

Chapter 8

Real Time Task Execution in Cloud Using MapReduce Framework

Sampa Sahoo

*National Institute of Technology Rourkela,
India*

Ashok Kumar Turuk

*National Institute of Technology Rourkela,
India*

Bibhudatta Sahoo

*National Institute of Technology Rourkela,
India*

Sambit Kumar Mishra

*National Institute of Technology Rourkela,
India*

ABSTRACT

Cloud Computing era comes with the advancement of technologies in the fields of processing, storage, bandwidth network access, security of internet etc. The development of automatic applications, smart devices and applications, sensor based applications need huge data storage and computing resources and need output within a particular time limit. Now users are becoming more sensitive towards, delay in applications they are using. So, a scalable platform like Cloud Computing is required that can provide huge computing resource, and data storage required for processing such applications. MapReduce framework is used to process huge amounts of data. Data processing on a cloud based on MapReduce would provide added benefits such as fault tolerant, heterogeneous, ease of use, free and open, efficient. This chapter discusses about cloud system model, real-time MapReduce framework, Cloud based MapReduce framework examples, quality attributes of MapReduce scheduling and various MapReduce scheduling algorithm based on quality attributes.

INTRODUCTION

Cloud Computing is an internet-based computing, where resources can be accessed as pay-as-you-go basis over the internet. It enables computing as a utility like electricity, gas, etc. Cloud Computing offers its users a new dimension of viewing the resources i.e. resources as services via internet. Virtualization is the base of Cloud Computing, which creates logical (virtual) machine consisting of all the resources

DOI: 10.4018/978-1-5225-1721-4.ch008

Real Time Task Execution in Cloud Using MapReduce Framework

(Operating System, server, storage, and network) similar to a physical machine. We can consider cloud as a network of large groups of servers where resources can be shared and accessed as virtual resources in a scalable and secure manner. The virtualization technique also eliminates the need for maintenance of computing hardware, dedicated space and software. The paradigm shifts from on premise computing to Cloud Computing allows resource sharing and isolation from the underlying hardware. Various advantages of Cloud Computing include on-demand scalability, low cost storage services, parallelization, high computing power, security, etc. On-demand scalability feature allows to scale-in/out resources depending on workload which reduces unnecessary resource usage. On-premise computing users spend a major part of their budget on hardware, software, recovery management, networking, power supply. But, Cloud Computing reduces it as accessing a Virtual Machine (VM) is cheaper than buying and installing a physical resource. Cloud Computing allows parallelization i.e. several VMs can work simultaneously without affecting each other and thus increasing the resource utilization and efficiency of the cloud system. Cloud Computing is an off-premise form of computing, which is changing the expectation of what, how and when computing, allocation and management of storage and networking resources. It is also referred to as ubiquitous (anywhere, anytime) computing which requires only a subscription to use the virtual resources through the internet. In cloud, VM instances are deployed in cloud service provider data centres from where users can use the computational resources. An agreement is signed between client and service provider known as Service Level Agreement (SLA) in terms of price, Quality of Service (QoS) level, penalties associated with SLA violation, etc.

The cloud service model can be categorized as Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS). SaaS model provides an interface that can be used to access applications managed by a third party through the internet. Users can use applications directly from a web browser without any downloads or installations, in some cases through plugins. Here everything (data, OS, servers, applications, storage and networking) is managed by third party. Popular SaaS examples are Google Apps, Salesforce, Citrix GoToMeeting, Concur, Gmail, Microsoft Office 365, DropBox, etc. A development environment is provided in PaaS model to build, host, and deploy an application. To run an application hardware and software maintenance is not required at company's workspace. PaaS examples include Google App Engine and Salesforce.com, etc. IaaS model gives infrastructure (computing, storing, networking resources, etc.) like an actual physical infrastructure as a service. It is a self-service model where users can purchase IaaS based on requirement like other utility (electricity, gas, etc.). This model gives users the ability to manage applications, data, OS, middleware, etc. IaaS examples are Microsoft Azure, Amazon Web Services (AWS), Joyent, Google Compute Engine (GCE). The Cloud Computing implementation models are public cloud, private cloud, hybrid cloud and community cloud. In private cloud the services and computing infrastructure are maintained by a particular organization. Here the required software and infrastructure are purchased and maintained in the organization, which offers high security and control. In public cloud services and infrastructure can be accessed over the internet. Here the user has no visibility and control over the computing infrastructure. It can be shared between any organizations, which makes it less secure than private cloud. Hybrid cloud uses the functionality of both private and public cloud features. An organization with the hybrid cloud model can use a private cloud for normal usage and public cloud for high load requirements to handle the sudden increase in computing requirements (Puthal, Sahoo, Mishra & Swain, 2015).

One of the research domains in the cloud is applications that must complete within the deadline (e.g. weather prediction, video streaming). Real-time applications need fast processing of tasks to maintain logical and timing correctness. The development in Internet of Things (IoT) applications, availability of

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/real-time-task-execution-in-cloud-using-mapreduce-framework/171353

Related Content

Comparative Study for Different Provisioning Policies for Load Balancing in Cloudsim

Layla Albdour (2017). *International Journal of Cloud Applications and Computing* (pp. 76-86).

www.irma-international.org/article/comparative-study-for-different-provisioning-policies-for-load-balancing-in-cloudsim/182254

On Developing Fair and Orderly Cloud Markets: QoS- and Security-Aware Optimization of Cloud Collaboration

Olga Wenge, Dieter Schuller, Christoph Rensing and Ralf Steinmetz (2019). *Cloud Security: Concepts, Methodologies, Tools, and Applications* (pp. 2097-2119).

www.irma-international.org/chapter/on-developing-fair-and-orderly-cloud-markets/224673

Using Value-Based Approach for Managing Cloud-Based Services

Salah Eddin Murad and Salah Dowaji (2017). *Resource Management and Efficiency in Cloud Computing Environments* (pp. 33-60).

www.irma-international.org/chapter/using-value-based-approach-for-managing-cloud-based-services/171347

MeghaOS: A Framework for Scalable, Interoperable Cloud Based Operating System

K. G. Srinivasa, Harish Raddi C. S., Mohan Krishna S. H. and Nidhi Venkatesh (2012). *International Journal of Cloud Applications and Computing* (pp. 53-70).

www.irma-international.org/article/meghaos-framework-scalable-interoperable-cloud/64635

A Multi-Objective Adaptive Upper Threshold Approach for Overloaded Host Detection in Cloud Computing

Rajeshwari Sissodia, ManMohan Singh Rauthan and Varun Barthwal (2022). *International Journal of Cloud Applications and Computing* (pp. 1-14).

www.irma-international.org/article/a-multi-objective-adaptive-upper-threshold-approach-for-overloaded-host-detection-in-cloud-computing/311038