Introduction to the 3rd International Workshop on Cloud Computing and Scientific Applications (CCSA'13)

Suraj Pandey[1] and Surya Nepal[2]

[1] IBM Research Australia
suraj.pandey@au.ibm.com
[2] CSIRO Computational Informatics
surya.nepal@csiro.au

CCSA workshop has been formed to promote research and development activities focused on enabling and scaling scientific applications using distributed computing paradigms, such as cluster, Grid, and Cloud Computing. With the rapid emergence of virtualized environments for accessing software systems and solutions, the volume of users and their data are growing exponentially. According to the IDC, by 2020, when the ICT industry reaches $5 billion - $1.7 billion larger than it is today - at least 80% of the industry's growth will driven by 3rd platform technologies, such as cloud services and big data analytics. Existing computing infrastructure, software system designs, and use cases will have to take into account the enormity in volume of requests, size of data, computing load, locality and type of users, and every growing needs of all applications. Cloud computing promises reliable services delivered through next-generation data centers that are built on compute and storage virtualization technologies. Users will be able to access applications and data from a Cloud anywhere in the world on demand. In other words, the Cloud appears to be a single point of access for all the computing needs of users. The users are assured that the Cloud infrastructure is robust and will always be available at any time. To address the growing needs of both applications and Cloud computing paradigm, CCSA brings together researchers and practitioners from around the world to share their experiences, to focus on modeling, executing, and monitoring scientific applications on Clouds. In this workshop, there were 20 submissions. The committee decided to accept 7 papers. The program also includes 1 invited talk as a keynote.

Summary of Papers Presented in the Workshop

The paper titled "SLA-Aware Load Balancing in a Web-Based Cloud System over OpenStack" presents an architecture that enables load balancing of web-application by distributing the load across virtual machines, while preserving the service-level-agreement. The paper maintains the SLA by expanding the computing capacity dynamically to avoid system overload by adding additional VMs when experiencing sudden increases in the number of users and requests in system.

The paper titled "Are Public Clouds Elastic Enough for Scientific Computing?" presents a review of solutions proposed by public cloud providers and points the open

A.R. Lomuscio et al. (Eds.): ICSOC 2013 Workshops, LNCS 8377, pp. 279–280, 2014.

issues and challenges in providing elasticity for scientific applications. It also describes initiatives that are being developed in that space.

The paper titled "A light-weight framework for bridge-building from desktop to cloud" describes a light-weight framework based on cloud and REST to address (i) the heavy weight and diversity of infrastructures that inhibits sharing and collaboration between services, (ii) the relatively complicated processes associated with deployment and management of web services for non-disciplinary specialists, and (iii) the relative technical difficulty in packaging the legacy software that encapsulates key discipline knowledge for web-service environments.

The paper titled "Planning and Scheduling Data Processing Workflows in the Cloud with Quality-of-Data Constraints" introduces a new scheduling criterion, Quality-of-Data (QoD), that specifically focuses on continuous data processing workflows, where the scheduler does not perform any reasoning about the impact new input data may have in the workflow final output. The authors have illustrated the viability of their research by developing a WaaS (Workflow-as-a-Service), a workflow coordinator system for the Cloud where data is shared among tasks via cloud-based columnar databases.

The paper titled "Galaxy + Hadoop: Toward a Collaborative and Scalable Image Processing Toolbox in Cloud" presents a cloud-based image processing tool- box by integrating Galaxy, Hadoop and CSIRO's proprietary image processing tools. The paper provides the integration architecture and technical details about the whole system. In particular, it investigates the use of Hadoop to handle massive image processing jobs.

The paper titled "SciLightning - a Cloud Provenance-based Event Notification for Parallel Workflows" presents a workflow event notification mechanism based on runtime monitoring of provenance data produced by parallel scientific workflow systems in clouds. The paper also evaluated their proposed mechanism by monitoring SciPhy, a large-scale parallel execution of a bioinformatics phylogenetic analysis workflow.

The paper titled "Energy Savings on a Cloud-based Opportunistic Infrastructure" presents the problem of virtual machines consolidation on the opportunistic cloud computing resources. It investigates four workload packing algorithms that place a set of virtual machines on the least number of physical machines to increase resource utilization and to transition parts of the unused resources into a lower power states or switching off. In addition, it empirically evaluates these heuristics on real workload traces collected from our experimental opportunistic cloud, called UnaCloud.

Acknowledgement. We would like to thank Dr Shiping Chen for chairing the event. We also thank EasyChair.org for providing the conference management system.