Introduction of **Apache Hama**

Edward J. Yoon, October 11, 2011
<edwardyoon@apache.org>
About Me

- Founder of Apache Hama.
- Committer of Apache Bigtop.
- Employee for KT.
- http://twitter.com/eddieyoon
What Is Hama?

- Apache Incubator Project.
- BSP (Bulk Synchronous Parallel) for massive scientific computations.
- Written In Java.
- Currently 2 releases, 3 main committers.
Hama Characteristics

- Provides a Pure BSP model.
  - Job submission and management interface.
  - Multiple tasks per node.
  - Checkpoint recovery.
- Supports to run in the Clouds using Apache Whirr.
- Supports to run with Hadoop nextGen.
Bulk Synchronous Parallel?

- Parallel programming model introduced by Valiant.
- Consist of a sequence of supersteps.
- Conceptually simple and intuitive from a programming standpoint.
- Used for a variety of applications e.g., scientific computing, genetic programming, …
Schematic diagram of a superstep

- **Local Computation**
- **Communication**
- **Idle**
- **Barrier Synchronization**
Internals

- Proactive Task Scheduling.
- Hadoop RPC is used for BSP tasks to communicate each other.
- Collection and bundling of messages as a technique to reduce network overheads and contentions.
- Zookeeper is used for Barrier Synchronization.
Pi Calculation

- Each task executes locally its portion of the loop a number of times.
- One task acts as master and collects the results through the BSP communication interface.
Structural Analysis of Network Traffic Flows

- Traffic flows in KT clouds.
  - traffic engineering, anomaly detection, traffic forecasting and capacity planning
- Currently BSP jobs are experimentally running on 512 multi-cores machines.
Random Communication Benchmarks

- Benchmarked on 16 1U servers using 10 tasks per server.
- X axis is the time (sec.) of BSP job execution (32 supersteps).
- Y axis is the number of messages to be sent to random BSP tasks in each superstep.
What’s Next?

- Support Input/Output Formatter like MapReduce.
- Message Compression for High Performance.
- Add some frameworks on top of Hama.
More Information

- [http://incubator.apache.org/hama](http://incubator.apache.org/hama)