Harnessing Next Generation Software Defined Infrastructure For Scientific Applications

Objectives. Infrastructure underutilization coupled with management complexity is leading an unprecedented transformation of modern datacenters through the adoption of hyper-converged server infrastructure utilizing software-defined solutions. These software defined data centers (SDDCs) leverage abstraction of hardware unilaterally by virtualization all layers of the system including network, disk, and compute resources. This research will investigate scientific application performance implications of hyperconverged virtualization architectures to characterize the tradeoff space. The primary goal is to identify how various virtual architecture configurations for compute, network, and storage resource impact performance based on the unique resource requirements for batch oriented scientific workloads. This research will enable the development of a recommendation system that classifies application workloads to suggest and drive SDDC configuration. Two levels of infrastructure configuration and management will be considered: (1) SDDC configuration - where characteristics of predominant applications are considered to guide the SDDC implementation, and (2) SDDC utilization – where intelligent load balancing will serve to maximize the utilization of the SDDC to optimize application performance while minimizing resource costs.

Methods. (*Resources*) This research will be supported through partnership with the NSF Chameleon and CloudLab public clouds where baremetal OpenStack cloud resources can be harnessed to define and study SDDCs for experimental work. Additionally industry collaboration will be sought to acquire access to enable investigation using one or more major hyperconverged server infrastructure platforms from vendors such as: Nutanix, Simplivity, Maxta, or Stratoscale.
(*Applications*) This research will leverage large-scale distributed science application workloads including University of Washington science applications such as the Rosetta protein folding research, and RNA-seq gene sequencing and analysis of mass spectrometry data and computational proteomics. Additionally hydrological and soil erosion modeling applications from the US Department of Agriculture provide additional scientifc workloads for measurement and analysis.
(*Experiments*) This research will outline a number of experiments where performance of science application workloads will be benchmarked across a variety of SDDC systems. Statistical analysis and regression will be used to characterize performance implications based on application workload resource requirements and SDDC configuration. Workload classification will be used to recognize characteristics implicating performance to improve SDDC configuration and load balancing approaches to optimize workload performance while minimizing resource requirements.

Intellectual Merit. While the implications of virtualization overhead have been investigated for scientific applications in depth at each individual layer, little work has been done at considering the overall implications of the use of new hyper-converged server infrastructure. This research will deploy, benchmark, analyze, and model the implications of software defined infrastructure for batch-oriented scientific application workloads to develop a greater understanding of how this new technology impacts performance. This research will develop insights on how hyper-converged infrastructure can be best serve the unique demands of diverse scientific application workloads: Rosetta (distributed-compute bound), RNA-seq (distributed compute & big data), and hydrological/soil erosion modeling (large number of concurrent transactions).

Broader Impacts. Hyper-converged server infrastructure is positioned to redefine how private data centers of the future are conceived and built. They leverage virtualization at every layer to abstract modular server infrastructure to simplify management and scalability. The costs and implications of this abstraction can negatively impact application performance and fine-grained infrastructure control leading to increased application overhead and under utilization of resources. Hyper-converged data center architectures must be studied to characterize how this full hardware abstraction implicates system analysts, scientists, and anyone who relies on the use of private clouds and local data centers in their daily lives.