Final Report

# Research Question 1 approach for obtain results:

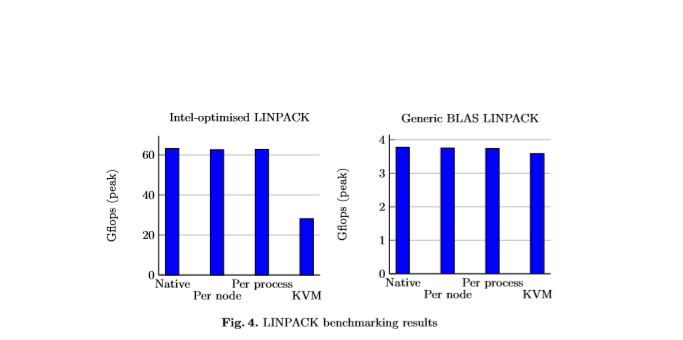
Since all papers related to HPC always benchmark their systems in the results we will demonstrate similar results, and we are offering a PaaS.

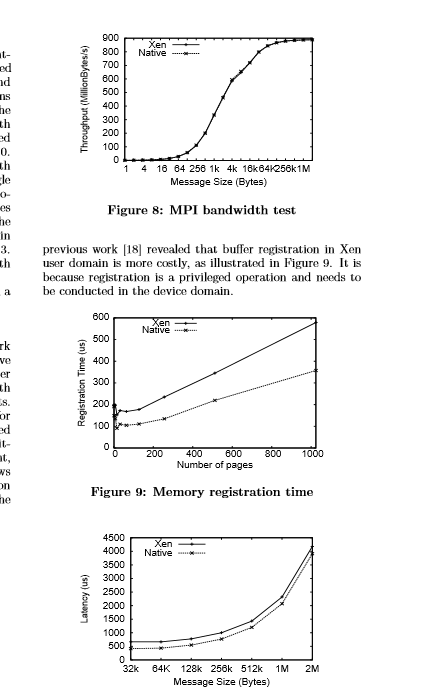
1)We will demonstrate how the system architecture was built and show why the clusters where setup in such a way following recommendations found from research papers.

2) Benchmark the performance of our cluster by using LINPACK and other libraries in terms of performance, latency, ram speed etc

As we are proposing a High-performance computing platform as a service for students. In this section we will be demonstrating diagrams and explain why I utilised docker instead of virtual machines and quote the research conducted by (Bernstein, 2014; Chung et al., 2016; Combe et al., 2016; Higgins et al., 2015; Jacobsen & Canon, 2015; Manu et al., 2016; Mauch et al., 2013; Nguyen & Bein, 2017; Yu & Huang, 2015). These papers explore the security issues, performance benefits and implementation in a HPC environment and by following their advice we created prototype following their recommendations.

Furthermore, when looking at research papers that are conducting studies in the HPC area, they build a cluster to be able to demonstrate their studies and use libraries like LINPACK intended to benchmark their systems in terms of processing power, RAM, internet speed and file storage transfer. Therefore, our system will also be benchmarked following the guidelines of other research and present data similarly as shown in the image below. (Taken from paper (Higgins et al., 2015))



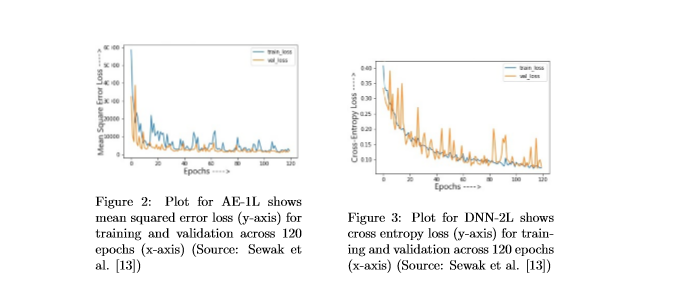
‘

# Research Question 2 approach to obtain results:

For research question two, we explore how researchers will benefits using our system we utilise the system for neural network research.

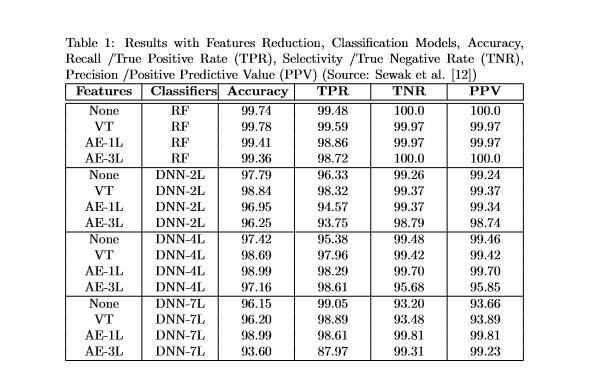
1) We create two types of neural network models (Convolutional neural networks and Deep neural networks) where one dataset contains images and the other contains textual data to see how they perform and find the differences when executed on our platform.

- we take timings to see how long it takes for these models to take to be trained on the system. We will also demonstrate the mean squared error loss for training and validation across the epoch amount set and cross entropy loss across epoch amount. As can be seen similarly in image below. (Taken from (Rathore et al., 2019))



- Produce Accuracy, TP, TN, FAR and precision for each model trained with the cluster and compare them to existing research done using the same dataset and model configurations.

2) We time these neural network models on one computer vs the created cluster by utilising four neural models ( 4 CNN and 4 DNN) with different model configurations ( Hidden layers, Max steps, optimizers and different data splitting configurations



# Research question 3 approach to obtain results:

In research question three, we will be utilising read made timetables and produce a small prototype that can extract data and check when the room will be available or in use. A program will take care of this issue by pausing the cluster when a classroom is in use and continue the training process when a classroom is free.

We will report how much resources are being used while the cluster is training the neural networks in contrast with the systems only operating necessary school programs. Cluster resource usage will be monitored when idle and when it is hosting necessary services for the distribution of the training process. This will be done by creating a linux bash script that will execute a program that will monitor cpu/ram/storage usage while a classroom is not in use and plot a timeline to demonstrate the difference.

# Bibliography:

Bernstein, D. (2014). Containers and Cloud: From LXC to Docker to Kubernetes. *IEEE Cloud Computing*, *1*(3), 81–84. https://doi.org/10.1109/MCC.2014.51

Chung, M. T., Quang-Hung, N., Nguyen, M., & Thoai, N. (2016). Using Docker in high performance computing applications. *2016 IEEE Sixth International Conference on Communications and Electronics (ICCE)*, 52–57. https://doi.org/10.1109/CCE.2016.7562612

Combe, T., Martin, A., & di Pietro, R. (2016). To Docker or Not to Docker: A Security Perspective. *IEEE Cloud Computing*, *3*(5), 54–62. https://doi.org/10.1109/MCC.2016.100

Higgins, J., Holmes, V., & Venters, C. (2015). Orchestrating Docker Containers in the HPC Environment. In J. M. Kunkel & T. Ludwig (Eds.), *High Performance Computing* (pp. 506–513). Springer International Publishing.

Jacobsen, D. W., & Canon, S. (2015). *Contain This , Unleashing Docker for HPC*.

Manu, A. R., Patel, J. K., Akhtar, S., Agrawal, V. K., & Murthy, K. N. B. S. (2016). A study, analysis and deep dive on cloud PAAS security in terms of Docker container security. *2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, 1–13. https://doi.org/10.1109/ICCPCT.2016.7530284

Mauch, V., Kunze, M., & Hillenbrand, M. (2013). High performance cloud computing. *Future Generation Computer Systems*, *29*(6), 1408–1416. https://doi.org/https://doi.org/10.1016/j.future.2012.03.011

Nguyen, N., & Bein, D. (2017). Distributed MPI cluster with Docker Swarm mode. *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*, 1–7. https://doi.org/10.1109/CCWC.2017.7868429

Rathore, H., Agarwal, S., Sahay, S. K., & Sewak, M. (2019). Malware Detection using Machine Learning and Deep Learning. *CoRR*, *abs/1904.0*. http://arxiv.org/abs/1904.02441

Yu, H.-E., & Huang, W. (2015). *Building a Virtual HPC Cluster with Auto Scaling by the Docker*.