



The Computerworld Honors Program

Honoring those who use Information Technology to benefit society

Final Copy of Case Study

Status:

Laureate

Year:

2013

Organization Name:

Frederick National Laboratory for Cancer Research

Organization URL:

<http://ncifrederick.cancer.gov/>

Project Name:

Biomedical Big Data Proof-of-concept

Please select the category in which you are submitting your entry:

Innovation

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.)

Oracle and the Frederick National Laboratory for Cancer Research (operated by SAIC on behalf of the National Cancer Institute) used Apache Hadoop, Hive, and Mahout to cross-reference the relationships between 17,000 genes and five major cancer types across 20 million medical publication articles. This Big Data project also included cross-referencing genes from 60 Million patients and miRNA for a simulated 900 Million patient population. The open source technologies enabled analysis of information at a scale, speed, and cost that could not have been previously accomplished.



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.)

This project was implemented in the summer of 2012, and last updated at the end of August 2012.

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.)

Oracle and the Frederick National Laboratory for Cancer Research have a three-year agreement to pursue research in this area. Phase 2 is expected to commence this summer.

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.)

Understanding additional layers of the pathways these genes operate in and the drugs that target them is expected to help cancer researchers in their work.

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.)

This project should be considered an innovation in applying open source technologies (Apache Hadoop, Hive, and Mahout) that was originally developed for web analytics and applying that innovation to analyzing medical information.