

Big Workflow: The Future of Big Data Computing

Robert Clyde, Adaptive Computing



How can organizations embrace — instead of brace for — the rapidly intensifying collision of public and private clouds, HPC environments and Big Data? The current go-to solution for many organizations is to run these technology assets in siloed, specialized environments. This approach falls short, however, typically taxing one datacenter area while others remain underutilized, functioning as little more than expensive storage space.

As larger and more complex data sets emerge, it becomes increasingly more difficult to process Big Data using on-hand database management tools or traditional data processing applications. To maximize their significant investments in these datacenter resources, companies must tackle Big Data with "Big Workflow," a term we've coined at Adaptive Computing to describe a comprehensive approach that maximizes datacenter resources and streamlines the simulation and data analysis process.

Big Workflow utilizes all available resources within the datacenter, including HPC environments, as well as other datacenter resources like private and public cloud, Big Data, virtual machines and bare metal. Under the Big Workflow umbrella, all datacenter resources are optimized, eliminating the logjam and turning it into an organized workflow that greatly increases throughput and productivity.

Looking at the industry, we are currently seeing state-of-the-art tools like OpenStack and Hadoop being used for Big Data processing. Late last year, we joined the OpenStack Community and announced an integration on OpenStack. In addition, we integrated Moab and Intel HPC Distribution for Apache Hadoop software, a milestone in the Big Data ecosystem that allows Hadoop workloads to run on HPC systems. Now, organizations have the ability to expand beyond a siloed approach and leverage both their HPC and Big Data investments together.

According to a Big Workflow survey we recently conducted, customized applications are primarily used to analyze Big Data. Among the approximately 400 survey takers — who were a mix of managers, administrators and users in a number of verticals, from education to technology and financial services — 83 percent believe Big Data analytics are important to their organization or department. However, 90 percent would have greater satisfaction from a better analysis process and 84 percent have a manual process to analyze Big Data.

The need for Big Workflow is best illustrated in comparing traditional IT workloads with Big Data workloads. Traditional IT workloads run forever, Big Data workloads run to completion. IT workloads require many apps per server, Big Data workloads require many servers per app. IT workloads do not need scheduling, scheduling is crucial for Big Data workloads. IT workloads only demand a light compute and data load, Big Data workloads are compute- and data-intensive.

The list goes on, but as is evident, siloed environments with no workflow automation to process simulations and data analysis fall short in their ability to extract game-changing information from data. Big Data analysis requires the supercomputing capabilities provided by HPC combined with scheduling and optimization software that can manage countless jobs over multiple environments simultaneously. This enables enterprises to leverage HPC while optimizing their existing diverse infrastructure.

With the wealth of data organizations are now accumulating, the key for any analytics application is to deliver results more rapidly and accurately. To achieve this, we forecast that more organizations will take a Big Workflow approach to Big Data and accelerate the time to discovery.

The Enterprise is advancing simulation and Big Data analysis to pursue game-changing discoveries, but these discoveries are not possible without Big Workflow optimization and scheduling software. Big Data, it's time to meet Big Workflow.

Robert Clyde is CEO of [Adaptive Computing](http://www.adaptivecomputing.com) [1]. He may be reached at editor@ScientificComputing.com.

Source URL (retrieved on 05/25/2016 - 7:05am):

<http://www.scientificcomputing.com/blogs/2014/03/big-workflow-future-big-data-computing>

Links:

[1] <http://www.adaptivecomputing.com/>