An Analysis of HPC, its Accessibility, and its Impact in Medical Data Processing

Audrey Smith

Assignment 1

High Performance Computing

Canterbury Christ Church University

Table of Contents

[Definition and Impact of HPC in Medical Data Processing 2](#_Toc87856096)

[A Brief History of HPC 2](#_Toc87856097)

[The Impact of HPC in Medicine 2](#_Toc87856098)

[Cloud Computing vs. In-House HPC Computing 4](#_Toc87856099)

[Cloud Computing Environment 4](#_Toc87856100)

[In-House HPC Environment 5](#_Toc87856101)

[Results of Comparison 6](#_Toc87856102)

[The Technology Better Suited for Innovation in Medicine 7](#_Toc87856103)

[References 9](#_Toc87856104)

# Definition and Impact of HPC in Medical Data Processing

High Performance Computing is defined as using the combined power of multiple computer nodes to perform tasks at a much higher load than a normal desktop PC can handle (University of Greenwich, n.d.). These tasks range from data analysis, large calculations, and data processing to weather forecasting, terrain modelling, and 3D design. These powerful clusters are not limited in availability to just global corporations, but can be used by small businesses, governments, and even individuals who may access them via cloud services.

## A Brief History of HPC

The first high performance computers were developed in the 1960s for the purpose of performing float point operations. It was not until the 1980s that these computers became more widely available and less expensive to produce, as processors began selling in large quantities, allowing for smaller corporations to enter the world of multi-processor computers. By the late 1990s, these computers had evolved from single units with multiple processors into clusters of x86 computer nodes and had larger amounts of memory and storage as well as higher speeds. In the early 2000s, clusters were dominating the global charts for top HPC systems. The desire for higher performance brought the increase in nodes per cluster as well as in performance of processors. This was met with limitations in memory speed, parallelism, and a power wall. The introduction of GPUs into clusters assisted in navigating these obstacles by taking the load off the processors. Today, HPC systems can perform billions of floating point operations per second and are relatively affordable, depending on the purpose and technology used (Inside HPC, 2016).

## The Impact of HPC in Medicine

In recent years, the medical field has experienced a shift into the use of HPC in data processing. Experts are making use of every resource available to push medicine further and more quickly, and one of the ways of doing this is through the analysis of patient data. This data can come from many sources – it can be reported by hospitals and medical professionals or collected through surveys of the public, wearable biometrics, and through patients discussing their health with their providers. Craven (2009) explains that the analysis of medical data can be advanced by HPC in identifying health and symptom trajectories for patients, which would help medical professionals detect risks such as a patient with desire to stop treatment or a symptom the patient may experience in the future. The phrasing of this data is not always the same, however, especially when provided by public surveys, which is a major limitation that will need to be addressed before the medical field can make better use of what HPC has to offer. This is just one example of a limitation of medical practice that has been identified by the introduction of HPC into the field, one that will eventually lead to the betterment and advancement of medicine and data processing.

The ever-widening availability of high-performance computers to individuals has inspired start-ups and smaller companies to create new technologies that will assist in resolving these limitations. The European Commission (2017) presents a recent example in which a footwear device has been introduced that uses real-time data analysis to detect signs of freezing attacks in patients suffering with Parkinson’s disease. This is a major success for the company that created the device.

# Cloud Computing vs. In-House HPC Computing

When planning for a new project or research area, it is important to ask partners a number of criteria before beginning work, these being what goals everyone has for the project as well as what these goals will cost in terms of time, resources, and technology. The question will arise of how any data used should be processed and stored, which leads to the discussion of cloud computing environments vs. in-house HPC systems. The options appear differently for every scenario, however at baseline, the consumer wants to use these environments to perform large-scale computations, process data, and store everything in a secure area. Some consumers prioritise data protection, where others will prioritise elasticity or accessibility. Understanding the difference in these technologies is crucial to the decision-making in the early days of projects, partly due to the obstacles consumers will be met with if they decide to switch from one to the other further down the line.

## Cloud Computing Environment

Defined by the US National Institute of Standards and Technology (2011), cloud computing is the delivery of on-demand resources such as networks, servers, storage, and security, to consumers via their network connection. This allows for the use of these resources without the physical space allocation that they would typically require. Cloud services are offered by numerous providers, including leading corporations Amazon, Microsoft, and Google, and come in many varying combinations of the available services for the purpose of targeting specific needs of the consumer. The consumer’s data and applications are stored and maintained not on their local devices, but on servers that can be found in data centres and server farms worldwide (Carter, A., 2019).

There are a few models of cloud computing that can be used when provisioning the service from a provider, including Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), and Recovery as a Service (RaaS). These models are intended for, respectively, consumers using virtualized computing resources from the provider, consumers designing and developing their own applications to be run on the provider’s platform, consumers accessing the provider’s web-based applications via a browser, and consumers using the provider’s platform for backup, archives, and recovery. The table below denotes the pros and cons of using a cloud computing environment.

|  |  |
| --- | --- |
| Pros | Cons |
| * Accessibility: Resources can be accessed at any time and location without the need for human interaction (NIST, 2011) * Cost: Pay for only the exact services needed with no direct investment in hardware (Chaturvedi, A., Rashid, A., 2019) * Elasticity: Real-time adjustments can be made in all models, as if the resources are infinitely available (Chaturvedi, A., Rashid, A., 2019) * Scalability: Cloud service providers have the freedom to change the number of nodes/servers available to consumers (Chaturvedi, A., Rashid, A., 2019) | * Security: Risk of data leakage that could give attackers access to encryption keys (Srinivasan, S., 2013) * Reliability: Service providers can experience outages, causing concern for consumers relying on the cloud for important data (Srinivasan, S., 2013) * Privacy: Consumers do not have total control over the data in the cloud, due to the infrastructure being maintained by a service provider. This can cause concern for the confidentiality and privacy of data (Chaturvedi, A., Rashid, A., 2019) |

Being an easily accessible and manageable service, cloud computing tends to be a highly appealing option for most businesses, especially considering the ability for consumers to choose exactly what services they need from providers and pay for only those products for the time they are in use. Unfortunately, with every advantage comes a drawback, and the security and privacy of data are a major priority for most consumers, so it may be deal-breaking for many. On the other hand, however, there is always potential for growth in the future that could mitigate these risks and make cloud computing even more efficient and worth consideration than it already is.

## In-House HPC Environment

Many consumers prefer to keep their computations and data secure and readily available by creating an HPC cluster environment in-house without the help of cloud services. An HPC cluster is a group of servers (nodes) connected into a network together via a fast interconnect. All nodes contain what would normally be found in a desktop PC: CPU cores, memory, and disk space, but they can have varying specialisations. Within a cluster there is a login node, which is the main point for a user to access the system, followed by a node that focuses on the transfer of data between nodes and external computers. The rest of the nodes are compute nodes, which perform all computations, however some may contain a GPU or extra memory (Iowa State University, n.d.). Alongside the nodes, there may also be a storage tower connected to the cluster network. Companies that use in-house HPC systems will have these components kept together in a secure facility that can be accessed in person by whoever is managing the system. The table below denotes the pros and cons of using an in-house HPC environment.

|  |  |
| --- | --- |
| Pros | Cons |
| * Control: The company has full control over every aspect of the system and can change anything at any time (Farrow, C. 2018) * Security: Low risk for potential attacks due to data being stored locally and in a secure facility (Farrow, C. 2018) | * Cost: Can cost thousands up front due to required upfront investment in hardware * Elasticity: Does not match the near-infinite storage abilities that cloud computing can provide due to limitations in cost, hardware, and space (Downing, C. 2018) |

While there will be many resources put into the creation of an in-house HPC environment, there are a few advantages this technology poses. For a consumer who prioritises data security, knowing exactly where all components are located and that they are safe and secure will be very appealing. Likewise, for a consumer who prioritises control and accessibility, they will be delighted to know that this path for the company will allow full control over every aspect of the environment. As with cloud computing, there are many advantages as there are drawbacks to this technology, making it worth consideration as well.

## Results of Comparison

Both technologies are useful and advantageous in their own ways and will be appealing to consumers for many different reasons. A cloud computing environment may be more suitable for those who need specific resources temporarily or require access to their systems from any location at any time, whereas an in-house HPC environment will be more ideal for consumers who are interested in knowing exactly where their data is stored and who has access to the hardware. The best way to go about making this decision is to see where the company’s priorities lie and to move forward from there.

# The Technology Better Suited for Innovation in Medicine

In recent years, the medical field has seen a substantial change in the way patient data and documents are stored, accessed, and processed due to widely available technology that allowed this data to be migrated into digital form. This implementation allowed for better organisation and collection of patient data, but it also opened professionals’ eyes to just how much data truly exists for not only patient data, but for medicine in general – The Global Innovation Index (2019) investigated this issue and found that every year, hospitals produce 50 petabytes of data, which is made up of clinical notes, lab tests, medical images, sensor readings, genomics, and operational and financial data, much of which is currently not in use due to lack of resources for processing. These vast amounts of information need to be sorted, stored, amended, and analysed repeatedly, which at this scale requires specific computational and algorithmic resources that can be very difficult and expensive to accumulate. The introduction of HPC into medicine more recently has allowed some solutions to be found regarding this issue, allowing those large amounts of data to be used in ways they could not be previously, and there has been a rise in smaller companies putting forward new innovations that can assist in adapting to this shift.

An example of this is the amount of rejected images that come from X-ray imaging, which is close to 25%. The University of Washington Medical Center assisted in piloting an application that assists clinicians in finding the reason for these rejected images while also automating a process that took nearly seven hours of work (Global Innovation Index, 2019). With this innovation, clinicians can now address the 25% of unusable imaging, understand where the problem is, and move towards reducing that rate.

With the provided analysis of a cloud computing environment versus an in-house HPC environment, it is important to consider all aspects of the problem before choosing between the two technologies. Manogaran, G. et al. (2018) introduced the Internet of Medical Things (IoMT), a global infrastructure that creates a link between medical devices and medical professionals’ workspaces. Patient data is taken from the medical devices once readings have been conducted and then transferred via remote cloud data centres to clinical databases. It is used for observing patient health, with the main priority typically being disease diagnosis and the prevention of future health-related issues. The lack of used data means there is space for new innovations that will create new ways to analyse and use incoming data. With this information and thinking about the future of this company, the best choice of technology will be a cloud computing environment.

All other justification aside, the main thing to consider is the potential growth that could occur once the company becomes well known in the medical field, which could bring an influx of sample data to be analysed. This will require an environment that can be stretched at short notice, which would not be feasible in an in-house HPC environment due to the potential lack of space and funding to expand. In a cloud computing environment, elasticity and scalability are two key features that will ensure all data is manageable and that will keep costs down, making it the more viable choice between the two technologies for this company.

# References

Amazon. n.d. *Getting Started with AWS.* Available at: <https://aws.amazon.com/getting-started/?nc2=h_ql_le> (Accessed 9 November 2021).

Carter, A. (2019). *Where is Cloud Data Physically Located?* Available at: <https://virtusdatacentres.com/item/381-where-is-cloud-data-physically-located> (Accessed 9 November 2021).

Chaturvedi, A., Rashid, A. (2019). *Cloud Computing Characteristics and Services: A Brief Review.* Available at: <https://www.researchgate.net/profile/Aaqib-Rashid/publication/331731714_Cloud_Computing_Characteristics_and_Services_A_Brief_Review/links/5c89f6c045851564fadca23f/Cloud-Computing-Characteristics-and-Services-A-Brief-Review.pdf> (Accessed 9 November 2021).

Craven, C. (2019). *High-Performance Computing Use Case: Medical Research.* Available at: <https://www.sdxcentral.com/data-center/high-performance-computing/definitions/hpc-use-case-medical-research/> (Accessed 5 November 2021).

Downing, C. (2018) *How the Cloud is Falling Short for HPC.* Available at: <https://www.hpcwire.com/2018/03/15/how-the-cloud-is-falling-short-for-research-computing/> (Accessed 11 November 2021).

Dube, N., Faraboschi, P., Milojicic, D., Roweth, D. (2021). *Future of HPC: The Internet of Workflows.* Available at: <https://ieeexplore.ieee.org/abstract/document/9522103?casa_token=YmaI1FjgswgAAAAA:jOPPSOHCmih6JrJZ_Bg9QdIPX-umgYh10Tx-_zRteeW4GJhxSiCUDl0oYKG63b4u___Csp3fx-s> (Accessed 11 November 2021).

Farrow, C. (2018) *The Cloud Vs Onsite Servers.* Available at: <https://www.singularitee.co.uk/2018/10/benefits-of-the-cloud-vs-onsite-servers/> (Accessed 11 November 2021).

Global Innovation Index (2019). *Four Ways Data is Improving Healthcare*. Available at: <https://www.globalinnovationindex.org/gii-blog/2019/four-ways-data-is-improving-healthcare--b195> (Accessed 12 November 2021).

Grance, T., Mell, P. (2011) *The NIST Definition of Cloud Computing.* Available at: <https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-145.pdf> (Accessed 9 November 2021).

Inside HPC (2016). *The Evolution of HPC.* Available at: <https://insidehpc.com/2016/08/the-evolution-of-hpc/> (Accessed 4 November 2021).

Iowa State University. n.d. *What is an HPC cluster.* Available at: <https://www.hpc.iastate.edu/guides/introduction-to-hpc-clusters/what-is-an-hpc-cluster#:~:text=An%20HPC%20cluster%20is%20a,node%2C%20where%20users%20log%20in> (Accessed 11 November 2021).

Manogaran, G., Chilamkurti, N., Hsu, CH. (2018). *Emerging Trends, Issues, and Challenges in Internet of Medical Things and Wireless Networks.* Available at: <https://doi.org/10.1007/s00779-018-1178-6> (Accessed 12 November 2021).

Microsoft. n.d. *What is Azure?* Available at: <https://azure.microsoft.com/en-gb/overview/what-is-azure/> (Accessed 9 November 2021).

Srinivasan, S. (2013). *Is Security Realistic in Cloud Computing?* Available at: <https://scholarworks.lib.csusb.edu/jitim/vol22/iss4/3> (Accessed 9 November 2021).

University of Greenwich. n.d. *What is HPC?* Available at: [https://www.gre.ac.uk/it-and-library/computing/hpc/what-is-hpc#](https://www.gre.ac.uk/it-and-library/computing/hpc/what-is-hpc) (Accessed 4 November 2021).