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 2 planned

- Environment setup
- Develop a comprehensive benchmark.
- Run K-Means model on PyTorch on a single node-setup.
Single node setup
Assigning CPU cores

```
SreeVad@node0 thk-mckernel] $ sudo ./sbln/mcreboot.sh -c 1,2,3,10,31 -m 512m
```
### Kmeans test-suite

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Dimension size</th>
<th>Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>50K</td>
<td>5000</td>
<td>10</td>
</tr>
<tr>
<td>100K</td>
<td>5000</td>
<td>10</td>
</tr>
<tr>
<td>500K</td>
<td>5000</td>
<td>10</td>
</tr>
<tr>
<td>1000K</td>
<td>5000</td>
<td>10</td>
</tr>
<tr>
<td>100K</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>10</td>
</tr>
<tr>
<td>100K</td>
<td>5000</td>
<td>10</td>
</tr>
<tr>
<td>100K</td>
<td>10000</td>
<td>10</td>
</tr>
</tbody>
</table>
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
Next Steps

• Generate datasets for other workloads
  - MNSIT
  - CIFAR-10
• Replicate setup on multiple nodes
• Use HOROVOD, a distributed deep learning framework
THANK YOU