A Scalable Data Chunk Similarity based Compression Approach for Efficient Big Sensing Data Processing on Cloud

Objective:

The objective of this system is to making big data smaller and reducing problems of storage and management and helping to overcome the problem of velocity(speed) in big data.

ABSTRACT:

Big sensing data is prevalent in both industry and scientific research applications where the data is generated with high volume and velocity. Cloud computing provides a promising platform for big sensing data processing and storage as it provides a flexible stack of massive computing, storage, and software services in a scalable manner. Current big sensing data processing on Cloud have adopted some data compression techniques. However, due to the high volume and velocity of big sensing data, traditional data compression techniques lack sufficient efficiency and scalability for data processing. Based on specific on-Cloud data compression requirements, we propose a novel scalable data compression approach based on calculating similarity among the

partitioned data chunks. Instead of compressing basic data units, the compression will be conducted over partitioned data chunks. To restore original data sets, some restoration functions and predictions will be designed. MapReduce is used for algorithm implementation to achieve extra scalability on Cloud.

INTRODUCTION:

Big data the large volumes of data that are now produced in many fields can present problems in storage, transmission, and processing, but their analysis may yield useful information and useful insights. In broad terms, the potential benefits of the big sensing system, as applied to big data, are in these areas.

- Overcoming the problem of variety in big data. Harmonizing diverse kinds of knowledge, diverse formats for knowledge, and their diverse modes of processing, via a universal framework for the representation and processing of knowledge.
- Learning and discovery. The unsupervised learning or discovery of 'natural' structures in data.

- Interpretation of data. The big sensing system has strengths in areas such as pattern recognition, information retrieval, parsing and production of natural language, translation from one representation to another, several kinds of reasoning, planning and problem solving.
- Velocity: analysis of streaming data. The big sensing system lends itself to an incremental style, assimilating information as it is received, much as people do.