

HADOOP: FUTURE TALK ABOUT BIG DATA STORAGE

KAILASH CHANDER¹ & POONAM²

¹ Research Associate, Department of Science & Technology, HARSAC, Government of Haryana, India

² Resource Person, Department of Computer Science, Government P.G. College, Jind, Haryana, India

ABSTRACT

This paper presents a review on modern technology called HADOOP, which is used for managing very large amount of data. Multi Peta-byte Data sets becomes challenge for companies to process effectively and efficiently. Conversation about Big Data for very long without running into the elephant in the room is not possible. It is complex to have the data at distributed locations to process. For this a solution is needed i.e. open Source Apache License: HADOOP. It stores enormous data sets across distributed clusters of servers and then running “distributed” analysis applications in each cluster. Data applications will continue to run even when individual servers or cluster fails. Hadoop is almost completely modular, that allow swap out almost any of its components for a different software tool due to the flexibility of architecture.

KEYWORDS: Hadoop, Mapreduce, HDFS, Petabyte, Hortonworks, Sqoop