Leveraging New-Age Computing to Advance Life Sciences Research

Abstract

The operational modules of pharmaceutical companies change with time, in line with evolving trends and business requirements. Rapid developments on the technology front have taken place in the last few years. Today, high throughput technologies drive modern drug research and development, generating large volumes of data, and consequently, requiring higher compute, bolstered data management and superior big data analytics. Moreover, adopting next-generation technologies is catalyzing collaborations to create synergies for faster time to market and improved quality and return on investment. This paper discusses the technology options for scientific computing that can be leveraged to streamline drug research, including analytics, complex bio-simulations, and high-performance data analytics.

Need for smarter research

Staying competitive requires pharmaceutical companies to recalibrate strategies to adapt to the changing business environment. As recently as 10-15 years ago, pharmaceutical companies worked in silos, preferred the blockbuster drug model, and ensured profitability through awarded patents which were not as challenging to get. However, with the tightening of regulations and the ever-increasing pressure to perform in a strained economic environment, pharmaceutical firms have had to make their research and development workflows more productive. The industry, therefore, witnessed several acquisitions and mergers, which were the quickest way for companies to grow and replenish the dwindling pipeline. That trend, though, was not sustainable, resulting in the need to innovate and carry out smarter research[1].

The operating models and footprint of pharmaceutical companies have changed over time. In the early 2000s and the few years preceding that, the revenue depended on patents acquired on small molecule drugs. Now, pharmaceutical giants have diversified and are generating parallel revenue from precision medicine products and biologicals[2]. Further, pharmaceutical companies are exploring the possibilities of harnessing the full potential of data within and across organizations, thus, moving away from the old blockbuster model and siloed approach to becoming more data-driven in the future[3]. Artificial intelligence and machine learning techniques are also being used to find inconspicuous patterns and correlations in data.

Conducting research powered by data requires niche skill sets and capabilities. The absence of these has resulted in acquisitions, strategic partnerships, and industry-wide collaboration. Building skill sets and capabilities has prompted pharmaceutical companies to embrace a collaborative approach. Pharmaceutical partnerships are now imperative and range from asset swap deals and combining promising candidates in the pipeline, all the way to co-marketing. Companies are collaborating with academia, promoting joint ownership of laboratories, providing access to industry resources, and facilitating sponsored research. Adopting such approaches would significantly challenge traditional operating models in the trillion-dollar pharmaceuticals industry, which is expected to touch $1,170 billion by 2021, according to a report by market intelligence firm The Business Research Company[4].

Ensuring optimal scientific computing in drug research

Scientists use hundreds of IT applications as well as multiple public and in-house databases. Traditionally, simulations and analytics for drug research have been run on desktop computers, high-performance workstations, and in-premise high-performance computing (HPC) clusters. Exploiting the full potential of an HPC cluster requires all the tools ported on it to run at their optimal efficiency, and achieving that depends on the availability of the appropriate resource requirement. The resource requirement, however, for each tool is very specific, encompassing compute, input/output (I/O), memory, and storage.

Pharmaceutical research requires implementing varied workloads. Sometimes accelerators like graphics processing units (GPUs) and field-programmable gate arrays (FPGAs) are used to improve the performance of a specific class of applications. The accelerators and storage devices are chosen based on the assessment of the workload and application areas.

Hence, research areas (choice of tools and type of workflows) play a significant role in deciding the choice of infrastructure between monolithic and hybrid clusters.

The role of new-age technologies

Conducting research in a smarter way by harnessing the full potential of available research data necessitates running simulations and big data analytics that would require high compute, often exceeding the on-premise cluster capacity. To overcome the compute limitation, the on-premise compute infrastructure can be augmented to address the surge requirement. However, this comes with a cost and will reduce the overall cluster utilization when the surge subsides.

Furthermore, data sharing and results reproducibility are critical requirements for fruitful scientific collaboration.

Researchers on a collaborative project are spread out across the globe. The preferred infrastructure to enable global access involves migrating the IT environment from the local, on-premise cluster to the cloud. Pharmaceutical companies are now adopting cloud technologies to complement the locally maintained traditional HPC clusters. Cloud computing not only addresses computational requirements in surges but also provides a flexible, scalable, secure, and transparent environment for collaborative work. Reproducibility of scientific findings is a primary requirement to be accepted by collaborators and policymakers, as complex biological problems need multiple rounds of data analysis using various tools to derive meaningful analysis.

Scientists routinely use computational workflows, employ data analytics, and share data. More importantly, conducting research in a streamlined and efficient manner makes it necessary for scientists to have gateways and collaborative platforms with global access and scalable computing resources to meet requirements during surges.

Challenges

Compute: The primary objective of cloud adoption in on-premise HPC clusters is not to invest in augmenting them but to manage compute surges. The scalability and elasticity of the cloud make it possible to assess computational needs. Cloud providers offer various services with different prices and computational capabilities. However, finding the right instance for a workload application is critical, as resource provisioning plays a vital role in the cost and performance optimization of scientific workflows. A wrong choice may result in escalated costs and diminish the benefits associated with the 'pay per use model' of the cloud approach.

Compute for varied workloads: Performance on the cloud versus on-premise HPCs is application-specific, and hence, it does not make sense to move all applications and workflows from an in-house HPC to the cloud. Humphrey et al[5] compared various cloud models – an all-cloud design, a version that runs in-house on an HPC cluster, and a hybrid cloud bursting version using both in-house and cloud resources. The application used for the study benefited from the hybrid version, both on execution time and cost. Hence, carrying out a feasibility analysis for each application before moving it to the cloud environment is of paramount importance.

Conclusion

Most pharmaceutical research data analytics include workloads such as bio-simulations, big data computations, and data-intensive AI/ML approaches. Such varied workloads must be executed on appropriate hybrid infrastructure – on-premise CPU/GPU with cloud or a combination of both CPU and GPU with cloud – to deliver optimal performance. The research areas (types of workflows) and the need to collaborate determine the choice of infrastructure to be used to implement the workloads – internal HPC cluster, a private cloud, or a combination of both. Depending on the sensitivity of the work, workloads can be migrated to the public cloud to meet compute requirements during surges through a cloud burst.

Furthermore, one can adopt a hybrid approach with both the in-house HPC cluster and the cloud, as and when required. However, as reproducibility is a significant concern, rapidly deploying solutions and tools to generate a consistent environment through a cloud burst would be a desirable solution. Companies that can overcome organizational inertia and traditional mindsets, and adopt new-age computing technologies stand a better chance at improving drug research and ensuring development efficiency.

References

[1] Pharmafile; The age of collaboration: Why pharma companies now have to work together; December 9, 2015; http://www.pharmafile.com/news/501725/age-collaboration-why-pharma-companies-now-have-work-together; Accessed on December 17, 2018

[2] BiopharmaTrend.com; Top 7 Trends In Pharmaceutical Research In 2018; April 25, 2018; https://www.biopharmatrend.com/post/60-top-7-trends-in-pharmaceutical-research-in-2018-and-beyond/

[3] Elsevier; R&D Solutions for Pharma & Life Sciences; From molecule to phenotype: Predictive modeling in Data-Driven Drug Discovery; https://www.elsevier.com/rd-solutions/industry-insights/pharma-and-life-sciences/predictive-modeling-in-ddd

[4] The Business Research Company; The Growing Pharmaceuticals Market: Expert Forecasts and Analysis; May 16, 2018; https://blog.marketresearch.com/the-growing-pharmaceuticals-market-expert-forecasts-and-analysis#targetText=The%20global%20pharmaceuticals%20market%20was,by%20The%20Business%20Research%20Company

[5] Proceedings - 2011 7th IEEE International Conference on eScience; Assessing the Value of Cloudbursting: A Case Study of Satellite Image Processing on Windows Azure; 5-8, December, 2011; https://dl.acm.org/citation.cfm?id=2116576

About The Author

Arundhati Saraph

Arundhati Saraph, who has over 23 years of experience in biomedical and life sciences research, specializes in discovery research. She holds a PhD in biochemistry from the National Chemical Laboratory, Pune. She has worked as a postdoctoral fellow at Rensselaer Polytechnic Institute, New York, USA, where she oversaw the design and validation of peptide-based therapeutics. She has also led research projects, collaborations, and partnerships with industry and research organizations that have resulted in several publications and the development of solutions for biomedical research and in the pharmaceuticals sector.

Contact

Visit the Life Sciences & Healthcare page on www.tcs.com

Email: ADD.Platforms@tcs.com

Subscribe to TCS White Papers

TCS.com RSS: http://www.tcs.com/rss_feeds/Pages/feed.aspx?f=w
Feedburner: http://feeds2.feedburner.com/tcswhitepapers

About Tata Consultancy Services Ltd (TCS)

Tata Consultancy Services is an IT services, consulting and business solutions organization that delivers real results to global business, ensuring a level of certainty no other firm can match. TCS offers a consulting-led, integrated portfolio of IT and IT-enabled, infrastructure, engineering and assurance services. This is delivered through its unique Global Network Delivery Model™, recognized as the benchmark of excellence in software development. A part of the Tata Group, India's largest industrial conglomerate, TCS has a global footprint and is listed on the National Stock Exchange and Bombay Stock Exchange in India.

For more information, visit us at www.tcs.com

Experience certainty.

IT Services
Business Solutions
Consulting

TCS Design Services I M I 09 I 19