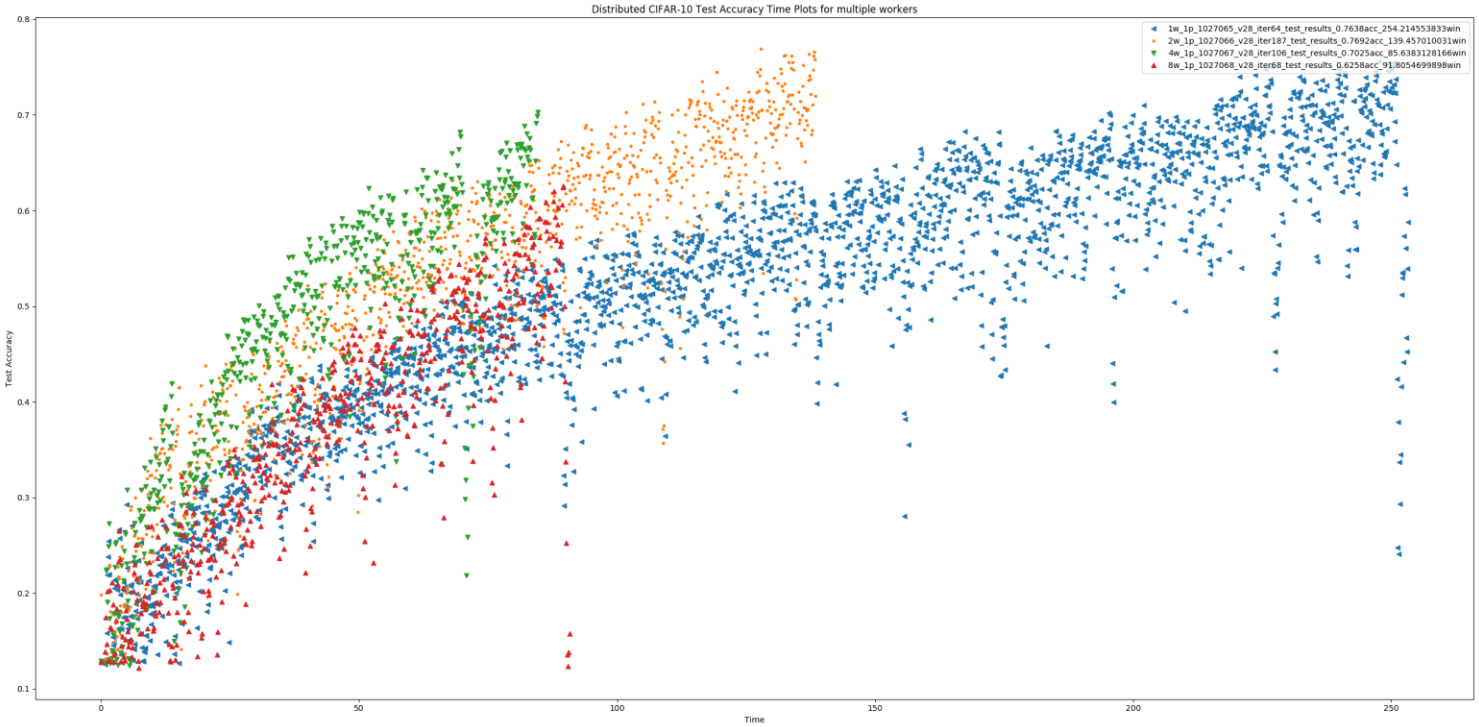
Name: *Souham Biswas* ***CWID***: *A20365242* ***Email***: *sbiswas7@hawk.iit.edu*

CS 597 – Report 6

Experimented and set up environment on cooley (Argonne's cluster) for commencement of experiments.

For benchmarking purposes, extended single node caffe to include support for multi-node training using MPI.

Implemented AlexNet on Caffe and performed distributed training experiments on CIFAR-10 dataset. The following plots show the test accuracies of the models as the training progresses. –



After a parameter sweep of batch size and learning rates for each configuration of number of workers the following batch sizes and learning rates were converged upon for each number of workers –

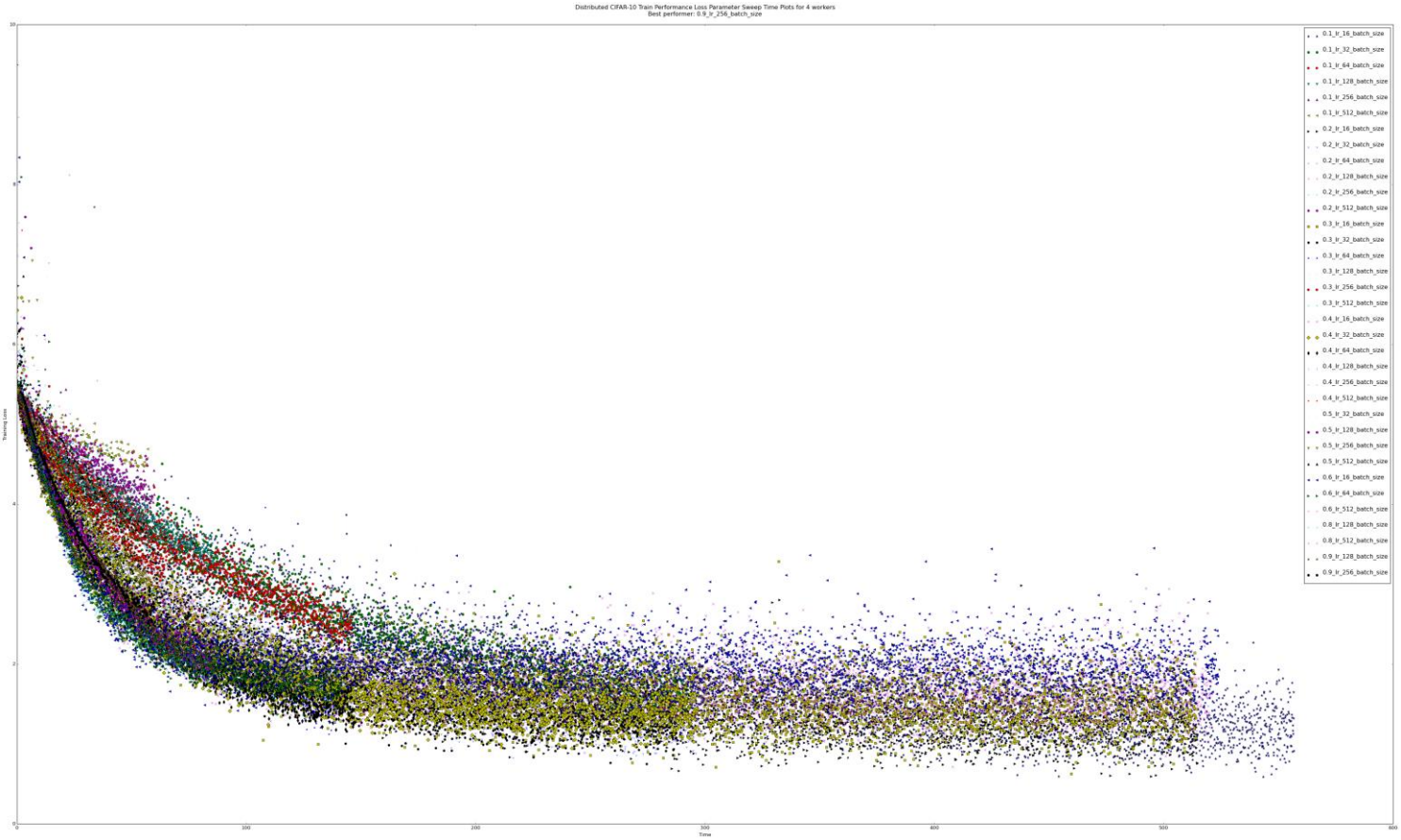
0.5_lr_128_batch_size_1w
0.5_lr_128_batch_size_2w
0.4_lr_128_batch_size_4w
0.6_lr_64_batch_size_8w

The criteria used to select the best configuration for a given number of workers takes in to consideration both, the time taken to converge and the final loss attained at convergence.

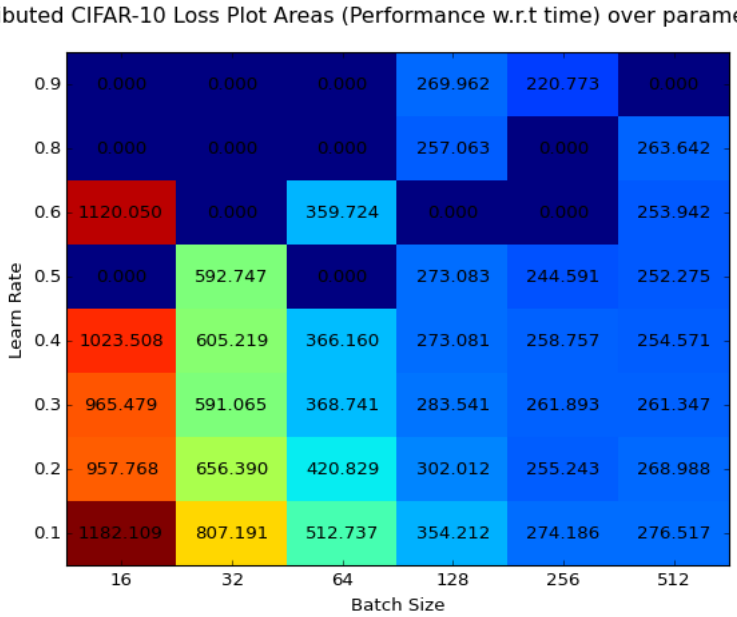
Name: *Souham Biswas* ***CWID***: *A20365242* ***Email***: *sbiswas7@hawk.iit.edu*

CS 597 – Report 6

An illustration is presented below for 4 workers –

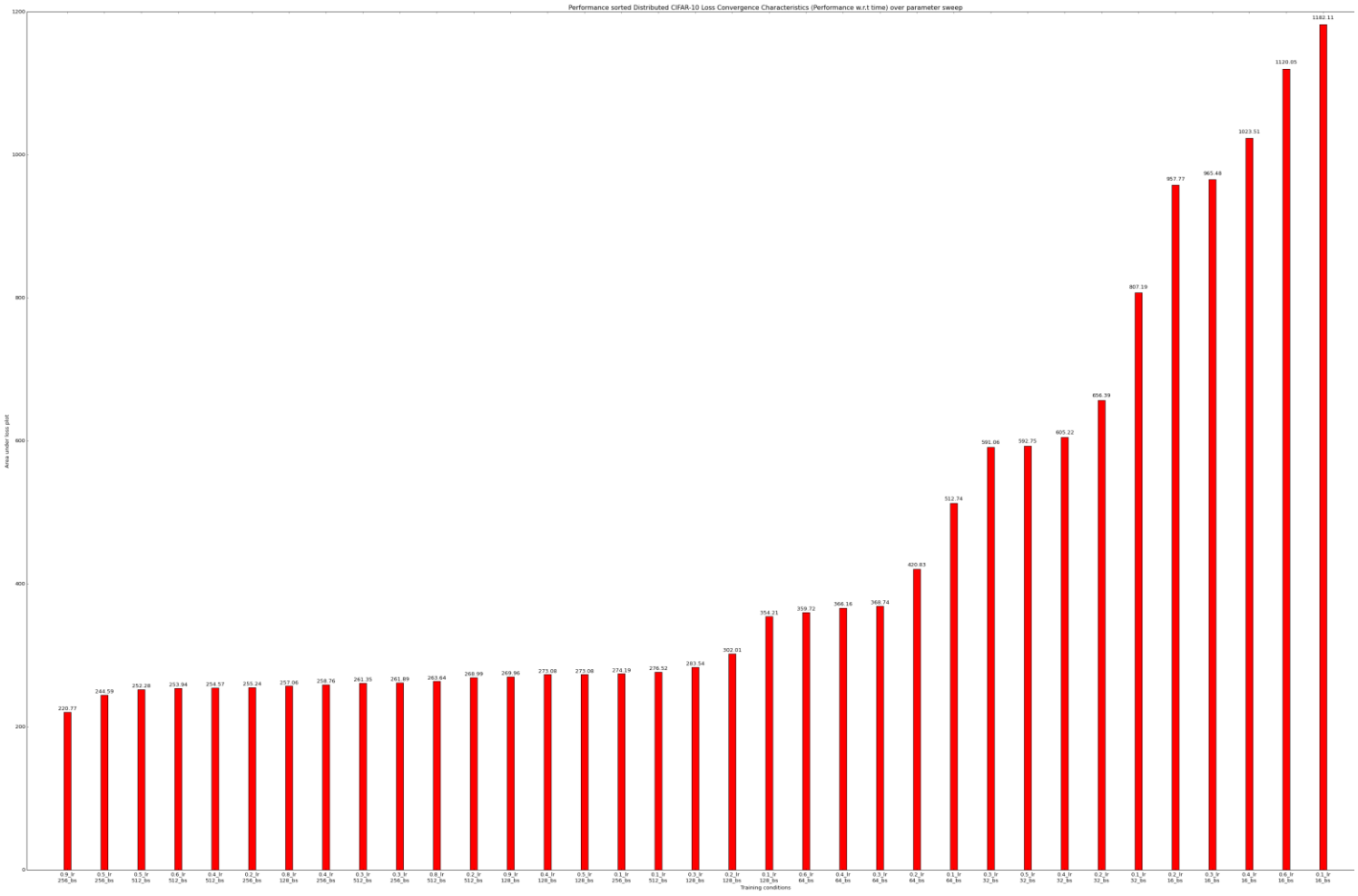


This represents the progression of L2 loss for 4 workers as training proceeds with respect to time-
The plot below maps the area under the loss curves for each experiment-



Name: *Souham Biswas* ***CWID***: *A20365242* ***Email***: *sbiswas7@hawk.iit.edu*

CS 597 – Report 6

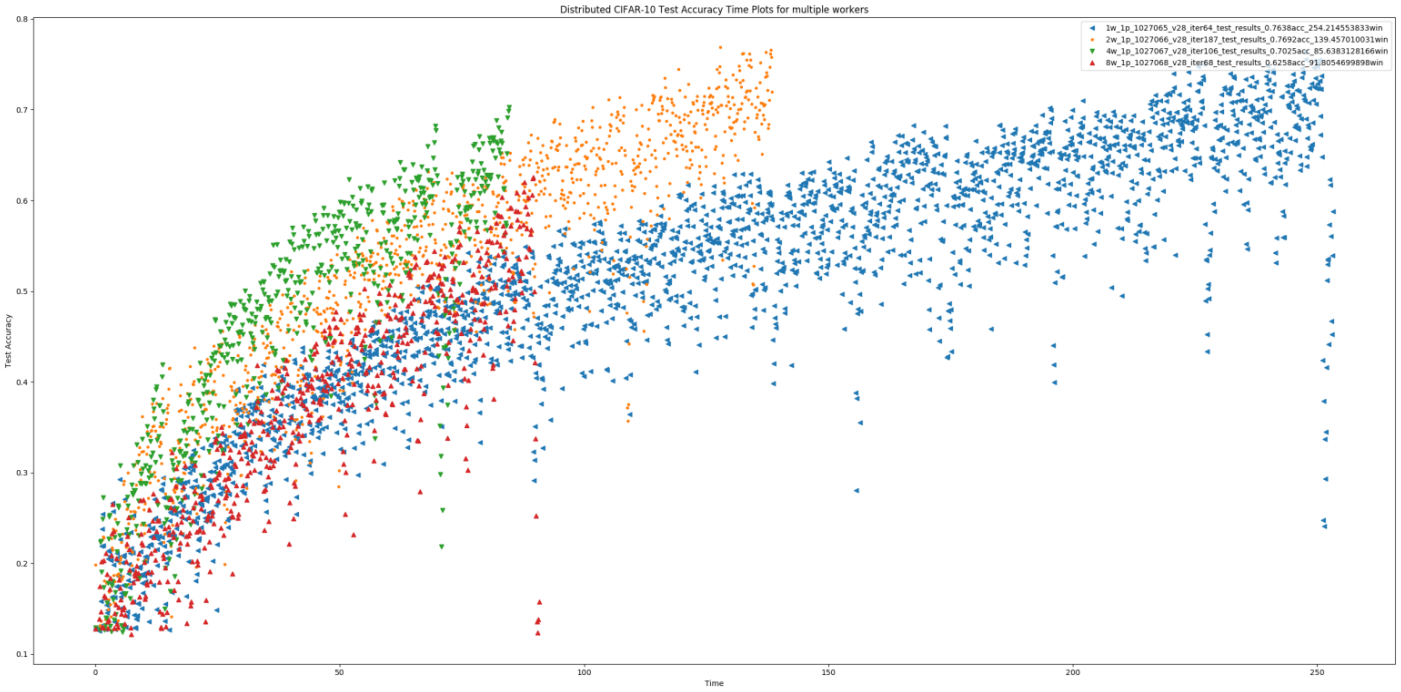


The histogram above visualizes and sorts each experiment according to the areas of the loss curves.

This whole experiment was repeated for each #worker configuration and plots were obtained for 1, 2, 4 & 8 workers over the CIFAR-10 dataset. The best of each were chosen and trained to obtain the following plot –

Name: *Souham Biswas* ***CWID***: *A20365242* ***Email***: *sbiswas7@hawk.iit.edu*

CS 597 – Report 6



As shown previously. Now, our objective is to obtain better performance for our library.

The selective compilation and environment setup has concluded on Cooley. The makefile looks like this –
all: program

```
program: cppcode.o
        g++ -std=c++11 -o convnet -L/soft/visualization/cuda-8.0.61/lib64 -lcuda
\
        -lcudart -lcudnn -lcublas -lcurand main.cpp FCLayerCU.o FCLayerCPP.o \
        ConvLayerCU.o ConvLayerCPP.o
        rm *.o

cppcode.o: cudacode.o
        g++ -c -std=c++11 -o FCLayerCPP.o FCLayer.cpp
        g++ -c -std=c++11 -o ConvLayerCPP.o ConvLayer.cpp

cudacode.o:
        nvcc -c -std=c++11 -arch=sm_37 -o ConvLayerCU.o ConvLayer.cu
        nvcc -c -std=c++11 -arch=sm_37 -o FCLayerCU.o FCLayer.cu

clean:
        rm -rf *o convent
```

However, it was observed that a simple hardware change has resulted in a couple of cuDNN primitives failing to execute resulting in NaN outputs which are currently being tackled.