable researchers to do their analysis without giving them access to the raw data. It allows data-owners to stay in control while still enabling researchers to perform their studies.

vantage6[1][2] is an open-source framework for privacy preserving analysis created by the Netherlands Comprehensive Cancer Organisation (IKNL).

Conceptually, vantage6 consists of the following parts:

- A (central) server that coordinates communication with clients and nodes. The server is in charge of processing tasks as well as handling administrative functions such as authentication and authorization.
- One or more node(s) that have access to data and execute algorithms
- Users (i.e. researchers or other applications) that request computations from the nodes via the client
- Organizations that are interested in collaborating. Each user belongs to one of these organizations.
- A Docker registry that functions as a database of algorithms

vantage6 has already been used in various projects [3][4][5].

2. Topic

In this demo we will go into the process of creating and running federated learning tasks in the vantage6 user-interface (UI). The UI is a web application that connects to the vantage6 server and enables researchers to execute federated analysis with little to no coding experience. The researchers can make use of readily available algorithms that can be retrieved from the accompanying *algorithm store*.

The algorithm store is a new addition to the vantage6 ecosystem that allows researchers to share federated learning algorithms. Moreover, this platform will eventually allow peer review of code to guarantee the security of these complex algorithms.

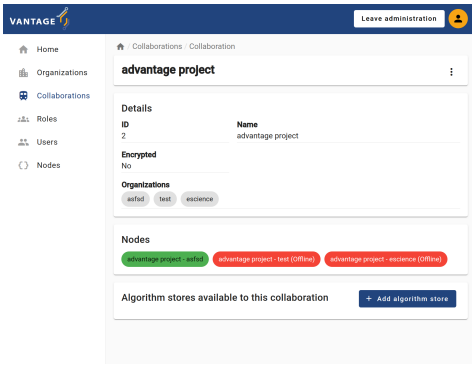

Figure 1. Interface to manage a collaboration

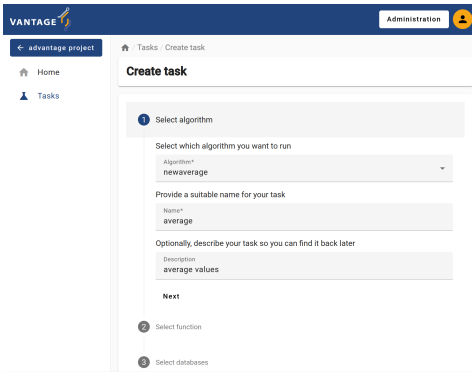

Figure 2. Interface to create a federated analysis task

3. Contents of the demonstration

In this demo we will start with a general introduction of the vantage6 federated learning framework. We will explain how federated learning can help researchers performing analyses on sensitive data while respecting privacy regulations. We will briefly go over the various components and tools that the vantage6 framework consists of.

Next, we will showcase the vantage6 UI. We will explain the concept of vantage6 *nodes* and how the UI gives an overview over which nodes are available. Consequently we will move on to the concept of organizations, that indicates the affiliation of both users and nodes.

After these concepts are clear, we can explain how collaborations link multiple nodes and users together (figure 1) so that federated analyses can be performed. We will show how to create a new collaboration within the UI.

Further on, we will demonstrate how to create an analysis task using our newly created collaboration (figure 2), and show how we can select available algorithm from the vantage6 *algorithm store*.

Lastly, we will showcase the algorithm store and how this encourages reusability and transparency of privacy preserving algorithms.

4. About the presenters

Djura Smits,Msc. is a research software engineer at the Netherlands eScience Center. She has a focus on machine learning and privacy preserving analysis. She contributed to the development of the vantage6 infrastructure [2], as well as in the implementation of federated learning algorithms.

Walter Baccinelli, PhD is research software engineer at the Netherlands eScience Center. He focuses on digital health technologies and biodata analysis.

5. Acknowledgements

This work was supported by the Netherlands eScience Center under the NWO joint call COMMIT2DATA: Big Data & Health

References

[1] Moncada-Torres A, Martin F, Sieswerda M, Van Soest J, Geleijnse G. VANTAGE6: an open source priVAcy preserviNg federaTed leArninG infrastructurE for Secure Insight eXchange. AMIA Annual Symposium Proceedings. 2021 Jan;2020:870-7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8075508/.

[2] Smits D, Van Beusekom B, Martin F, Veen L, Geleijnse G, Moncada-Torres A. An Improved Infrastructure for Privacy-Preserving Analysis of Patient Data. Studies in Health Technology and Informatics. 2022 Jun;295:144-7.

[3] Choudhury A, Theophanous S, Lønne PI, Samuel R, Guren MG, Berbee M, et al. Predicting outcomes in anal cancer patients using multi-centre data and distributed learning – A proof-of-concept study. Radiotherapy and Oncology. 2021 Jun;159:183-9. Available from: https://www.sciencedirect.com/science/article/pii/S0167814021061399.

[4] Scheenstra B, Bruninx A, van Daalen F, Stahl N, Latuapon E, Imkamp M, et al. Digital Health Solutions to Reduce the Burden of Atherosclerotic Cardiovascular Disease Proposed by the CARRIER Consortium. JMIR Cardio. 2022 Oct;6(2):e37437. Company: JMIR Cardio Distributor: JMIR Cardio Institution: JMIR Cardio Label: JMIR Cardio Publisher: JMIR Publications Inc., Toronto, Canada. Available from: https://cardio.jmir.org/2022/2/e37437.

[5] Geleijnse G, Chiang RCJ, Sieswerda M, Schuurman M, Lee KC, van Soest J, et al. Prognostic factors analysis for oral cavity cancer survival in the Netherlands and Taiwan using a privacy-preserving federated infrastructure. Scientific Reports. 2020 Nov;10(1):20526. Number: 1 Publisher: Nature Publishing Group. Available from: https://www.nature.com/articles/s41598-020-77476-2.