War on Big Data

Ritika Mehra
Professor, School of Computing, DIT University

Mayank Upadhyay and Vishakha Arya
M.Tech CSE Student, DIT University

"The data protection law will be like a new shoe, tight in the beginning but comfortable eventually"

– Justice B.N. Srikrishna

The big data mania has taken the market by storm. Organizations are investing more to store data for a prolonged period, as storage technologies (like Amazon S3) become more economical. To safeguard the personally identified information there is a need for strict data protection laws and guideline [1].

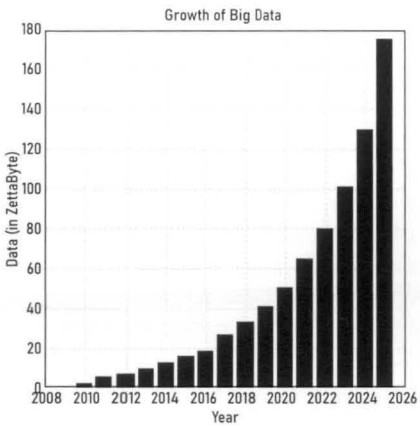

Fig. 1: Growth in Big Data Generation

Introduction

Big data is gathered from disparate sources (like defense, stock exchange) in the form of structured, semi-structured or unstructured. In recent years there has been a boom in digital technologies leading to an exponential growth of data. IDC (International Data Corporation) reports shows that the cumulative data of the world will reach around 40 ZB (zettabyte) in 2020.

ML, predictive analytics, product development, fraud and compliance and many more depend on data analysis. They require bigger data to generate better results. Data regulation laws (like General Data Protection Regulation, California Consumer Privacy Act, Personal Data Protection Bill and Lei Geral de Proteção de Dados Pessoais (LGPDP)) ensure individual's right to privacy. In this article, we will discuss various concerns regarding data security and privacy of big data.

Big Data

Extracting valuable information from big data that is generated from multiple sources involves collecting the data, processing the data, analyzing the data and utilizes the result.

Fig 2: Big Data Life cycle

This valuable information is used in many fields (like education, healthcare, mining sector). The huge amount of data generation leads to various privacy and security concerns. Many organizations recognize that there is a need for better safety measures to prevent data breach. Many countries have taken steps to protect its citizen's right to privacy.

Why protect it?

Tech companies not only use big data to know the easily available information like your recent browsing history, your location, amount of time you spent online, etc., but use data analysis to infer things like your political opinion, and numerous other information which you have not provided directly to the company.

One of the famous examples was of Target's advanced advertisement system which analyses the shopping patterns of its customer to provide coupons to increase sales. In one of the incident, it was able to predict the pregnancy of a teenage girl before her parents knew about it [2].

Corporation Vs Corporation

With the increasing popularity of Globalization, it is also becoming much easier for personal data to flow across the border.

Big tech giants like Facebook, Google, are built on the model of freely exchanging individuals data in return for free services[3].

So that they can analyze and predict consumer behaviour to increase sales and stay ahead of their competitors. Or to simply sell the data to a third party organization.

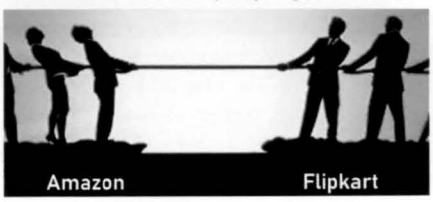

Fig. 3 : Tug Of War Between E-Commerce Giants For Consumers Data

Corporation Vs Country

There has been a tussle between giant tech companies and government regarding the degree to which personal data is collected and used. The EU's GDPR (General Data Protection Regulation), California's CCPA (California Consumer Privacy Act) and LGDP (Lei Geral de Protecao de Dados) are regulations to protect consumer private information.

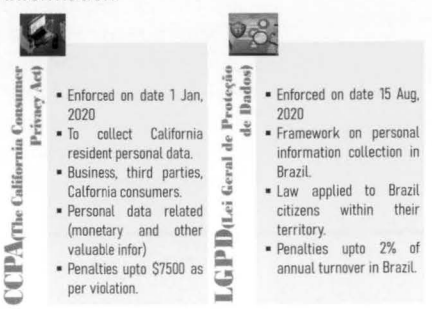

Fig. 4: CCPA and LGPD Regulation

The Indian government in their part has also taken steps to protect the sensitive data of its citizens. The Personal Data Protection Bill (PDPB) mandates that personal data characterized as sensitive (financial, health, religious belief) will be stored in India [4].

Company	Government Orgainisation	Fine
▪ Google	▪ European Union	▪ $9 in total, for unfair advertising rules and forcing smartphone manufacturers to pre-install google chrome.
▪ Facebook	▪ Federak Trade Commission (FTC)	▪ $5 billion for mishandling users personal data
▪ Youtube	▪ FTC	▪ $170 million for violating children's privacy
▪ Facebook, Twitter	▪ Russian Court	▪ $63 thousand each for refusing to store Russian citizens data on servers in Russia.

Fig. 5 : Fine on Tech-Giants

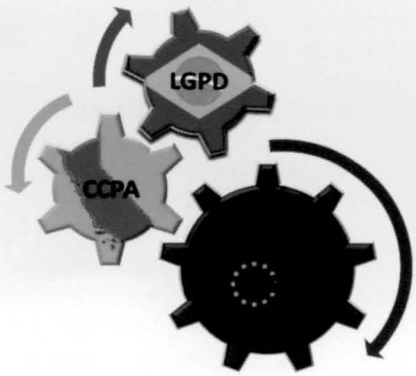

Fig. 6: Data Regulations

GDPR (European Union)

On 25 May 2018, the General Data Protection Regulation (GDPR) was adopted. It outlines rules to protect personal information and privacy of EU citizens. This provision is applicable to all 28 EU countries.

Key Principles of GDPR:

- Consumers consent must be needed for data collection.
- Companies can collect only relevant data.
- Minimization rule and process must implement for relevant data.
- Inaccurate personal data must be rectified.
- Only task-relevant personal data be stored for a longer duration.
- Secure personal data from theft and destruction.
- Company failing to comply with these principles will have to pay 4% of global revenue as a penalty.

PDPB (India)

On 11 Dec 2019, the Personal Data Protection Bill (PDPB) was initiated in Lok Sabha. The bill asks for personal data protection of its citizens. The PDPB bill inherits its core structure from GDPR.

Key Areas of PDPB:

- **Data localization-** Sensitive personal information should be stored within Indian boundaries.
- **Accountability –** Data fiduciaries to conduct annual audits of processing activities.
- **Identity verification –** Social media mediators must enable voluntarily identity verification to Indian Citizens.
- **Data infringement notification –** In case of data breach data fiduciaries should notify DPA (Data Protection Authorities).
- In case of non-compliance of PDPB penalty up to 15 cr or 4% of global revenue.

Conclusion

Big Data have used across several domains making it a valuable resource. It is generated from casual web surf to online transaction, so it is crucial to protect personal data. The government needs to enact data regulation and to educate their citizens regarding their data rights. The organizations collecting customer data needs consumer consent, and inform them if their data is being sold to a third party, and ensure the safety of their personal data like "Digital Safe" bank is one of the method for guarding consumers personal data providing digital password. Thanks to wide range of Anti-surveillance companies, freely available security oriented sofware tools that are working to guard consumers data by limiting its access to the service providers.

References

[1] http://www.isacajournaldigital.org/isacajournal/2014_volume_3/MobilePagedArticle.action?articleId=1077841

[2] https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

[3] https://www.wired.com/story/opinion-new-data-cold-war/

[4] https://www.thehindu.com/sci-tech/technology/what-is-indias-stand-on-data-storage/article27172841.ece

■

About the Authors

Dr. Ritika Mehra (CSI membership no: 00128568) is working as an Professor and Head in School of Computing, DIT University. She received her Ph.D. degree in Computer Science from Gurukul Kangri University, Haridwar in the year 2010 and M.Tech degree in Computer Science and Engineering from Uttarakhand Technical University, Dehradun She specializes in core areas of computer science and holds experience of more than 18 years. She is an innovative person with deep knowledge of Machine Learning, Data Mining, Bigdata Analytics.

Mr. Mayank Upadhyay is currently pursuing M.Tech in D.I.T. University. His areas of interest include M.L. and Big Data.

Ms. Vishakha Arya is currently pursuing M.Tech in D.I.T University. Her areas of interest include Big Data, Machine Learning.