

INFORMATION COMPRESSION OF BIG DATA USING THE SP THEORY OF INTELLIGENCE

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 data includes large quantity of data that results drawbacks both in accessing and managing the data. These drawbacks are overcome by the introduction of the SP theory of Intelligence. The term 'SP' indicates 'Simplicity' and 'Power'. The central theme used in the theory is lossless information compression. This helps in making the big data small, thereby provides benefits both in accessing and management. The purpose of this project is to overcome the problems in big data using the SP Theory of Intelligence. In order to achieve this goal, big data is subjected to clustering and compression techniques. Compression of information is achieved by pattern matching. Using such a system leads to the improvement in the processing of big data. The SP Theory provides pattern recognition, information storage, retrieval and information compression. Although this theory leads in faster information retrieval, the integrity of the original information is maintained. Future work has to be done in this area to work with patterns in two dimensions.

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 SP 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 SP 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 SP system lends itself to an incremental style, assimilating information as it is received, much as people do.

EXISTING SYSTEM

In the bigdata palatform we faced many problems like storing the data, managing the data and speed of data processing. While storing the data into the big data, duplication of the file is not checked and speed of data processing is less while uploading to the cloud.