The Computerworld Honors Program

Honoring those who use Information Technology to benefit society

Final Copy of Case Study

Status:
Laureate

Year:
2013

Organization Name:
Morgridge Institute for Research

Organization URL:
http://discovery.wisc.edu/morgridge/

Project Name:
Gene Expression Analysis for Stem Cell Differentiation Using Utility Supercomputing

Please select the category in which you are submitting your entry
Emerging Technology

Please provide an overview of the nominated project. Describe the problem it was intended to solve, the technology or approach used, how it was innovative and any technical or other challenges that had to be overcome for successful implementation and adoption. (In 300 words or less.)

Stem cell researchers have been working toward creating therapies for many diseases, but they do so without straightforward or universal guidelines. Victor Ruotti, a computational biologist at The Morgridge Institute for Research, entered Cycle Computing's Big Science Challenge with the hopes of creating an indexing system of the stem cells to allow researchers to quickly classify cells based on their expression pattern and identify genes and regions of the genome, which are critical for establishing and maintaining cell states that have potential for clinical applications. Ruotti, the winner of Cycle's challenge, harnessed the power of high-performance utility supercomputing to run over one million hours of compute a compute century -- in just one week. By using Cycle Computing's CycleCloud utility supercomputing software and infrastructure from Amazon Web Services, Ruotti ran 115 years of computation, to garner the components needed to build an index to help identify cells in a laboratory setting, in 1 week for only $19,555. Utility supercomputing as an affordable research tool for scientists has transformed the class of problems that are solvable, enabling larger breakthroughs not

previously thought possible. The results of this run have wide impacts in improving the speed with which disease cells can be replicated in the petri dish, and opens the door for a wide variety of treatment developments in many diseases.

When was this project implemented or last updated? (Please specify month and year.) Has it incorporated new technologies and/or other innovations since its initial deployment? (In 300 words or less.)

Ruotti's run was completed in September of 2012, and the completion of the index compiled from computations is ongoing. The core technologies used include genomic sequencing, utility supercomputing, and cloud computing.

Is implementation of the project complete? If no, please describe the project's phases and which phase the project is now in. (In 300 words or less.)

Phase 1 of Victor Ruotti's research is complete (see Appendix 1). Subsequent phases involve taking the results from the 115 years of compute and compiling a knowledge base that other researchers can use to identify the state of induced pluripotent stem cells in the petri dish. This will help researchers in being able to quickly identify whether they have successfully replicated disease cells in a way that allows testing prospective treatments without impacting patients until a viable one is found.

Please provide at least one example of how the technology project has benefited a specific individual or organization. Feel free to include personal quotes from individuals who have directly benefited from the work. (In 300 words or less.)

This project will benefit society as a whole, and specifically people who are afflicted with diseases that stem cell therapy may be able to treat in the future. By creating a knowledge base that ties expressed genes to different states of induced pluripotent stem cells, Ruotti's research will make it easier for scientists to replicate disease in the petri dish. Because researchers will be able to test potential new treatments against the petri dish and not potential patients, this will overall lead to a much faster, streamlined mechanism to develop treatments for diseases that currently are difficult to treat. The project is ongoing, and once the index is built, it has potential to expedite stem cell research significantly, leading to various clinical application of stem cell therapies.

Would this project be considered an innovation, a best practice or other notable advancement that could be adopted by or tailored for other organizations and uses? If yes, please describe that here. (In 300 words or less.)

Yes this project's end result, the stem cell index, is a notable advancement that was meant to be adopted by stem cell researchers in order to facilitate research leading to disease therapy. In addition, using utility supercomputing software to compute massive amounts of data quickly in the cloud to further research is a best practice that has become increasingly affordable and accessible to scientists and researchers. We have many examples of the vast quantity of genomic data that today's sequencers generate now being able to be analyzed in totality. Mr. Ruotti's use case is important because it highlights the ability to inexpensively analyze big data, in this case 78 terabytes of

genome data, with big compute totally 115 years of computation. Hopefully examples like this stem cell research will only increate the pace of innovation as researchers like Ruotti strive to make stem cells a widely usable tool for developing treatments for disease.

If there are any other details that the judges should know about this project, please note them here. (In 300 words or less.)

The compute workload Victor was suggesting was the largest RNA Sequencing run done, especially in the cloud, up until that time. He took the 2012 CycleCloud BigScience Challenge to heart, and asked the right question to help with the use of pluripotent stem cells, even though the amount of computing seems impossibly large. This computationally liberated thinking will likely mean a much more streamlined use of stem cells in replicating disease, and also serve as an example for other genomic researchers because of its use of over a century of computing (1,003,404 hours to be precise) against 78 terabytes of data, that completed in a week for $19,555.