

Workshop on Big Data Analytics (8th Dec, 2022)

Big Data Analytics refers to the process of extracting valuable insights and patterns from large and complex datasets. It involves collecting, organizing, and analyzing vast amounts of data to uncover hidden patterns, trends, and correlations. By leveraging advanced algorithms and technologies, organizations can gain actionable intelligence that can drive decision-making, optimize operations, and uncover new opportunities. Big Data Analytics enables businesses to understand customer behaviour, enhance marketing strategies, improve operational efficiency, and mitigate risks. It has applications across various industries, including finance, healthcare, retail, and manufacturing. With the exponential growth of data, Big Data Analytics plays a crucial role in unlocking the potential of data and driving innovation in the digital era. The Eureka Prize Problems society, under the mentorship of Dr. Souvik Ganguli and Dr. Amit Kumar organized the online workshop on Big Data Analytics on 8th Dec, 2022 in association with Pantech e Learning. The program link is given as

<https://www.youtube.com/live/gr6ZIA3LEOw?feature=share>

The collage consists of six video thumbnails from a YouTube live stream. The thumbnails are as follows:

- Thumbnail 1 (Top Left):** Workshop title slide for 'BIG DATA ANALYTICS' on Thursday, 8th December 2022, featuring logos for Thapar Institute, Eureka, and Pantech e Learning.
- Thumbnail 2 (Top Right):** A hand writing 'What is Big Data?' on a whiteboard.
- Thumbnail 3 (Middle Left):** Text defining big data: 'Big data is the term used for such huge collection of data that requires technological services to be sorted out. Managing these data in the traditional on hand management techniques have also proved ineffectual.'
- Thumbnail 4 (Middle Right):** A timeline from 2002 to 2009 showing key milestones in big data technology, including Google's GFS and MapReduce, Yahoo! hiring Doug Cutting, and the founding of Hadoop and Cloudera.
- Thumbnail 5 (Bottom Left):** A slide titled 'The six Vs of big data' with a table defining Volume, Variety, Velocity, Veracity, Value, and Variability.
- Thumbnail 6 (Bottom Right):** A slide titled 'SOURCES OF BIG DATA' listing Social Networking Sites and E-Commerce Sites as primary sources.

VOLUME	VARIETY	VELOCITY	VERACITY	VALUE	VARIABILITY
The amount of data from myriad sources.	The types of data: structured, semi-structured, unstructured.	The speed at which big data is generated.	The degree to which big data can be trusted.	The business value of the data collected.	The ways in which the big data can be used and formatted.

- SOURCES OF BIG DATA:**
 - Following are the various sources of Big Data:
 - Social Networking Sites:** Facebook, Google, LinkedIn and other sites generate huge amounts of data. Everyday, they generate huge amounts of data as they have billions of users worldwide.
 - E-Commerce Site:** Many e-commerce sites like Amazon and Flipkart generate huge amounts of logs. The logs can be analyzed and users buying trends can be traced.