An Uninformed View on GraphLab

ExaScience life lab

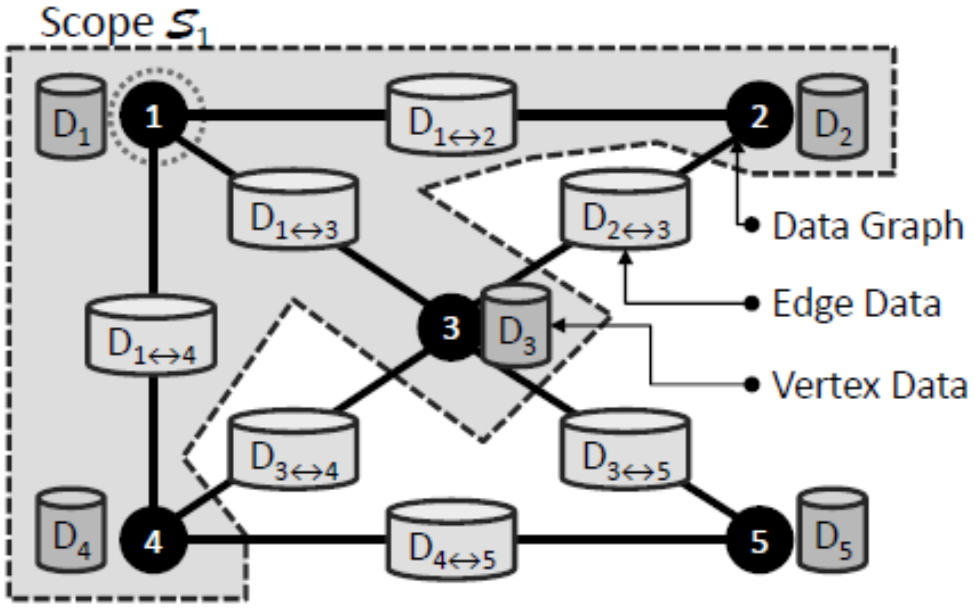
Tom Vander Aa and Tom Ashby

EuroMPI Tutorial on Machine Learning at Scale

janssen PHARMACEUTICAL COMPANIES OF Johnson & Johnson intel iwt imec UNIVERSITEIT GENT universiteit hasselt KU LEUVEN

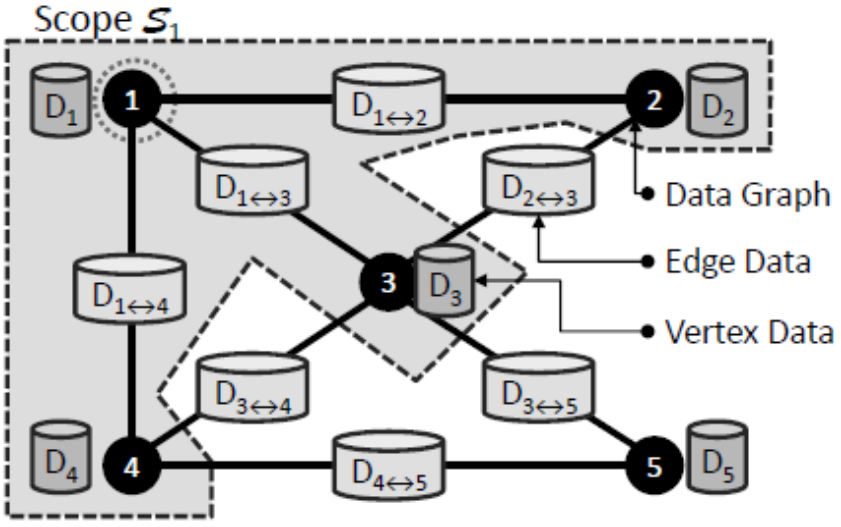
Basic GraphLab

- Operations
  - Update: modify the vertex and edges state
  - Merge/Fold/Update: for shared key/value pairs



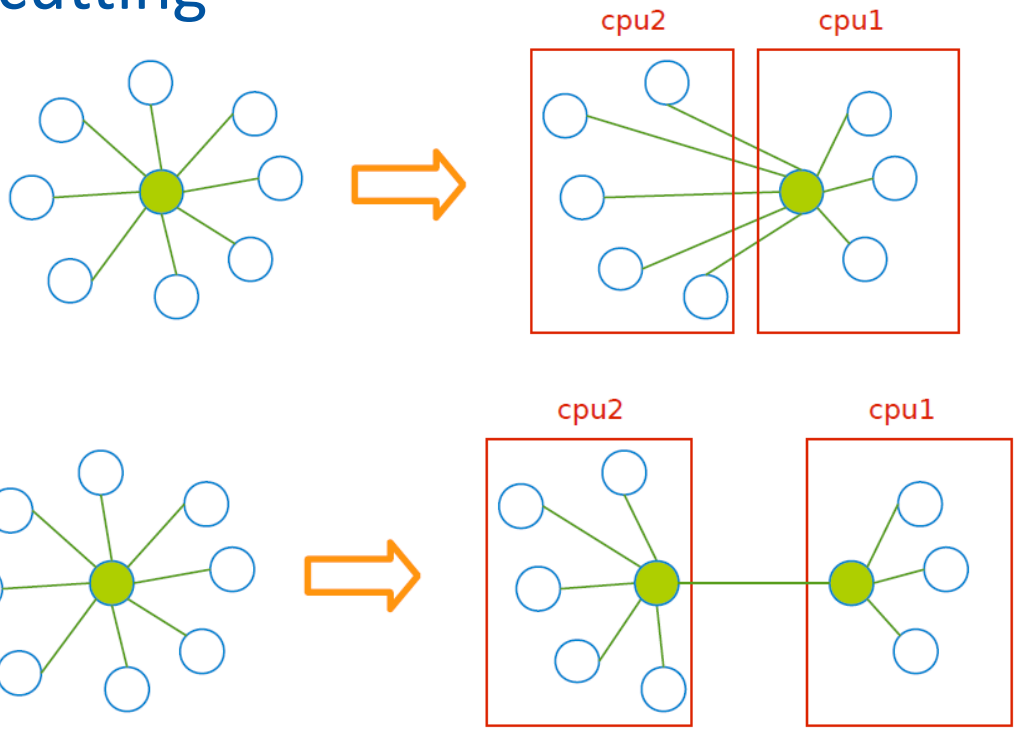
Basic GraphLab

- Scheduling:
  - Parallel processing of non-overlapping scopes
  - Asynchronous graph update as opposed to block synchronous
    - Using latest value of edges and vertices
- Distribution
  - Distributed using synchronous MPI



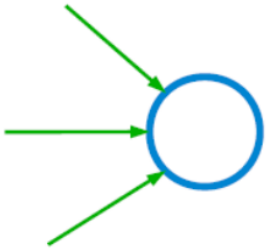
PowerGraph

- PowerGraph == GraphLab v2.1
  - Power law in natural graphs
  - From edge-cutting to vertex-cutting
  - Duplicated Halo Points

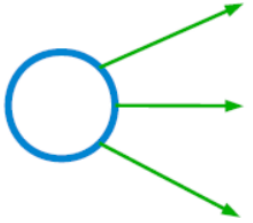


PowerGraph

- Update operation replaced by:
  - Gather
    - Accumulate information about neighborhood through a generalized sum.



  - Apply
    - Apply the accumulated value to center vertex.
  - Scatter
    - Update adjacent edges and vertices.



GraphChi

- Only Shared Memory Parallelism
  - Large graphs are efficiently read from/written to disk
  - Parallel sliding window
- Simplified Operations
  - Only vertex update function