In the cloud, billing is based on wall-clock time rather than resource utilization. Efficiency matters a lot, time is money in the cloud. Our goal is to provide an efficient solution for complex analytics in the cloud that utilizes the hardware to the maximum.

**Incremental Gradient Descent**

We focus on convex problems with separable objective functions:

\[
\min_{w \in \mathbb{R}^d} \sum_{i=1}^{N} f(w, z_i) + P(w)
\]  

(1)

\[w^{(k+1)} = w^{(k)} - \alpha_k \nabla f_i^{(k)} (w^{(k)})\]  

(2)

\[\alpha_k \geq 0\] is the step size and \(\nabla f_i^{(k)} (w)\) is the approximation to the gradient \(\nabla f(w)\) based on a single term \(f_i^{(k)}(w)\) at iteration \(k\), respectively.

**Datasets & Tasks**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Dimension</th>
<th># Examples</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classify300M</td>
<td>50</td>
<td>300M</td>
<td>135G</td>
</tr>
<tr>
<td>Matrix10B</td>
<td>1M x 1M</td>
<td>10B</td>
<td>200G</td>
</tr>
<tr>
<td>DBLP</td>
<td>600M</td>
<td>2.3M</td>
<td>7.2G</td>
</tr>
</tbody>
</table>

**IGD in GLADE**

GLADE is a scalable and efficient parallel framework for Big Data analytics.

**Experimental Results**

Figure: Multi-thread and multiple nodes speed-up of LMF model executed over the matrix10b dataset.

**Effect of Randomization**

Figure: The effect of randomization on convergence rate for the LR model over the classify300M dataset.

**Effect of Merging**

Figure: The effect of merging on convergence for the LR model over the classify300M dataset. 

University of California, Merced

Mail: Chengjie Qin cqin3@ucmerced.edu, Florin Rusu frusu@ucmerced.edu