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 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.
Milestone 2 outcomes

- Completed the setup of nodes on Cloudlab
  - Installed required libraries
  - Installed the IHK/McKernel
- Executed K-Means workloads on single-node
  - Varying sample size
  - Varying number of clusters
  - Varying dimensionality of the data
Milestone 3

- i) Benchmark for Kmeans
  - Vary dimensions
  - Vary cluster-size
  - Vary data size
- ii) Analyse performance
Benchmarking Framework

- **Inputs**
  - i) Workloads/Tasks
  - ii) Dataset parameters
  - iii) Number of CPU cores

**Execute tasks on**
- i) IHK/McKernel
- ii) Vanilla Linux
Performance

Kmeans Batch Processing - Varying Feature Size

Time in sec

Dataset Size (in 1000)

Linux
LWK
CPU utilization
Next Steps

- Evaluate other workloads
  - MNSIT
  - CIFAR-10
- Use tools like perf and flame-graphs to dig deep into performance analysis
• At the time Linux seems to be performing better than LWK
• IHK/McKernel doesn’t seem to be a well suited for these workloads
• Linux seems to better utilize the CPU cores than IHK/McKernel
Future Work

- Use HOROVOD, a distributed deep learning framework
- Extend the research to run workloads on Tensorflow
- More extensive benchmark suites
- Modify IHK/McKernel code to improve performance
THANK YOU