Multi-Layered Security Model for Hadoop Environment:
Security Model for Hadoop

P. Victor Paul, Department of Computer Science and Engineering, Vignan’s Foundation for Science, Technology and Research, Prades, India
D. Veeraiah, Department of Computer Science and Engineering, Vignan’s Foundation for Science, Technology and Research, Prades, India

ABSTRACT

In this article, a novel security model for the Hadoop environment has been developed to enhance security credentials of handheld systems. The proposed system deals with enabling Hadoop security in terms of a dataset and a user which is willing to access the content inside the Hadoop system. It deals with security in terms of three different features: encryption, confidentiality and authentication. The significance of the proposed model is it provides protection against malicious intent which allows only valid content into the Big data system; it enables authenticated users and people to enter into the system and make the dataset more secure; and if authentication is enhanced, then authorization can be easily gained in the Hadoop system which provides access control and access rights to resource which the user is willing to perform its function or operation. This model is implemented, and the performance has been validated using existing security variants.

KEYWORDS

Big Data, Cryptography Model, Hadoop, Security

1. INTRODUCTION

Big Data is a new emerging domain in which the various types of datasets are grouped into single unit which can be used to perform several tasks and operation in an organized manner. It is also a storage mechanism in which complex and huge size data can be stored in a larger manner at the same time it can be accessed and processed in an efficient way. This mechanism brings a drastic change in the Database storage and management where most of the corporate and user prefer this domains for storing and other purposes. It also has various tools to improvise its importance more and increase the need of the user and satisfy the wish of the customer as per their needs. This technology is being implemented in various platforms and various other domains such as cloud computing, web services etc.

DOI: 10.4018/IJHCR.2017100106

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
The various characteristics of big data are grouped into five categories i.e. (5V’s) which are:

- **Volume**: It is used to specify the capacity or amount of space need for a user or system to store the data in a memory or disk;
- **Variety**: The various types of data type and data analysis are being done in this type of characteristic to check the proper flow of data;
- **Velocity**: This specify the speed in which the dataset takes to process and access in a big data storage area;
- **Veracity**: It is used to provide the various data inconsistency and problems which are present inside the system;
- **Value**: This is used to determine the purpose and the need of using big data usage in an optimized way.

Big data is highly needed to make the companies and the business peoples more ease and help in discovering new aspects in the business field. It also used in transaction since large storage is being provided by big data in terms of storing, accessing and updating of information in the dataset. Its importance is in different fields such as scientists, social media, webservice, data analyzers and others business related fields. This also enhances several forms of data such as structured, unstructured and semi structured so that the user can manipulate any forms of data which is used to yield a greater performance.

### 1.1. Tools for Bigdata

Some of the tools which are used in big data to improve the efficiency and capability of the systems are:

- **Hadoop**: The tool which is delivered by the Apache foundation to run in a parallel platform and cluster based in big data is the Hadoop tool which is an open source software and free licensed tool. It is used for large number of storage and processing of complex dataset to improve the performance and speed of the system;
- **MongoDB**: This is used for mapping and merging the data which provides high data processing speed in terms of huge datasets. It’s a Google product which is highly used in website to improve the searching and manipulation speed;
- **Grid gain**: The alterative tool which is used in terms of map reduce is this tool in which it used in memory processing in real time systems for fast analysis;
- **HPCC**: This high-performance tool which is used in cluster computing which yields greater efficiency similar to Hadoop. It is free software which is operated in LINUX operating system since it a high security system (Thamizh selvan et al., 2015; Thamizh selvan et al., 2015; Baskaran et al., 2012);
- **Storm**: This tool is highly used for scalable, robust, fault-tolerant which can be used in all programming languages to increase the ease of the users. It is used in real time distributed system to run data parallel.

### 1.2. Hadoop

Hadoop is an open framework which is being developed by Apache foundation in order to process large number of system in a reliable manner. This is highly used in Distributed and cluster-based system in which the need of data is very high and most of the parallel operation is carried out in both
The Design of Mobile Television in Europe
www.igi-global.com/chapter/design-mobile-television-europe/26577?camid=4v1a

Design of Compensators for Comb Decimation Filters
www.igi-global.com/chapter/design-of-compensators-for-comb-decimation-filters/214664?camid=4v1a