Chapter 6

Distributed Multi-Cloud Based Building Data Analytics

Ioan Petri
Cardiff University, UK

Javier Diaz-Montes
Rutgers University, USA

Mengsong Zou
Rutgers University, USA

Ali Reza Zamani
Rutgers University, USA

Thomas H Beach
Cardiff School of Engineering, UK

Omer F. Rana
Cardiff University, UK

Manish Parashar
Rutgers University, USA

Yacine Rezgui
Cardiff University, UK

ABSTRACT

Cloud computing has emerged as an attractive platform for computing data intensive applications. However, efficient computation of this kind of workloads requires understanding how to store, process, and analyse large volumes of data in a timely manner. Many “smart cities” applications, for instance, identify how data from building sensors can be combined together to support applications such as emergency response, energy management, etc. Enabling sensor data to be transmitted to a cloud environment for processing provides a number of benefits, such as scalability and on-demand provisioning of computational resources. In this chapter, we propose the use of a multi-layer cloud infrastructure that distributes processing over sensing nodes, multiple intermediate/gateways nodes, and large data centres. Our solution aims at utilising the pervasive computational capabilities located at the edge of the infrastructure and along the data path to reduce data movement to large data centres located “deep” into the infrastructure and perform a more efficient use of computing and network resources.

1. INTRODUCTION

Cloud computing has generally involved the use of specialist data centres to support computation and data storage at a central site (or a limited number of sites). The motivation for this has come from the need to provide economies of scale (and subsequent reduction in cost) for supporting large scale computation for multiple user applications over (generally) a shared, multi-tenancy infrastructure. The use of

DOI: 10.4018/978-1-5225-0153-4.ch006
such infrastructures requires moving data to a central location, undertaking processing on the data, and subsequently enabling users to download results of analysis. However, when applications need data to be captured and processed/analysed in real time, migrating all the data to a central site prior to analysis can create significant overhead. Examples include sensor network-based applications, such as smart buildings and smart grids, where sensors interface with real world artefacts and must respond to physical phenomenon that cannot be predicted a priori. The amount of data likely to be generated by a sensor and processing requirements in such applications cannot be pre-determined - they are often dependent on the rate of change of the physical phenomenon being measured and potential occurrence of “trigger events” which are non-deterministic.

Recently, there has been significant interest in creating “multi-clouds” or Cloud-of-Clouds federation to aggregate capabilities and capacities offered by a variety of cloud providers. Some of the efforts are focused on cloud interoperability. For example, the Open Cloud Computing Interface (OCCI) effort (OCCI), which defines a common interface for cloud providers; and the European FP7 “UNIFY” project (UNIFY EU FP7 project), which develops a Cloud Operating System (CloudOS) to connect distributed clouds and make use of in-network capabilities to process data (GENICloud project (GENICloud project)). Similarly, on-line sites such as CloudHarmony (CloudHarmony) report over 100+ cloud providers that offer capability ranging from storage and computation to complete application containers that can be acquired at a price, primarily using service-based access models. From the user’s perspective these environments bring a variety of benefits: (i) reduced reliance on a single vendor’s infrastructure; (ii) improved fault tolerance, as failure in one cloud system does not render the entire infrastructure inoperable; (iii) improved security – similar argument to fault tolerance, i.e. a breach in one cloud system does not impact the entire infrastructure; and (iv) the ability to utilise capability (and data) that may only be available in one cloud system and may not be transferable due to volume or legal constraints.

Resource elasticity and scalability offered through cloud computing enables researchers to explore complex problems in energy optimisation that are otherwise impractical or impossible to address (Perez-Lombard, Ortiz, & Pout, 2010; Thain, Tannenbaum, & Livny, 2008). In this chapter we describe how a distributed cloud system can be used to perform a number of different data analysis operations (in-situ vs. in-transit) based on pre-defined application requirements. We describe and evaluate the establishment of such a sensor-based application using a CometCloud (Diaz-Montes, AbdelBaky, Zou, & Parashar, 2015; Diaz-Montes, Zou, Singh, Tao, & Parashar, 2014) implementation with data collection from real building pilots and determine how processing can be distributed across multiple data centre locations to achieve QoS and cost targets. Building sensor data is analysed using EnergyPlus. EnergyPlus is a time-step based energy simulation package that can be used to model heating, cooling, lighting, ventilation and other energy flows within a building. Our solution enables resources to be utilised across geographically distributed cloud environments connected over the Internet. In this paper we propose a cost based multi-cloud framework for supporting real application deployment over three federated sites Cardiff University in Cardiff (UK), Rutgers University in New Jersey (USA), and Indiana University in Indiana (USA) – all hosting EnergyPlus; and a performance analysis of the application scenarios to determine how task submission could be supported across these three sites, subject to particular revenue targets. The reminder of this chapter is organised as follows: Section 2, outlines the development and use of distributed clouds, providing a key motivation for our research and analysing several related approaches. Section 3 presents our approach for cloud analytics. Section 4 presents our application sce-
Related Content

Chemometrics: From Data Preprocessing to Fog Computing
www.igi-global.com/article/chemometrics/219359?camid=4v1a

Domain-Based Dynamic Ranking
www.igi-global.com/chapter/domain-based-dynamic-ranking/138509?camid=4v1a

A Cloud-Oriented Reference Architecture to Digital Library Systems
www.igi-global.com/chapter/a-cloud-oriented-reference-architecture-to-digital-library-systems/88043?camid=4v1a