Protection of Big Data Privacy

Aim:

The aim of this project to provide a comprehensive overview of the privacy preservation mechanisms in big data.

Objective:

To develop the application to provide the privacy protection for the big data stored in hadoop storage.

Abstract:

In recent years, big data have become a hot research topic. The increasing amount of big data also increases the chance of breaching the privacy of individuals. Since big data require high computational power and large storage, distributed systems are used. As multiple parties are involved in these systems, the risk of privacy violation is increased. There have been a number of privacy-preserving mechanisms developed for privacy protection at different stages (e.g., data generation, data storage, and data processing) of a big data life cycle. The goal of this paper is to provide a comprehensive overview of the privacy preservation mechanisms in big data and present the challenges for existing mechanisms. In particular, in this paper, we illustrate the infrastructure of big data and the state-of-the-art privacy-preserving mechanisms in each stage of the big data life cycle. Furthermore, we discuss the challenges and future research directions related to privacy preservation in big data.

Introduction:

All the huge amount of data generated from different sources in multiple formats with very high speed is referred as big data. Big data has become a very active research area for last couple of years. The data generation rate is growing so rapidly that it is becoming extremely difficult to handle it using traditional methods or systems . Meanwhile, big data could be structured, semi-

structured, or unstructured, which adds more challenges when performing data storage and processing tasks. Therefore, to this end, we need new ways to store and analyze data in real time.

Big data, if captured and analyzed in a timely manner, can be converted into actionable insights which can be of significant value. It can help businesses and organizations to improve the internal decision making power and can create new opportunities through data analysis. It can also help to promote the scientific research and economy by transforming traditional business models and scientific values.

Users' privacy may be breached under the following circumstances :

- Personal information when combined with external datasets may lead to the inference of new facts about the users. Those facts may be secretive and not supposed to be revealed to others.
- Personal information is sometimes collected and used to add value to business. For example, individual's shopping habits may reveal a lot of personal information.
- The sensitive data are stored and processed in a location not secured properly and data leakage may occur during storage and processing phases.

Protecting privacy in big data is a fast growing research area. Although some related papers have been published but only few of them are survey/review type of papers . Moreover, while these papers introduced the basic concept of privacy protection in big data, they failed to cover several important aspects of this area. For example, neither nor provides detailed discussions regarding big data privacy with respect to cloud computing. Besides, none of the papers discussed future challenges in detail.