CLOUD COMPUTING APPLICATIONS

Pregel - Part 3

Roy Campbell & Reza Farivar

System Architecture

- Pregel system uses the master/worker model
  - Master
    - Maintains worker
    - Recovers faults of workers
    - Provides Web-UI monitoring tool of job progress
  - Worker
    - Processes its task
    - Communicates with the other workers
- Persistent data is stored as files on a distributed storage system (such as GFS, HDFS, or BigTable)
- Temporary data is stored on local disk

Execution of a Pregel Program

1. Many copies of the program begin executing on a cluster of machines
2. The master assigns a partition of the input to each worker
   - Each worker loads the vertices and marks them as active
3. The master instructs each worker to perform a superstep
   - Each worker loops through its active vertices and computes for each vertex
   - Messages are sent asynchronously, but are delivered before the end of the superstep
   - This step is repeated as long as any vertices are active, or any messages are in transit
4. After the computation halts, the master may instruct each worker to save its portion of the graph

Fault Tolerance

- Checkpointing
  - The master periodically instructs the workers to save the state of their partitions to persistent storage
    - e.g., Vertex values, edge values, incoming messages
- Failure detection
  - Using regular “ping” messages
- Recovery
  - The master reassigns graph partitions to the currently available workers
  - The workers all reload their partition state from most recent available checkpoint