

ARMS-CC-2014

Workshop on Adaptive Resource Management
and Scheduling for Cloud Computing

Held in conjunction with PODC 2014
(ACM Symposium on Principles of Distributed Computing)

Paris, France, on July 15th, 2014



<http://arms-cc.hpc.pub.ro>

<http://www.podc.org>

Welcome

We have a great pleasure to invite you to participate to ARMS-CC 2014, Workshop on Adaptive Resource Management and Scheduling for Cloud Computing.

The workshop will be held on July 15th 2014 at Laboratoire d'Informatique de Paris 6 (LIP6) and Laboratoire Jacques Louis Lions (LJLL) at Tower 26, Tower 15, Campus de Jussieu, 4 place Jussieu, 75005, Paris

ARMS-CC-2014 is organized by University Politehnica of Bucharest, Romania, and UPMC Sorbonne Universités, LIP6, Paris, France

Cloud Computing concerns large-scale interconnected systems and it has the main purpose to aggregate and to efficient exploit the power of widely distributed resources. Resource Management and Task Scheduling play an essential role, in cases where one is concerned with optimized use of resources. The ubiquitous networks are highly dynamic distributed systems so the changes in overlay are frequent. On the other hand, the Cloud systems are highly dynamic in its structure because the user requests must be respected as an agreement rule. When ubiquitous networks become clients for Cloud systems new algorithm for events and tasks scheduling and new methods for resource management should be designed in order to increase the performance of such systems. The adaptive methods used in context are oriented on: self-stabilizing, self-organizing and autonomic systems; dynamic, adaptive and machine learning based distributed algorithms; fault tolerance, reliability, availability of distributed systems.

The mission of the first edition of ARMS-CC workshop is to bring together researchers, engineers, and practitioners involved in the development of new resource management methods, scheduling algorithms, and middleware tools for Cloud Computing. The goal is to provide an interactive and friendly yet professional forum for original research contributions describing novel ideas, groundbreaking results or quantified system experiences, in the context of PODC Symposium. The work presented in ARMS-CC workshop can report and summaries on previous work, present early new results or put forward new and outrageous ideas.

This is the first year when ARMS-CC workshop is organized. The organization institutions are UPMC Sorbonne Universités, LIP6, Paris, France and University Politehnica of Bucharest, Romania. ARMS-CC workshop is organized in the context of PHC Bilateral Research Project: *SideSTEP – Scheduling Methods for Dynamic Distributed Systems: a self-* approach*, ID: PN-II-CT-RO-FR-2012-1-0084.

ARMS-CC Highlights

ARMS-CC highlights the consideration of self-* improvements, which is explicitly focused on resource utilization and combines local and global enforcement solutions, supports the estimation of resources needed, scalability, mobility, reliability, and fault tolerance.

- ARMS-CC covers topics about Foundational Models for Resource Management in Cloud, Distributed Scheduling Algorithms, Load-Balancing and Co-Allocation, Dynamic, Adaptive and Machine Learning based Distributed Algorithm, Many-Task Computing in the Cloud, Self-* and Autonomic Cloud Systems, Cloud Composition, Federation, Bridging, and Bursting, Cloud Resource Virtualization and Composition, Fault Tolerance, Reliability, Availability of Cloud Systems, Cloud Workload Profiling and Deployment Control, Cloud Quality Management and Service Level Agreement (SLA), High Performance Cloud Computing, Mobile Cloud Computing, Green Cloud Computing, Cloud Computing Techniques for Big Data, Scheduling Algorithms for Big Data.
- We are proud to announce that the first edition of ARMS-CC will publish all accepted papers post-proceeding with the support of Springer LNCS. All accepted papers to ARMS-CC will be invited to be extended for Special Issue on “Advanced Topics in Resource Management for Ubiquitous Cloud Computing: an Adaptive Approach (ARMCO)” organized by Future Generation Computer Systems (FGCS).
- We also express our gratitude and thank to all of the members of the program committee, to all of the reviewers, for their hard work in finalizing the reviews on time, as well as the authors for submitting their papers to ARMS-CC-2014.
- This edition of ARMS-CC workshop received 29 paper submissions from all over the world (Austria, Brazil, China, France, Greece, India, Iran, Ireland, Italy, Korea, Netherlands, Pakistan, Romania, Spain, Sweden, Switzerland, United Kingdom, United States). Each submission was peer-reviewed by program committee members or invited external reviewers. Finally, 14 high quality papers were selected (about 48% acceptance ratio) for publishing in the LNCS Post-Proceedings and presented during the conference.
- We hope you will enjoy the workshop together with us and have a great time in Paris and also in France.

ARMS-CC Chairs

Florin Pop, University Politehnica of Bucharest, Romania, Email: florin.pop@cs.pub.ro

Maria Gradinariu Potop-Butucaru, UPMC Sorbonne Universités, LIP6, Paris, France, Email: maria.gradinariu@lip6.fr

General Chairs

Florin Pop

University Politehnica of Bucharest, Romania, Email: *florin.pop@cs.pub.ro*

Maria Gradinariu Potop-Butucaru

UPMC Sorbonne Universités, LIP6, Paris, France, Email: *maria.gradinariu@lip6.fr*

Program Committee

Ajith Abraham, Machine Intelligence Research Labs (MIR Labs), USA

Marcos Aguilera, Microsoft Research Silicon Valley, USA

Silvia Bonomi, Università degli studi di Roma “La Sapienza”, Italy

Alexandru Costan, INRIA / INSA Rennes, France

Valentin Cristea, University Politehnica of Bucharest, Romania

Ciprian Dobre, University Politehnica of Bucharest, Romania

Pascal Felber, Université de Neuchâtel, Switzerland

Adriana Iamnitchi, University of South Florida, USA

Thilo Kielmann, VU University (Vrije Universiteit) Amsterdam, The Netherlands

Joanna Kolodziej, Institute of Computer Science, Cracow University of Technology, Poland

Young Choon Lee, The University of Sydney, Australia

Vivien Quéma, Grenoble INP / ENSIMAG, France

Iosif Charles Legrand, California Institute of Technology, USA / CERN, Switzerland

Guillaume Pierre, IRISA / Université de Rennes 1, France

Ioan Raicu, Illinois Institute of Technology (IIT) / Argonne National Laboratory (ANL), USA

Marc Shapiro, UPMC Sorbonne Universités, LIP6, Paris, France

Nicolae Tapus, University Politehnica of Bucharest, Romania

Albert Zomaya, The University of Sydney, Australia

ARMS-CC-2014 Programme

July 15th, 2014

08.45 – 09.00 Opening

Maria Gradinariu Potop-Butucaru and Florin Pop, ARMS-CC Chairs

09:00-10:00 Invited Paper

Session Chair: Florin Pop, UPB

Overcoming Data Locality: an In-Memory Runtime File System with Symmetrical Data Distribution

Thilo Kielmann, VU Amsterdam

10:00-10:30 Coffee break

10.30 – 12.00 Session 1: Scheduling Algorithms

Session Chair: Thilo Kielmann, VU Amsterdam

01. A Multi-Capacity Queuing Mechanism in Multi-Dimensional Resource Scheduling

Mehdi Sheikhalishahi, Richard M. Wallace, Lucio Grandinetti, Jose Luis Vazquez-Poletti, Francesca Guerriero

02. A Green Scheduling Policy for Cloud Computing

Jordi Vilaplana, Francesc Solsona, Jordi Mateo, Ivan Teixidó, Josep Rius and Josep Rius

03. A Framework for Speculative Scheduling and Device Selection for Task Execution on a Mobile Cloud

Ansuman Banerjee, Himadri Sekhar Paul, Arijit Mukherjee, Swarnava Dey and Pubali Datta

04. An Interaction Balance Based Approach for Autonomic Performance Management in Cloud Computing Environment

Rajat Mehrotra, Srishti Srivastava, Ioana Banicescu and Sherif Abdelwahed

05. Power-efficient Assignment of Virtual Machines to Physical Machines

Jordi Arjona Aroca, Antonio Fernandez Anta, Miguel A. Mosteiro, Christopher Thraves and Lin Wang

12:00-13:30 Lunch

13:30-14:30 Invited Paper

Session Chair: Maria Gradinariu Potop-Butucaru, UMPC-LIP6

A principled approach to the performance-vs.-programmability trade-off in large-scale distributed data stores

Marc Shapiro, INRIA & UMPC-LIP6

14.30 – 15.30 Session 2: Services and Applications

Session Chair: Mark Shapiro

06. Simulation of Multi-Tenant Scalable Cloud-Distributed Enterprise Information Systems

Alexandru-Florian Antonescu and Torsten Braun

07. Towards Type-based Optimizations in Distributed Applications using ABS and JAVA 8

Vlad Serbanescu, Chetan Nagarajagowda, Frank De Boer, Keyvan Azadbakht and Behrooz Nobakht

08. A Parallel Genetic Algorithm Framework for Cloud Computing Applications

Elena Apostol, Iulia Balauta, Alexandru Gorgoi and Valentin Cristea

09. Analysing Scalability Strategies for Service Choreographies on Cloud Environments

Raphael Gomes, Fabio Costa and Ricardo C. A. Da Rocha

15:30-16:00 Coffee break**16.00 – 17.30 Session 3: Foundational Models for Resource Management in Cloud**

Session Chair: Alexandru Costan, INRIA Rennes

10. Towards Efficient Power Management in MapReduce: Investigation of CPU-Frequencies Scaling on Power Efficiency in Hadoop

Shadi Ibrahim, Diana Moise, Houssem-Eddine Chihoub, Alexandra Carpen-Amarie, Luc Bouge, and Gabriel Antoniu

11. Self-management of Live Streaming Application in Distributed Cloud Infrastructure

Patricia Takako Endo, Marcelo Santos, Jônatas Vitalino, Glauco Gonçalves, Moisés Rodrigues, Djamel F H Sadok, Judith Kelner and Azimeh Sefidcon

12. Towards the Impact of Design Flows on the Resources used by an Application

Cristina Marinescu, Serban Stoenescu and Teodor-Florin Fortis

13. Policy-based Cloud Management Through Resource Usage Prediction

Catalin Leordeanu, Silviu Grigore, Octavian Moraru and Valentin Cristea

14. An Inter-Cloud Architecture for OpenStack Infrastructures

Stelios Sotiriadis, Nik Bessis and Euripides Petrakis

17.30 - Closing, Discussion, ARMS-CC-2015

More information on: <http://arms-cc.hpc.pub.ro>

Invited Talks



Thilo Kielmann

Overcoming Data Locality: an In-Memory Runtime File System with Symmetrical Data Distribution

Abstract: In many-task computing (MTC), applications such as scientific workflows or parameter sweeps communicate via intermediate files; application performance strongly depends on the file system in use. The state of the art uses runtime systems providing in-memory file storage that is designed for data locality: files are placed on those nodes that write or read them. With data locality, however, task distribution conflicts with data distribution, leading to application slowdown, and worse, to prohibitive storage imbalance. To overcome these limitations, we develop MemFS, a fully symmetrical, in-memory runtime file system that stripes files across all compute nodes, based on a distributed hash function. Our experiments with Montage and BLAST workflows, using up to 512 cores, show that MemFS has both better performance and better scalability than the state-of-the-art, locality-based file system, AMFS.

Marc Shapiro

A principled approach to the performance-vs.-programmability trade-off in large-scale distributed data stores

Abstract: Distributed systems face a fundamental trade-off. Strong consistency is easy to understand but is slow, expensive, and is unavailable when the system partitions. Eventual consistency (EC) can be cheaper, faster, and more scalable, but is hard to understand and get correct. We explore the spectrum between those two end-points, from the perspective of parallelism, availability and responsiveness on the one hand, and the ease of programming and maintaining invariants on the other. When concurrent updates to replicated data are allowed, ensuring convergence is complex and often ad-hoc. We propose a simple, scalable, but somewhat restrictive programming model, Conflict-Free Replicated Data Types (CRDTs). Some simple sufficient conditions ensure that all replicas of a CRDT object converge to a correct state, without requiring synchronization or rollback. CRDTs are safe by construction, and remain responsive, available and scalable despite high network latency, faults, or disconnection.



Joint work with

Marek Zawirski, UMPC-LIP6 & INRIA

Masoud Saeida-Ardekani, UMPC-LIP6 & INRIA

Nuno Preguiça, U. Nova de Lisboa

Book of Abstracts

A Multi-Capacity Queuing Mechanism in Multi-Dimensional Resource Scheduling

Mehdi Sheikhalishahi, Richard M. Wallace, Lucio Grandinetti, Jose Luis Vazquez-Poletti, Francesca Guerriero

Abstract: With the advent of new computing technologies, such as cloud computing and contemporary parallel processing systems, the building blocks of computing systems have become multi-dimensional. Traditional scheduling algorithms based on a single-resource optimization like processor fail to provide near optimal solutions. The efficient use of new computing systems depends on the efficient use of all resource dimensions. Thus, the scheduling algorithms have to fully use all resources. In this paper, we propose a queuing mechanism based on a multi- resource scheduling technique. For that, we model multi-resource scheduling as a multi-capacity bin-packing scheduling algorithm at the queue level to reorder the queue in order to improve the packing and as a result improve scheduling metrics. The experimental results demonstrate performance improvements in terms of wait time and slowdown metrics.

A Green Scheduling Policy for Cloud Computing

Jordi Vilaplana, Francesc Solsona, Jordi Mateo, Ivan Teixidó, Josep Rius and Josep Rius

Abstract: This paper presents a power-aware scheduling policy algorithm called Green Preserving SLA (GPSLA) for cloud computing systems with high workload variability. GPSLA aims to guarantee the SLA (Service-Level Agreement) by minimizing the system response time and, at the same time, tries to reduce the energy consumption. We present a formal solution, based on linear programming, to assign the system load to the most powerful Virtual Machines, while respecting the SLA and lowering the power consumption as far as possible. GPSLA is thought for one node load-aware and jobs formed by embarrassingly parallel heterogeneous tasks. The results obtained by implementing the model with the IBM CPLEX prove the applicability of our proposal for guaranteeing SLA and saving energy. This also encourages its applicability in High Performance Computing due to its good behavior when scaling the model and the workload. The results are also highly encouraging for further research into this model in real federated clouds or cloud simulation environments, while adding more complexity.

A Framework for Speculative Scheduling and Device Selection for Task Execution on a Mobile Cloud

Ansuman Banerjee, Himadri Sekhar Paul, Arijit Mukherjee, Swarnava Dey and Pubali Datta

Abstract: In this paper, we study the problem of opportunistic task scheduling and workload management in a mobile cloud setting considering computation power variation. We gathered mobile usage data for a number of persons and applied supervised clustering to show that a pattern of usage exists and that follows a state-based model. Based on this model, we present a strategy to choose and offload work on a mobile device. We present a framework and experimental results showing the efficacy of our proposed approach.

An Interaction Balance Based Approach for Autonomic Performance Management in Cloud Computing Environment

Rajat Mehrotra, Srishti Srivastava, Ioana Banicescu and Sherif Abdelwahed

Abstract: In this paper, a performance management approach is introduced that provides dynamic resource allocation for deploying a general class of services over a federated cloud computing infrastructure. This performance management approach is based on distributed control, and is developed by using an interaction balance methodology, which has previously been successfully used in developing management solutions for traditional large-scale industrial systems. The proposed management approach is used to derive an optimal resource allocation for hosting a set of services in a cloud infrastructure by considering both, the availability and the demand of the cloud computing resources. The primary goals of this performance management approach are to maintain the service level agreements, maximize the profit, and minimize the operating cost for both, the service providers and the cloud brokers. The cloud brokers are considered third party organizations, which work as intermediaries between the service providers and the cloud providers to sublet the rented cloud resources for fixed time duration and deploy respective services on these resources at a profitable rate. Further, the allocated computing resources are utilized by the service providers to maintain the service level agreements of hosted services by using model-based predictive controllers.

Power-efficient Assignment of Virtual Machines to Physical Machines

Jordi Arjona Aroca, Antonio Fernandez Anta, Miguel A. Mosteiro, Christopher Thraves and Lin Wang

Abstract: Motivated by current trends in cloud computing, we study a version of the generalized assignment problem where a set of virtual processors has to be implemented by a set of identical processors. For literature consistency, we say that a set of virtual machines (VMs) is assigned to a set of physical machines (PMs). The optimization criterion is to minimize the power consumed by all the PMs. We term the problem Virtual Machine Assignment (VMA). Crucial differences with previous work include a variable number of PMs, that each VM must be assigned to exactly one PM (i.e., VMs cannot be implemented fractionally), and a minimum power consumption for each active PM. Such infrastructure may be strictly constrained in the number of PMs or in the PMs' capacity, depending on how costly (in terms of power consumption) it is to add a new PM to the system or to heavily load some of the existing PMs. Low usage or ample budget yields models where PM capacity and/or the number of PMs may be assumed unbounded for all practical purposes. We study four VMA problems depending on whether the capacity or the number of PMs is bounded or not. Specifically, we study hardness and online competitiveness for a variety of cases. To the best of our knowledge, this is the first comprehensive study of the VMA problem for this cost function.

Simulation of Multi-Tenant Scalable Cloud-Distributed Enterprise Information Systems

Alexandru-Florian Antonescu and Torsten Braun

Abstract: Cloud Computing is an enabler for delivering large-scale, distributed enterprise applications with strict requirements in terms of performance. It is often the case that such applications have complex scaling and Service Level Agreement (SLA) management requirements. In this paper we present a simulation approach for validating and comparing SLA-aware scaling policies using the CloudSim simulator, using data from an actual Distributed Enterprise Information System (dEIS). We extend CloudSim with concurrent and multi-tenant task simulation capabilities. We then show how different scaling policies can be used for simulating multiple dEIS applications. We present multiple experiments depicting the impact of VM scaling on both datacenter energy consumption and dEIS performance indicators.

Towards Type-based Optimizations in Distributed Applications using ABS and JAVA 8

Vlad Serbanescu, Chetan Nagarajagowda, Frank De Boer, Keyvan Azadbakht and Behrooz Nobakht

Abstract: In this paper we present an API to support modeling applications with Actors based on the paradigm of the Abstract Behavioural Specification (ABS) language. With the introduction of JAVA 8, we expose this API through a JAVA library to allow for a high-level actor-based methodology for programming distributed systems which supports the programming to interfaces discipline. We validate this solution through a case study where we obtain significant performance improvements as well as illustrating the ease with which simple high and low-level optimizations can be obtained by examining topologies and communication within an application. Using this API we show it is much easier to observe drawbacks of shared data-structures and communications methods in the design phase of a distributed application and apply the necessary corrections in order to obtain better results.

A Parallel Genetic Algorithm Framework for Cloud Computing Applications

Elena Apostol, Iulia Balauta, Alexandru Gorgoi and Valentin Cristea

Abstract: Genetic Algorithms (GA) is a subclass of evolutionary algorithms that use the principle of evolution in order to search for solutions to optimization problems. Evolutionary algorithms are by their nature very good candidates for parallelization, and genetic algorithms do not make an exception. Moreover, researchers have stated that genetic algorithms with larger populations tend to obtain better solutions with faster convergence. These are the main reasons why they can benefit from a MapReduce implementation. However, research in this area is still young, and there are only a few approaches for adapting genetic algorithms to the MapReduce model. In this article we analyze the use of subpopulations for the GA MapReduce implementations. MapReduce naturally creates subpopulations, and if this characteristic is properly explored, we can find better solutions for genetic algorithm parallelization. In this context, we propose new models for two well know genetic algorithm implementations, namely island and neighborhood model. Our solutions are using the island model, with isolated subpopulations, and the neighborhood model, with overlapping subpopulations. We incorporate these solutions in a framework, which makes the development of Cloud applications using Genetic Algorithm easier.

Analysing Scalability Strategies for Service Choreographies on Cloud Environments

Raphael Gomes, Fabio Costa and Ricardo C. A. Da Rocha

Abstract: Scalability is one of the major advantages brought by cloud computing environments. This advantage can be even more evident when considering the composition of services through choreographies. However, when dealing with applications that have quality of service concerns scalability needs to be performed in an efficient way considering both horizontal scaling - adding new virtual machines with additional resources, and vertical scaling - adding/removing resources from existing virtual machines. By efficiency we mean that non-functional properties must be offered in the choreographies while is made effective/improved resource usage. This paper discusses scalability strategies to enact service choreographies using cloud resources. We present efforts at the state of the art technology and an analysis of the outcomes in adopting different strategies of resource scaling. We also present experiments using a modified version of CloudSim to demonstrate the effectiveness of these strategies in terms of resource usage and the non-functional properties of choreographies.

Towards Efficient Power Management in MapReduce: Investigation of CPU-Frequencies Scaling on Power Efficiency in Hadoop

Shadi Ibrahim, Diana Moise, Housseem-Eddine Chihoub, Alexandra Carpen-Amarie, Luc Bouge, and Gabriel Antoniu

Abstract: With increasingly inexpensive cloud storage and increasingly powerful cloud processing, the cloud has rapidly become the environment to store and analyze data. Most of the large-scale data computations in the Cloud heavily rely on the MapReduce paradigm and its Hadoop implementation. These MapReduce-based cloud solutions have attracted massive interest both from industry and academia. Nevertheless, this exponential growth in popularity has significantly impacted power consumption in cloud infrastructures. While various energy-saving mechanisms have been devised for large-scale infrastructures, not all of them are suitable in a cloud context, as they might impact the performance of the executed workloads. In this paper, we focus on MapReduce and we investigate the impact of dynamically scaling the frequency of compute nodes on the performance and energy consumption of a Hadoop cluster. MapReduce systems span over a multitude of computing nodes that are frequency and voltage-scalable. Furthermore, many MapReduce applications show significant variation in CPU load during their running time. Thus, there is a significant potential for energy saving by scaling down the CPU frequency. Some power-aware data-layout techniques have been proposed to save power but at the cost of performance. Taking into account the nature of a MapReduce application (CPU-intensive, I/O-intensive, or both) and the fact that its subtasks execute different workloads (disk read, computation, network access), there is significant potential for reducing power consumption by scaling down the CPU frequency when peak CPU performance is not needed. To this end, a series of experiments are conducted to explore the implications of Dynamic Voltage Frequency scaling (DVFS) settings on power consumption in Hadoop-clusters: benefiting from the current maturity in DVFS research and the introduction of governors (e.g., performance, power save, on-demand, conservative and user space). By adapting existing DVFS governors in the Hadoop cluster, we observe significant variation in performance and power consumption of the cluster with different applications when applying these governors: the different DVFS settings are only sub-optimal for different MapReduce applications. Furthermore, our results reveal that the current CPU governors do not exactly reflect their design goal and may even become ineffective to manage the power consumption in Hadoop cluster. This study aims at providing more clear understanding of the interplay between performance and power management in Hadoop cluster and therefore offers useful insight into designing power-aware techniques for Hadoop system.

Self-management of Live Streaming Application in Distributed Cloud Infrastructure

Patricia Takako Endo, Marcelo Santos, Jônatas Vitalino, Glauco Gonçalves, Moisés Rodrigues, Djamel F H Sadok, Judith Kelner and Azimeh Sefidcon

Abstract: Currently, live streaming traffic is responsible for more than half of aggregated traffic from fixed access networks in North America. But, due to traffic redundancy, it does not suitably utilize bandwidth and network resources. To cope with this problem in the context of Distributed Clouds (DClouds) we present RBSA4LS, an autonomic strategy that manages the dynamic creation of reflectors for reducing redundant traffic in live streaming applications. Under this strategy, nodes continually assess the utilization level by live streaming flows. When necessary, the network nodes communicate and self-appoint a new reflector node, which switches to multicasting video flows hence alleviating network links. We evaluated RBSA4LS through extensive simulations and the results showed that such a simple strategy can provide as much as 40% of reduction in redundant traffic even for random topologies and reaches 85% of bandwidth gain in a scenario with a large ISP topology.

Towards the Impact of Design Flaws on the Resources used by an Application

Cristina Marinescu, Serban Stoenescu and Teodor-Florin Fortis

Abstract: One major research direction in cloud computing deals with the reduction of energy consumption. This can be seen as an optimization problem that must be addressed both at the hardware and the software (i.e., application) level. At the software level, optimizing energy consumption is usually related with scaling down the resources (e.g., memory, CPU usage) required for running an application. In this context, we can make the assumption that the presence of design flaws in the implementation of a software system may lead to a suboptimal resource usage. Our investigations on the impact of several design flaws on the amount of resources used by an application indicate that the presence of design flaws has an influence on memory consumption and CPU time, and that proper refactoring can have a beneficial influence on resource usage.

Policy-based Cloud Management Through Resource Usage Prediction

Catalin Leordeanu, Silviu Grigore, Octavian Moraru and Valentin Cristea

Abstract: Cloud computing services are becoming increasingly more widespread, mainly because they offer a convenient way of using remote computational resources at any time. Constantly satisfying client needs is a difficult task due to the limited nature of the physical resources. Careful handling of computing capabilities is critical. Cloud systems offer resource elasticity, which is essential for respecting Service Level Agreements (SLAs) or other types of contracts. This paper proposes a novel solution, which offers an efficient resource management mechanism for Clouds. The solution is based on monitoring hosts belonging to the Cloud in order to obtain load data. A policy-based system uses the monitoring information to make decisions about deployment of new virtual machines and migration of already running machines from overloaded hosts. The policy-based solution is enhanced by prediction algorithms to optimize the resource usage and to make sure that the available hosts are capable of handling the increased load before it happens. This leads to more efficient resource usage and can help fulfill the SLA requirements even under heavy loads.

An Inter-Cloud Architecture for OpenStack Infrastructures

Stelios Sotiriadis, Nik Bessis and Euripides Petrakis

Abstract: In latest years, the concept of interconnecting clouds to allow common service coordination has gained significant attention mainly because of the increasing utilization of cloud resources from Internet users. An efficient common management between different clouds is essential benefit, like boundless elasticity and scalability. Yet, issues related with different standards led to interoperability problems. For this reason, the definition of the open cloud-computing interface defines a set of open community- lead specifications along with a flexible API to build cloud systems. Today, there are cloud systems like OpenStack, OpenNebula, Amazon Web Services and VMWare VCloud that expose APIs for inter-cloud communication. In this work we aim to explore an inter-cloud model by creating a new cloud platform service to act as a mediator among OpenStack, FI-WARE datacenter resource management and Amazon Web Service cloud architectures, therefore to orchestrate communication of various cloud environments. The model is based on the FI-WARE and will be offered as a reusable enabler with an open specification to allow interoperable service coordination.

Your Notes



ARMS-CC-2014

Workshop on Adaptive Resource Management and Scheduling for Cloud Computing, Held in conjunction with PODC 2014 (ACM Symposium on Principles of Distributed Computing)

Paris, France, on July 15th, 2014

Contact:

ARMS-CC Chairs

Florin Pop, University Politehnica of Bucharest, Romania, Email: florin.pop@cs.pub.ro

Maria Gradinariu Potop-Butucaru, UPMC Sorbonne Universités, LIP6, Paris, France, Email: maria.gradinariu@lip6.fr



<http://arms-cc.hpc.pub.ro>



<http://www.podc.org>