Iterative MapReduce Enabling HPC-Cloud Interoperability

Judy Qiu, Indiana University

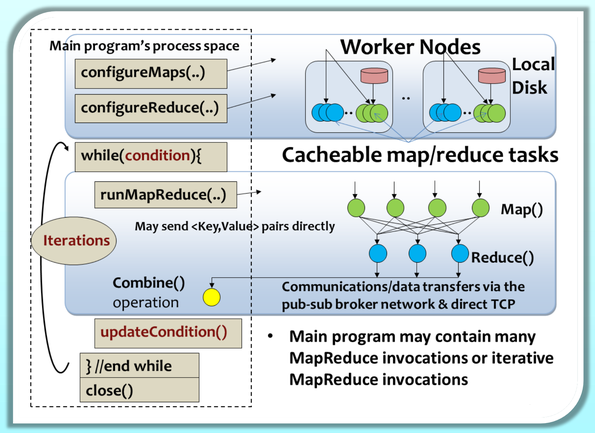
Clouds provide both cost effectiveness and powerful parallel programming paradigms. On the other hand HPC clusters spanning up to exascale capability will continue to be critical. A key challenge is portability not just as systems scale up in size but also between HPC and Cloud systems, in addition to common challenges in fault tolerance and storage. Iterative MapReduce (as illustrated by Pregel and Twister) are a basis to address the above challenges as they interpolate between the traditional tightly coupled MPI jobs typical of supercomputers, and the more loosely coupled information retrieval and pleasingly parallel ("map only") applications typical of clouds and high throughput systems. It will provide an unified programming paradigm and prove very important for data analysis of science applications.

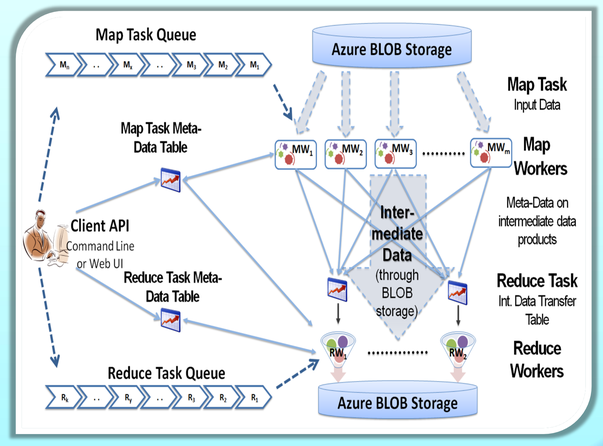
Scientific Applications

Iterative MapReduce Runtime

Twister
Cross Platform Iterative MapReduce

Twister4Azure
Iterative MapReduce for Azure Cloud





Large Scale Infrastructure

HPC Cluster

Windows Azure

amazon web services

INDIANA UNIVERSITY BLOOMINGTON

SALSA HPC