Evaluating Machine Learning and Big Data Analytics Models on the IHK/McKernel Lightweight Kernel

Student: Sreevatsa Vadiraj
Capstone Advisor: Prof M Mustafa Rafique
Colloquium Advisor: Prof Matthew Fluet
Project overview

- Use IHK/McKernel to run datacenter workloads like Deep learning and Big data analytics models and see if performance is better than a vanilla Linux

- Use horovod to run distributed deep learning models

- Might have to tweak Tensorflow or McKernel to obtain the required gains
Milestone 1 planned

- Read research papers to understand the working of LWKs and multi-kernels.
- Identify the PyTorch workloads to be used to benchmark.
- Generate synthetic datasets for the workloads.
- Environment setup
Outcomes

- Read research papers
  - Understood the basic principles and working of a LWK
  - Understood the architecture and functionalities of IHK/McKernel
- Identify the datacenter workloads
  - K-Means
  - MNSIT
  - CIFAR image classification
- Generate synthetic data
  - Generated synthetic data for K-Means using Scikit learn having various parameters.
Next Steps

• Setup the environment required to run these experiments.
  – Setup instances of CentOs on cloudlab
  – Install required dependencies
  – Install McKernel
• Generate datasets for other workloads
• Run the models on these datasets
• Create multi-node setups
THANK YOU