Apache Giraph

*Functional Area:* Parallel Horizontally Scalable Data Processing:Graph

*Overview:* Apache Giraph is an iterative graph processing system built for high scalability. In other words, Apache Giraph is an [Apache](http://en.wikipedia.org/wiki/Apache_Software_Foundation) project to perform [graph processing](http://en.wikipedia.org/wiki/Graph_(computer_science)) on big data. It utilizes [Apache Hadoop](http://en.wikipedia.org/wiki/Apache_Hadoop)'s MapReduce implementation to process graphs. For example, it is currently used at Facebook to analyze the social graph formed by users and their connections.

Giraph originated as the open-source counterpart to Pregel, the graph processing architecture developed at Google and described in a 2010 [paper](http://dl.acm.org/citation.cfm?id=1807184). Both systems are inspired by the [Bulk Synchronous Parallel](http://en.wikipedia.org/wiki/Bulk_synchronous_parallel) model of distributed computation introduced by Leslie Valiant. Giraph adds several features beyond the basic Pregel model, including master computation, sharded aggregators, edge-oriented input, out-of-core computation, and more. With a steady development cycle and a growing community of users worldwide, Giraph is a natural choice for unleashing the potential of structured datasets at a massive scale.