Performance Measurement of a Hierarchical File System for Distributed Deep Neural Network Training

Takaki Fukai
RIKEN Center for Computational Science
Kobe, Hyogo, Japan
takaaki.fukai@riken.jp

Kento Sato
RIKEN Center for Computational Science
Kobe, Hyogo, Japan
kento.sato@riken.jp

1 INTRODUCTION

Today, deep learning is an essential technology for our life. To solve more complex problems with deep learning, both sizes of training datasets and neural networks are increasing. To train a model with large datasets and networks, a single computer is not enough. Therefore, distributed deep neural network (DDNN) training, using multiple computers for training a model, is necessary. For large scale DDNN training, HPC clusters are a promising computation environment. Therefore, DDNN learning application will be an important target application on designing HPC clusters.

In large-scale DDNN, I/O performance is a critical matter because it is a bottleneck in some training workloads [3]. Therefore, the storage system performance will become more crucial in HPC clusters. For designing the storage systems, it is necessary to reveal the impact on the training performance of the storage system. Existing research for the file system performance for DDNN in HPC clusters [2] analyzed the performance in detail, however, it did not assume hierarchical storage systems, which are mainstream in current HPC systems.

To reveal it, we study on the performance measurement and analysis of the hierarchical storage system for the DDNN training workload. We plan to evaluate the impact on the training performance by the difference in the throughput and volume size between the local file system and the global file system. In this poster presentation, we report the results of the preliminary experiments for the measurement. One of the preliminary experiments is for changing the throughput of the global file system. To achieve the different speed global file systems for evaluating the impact of throughput difference, we change its performance by changing the number of object storage targets (OST). The results shows that we can change the throughput of the global file system even in the access pattern by controlling the number of OSTs.

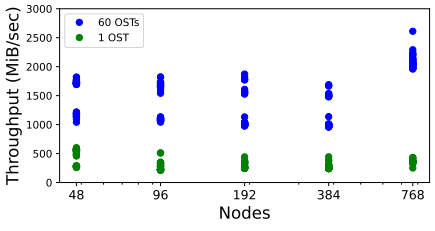

Figure 1: Fio benchmark results

2 METHODOLOGY

As a preliminary experiment, we measure the file system throughput on the global file system of supercomputer Fugaku [4] with 60 OSTs and 1 OST. For the measurement, we use the fio benchmark [1]. In our measurement, we configure fio such that each process selects files in random order and reads each file sequentially because the actual training processes access files in a similar way. We execute fio on 48, 96, 192, 384 ,and 768 nodes simultaneously. On each node, 8 fio processes are executed. We prepare the directories on the global file system for each node.

3 RESULTS

Figure 1 shows the results of the preliminary experiment. The horizontal axis indicates the number of nodes. The vertical axis indicates the throughput. We execute jobs 35 times that run the benchmark 5 times for each setup and each point on the graph indicates the mean throughput in each job.

The results show that the throughput of the file system with 60 OSTs is 2-8 times higher than that with 1 OST. With 60 OSTs, the throughput does not much change from 48 nodes to 384 nodes. On the other hand, the throughput with 768 nodes is higher than that with 384 nodes. We expect that the reason is the difference in the number of compute nodes connected with the global file system because the nodes are provided for every 384 nodes in Fugaku. With 1 OST, the throughput with 768 nodes is almost the same as that with less than 768 nodes. Therefore, we think that the throughput with 1 OST is saturated by the OST.

4 CONCLUSION

In this extended abstract, we have presented the measurement results of file system performance toward analyzing hierarchical storage system performance for DDNN. The results show that the number of OSTs of the global file system strongly affects the throughput in a similar access pattern to deep neural network training workload.

Our future works are measurements of training performance with various volume and throughput ratio, analyzes the results, and indicate a performance model for designing file systems in the next-generation HPC clusters.

REFERENCES

[1] 2012. FIO. https://github.com/axboe/fio
[2] Hariharan Devarajan et al. 2021. DLIO: A Data-Centric Benchmark for Scientific Deep Learning Applications. In *2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*. 81–91. https://doi.org/10.1109/CCGrid51090.2021.00018
[3] Jayashree Mohan et al. 2021. Analyzing and Mitigating Data Stalls in DNN Training. *Proc. VLDB Endow.* 14, 5 (Jan. 2021), 771–784. https://doi.org/10.14778/3446095.3446100
[4] Mitsuhisa Sato et al. 2020. Co-Design for A64FX Manycore Processor and "Fugaku". In *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis*. 1–15. https://doi.org/10.1109/SC41405.2020.00051