Abstract

Due to the advent of new technologies, devices, and communication means like social networking sites, the amount of data produced by mankind is growing rapidly every year. Traditional computing techniques are not enough to process that much large amount of data. Hadoop is a bunch of technology & have capacity to store large amount of data on Data nodes. Hadoop uses MapReduce algorithm to process and analyze large scale datasets over large clusters. MapReduce is essential for Big Data processing. This algorithm divides the task into small parts and assigns those parts to many computers connected over the network, and collects the results to form the final result dataset. Bloom filter technique is probabilistic data model which is used to make processing of data more efficient. Implementation of this filter with mapper can reduce the amount of data travel. In this paper we implemented Bloom filter in Hadoop architecture. This help to reduce network traffic over network which save bandwidth as well as data storage.


Index Terms
Computer Science
Databases

Keywords
Big Data, Hadoop, MapReduce, Bloom filter.