Calvin Li

Professor Stacey Suver

CSE300

Due: 30 March 2018

The Impact of Cloud Computing on Cancer Research

INTRODUCTION

As the Internet boom arrived during the early 2000s, tech companies followed suit. This marked the era of new applications, ones that were interactive and available on the web, and infrastructure needs shifted. Cloud computing was the promised solution; quick and easy platforms or other online computing services became widespread, starting with the launch of Amazon Web Services in 2006. However, as related industries began to recognize the usefulness and scalability of cloud services, they began to expand into this new field as well; genomics, a notoriously computationally-heavy field, has not been an exception, and cancer researchers have recently begun employing these services. This review seeks to discuss how cloud computing has affected the state of cancer research.

REVIEW

Data Services

For cancer researchers and clinicians, major concerns include availability of data, ease of use, and decentralization; that is, to have patient information abstracted enough so as to be anonymous, while retaining all metadata that may be relevant to findings. Melanoma is one area

where data concerns prevail- while information is increasingly available, the standards are too loose to gather proper study data. MelanomaDB, a publicly available database that serves a web application platform, is a shining example of a cloud-based service that offers a unique, powerful way to approach cancer research. By providing both an application interface as well as integration techniques, Trevarton's team have shown that there is practicality in database applications, a popular subset of cloud services, to serve researcher needs.

Another approach to servicing research needs through data is detailed in "A Network Approach for Managing and Processing Big Cancer Data in Clouds." Combining cloud computing with another hot field known as "Big Data", Xing et al. have developed a Cancer Data Network (CDN) that manages biological and clinical data, as well as patient care information. This architecture claims to reinvent the wheel as far as computing resources for cancer researchers go; it promises to offer a dynamic ecosystem for researchers to use, that organically provides new and relevant data while discovering relationships between data points. All of this is supported by the CELAR cloud platform, with initial tests showing positive results. These are quite obviously bold claims, but the technology design certainly supports the aforementioned goals. Future goals include adding more algorithms to better expand the capabilities of data aggregation in the platform, as well as evaluating further usage of the platform itself.

Cancer research is done in many ways, and one main source of understanding the mechanics behind cancer is known as cancer genomics. The human genome is defined as "the complete set of genes or genetic material present in a cell or organism", and by studying the genome of people with cancer, geneticists can understand the sequence changes that affect the

development of cancer. As sequencing technologies have become more powerful, clinical databases have grown immensely, spurring the need for tools to be developed such that non-technical specialists can still analyze the content. However, because of the sheer size of data that needs to be stored (upwards of millions of terabytes), it is impractical for smaller research teams to participate in discovery, as it would require major funding in high performance computing. A solution by Paul Walsh proposes a new platform, called “Simplicity”, leverages the high performance computing capability of cloud-based platforms. Because its infrastructure is based in Microsoft Azure and Amazon EC2 (two leading cloud providers), smaller teams can scale their needs on a per-use-basis. Not only does the platform remove major overhead and upfront costs in terms of IT infrastructure, it can integrate large amounts of data and provide visualizations through the platform’s graphical user interface. Internal research conducted on users shows that the multitude of visualization services were well-used, and that the simple interface results in high usability, even for larger research team needs.

MelanomaDB, the CELAR CDN, and Simplicity all share the common goal of improving cancer research through the power of cloud computing. By alleviating concerns of computing cost, interpretation, filtering, and aggregating cost, as well as providing simple and clean user interfaces, the cloud platforms seem to hold much potential. Their efficiency can not yet be assessed confidently, as these technologies themselves are very recent. However, internal surveying and test sampling on actual researchers have shown positive results, showing that cloud computing will likely continue to pervade the field of cancer research.

Legal and Ethical Issues

Since the introduction of cloud computing into the field of biomedical research, genomic and clinical data have been more available and more widespread through the usage of new platforms and services. While most of the appeal and power of cloud computing extends to public platforms, such as Amazon AWS, Microsoft Azure, and Google Cloud Platform, there are fundamental security risks when hosting data off-site. Furthermore, they extend the possibility of groups of study due to increased selection and cohort size. This increased range of study does come with its own share of issues. Namely, there are regulation and security concerns with using sensitive patient information in a public cloud platform. Molnár-Gábor discusses these concerns as well as alternatives to public clouds, such as federalized or hybrid (a mix of private/public) cloud platforms. Other fields that employ cloud services also take this approach; it is already in practice to employ public cloud computing power for strictly anonymized data, while the computed and processed data is sent to a private network for further analysis.

The culmination of previously discussed platforms, alongside ethical and legal concerns, is further explored by Kathleen Charlebois in "The Adoption of Cloud Computing in the Field of Genomics Research: The Influence of Ethical and Legal Issues." By using the Diffusion of Innovation (DOI) Theory, Charlebois and her team conducted a structured study on genomics researchers and various aspects of cloud computing technology. The results largely point to researchers being careful and slow to adopt cloud tech, with their main concerns being cost, security, and lack of familiarity. Despite the available services and technologies that circumvent issues such as cost and lack of familiarity, as well as major cloud providers having multiple

layers of security in place for their cloud instances, genomics researchers are legally responsible for maintaining their data and thus, the cost and complexity of securing data in the cloud increases. Because of these concerns, most researchers have not adopted cloud technology at the time of the article (10/18/2016), although major/international projects have begun using cloud services. Those who have begun stepping inside the realm of cloud computing are mostly using hybrid services as previously explained- private data storage, and public computation. This is in part due to legal issues with non-local hosting of data, as well as consent laws, needing to "catch up" to the newness of cloud technology.

The future of cloud computing may need to shift towards these hybrid models, or completely novel solutions, as security and privacy concerns continue to be explored. The literature involved is inconclusive so far, which is in large part due to the relative youngness of the field.

CONCLUSION

With the emergence of cloud platforms, computational provisions of countless industries have shifted; genomics among them. The future may be promising, as familiarity and trust with cloud providers is expected to come with time, but the current status of cloud usage in genomics researchers is mainly for the largest scale projects that involve multinational teams. Furthermore, legal jurisdiction and consent laws will take time to provide proper regulation on these new research methods. There is much research and study to be done in this field, and in the years to come, the usage of cloud in genomics should continue to be monitored.

Works Cited

"Amazon Web Services." *Wikipedia*, Wikimedia Foundation, 28 Mar. 2018, en.wikipedia.org/wiki/Amazon_Web_Services.

Trevarton, Alexander Joseph, et al. "MelanomaDB: a Web Tool for Integrative Analysis of Melanoma Genomic Information to Identify Disease-Associated Molecular Pathways." *Frontiers in Oncology*, Frontiers Media S.A., 1 July 2013, doaj.org/article/99ec3686bacd40afbfb6e76661631e3a.

Walsh, Paul, et al. "Visualizing Next-Generation Sequencing Cancer Data Sets with Cloud Computing." *SpringerLink*, Springer, Cham, 7 June 2016, link.springer.com/chapter/10.1007/978-3-319-50070-6_4.

"What Is a Genome? - Genetics Home Reference." *U.S. National Library of Medicine*, National Institutes of Health, ghr.nlm.nih.gov/primer/hgp/genome.

Xing, Wei, et al. "A Network Approach for Managing and Processing Big Cancer Data in Clouds." *SpringerLink*, Springer US, 21 Apr. 2015, link.springer.com/article/10.1007/s10586-015-0456-6.

Molnár-Gábor, Fruzsina. "Computing Patient Data in the Cloud: Practical and Legal Considerations for Genetics and Genomics Research in Europe and Internationally." Genome Medicine, vol. 9, 20 June 2017, pp. 1–12.

Charlebois, Kathleen. "The Adoption of Cloud Computing in the Field of Genomics Research: The Influence of Ethical and Legal Issues." PLoS ONE, vol. 11, no. 10, 18 Oct. 2016, pp. 1–33.