



CLOUD COMPUTING APPLICATIONS

Giraph Introduction

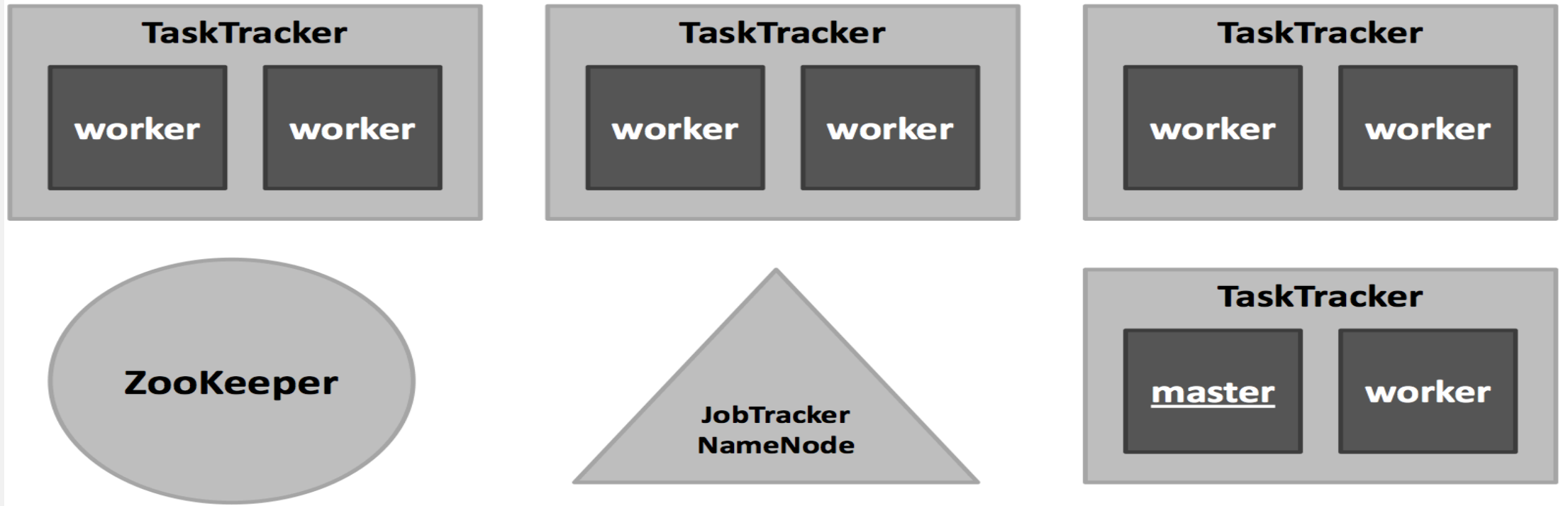
Roy Campbell & Reza Farivar

Apache Giraph

- Open source implementation based on Pregel
- Giraph is currently in Apache incubator
- Modifications are continuous



Giraph Framework



Task Assignment

- **ZooKeeper**: responsible for **computation state**
 - Partition/worker mapping
 - Global state: #superstep
 - Checkpoint paths, aggregator values, statistics
- **Master**: responsible for **coordination**
 - Assigns partitions to workers
 - Coordinates synchronization
 - Requests checkpoints
 - Aggregates aggregator values
 - Collects health statuses
- **Worker**: responsible for **vertices**
 - Invokes active vertices compute() function
 - Sends, receives, and assigns messages
 - Computes local aggregation values

Anatomy of an Execution

