International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 10
October 2017

Containers On Cloud for Machine Learning Applications

S.H. Chandak, S.C. Dharmadhikari, R.R. Chhajed
Pune Institute of Computer Technology, Pune

ABSTRACT :

Containers are a lightweight virtualization solution to replace virtual machines for deploying cloud applications as they are less resource and time consuming. Machine learning paradigms facilitate pattern recognition and computational learning for automation in systems beneficial to several organizations using computationally expensive resource dependent algorithms. Containers are used to create portable machine learning applications in resource constraint scenarios. Easy deployment of applications on containers is done using FlexTuner which helps user to analyze applications with various network topologies improving transfer rate of large data sets which are required for machine learning models. CryptoML framework uses dimension reduction technique for optimized processing of such large data sets and Shamir's secret sharing for preserving the privacy and security of client's data. Deployment of applications incorporating machine learning on containers is an effective approach for coping up with acceleration in the amount of data collected and business demand for automation with resource constraint.

KEYWORDS :

containers; virtualization; machine learning; cloud; big data; FlexTuner; CryptoML; Shamir's secret sharing

I. INTRODUCTION

Virtual machines (VMs) are an infrastructure as a service (IaaS) focusing on hardware virtualization whose techniques have been used at the infrastructure layer. To achieve sharing and elasticity of resources, the cloud makes use of such virtualization techniques for administering scheduling, provisioning and security. However, VMs suffer from slow start up time and exhibit lower densities even when full. Applications occupy entire host operating system (OS) in VMs and there are also several limitations on storage and

Containers offer as a lightweight virtualization solution to deploying applications across various domains and sectors. They allow infrastructure and platform to be shared in a secure and portable manner, along with application packaging and management. They facilitate faster deployment of applications and have faster start up times. They are less resource and time consuming and can be scaled up or down providing higher density levels than full VMs. They facilitate easier and portable across infrastructure deployment of applications in an interoperable way. Using containers will accelerate agile application development of distributed applications providing an additional layer of protection by isolating applications and the host, without using incremental resources. They also allow easy updates to applications. Fig. 1 shows the difference between traditional hypervisor and container-based architectures.

This paper tried to convey the importance of container over virtualization   on cloud environment and improvement in the processing of Machine learning algorithms with containers. In this paper, we will start with introduction to container over virtualization, Literature review discuss relevance of container technology on cloud , how to deploy application on container through flextuner , machine learning and container, some of the security issues with machine learning and containers and CryptoML framework.

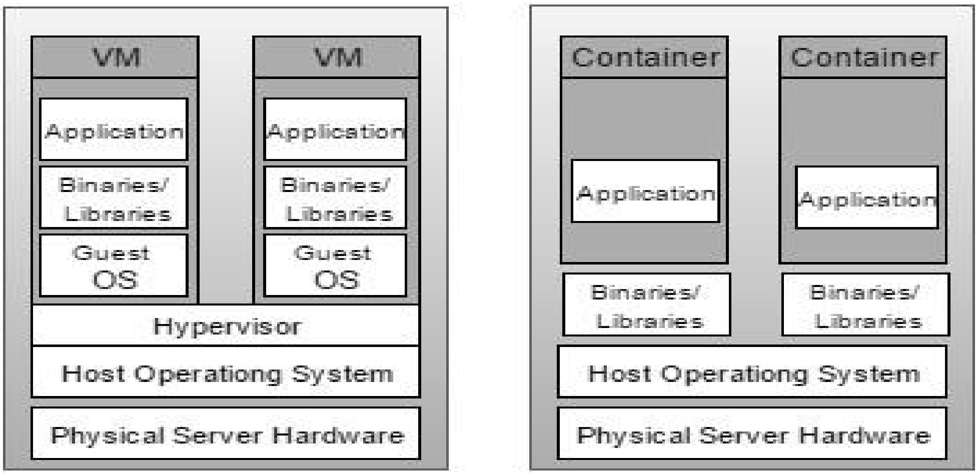

Fig.1: Traditional hypervisor architecture on the left and a container-based architecture on the right.

II. LITERATURE REVIEW

Claus Pahl [1] discusses the relevance of the new container technology for PaaS clouds and its ability to change the PaaS cloud as a virtualization technique along with the virtualization principles behind containers, particularly in comparison with VMs. It goes beyond Bernstein [2], addressing what's needed to evolve PaaS as a distributed cloud software platform and limitations of the state of the art. David Bernstein discusses the importance of container-based application deployment and cluster management for the cloud computing infrastructure.

Container migration algorithms using the Containers as a Service (CaaS) [3] framework decreases the number of running servers with the objective of minimizing the overall energy consumption of cloud data centers while meeting the Service Level Agreement (SLA). Virtual cells as a service (vCAAS) model [4] is a container-based model of cloud computing where all resources (virtual machines, storage, and interconnecting networks) and the logic to manage these resources are packaged in a virtual container and delivered to users and is similar to that presented by Musa IK and Stuart W [5].

The stable matching theory [6] is used to model the container placement problem and presents resource stable placement algorithm for solving the resources scheduling problem on container level using the stable matching theory. The flexible network addressing architecture [7] which performs M-to-N mapping between network addresses and containers using software defined networking (SDN) approach in order to simplify the network setup and configuration is contrasted to the existing network architectures like dotCloud.

FlexTuner [8] is discussed as a flexible tuning system for cloud applications on containers. It eases deployment of distributed applications on containers in clouds and helps to find appropriate routing algorithm and network topology for big data applications whose data sets are processed by machine learning algorithms by analyzing the communication cost. CryptoML [9] is discussed as a framework for optimizing the performance of machine learning algorithms on large data sets of applications deployed on clouds by making use of dimension reduction on matrices while preserving privacy and security of client's data with Shamir's secret sharing.

III. EASY DEPLOYMENT OF APPLICATION THROUGH CONTAINERS WITH FLEXTUNER

Popularity and trend of big data applications is significantly increased in recent times, due to rapid growth of datasets. Compared to traditional applications, big data applications spend a lot of time to transfer data among computing nodes. Once the system is deployed, network topology is fixed and routing protocols are preinstalled hence becomes immutable for user. Moreover, it becomes hard for application developers to identify the optimal network configuration for their applications with distinct communication patterns.

Flexible container-based tuning system (FlexTuner) [8] allows users to create a farm of containers on host machines. Additionally, SDN technique is used to connect and direct the network traffic among these containers. FlexTuner mainly consists of Linux container and SDN techniques. Linux container provides freedom of deployment of substantial number of guest OSes on single machine and virtual links and network is used to connect these OSes together. SDN act as an interface for developing personalized routing algorithms to limit the network traffic. As a result, FlexTuner provides user a facility to evaluate the applications' efficiency on the same system with different network topology and routing protocols.

Processes of application within different containers created by FlexTuner uses network stack rather than using shared memory or local queue to exchange data. Heavy middle-wares frameworks such as MapReduce and workflow, defines their own resource management and scheduler, hence application itself cannot finer tune the data movements. However, specific communication patterns can be created easily using MPI applications due to their full control over computing nodes. Therefore, FlexTuner systems consists of four components: network, naming service, container agent, and program launcher.

Containers are attached to virtual switches through virtual links. The virtual switches connect all the virtual links together to form a tree network topology. For each packet received from the attached virtual machines, the virtual switch first checks whether the packet matches an existing rule. If such rule is found, it delivers the packet according to that rule. Otherwise, the head of the packet will be sent to the controller, and the controller makes a decision based on this information and informs the switch about the decision through open-flow message. Then, the switch forwards the packet accordingly. POX is a platform for users to write and prototyping the network control programs using python.

To study the communication cost for various network topologies and routing algorithms, FlexTuner uses MPI program. MPICH2 is a widely used MPI implementation of the MPI standard. MPICH2 uses "Hydra" engine instead of "mpd ring" to start parallel MPI processes on different computing nodes. Hydra can use ssh, rsh, pbs, slurm and sge as the launcher to start MPI processes. FlexTuner containers must have ssh server running in the backend as MPI launcher. It is essential to allow other containers to access it through ssh client. In FlexTuner, each VM (container) starts "sshd" deamon in the backend to allow Hydra PM to manage MPI processes running on it via ssh launcher.

IV. MACHINE LEARNING APPLICATIONS AND CONTAINERS :

Machine learning (ML) has become a rudimentary tool for extracting structured/unstructured information and knowledge from contemporary massive dataset/raw data, which is then used for automatic predictions and remarkable hypotheses for different applications. With time, the amount of data collected and the complexity of tasks and problems is continually increasing. Optimization problems are one of the most difficult ones to solve. Machine learning is handy to solve such big issues. Only downside about the machine learning algorithms is that they are computationally expensive and heavily dependent on resources . Such high computational power is not available easily and requires scalable approaches in terms of performance and storage. Hence, to deploy ML algorithms we can make use of the cloud environment in order to provide solutions to resource dependent problems in resource constraint

Containers are light weight virtualization solution to solve these issues of very high computational needs. As containers allow easy deployment and development of applications, more complex architectures can be used to solve modern optimization problems

By combining the advanced technology of machine learning systems with the deployment capabilities of containers, you can make machine learning systems much more useful and shareable.

V. MACHINE LEARNING AND SECURITY IN CLOUD SERVERS : CryptoML

Though deployment of applications on cloud containers abstracts it from the underlying OS, they add additional complexity to it in terms of security and vulnerabilities. Applications which require processing of large data sets are heavily dependent on computational resources and there is an increasing need of applying machine learning paradigms to optimize the working of applications in a resource – constrained environment. In order to resolve the issues of data security and privacy along with its efficient resource optimization, CryptoML [9] is a framework for secure and efficient delegation of contemporary matrix-based ML applications processing large data sets. CryptoML uses flexible ML algorithms to reduce the computational workload providing enhanced efficiency and resource awareness. It secures the privacy of the client's data by not revealing the original large database and the solution to the ML problem to cloud service providers (CSPs).

In CryptoML, a client with resource constraints wishes to use cloud servers for storing ML computations on cloud while preserving the privacy of its data. CryptoML uses Delegation – optimized sparse sketching (DSS) as a scalable approach for preprocessing the large data sets of resource limited clients. Fig. 2 shows overview of CryptoML for preserving security.

This leads to producing of low-dimensional sparse factors by trading exact output solutions to ML problems with improvement in resource cost. To work around a solution to the limitation of resources while preserving security, client's data is uploaded on N different CSPs, where the client authenticates itself to each independent server with its separate personal account on each one of them before downloading or uploading the data. There is no need for CSPs to communicate with each other and security is ensured such that the adversary is not able to infer information from the client's delegated data or computations even with unbounded computing power. The secure outsourcing protocol is built on Shamir's secret sharing which reconstructs the information of the client's data from shares expressed by values with the help of polynomials such that the data is retrieved by the client privately.

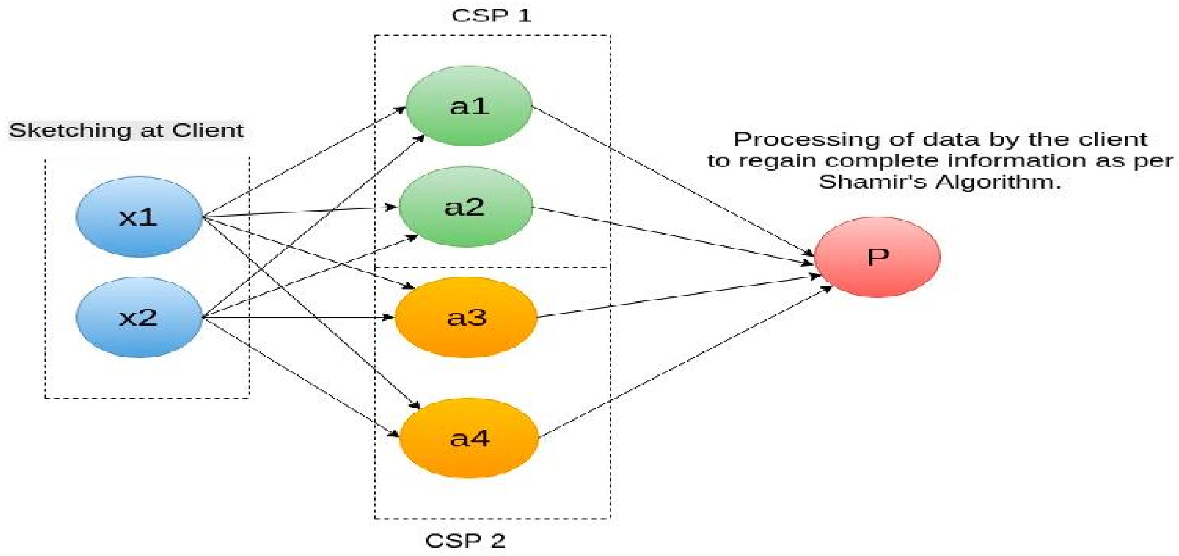
CSP 1
Sketching at Client
x1
x2
a1
a2
a3
a4
Processing of data by the client to regain complete information as per Shamir's Algorithm.
P
CSP 2

Fig.2: CryptoML applies DSS on ML algorithm data sets of the client which is uploaded on different CSPs. Client authentication is required to retrieve data privately by securely outsourcing it from these CSPs using Shamir's sharing security

VI. CONCLUSION

Container technology has huge potential to advance technology towards parallel and distributed heterogeneous clouds and improving the energy efficiency of cloud data centers for achieving power optimization. Optimization problems are one of the most difficult ones to solve with increasing size of data. Machine learning is handy to solve such big issues but with downside is that they are computationally expensive and heavily dependent on resources. Such high computational power is not available easily and requires scalable approaches in terms of performance and storage. Hence, to deploy ML algorithms we can make use of the cloud environment with containers in order to provide solutions to resource dependent problems in resource constraint scenarios and very high computational needs.. As containers allow easy deployment and development of applications, more complex architectures can be used to solve modern optimization problems. To protect large data sets of machine learning algorithm, CryptoML framework is helpful which uses Shamir's secret sharing to keep the data secured. Data flow between the various containers can be easily done by using FlexTuner.

REFERENCES

[1] C. Pahl, “Containerisation and the PaaS cloud,” Cloud Computing Magazine, vol. 2, no. 3 May/June, 2015.

[2] D. Bernstein, “Containers and Cloud: From LXC to Docker to Kubernetes,” IEEE Cloud Computing, vol. 1, no. 3, 2014, pp. 81–84.

[3] Sareh Fotuhi Piraghaj, Amir Vahid Dastjerdi, Rodrigo N. Calheiros, and Rajkumar Buyya, "A Framework and Algorithm for Energy Efficient Container Consolidation in Cloud Data Centers," 2015.

[4] Musa et al.: Self-service infrastructure container for data intensive application. Journal of Cloud Computing: Advances, Systems and Applications 2014 3:5, 2014.

[5] Musa IK, Stuart W (2014) Multi objective optimization strategy suitable for virtual cells as a service. In: Innovations in bio-inspired computing and applications. Springer, pp 49–59

[6] Xin Xu, Huiqun Yu, Xin Pei, "A Novel Resource Scheduling Approach in Container Based Clouds," IEEE, 2014.

[7] Kyung-Hwa Kim, Jae Woo Lee, Michael Ben-Ami, Hyunwoo Nam, Jan Janak and Henning Schulzrinne, "Flexible Network Address Mapping for Container-based Clouds," IEEE, 2015.

[8] Y. Yu, H. Zou, W. Tang, L. Liu, and F. Teng, FlexTuner: A Flexible Container-based Tuning System for Cloud Applications, IEEE International Conference on Cloud Engineer (IC2E 2015), 2015.

[9] A. Mirhoseini, A. Sadeghi and F. Koushanfar, "CryptoML: Secure outsourcing of big data machine learning applications", 2016 IEEE International Symposium on Hardware Oriented Security and Trust (HOST), 2016.

[10] "Use containers and machine learning to deploy portable, smart apps", TechBeacon, 2016 [Online].

[11] Eric P. Xing Qirong Ho Pengtao Xie Dai Wei "Strategies and principles of distributed machine learning on big data" <em>Engineering</em> vol. 2 no. 2 pp. 179 2016.