Scalable I/O-Bound Parallel Incremental Gradient Descent for Big Data Analytics in GLADE

Chengjie Qin & Florin Rusu
EECS, University of California, Merced

Motivation & Goal

- 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:

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

z_i represents a tuple

$$w^{(k+1)} = w^{(k)} - \alpha_k \nabla f_{\eta(k)} \left(w^{(k)} \right) \quad (2)$$

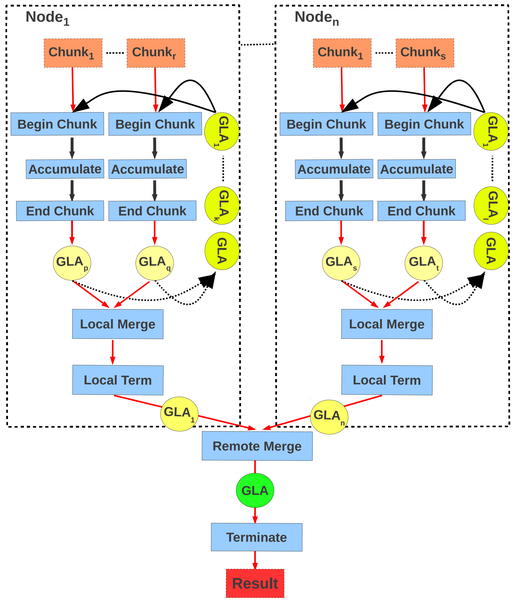
$\alpha_k \geq 0$ is the step size and $\nabla f_{\eta(k)}(w)$ is the approximation to the gradient $\nabla f(w)$ based on a single term $f_{\eta(k)}(w)$ at iteration k, respectively

Datasets & Tasks

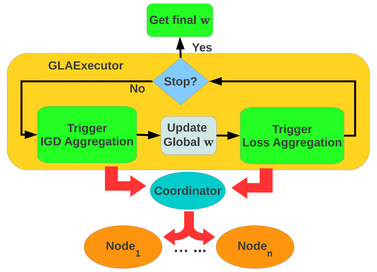
Dataset	Dimension	# Examples	Size
`Classify300M`	50	300M	135G
`Matrix10B`	1M×1M	10B	200G
`DBLP`	600M	2.3M	7.2G

IGD in GLADE

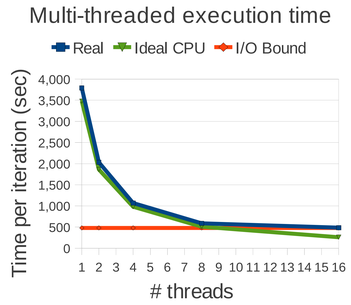
- GLADE is a scalable and **efficient** parallel framework for Big Data analytics



- IGD execution flow in GLADE



Experimental Results



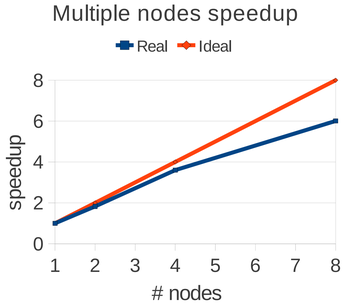


Figure: Multi-thread and multiple nodes speed-up of LMF model executed over the `Matrix10B` dataset.

Effect of Randomization

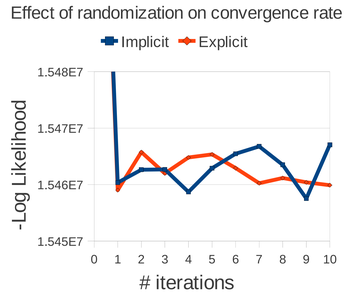


Figure: The effect of randomization on convergence for the LR model over the `classify300M` dataset.

Effect of Merging

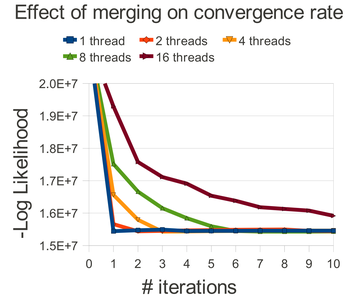


Figure: The effect of merging on convergence for the LR model over the `classify300M` dataset.

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