

Chapter 9

Architecture for Big Data Storage in Different Cloud Deployment Models

Chandu Thota

Infosys Ltd., India

Gunasekaran Manogaran

Vellore Institute of Technology, India

Daphne Lopez

 <https://orcid.org/0000-0003-1452-2144>

Vellore Institute of Technology, India

Revathi Sundarasekar

Priyadarshini Engineering College, India

ABSTRACT

Cloud Computing is a new computing model that distributes the computation on a resource pool. The need for a scalable database capable of expanding to accommodate growth has increased with the growing data in web world. More familiar Cloud Computing vendors such as Amazon Web Services, Microsoft, Google, IBM and Rackspace offer cloud based Hadoop and NoSQL database platforms to process Big Data applications. Variety of services are available that run on top of cloud platforms freeing users from the need to deploy their own systems. Nowadays, integrating Big Data and various cloud deployment models is major concern for Internet companies especially software and data services vendors that are just getting started themselves. This chapter proposes an efficient architecture for integration with comprehensive capabilities including real time and bulk data movement, bi-directional replication, metadata management, high performance transformation, data services and data quality for customer and product domains.

DOI: 10.4018/978-1-7998-5339-8.ch009

INTRODUCTION

Virtualization is playing a major role in implementation of both Big Data and Cloud Computing. It provides lots of information such as storing, accessing, analyzing and managing the Distributed Computing components in Big Data Analytics. In other words, Virtualization is also used to increase IT resource utilization, scalability and efficiency. The main goal of Virtualization is to increase the consumption of physical servers and efficiently saving on infrastructure costs.

Cloud Computing is one of the most powerful techniques used to perform large scale computing, parallel processing, complex computing, security and data service integration with scalable data storage. In the recent years, cloud application integration is growing at a high speed. Organizations used variety of integrations in cloud services especially for mobile apps that needs to exchange messages and huge data. The software deployed in Cloud Computing environment and the data stored in cloud data centers are connected virtually to communicate each other asynchronously or synchronously by fetching, transferring and storing the data. There is a need to develop efficient techniques for integrating cloud and Big Data.

The goal of this chapter is to describe the storage of Big Data in different Cloud Computing environments. The huge data of organizations has been stored in different cloud deployment models as Private, Public and Hybrid to achieve the characteristics, definition and classification of Big Data. In addition, this chapter also discusses the integration of Big Data with Cloud Computing, various challenges and problems, and uses with Big Data storage systems in cloud environment. The applications deployed in cloud datacenters and cloud service models, access control of the cloud data centers are also discussed in this chapter.

This chapter is organized as follows: the initial sections discuss about the needs for architecture patterns and data sources for Big Data storage in cloud deployment models, comparison of Big Data to other types of data and quality of Big Data. Second section discusses about the intersection of Big Data and Data Virtualization, Big Data and Graphics Processing Unit (GPU) Computing, Big Data Tools for Integration of Clouds, Clusters and Grids and Big Data Analysis using Grid Supercomputing. Third section proposes the architecture for integration of Big Data and Cloud Computing. Final section concludes the chapter.

CLOUD COMPUTING

Cloud Computing is the practice of using a network of remote servers hosted on Internet to manage, store and process data rather than a personal computer and local server. In other words, Cloud Computing is a type of computing and is used for the delivery of hosted services over the Internet to manage real time applications (Manogaran, Thota & Kumar, 2016).

Big Data Solutions for Cloud Applications

Cloud Computing and Big Data are combined together to achieve many tasks. Big Data provides many techniques and technologies to process distributed queries across multiple datasets and compute the results in a timely manner.

29 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/architecture-for-big-data-storage-in-different-cloud-deployment-models/275285

Related Content

A Novel Task Scheduling Algorithm in Heterogeneous Cloud Environment Using Equi-Depth Binning Method

Roshni Pradhan and Amiya Kumar Dash (2021). *Research Anthology on Architectures, Frameworks, and Integration Strategies for Distributed and Cloud Computing* (pp. 1303-1316).

www.irma-international.org/chapter/a-novel-task-scheduling-algorithm-in-heterogeneous-cloud-environment-using-equi-depth-binning-method/275340

Efficient Fault Tolerance on Cloud Environments

Sam Goundar and Akashdeep Bhardwaj (2021). *Research Anthology on Architectures, Frameworks, and Integration Strategies for Distributed and Cloud Computing* (pp. 1231-1243).

www.irma-international.org/chapter/efficient-fault-tolerance-on-cloud-environments/275336

Big Data Processing on Cloud Computing Using Hadoop Mapreduce and Apache Spark

Yassir Samadi, Mostapha Zbakh and Amine Haouari (2021). *Research Anthology on Architectures, Frameworks, and Integration Strategies for Distributed and Cloud Computing* (pp. 824-845).

www.irma-international.org/chapter/big-data-processing-on-cloud-computing-using-hadoop-mapreduce-and-apache-spark/275316

Fog Computing Qos Review and Open Challenges

R. Babu, K. Jayashree and R. Abirami (2021). *Research Anthology on Architectures, Frameworks, and Integration Strategies for Distributed and Cloud Computing* (pp. 1147-1157).

www.irma-international.org/chapter/fog-computing-qos-review-and-open-challenges/275331

Queuing Analysis of Cloud Load Balancing Algorithms

Santosh Kumar Majhi, Shankho Subhra Pal, Shweta Bhuyan and Sunil Kumar Dhal (2021). *Research Anthology on Architectures, Frameworks, and Integration Strategies for Distributed and Cloud Computing* (pp. 729-750).

www.irma-international.org/chapter/queuing-analysis-of-cloud-load-balancing-algorithms/275311