**Extending Twister4Azure and Integration with Apache Big Data Stack**

**Microsoft Azure Request from SALSA Group at Indiana University**

*Judy Qiu, Assistant Professor in Computer Science at Indiana University Bloomington*

In previous research, we have shown that Iterative MapReduce is a powerful programming model for data intensive applications on both cloud and HPC platforms [[1](#_ENREF_1), [2](#_ENREF_2)]. With a previous Azure grant, we have developed Twister4Azure [[3](#_ENREF_3)] that is an implementation of these ideas on Azure. In a recent report [[4](#_ENREF_4)] illustrated in fig. 1, we showed this performed significantly better than Hadoop on several data mining applications including clustering and multi-dimensional scaling MDS. Our recent work has focused on a generalization of MapReduce to a Map-Collective model [[5](#_ENREF_5)] where the “reduce” phase in MapReduce is supported by a library of powerful optimized collective communication routines covering operations like (all)reduce, scatter, gather, broadcast, regroup, combine, merge which cover the key primitives in MapReduce and MPI. In the last report, we presented optimal Azure collectives for reduce and gather and successfully applied to clustering and MDS. We also showed that the same collectives could be added to Hadoop with significant performance increase. Separately we added optimized broadcast collectives to the HPC implementation of Twister to enable clustering with millions of centers. We have recently integrated these ideas and request Azure time to perform the cloud based Map-Collective research.

*Fig. 1: Comparison of new Collectives in Twister4Azure compared original version with partial optimization*

Our new approach [[6](#_ENREF_6)] is motivated by the research described above combined with the growing importance of the Apache Big Data Stack (ABDS) illustrated in fig. 2. This open source activity (integrated by the Apache Bigtop project) has both a broad and deep set of capabilities shown in the figure spanning the deployment to the analytics layer. Our new work Harp adds a collective layer to the stack and we have already successfully integrated this with Hadoop so we can now take our successful work with Twister and offer it with a Hadoop interface. This figure shows another aspect of our architecture which is to span cloud and HPC environments. We show this at the file system, scheduling and collective layers.

In more detail, Harp is a collective communication library which can be used from Hadoop and with the Harp plug-in, Map-Reduce jobs can be transformed to Map-Collective jobs and users can invoke efficient in-memory collective communication directly in Map tasks. Five key features of Harp are:

1. Hierarchal data abstraction
2. Collective communication model
3. Resource pool based Memory management
4. BSP style computation parallelism with multi-thread support
5. Fault tolerance with check-pointing relying on HDFS

We expect our research to have greater impact if integrated with ABDS plus we can tackle many applications that assemble solutions from multiple ABDS components. We initially want to take our Harp work on HPC and research how best to use on Azure. Previously we found aspects of Azure Platform as a Service (queues, table, Windows communication) foundation), helpful in Twister4Azure. In general one expects such “polymorphic implementations” in Harp with different platforms needing different implementation strategies. We will continue our application oriented research testing ideas on K-means Clustering, Multi-Dimensional Scaling, Page-Rank
and Fruchterman-Reingold graph layout.

*Fig. 2: Apache Big Data Stack ABDS with enhanced layers for HPC-Cloud integration*

**Time Request for 12 months**

We presentthe Azure usage estimate in table below divided by usage mode based on our previous Azure use. **Large scale performance** tests include the periodic experiments to evaluate the performance, scalability and reliability of the Azure runtime frameworks and the associated applications we develop (Twister4Azure, MRRoles4Azure, Harp) as well as HDInsight in large deployments. These tests also include the experiments we perform for the publication purposes. **Unit tests** include the functionality tests after each major change and the experiments we plan on performing to evaluate the different implementation options/alternatives to improve performance, reliability and the scalability of our Azure runtime frameworks. **Application execution** include the execution of applications implemented using the above mentioned Azure runtime frameworks for science purposes. **Broader impact** activities include the Azure usage in tutorials, training and coursework that we plan on conducting to introduce the above mentioned frameworks to the users. Users will be given hands on tutorials using the compute time allocated under this category.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Usage Mode** | **Number of Cores** | **Hours Per Occurrence** | **Occurrences Per Month** | **Total core-Hours** |
| **Large Scale Performance Tests (Twister4Azure, Harp)** | 512 | 12 | 1 | 73728 |
| **Unit Tests (for each major change)** | 256 | 10 | 1 | 30720 |
| **Application Execution** | 128 | 2 | 8 | 24576 |
| **Development & Testing** | 64 | 2 | 20 | 30720 |
| **Broader impact activities** | 200 | 10 | 1 | 24000 |
| **Total Azure Compute Hours** |   |   |  | 183924 |

We also need 200GB of “out” data transfer per month, 1 terabyte of stored data and 50,000 Storage Transactions (Queue, Table and Blob operations) measured in units of 100K.

**References**

1. J.Ekanayake, H.Li, B.Zhang, T.Gunarathne, S.Bae, J.Qiu, and G.Fox, *Twister: A Runtime for iterative MapReduce*, in *Proceedings of the First International Workshop on MapReduce and its Applications of ACM HPDC 2010 conference June 20-25, 2010*. 2010, ACM. Chicago, Illinois. <http://grids.ucs.indiana.edu/ptliupages/publications/hpdc-camera-ready-submission.pdf>.

2. Bingjing Zhang, Yang Ruan, Tak-Lon Wu, Judy Qiu, Adam Hughes, and Geoffrey Fox, *Applying Twister to Scientific Applications*, in *CloudCom 2010*. November 30-December 3, 2010. IUPUI Conference Center Indianapolis. <http://grids.ucs.indiana.edu/ptliupages/publications/PID1510523.pdf>.

3. Thilina Gunarathne, Bingjing Zhang, Tak-Lon Wu, and Judy Qiu, *Scalable Parallel Computing on Clouds Using Twister4Azure Iterative MapReduce*  Future Generation Computer Systems 2013. 29(4): p. 1035-1048. <http://grids.ucs.indiana.edu/ptliupages/publications/Scalable_Parallel_Computing_on_Clouds_Using_Twister4Azure_Iterative_MapReduce_cr_submit.pdf>

4. Thilina Gunarathne, Thomas Wiggins, and Judy Qiu, *Collective Communication Patterns for Iterative MapReduce Report to Microsoft on Azure use*. October 11, 2013. <http://grids.ucs.indiana.edu/ptliupages/publications/MicrosoftReport_Collective_Communication.pdf>.

5. Bingjing Zhang, Judy Qiu, Stefan Lee, and David Crandall, *Large-Scale Image Classification using High Performance Clustering*. November 18, 2013. <http://grids.ucs.indiana.edu/ptliupages/publications/CCGRID2014_TLarge-Scale%20Image%20Classification%20using%20High%20Performance%20Clustering.pdf>.

6. Thilina Gunarathne, Judy Qiu, and Dennis Gannon, *Towards a Collective Layer in the Big Data Stack*. November 18, 2013. <http://grids.ucs.indiana.edu/ptliupages/publications/CCGRID2014_Towards_a_Collective_Layer_in_the_Big_Data_Stack.pdf>.