Building the RNAMake Gateway on PATh: a Student-Led Design Project

Dinuka De Silva
Indiana University Bloomington
Bloomington, IN, USA
dinsdesi@iu.edu

Derek Weitzel
University of Nebraska-Lincoln
Lincoln, NE, USA
dweitzel@unl.edu

Joseph Yesselman
University of Nebraska-Lincoln
Lincoln, NE, USA
jyesselm@unl.edu

Marcus Christie
Indiana University Bloomington
Bloomington, IN, USA
machrist@iu.edu

Robert Quick
Indiana University Bloomington
Bloomington, IN, USA
rquick@iu.edu

Eroma Abeysinghe
Indiana University Bloomington
Bloomington, IN, USA
eabeysin@iu.edu

Marlon Pierce
Indiana University Bloomington
Bloomington, IN, USA
marpierc@iu.edu

ABSTRACT

We summarize student-led work to build a science gateway for RNAMake, which is software for modeling the three-dimensional structure of RNA molecules. The gateway uses Apache Airavata, which has been extended to support HTCondor submissions. The students also extended the Airavata Django Portal to provide customized user interfaces. In the process, the students learned open source software and open governance practices.

CCS CONCEPTS

• **Applied computing** → **Bioinformatics**; • **Human-centered computing** → **Open source software**.

KEYWORDS

Science Gateways, RNA Nanotechnology, Bioinformatics, High Throughput Computing, Cyberinfrastructure, Distributed Systems, Open Source Software

ACM Reference Format:
Dinuka De Silva, Derek Weitzel, Joseph Yesselman, Marcus Christie, Robert Quick, Eroma Abeysinghe, and Marlon Pierce. 2022. Building the RNAMake Gateway on PATh: a Student-Led Design Project. In *Practice and Experience in Advanced Research Computing (PEARC '22), July 10–14, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3491418.3535184

PEARC '22, July 10–14, 2022, Boston, MA, USA

ACM ISBN 978-1-4503-9161-0/22/07.
https://doi.org/10.1145/3491418.3535184

1 RNAMAKE, HIGH THROUGHPUT COMPUTING, AND SCIENCE GATEWAYS

RNA nanotechnology has the potential to create custom nanomachines to revolutionize the engineering of materials, catalysts, biomedical diagnostics, and therapeutics. One of the main things people want to improve for RNA vaccines is stability, both of the vaccine shelf life and of the vaccine inside the cell. A popular research avenue is to chemically modify the RNA, which is being done extensively, but this goal could also potentially be achieved by adding more three dimensional (3D) structure, as RNAs with strong 3D structure are generally more stable. RNAMake is a suite of tools developed by the Yesselman Lab [4] for designing and optimizing RNA 3D structures, allowing researchers to explore structures that would address the stability problem.

RNAMake is a good fit for a High Throughput Computing (HTC) environment. The inputs and outputs are comparatively small, and the memory utilization is less than 2GB. Also, the jobs can be split into many small pieces and recombined post processing. Jobs usually take hours to execute, and so do not need interactive user interfaces.

RNAMake is thus a good candidate for a software as a service (SaaS) science gateway, in which the software can be deployed on cyberinfrastructure such as the Open Science Grid/PATh, which supports HTC submissions. End users can then access the deployed software through a web-based front end that simplifies the job submission process (users do not need to understand complex command line tools or the complexities of batch managers) and also integrates visualization or analysis tools to create a research environment.

2 BUILDING THE RNAMAKE GATEWAY

We implemented the RNAMake gateway using the open source Apache Airavata [2] framework and its Science Gateways Platform as a service (SciGaP) deployment at Indiana University. Apache Airavata middleware is well-suited for SaaS gateways that interact with schedulers. Apache Airavata's job execution management

system is extensible but did not include support for HTCondor [3], so making this new extension and contributing it back to the open source code base was a development opportunity for the students.

The Airavata Django Portal [1] provides a turnkey web front end to the middleware in order to provide a configurable basic gateway with no additional coding. The Airavata Django Portal can also be extended at various levels of sophistication, ranging from look-and-feel changes using the embedded Wagtail Content Management System to low-level JavaScript and Python API changes, allowing gateway developers to create end user environments that are tailored to the scientific user communities who will use the gateway. Examples include integrating third-party plotting and analysis tools for the generated outputs from the simulation.

3 CONCLUSIONS AND FUTURE WORK

Integration of HTCondor into the Apache Airavata software stack will enable other PATh projects and the larger HTCondor community to develop science gateways with the Apache Airavata software stack. The projects gave both of the student groups experience working with open source software developers and the Apache Software Foundation governance principles. Students learned to work and contribute to existing code bases and interact with open source communities, increasing the likelihood that their efforts fit within larger project goals and will be sustained beyond the specific student capstone projects.

ACKNOWLEDGMENTS

This work is supported by NSF under Cooperative Agreement OAC-2030508 as part of the PATh Project and OAC-1548562 for XSEDE 2.0. Apache Airavata was developed through NSF Awards 1339774, 1827641, and 1840003. This work was completed utilizing the Holland Computing Center of the University of Nebraska, which receives support from the Nebraska Research Initiative.

REFERENCES

[1] Marcus Christie, Suresh Marru, Eroma Abeysinghe, Dimuthu Upeksha, Sudhakar Pamidighantam, Stephen Paul Adithela, Eldho Mathulla, Aarushi Bisht, Shivam Rastogi, and Marlon Pierce. 2020. An extensible Django-based web portal for Apache Airavata. In *Practice and Experience in Advanced Research Computing*. 160–167.

[2] Suresh Marru, Lahiru Gunathilake, Chathura Herath, Patanachai Tangchaisin, Marlon Pierce, Chris Mattmann, Raminder Singh, Thilina Gunarathne, Eran Chinthaka, Ross Gardler, et al. 2011. Apache Airavata: a framework for distributed applications and computational workflows. In *Proceedings of the 2011 ACM workshop on Gateway computing environments*. 21–28.

[3] Douglas Thain, Todd Tannenbaum, and Miron Livny. 2005. Distributed computing in practice: the Condor experience. *Concurrency and computation: practice and experience* 17, 2-4 (2005), 323–356.

[4] Joseph D Yesselman, Daniel Eiler, Erik D Carlson, Michael R Gotrik, Anne E d'Aquino, Alexandra N Ooms, Wipapat Kladwang, Paul D Carlson, Xuesong Shi, David A Costantino, et al. 2019. Computational design of three-dimensional RNA structure and function. *Nature nanotechnology* 14, 9 (2019), 866–873.