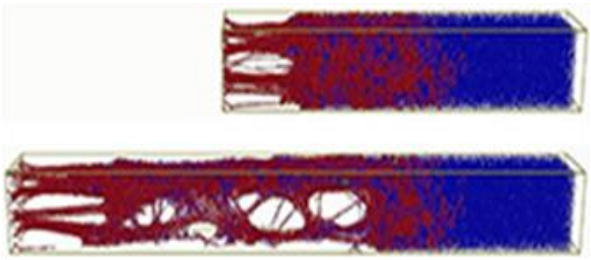
HPC Cloud Performance of Peptide Benchmark Using LAMMPS Molecular Dynamics

An UberCloud Experiment



TECH-X
SIMULATIONS EMPOWERING YOUR INNOVATIONS

With Support From:

This project has been sponsored by the Government's Small Business Innovation Research (SBIR) Program under Contract Number: DE-SC0017111 DOE-SBIR ASCR Topic: 13.a

UberCloud Case Study 199

http://www.TheUberCloud.com

March 03, 2018

Welcome!

The UberCloud* Experiment started in July 2012, with a discussion about cloud adoption in technical computing and a list of technical and cloud computing challenges and potential solutions. We decided to explore these challenges further, hands-on, and the idea of the UberCloud Experiment was born, then also due to the excellent support from INTEL generously sponsoring these experiments since the early days!

We found that especially small and medium enterprises in digital manufacturing would strongly benefit from technical computing in HPC centers and in the cloud. By gaining access on demand from their desktop workstations to additional and more powerful compute resources in the cloud, their major benefits became clear: the **agility** gained by shortening product design cycles through shorter simulation times; the superior **quality** achieved by simulating more sophisticated geometries and physics and by running many more iterations to look for the best product design; and the **cost** benefit by only paying for what is really used. These are benefits that obviously increase a company's innovation and competitiveness.

Tangible benefits like these make computing - and more specifically technical computing as a service in the cloud - very attractive. But how far are we from an ideal cloud model for engineers and scientists? At first we didn't know. We were facing challenges like security, privacy, and trust; traditional software licensing models; slow data transfer; uncertain cost & ROI; lack of standardization, transparency, cloud expertise. However, in the course of this experiment, as we followed each of the 199 teams closely and monitored their challenges and progress, we've got an excellent insight into these roadblocks, how our teams have tackled them, and how we are now able to reduce or even fully resolve them.

UberCloud Team 199 studied adhesion dynamics for surface-tethered chains entangled in a polymer melt, using the molecular dynamics package LAMMPS which is widely used in academia and some industries. LAMMPS has potentials for solid-state materials (metals, semiconductors) and soft matter (biomolecules, polymers) and coarse-grained or mesoscopic systems. It can be used to model atoms or, more generically, as a parallel particle simulator at the atomic, meso, or continuum scale. The simulations have been performed on up to eight 16-core high-performance computing nodes in the Amazon AWS Cloud.

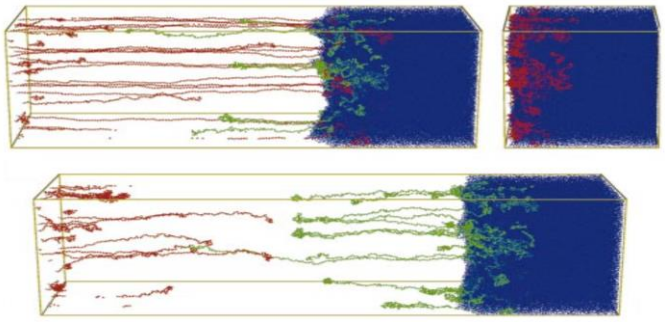
We want to thank all team members for their continuous commitment and contribution to this project. And we want to thank our main Compendium sponsors **Hewlett Packard Enterprise** and **INTEL** for generously supporting all 199 UberCloud experiments.

Now, enjoy reading!
Wolfgang Gentzsch and Burak Yenier

**) UberCloud is the online community & marketplace where engineers and scientists discover, try, and buy Computing Power as a Service, on demand. Engineers and scientists can explore and discuss how to use this computing power to solve their demanding problems, and to identify the roadblocks and solutions, with a crowd-sourcing approach, jointly with our engineering and scientific community. Learn more at: http://www.TheUberCloud.com.*

Team 199

HPC Cloud Performance of Peptide Benchmark Using LAMMPS Molecular Dynamics Package



"HPC software container-based cloud computing is an easy process compared to building and maintaining your own cluster in the cloud."

Figure 1: *Simulation snapshots using LAMMPS, studying adhesion dynamics for surface-tethered chains entangled in a polymer melt.*

MEET THE TEAM

End User – National Renewable Energy Lab (NREL), Tech-X Research
Software Provider – LAMMPS open source software and Steven J. Plimpton (Sandia National Lab)
Resource Provider – Amazon Web Services (AWS)
HPC Expert – Dr. Scott W. Sides, Senior Scientist, Tech-X Research Boulder, CO.

USE CASE

In order to address realistic problems in the nanomaterials and pharmaceutical industries, large-scale molecular dynamics (MD) simulations must be able to fully utilize high-performance computing (HPC) resources. Many small- and medium-sized industries that could make use of MD simulations do not use HPC resources due to the complexity and expense of maintaining in-house computing clusters.

Cloud computing is an excellent a way of providing HPC resources to an underserved sector of the simulation market. In addition, providing HPC software containers with advanced application software can make the use of these codes more straightforward and further reduce the barriers for entry to small- and medium-sized businesses.

The molecular dynamics package LAMMPS is widely used in academia and some industries. LAMMPS has potentials for solid-state materials (metals, semiconductors) and soft matter (biomolecules, polymers) and coarse-grained or mesoscopic systems. It can be used to model atoms or, more generically, as a parallel particle simulator at the atomic, meso, or continuum scale.

The cloud computing service provider, Amazon Web Services, provided a number of virtual machines each with up to 16 cores for this experiment with different levels of network communication performance.

Technical Details of the Simulation

Figure 2 shows the parallel scaling performance of LAMMPS containers running on an AWS multi-node cluster with each of the nodes having 16 cores available. A simple peptide chain model that is included in the tests for LAMMPS was used for performance scaling. The initial peptide input file only contains 2004 particles, but using the 'replicate' keyword available in LAMMPS the initial simulation cell may be copied in the x,y,z directions an arbitrary number of times.

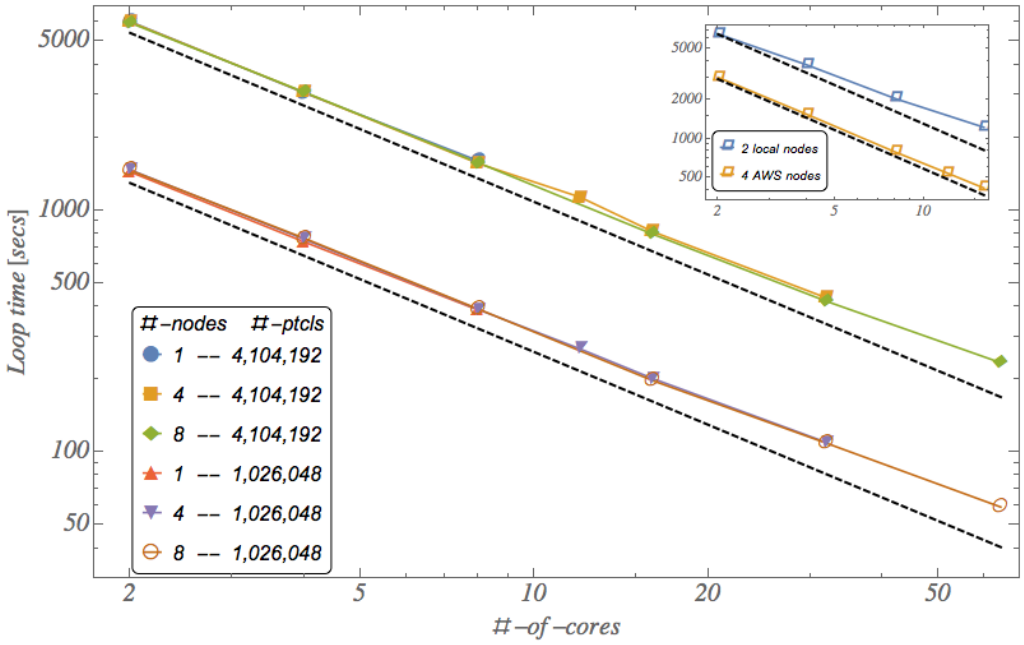


Figure 2: *LAMMPS parallel scaling performance on an AWS multi-node cluster with each of the nodes having 16 cores available. Inset upper right: Comparison of the parallel scaling performance between LAMMPS running on the bare-metal 2-node test cluster at Tech-X and LAMMPS containers running on a 4-node remote AWS cluster. The dotted lines indicate the optimal scaling behavior, showing that the performance of the LAMMPS containers running in the cloud is excellent.*

The simulations in Figure 2 show two system sizes using ≈ 10^6 and ≈ $4.1*10^6$ particles run for 300 update steps for reasonable timing statistics. The inset in the upper right shows a comparison of the parallel scaling performance for a system with ≈ $2.0*10^6$ particles between LAMMPS running on the bare-metal 2-node test cluster at Tech-X and LAMMPS containers running on a 4-node remote AWS cluster. The dotted line in the main figure and inset is the optimal scaling trend. The main figure shows that the LAMMPS multi-node container performance persists as the number of nodes in the cloud cluster increases. There was degraded performance when the number of processors/node reaches the maximum number of cores available as listed by AWS and is due to hyper-threading. But, there appears to be no degradation of performance as the size of the cluster increased, suggesting that an arbitrary number of processors can be used for HPC molecular dynamics simulations using LAMMPS in the cloud.

Summary of the SBIR project

This cloud experiment was initially funded as part of a Small Business Innovation Research (SBIR) grant. The solicitation called for enabling modern materials simulations in a larger sector of the industrial research community. High performance computing (HPC) is a technology that plays a key role in materials science, climate research, astrophysics, and many other endeavors. Numerical simulations can provide unique insight to physical phenomena that cannot be easily obtained by other means. Numerical simulations complement experimental observations, help in validating models, and advance our understanding of the world. Advances in HPC software development and algorithms are becoming increasingly important in materials science and for industries developing novel materials. According to a recent survey by the US Council on Competitiveness, faster time to market, return on investment, and enabling work that could not be performed by any other means are cited as the most common justifications for using HPC in industry. For instance, Goodyear was able to significantly reduce the time to bring new tires to market through a collaboration with Sandia National Laboratory by leveraging high performance clusters. The oil, aeronautic, and automobile industries are examples of big industries where HPC technologies have been leveraged for decades. The growing penetration of HPC into engineering fields has been fueled by the continued performance improvements of computer chips as well as the emergence of hardware accelerators such as general-purpose graphics processing units (GPUs) and the Intel Xeon Phi co-processor (also known as many integrated core architecture, or MIC).

However, one of most striking features of the US Council on Competitiveness survey, is how underrepresented are the companies that would be most likely to take advantage of soft materials simulations. The biosciences sector accounted for only 5.9% and the chemical engineering sector accounted for only 4.0% of respondents on their use of HPC resources. The Phase I SBIR proposal granted to Tech-X addresses this call and the two issues outlined above, by using an extensible object-oriented toolkit (STREAMM) for linking quantum chemistry (DFT) and classical molecular dynamics (MD) simulations and making this code suite available to take advantage of HPC cloud computing.

Process Overview

1. Kickoff team meeting of the experiment using WebEx.
2. Organization of project tasks, communication and planning through RedMine.
3. The end user, Scott Sides, obtained an AWS account and provided ssh-keys to UberCloud in order to setup a project specific security group that is used to configure the multi-node multi-container environment.
4. A specialized installer was created for LAMMPS and made available to the team.
5. The end user performed an MD scaling study on a 1-node, 4-node, and 8-node cluster.
6. The end user analyzed performance data and communicated the results to the rest of the team.

CHALLENGES

End user perspective - The cloud computing service at Amazon Web Services (AWS) provided high-quality compute nodes with efficient communication networks that enabled the good scaling seen in Figure 2. There is quite a bit of manual setup that needs to be performed by the end-user for AWS. For any cloud computing project, the first step is to create the remote compute instances. One must apply for an account at AWS, and use the AWS web interface to navigate to the services for the Elastic Compute Generation 2 (EC2). The 'elastic' refers to the ability to expand or shrink the hardware usage for a particular task at a given time. Then the desired number, type and security settings for the EC2 instances must be selected. For a first-time setup, an ssh-key pair is generated and stored within the user's account

information. The web interface instructs the user how to setup their local ssh configuration so that access to any remote AWS instance can be obtained. This procedure is straightforward but again, must currently be done manually. The security group must also be specified manually, and is one that is configured by UberCloud in order for the networking modules to function. Now the separate instances must be assembled and configured into a multi-node cluster.

The next steps are to copy setup applications, scripts and configuration files needed to install Docker, pull all needed Docker images, and start the computational images with all of the appropriate network configuration settings. The remote copy requires the DNS addresses generated by the AWS instance startup outlined above and must currently be performed manually. Then one of the compute instances must be designated as the 'Master' node which has two main purposes: (i) to run the 'Consul' container which is part of the framework that manages the network setup for all of the cluster instances and (ii) to provide a remote entry access point for the cluster. When launching simulations on this remote cloud cluster a user executes an SSH login command using the public IP address for the master node (again obtained manually through the AWS web tool) and a password that is automatically generated within the secure container and emailed to the user. These security measures are all part of the networking image layer in the UberCloud simulation containers. However, once these steps are in place, then running on a cloud cluster is much the same as running on an HPC cluster at a university or national lab.

BENEFITS

End user perspective

- Gained an understanding of the cloud computing philosophy and of what is involved in using a cloud-based solution for computational work.
- Cloud computing using novel HPC software containers based on Docker is an easy process compared to building and maintaining your own cluster and software environment.
- Developed an effective workflow for constructing additional HPC cloud containers.

CONCLUSIONS AND RECOMMENDATIONS

For the Phase II proposal based on this case study, Tech-X will add additional codes to the UberCloud marketplace for targeted industries and applications including those in nanotech and the pharmaceutical industries. We will also investigate ways to add functionality to our STREAMM framework to streamline the setup steps described in the ‘end-user perspective’ section. We will also check all our current scaling results on the Microsoft Azure cloud platform and compare with AWS and bare-metal. The Azure setup is reported to have ways of streamlining the setup process to make utilizing cloud HPC resources even easier.

Case Study Authors – Dr. Scott W Sides and Wolfgang Gentzsch

Thank you for your interest in the free and voluntary UberCloud Experiment

If you, as an end-user, would like to participate in an UberCloud Experiment to explore hands-on the end-to-end process of on-demand Technical Computing as a Service, in the Cloud, for your business then please register at: http://www.theubercloud.com/hpc-experiment/.

If you, as a service provider, are interested in building a SaaS solution and promoting your services on the UberCloud Marketplace then please send us a message at https://www.theubercloud.com/help/.

2013 Compendium of case studies: https://www.theubercloud.com/ubercloud-compendium-2013/
2014 Compendium of case studies: https://www.theubercloud.com/ubercloud-compendium-2014/
2015 Compendium of case studies: https://www.theubercloud.com/ubercloud-compendium-2015/
2016 Compendium of case studies: https://www.theubercloud.com/ubercloud-compendium-2016/

The UberCloud Experiments received several international Awards, among other:

- HPCwire Readers Choice Award 2013: http://www.hpcwire.com/off-the-wire/ubercloud-receives-top-honors-2013-hpcwire-readers-choice-awards/
- HPCwire Readers Choice Award 2014: https://www.theubercloud.com/ubercloud-receives-top-honors-2014-hpcwire-readers-choice-award/
- Gartner Cool Vendor Award 2015: http://www.digitaleng.news/de/ubercloud-names-cool-vendor-for-oil-gas-industries/
- HPCwire Editors Award 2017: https://www.hpcwire.com/2017-hpcwire-awards-readers-editors-choice/
- IDC/Hyperion Research Innovation Excellence Award 2017: https://www.hpcwire.com/off-the-wire/hyperion-research-announces-hpc-innovation-excellence-award-winners-2/

If you wish to be informed about the latest developments in technical computing in the cloud, then please **register at http://www.theubercloud.com/** and you will get our free monthly newsletter.

TECH-X
SIMULATIONS EMPOWERING YOUR INNOVATIONS

With Support From:

UberCloud
Hewlett Packard Enterprise
intel
amazon web services™