Ragini Gupta, NETID: raginig2
Yinfang Chen, NETID: yinfang3
Team Number: **42**
CS 425 MP-4

Design

We have used the **MNIST** dataset. The two jobs (models) for processing the data is; **CNN and Multi-Layer Perceptron (MLP) Model**. We choose 70% of the data (60,000 images) as the training set and the remaining 30% (21,000 images) is chosen as the testing data. The test and training subsets are saved in the sdfs directory for each of the VMs.
The client submits the queries to the coordinator node (Node-1) which will then schedule the jobs (i.e. the DL models) across the cluster nodes. Each submitted query has a batch size that represents the number of files that need to be processed by a given job per unit of time. We fix the size of each batch for a query as 100 images.

Training Phase: The Models (CNN and MLP) are loaded on to the sdfs.The training dataset is saved on the sdfs.
Inference Phase: Queries are submitted to the sdfs to give inference on the queried files (which is a part of the test data already stored on the sdfs). The results statistics including accuracy and inference time are written back into the sdfs.

Packages used: Keras, TensorFlow

The average inference time per query for the CNN model is around 1.519 minutes and is much faster than the MLP model which has an average inference time of 2.877 minutes. Additionally, the CNN model has better accuracy (98.09%) than the MLP model (96.28%) which makes it an overall preferable choice for processing the MNIST data. This might be due to the fact that MLP is a fully connected network where each node is connected to every node in the next and previous layer, which makes it a very dense network with large parameters that lead to higher model inference time. In our case, Job-1 is CNN and Job-2 is MLP.
Query implies the number of files (from the testing set) that have to be processed during the inference phase. Note that we are using batch_queries with a prefixed batch size for each query.

We used Socket and SCP for communication between the coordinator and the cluster nodes (worker nodes). The coordinator is used to submit and track the status of the running jobs on the worker nodes. When a job is added to the cluster coordinator, it first checks for the nodes that are up and running, followed by whether those nodes are currently running a previous job or not. If not, the new job and query is split between the idle nodes to ensure a fair-inference time.
A global membership list is maintained across cluster to keep track of nodes that are up and running, the job running on a node, and number of queries processed.

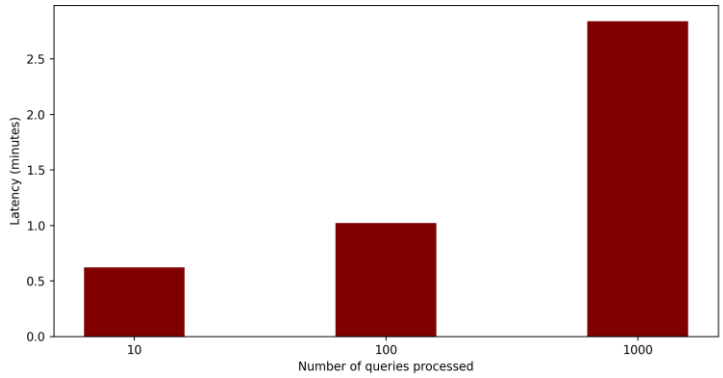
We used MP-3 code for implementing SDFS as well as for failure detection in SDFS . MP-1 was used for debugging the code and logging.

Ragini Gupta, NETID: raginig2
Yinfang Chen, NETID: yinfang3
Team Number: **42**
CS 425 MP-4

Performance:

Latency Vs. # of Queries Processed:

Number of queries processed (each query with batch size=100)	Average Latency (min)	Standard Deviation
10	0.62	0.2918
100	1.021	0.710
1000	2.835	0.6518



As evident from the above chart, as the number of queries increase (from 10→ 1000), there is only a small increase in the total latency during the inference phase. This is because as the number of queries increase, each VM will fetch a sharded set from the queries, thereby increasing the average querying rate across the cluster (instead of exponential increase in latency for processing on a single node).

Q.1. a. Since the MLP model (i.e. Job-2) is a much denser network in contrast to the CNN model (and it takes longer inference time), therefore the ratio of resources the system decides across Job-1: Job-2 is 1:2 (i.e. for the same batch size and query, the number of VMs be required to run the MLP model is twice that of the CNN model).

Q. 1. b. When a second job is added to the cluster, it takes an average of 4.91 seconds to start executing queries from the second job. This is because it has to take into account the current status of other nodes in the cluster before assigning new (idle) resources (i.e. VMs) to the second job.

Q. 2.

Time taken for resuming normal operation (min) for 100 queries

Ragini Gupta, NETID: raginig2
Yinfang Chen, NETID: yinfang3
Team Number: **42**
CS 425 MP-4

1.3876
1.09272
1.2234
1.81103
Average: 1.3786
Standard Deviation: 0.2709

This time is proportional to the time taken due to the SCP communication.

Q. 3. After the coordinator fails, the standby coordinator has to be revoked as the main coordinator which will have access to the global membership dictionary to monitor the job and query status across the cluster. It is worth mentioning that a failed node is determined if it does not respond to the periodic ping requests from it's neighbors (Fmapper.py and FServer.py). It takes an average of 0.87 mins to resume the operations.

GitLab: https://gitlab.engr.illinois.edu/yinfang3/cs425-mp/-/tree/main/mp4