



BIG DATA COMPUTING

Prof. Rajiv Misra

IIT Patna

ABOUT THE COURSE:

In today's fast-paced digital world, the incredible amount of data being generated every minute has grown tremendously from sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and GPS signals from cell phone to name a few. This amount of large data with different velocities and varieties is termed as big data and its analytics enables professionals to convert extensive data through statistical and quantitative analysis into powerful insights that can drive efficient decisions. This course provides an in-depth understanding of terminologies and the core concepts behind big data problems, applications, systems and the techniques, that underlie today's big data computing technologies. It provides an introduction to some of the most common frameworks such as Apache Spark, Hadoop, MapReduce, Large scale data storage technologies such as in-memory key/value storage systems, NoSQL distributed databases, Apache Cassandra, HBase and Big Data Streaming Platforms such as Apache Spark Streaming, Apache Kafka Streams that has made big data analysis easier and more accessible. And while discussing the concepts and techniques, we will also look at various applications of Big Data Analytics using Machine Learning, Deep Learning, Graph Processing and many others. The course is suitable for all UG/PG students and practicing engineers/ scientists from the diverse fields and interested in learning about the novel cutting edge techniques and applications of Big Data Computing.

ABOUT INSTRUCTOR:

Dr. Rajiv Misra is working in Department of Computer Science and Engineering at Indian Institute of Technology Patna, India. He obtained his Ph.D degree from IIT Kharagpur, M.Tech degree in Computer Science and Engineering from the Indian Institute of Technology (IIT) Bombay, and Bachelor's of engineering degree in Computer Science from MNIT Allahabad. His research interests spanned a design of distributed algorithms for Mobile, Adhoc and Sensor Networks, Cloud Computing and Wireless Networks. He has contributed significantly to these areas and published more than 70 papers in high quality journals and conferences, and 2 book chapters. His h-index is 10 with more than 590 citations. He has authored papers in IEEE Transactions on Mobile Computing, IEEE Transaction on Parallel and Distributed Systems, IEEE Systems Journal, Adhoc Networks, Computer Network, Journal of Parallel and Distributed Computing. He has edited a book titled as "Smart Techniques for a Smarter Planet: Towards Smarter Algorithms" for the "Studies in Fuzziness and Soft Computing" book series, Springer (2018). He has supervised four Phd students and currently four Phd students working under his supervision in the area of big data, cloud computing, distributed computing, and sensor networks. He is a senior member of the IEEE and fellow of IETE. He has completed as the Principal Investigator of R&D Project Sponsored by DeITY entitled as "Vehicular Sensor and Mesh Networks based Future ITS". He has mentored the online courses on Cloud Computing, Advanced Graph Theory and Distributed Systems in the platform of NPTEL.

PREREQUISITES : Data Structure & Algorithms, Computer Architecture, Operating System, Database Management Systems

INDUSTRY SUPPORT : Companies like Amazon, Microsoft, Google, IBM, Facebook

Course layout

Week 1 : Introduction to Big Data

Week 2 : Introduction to Enabling Technologies for Big Data

Week 3 : Introduction to Big Data Platforms

Week 4 : Introduction to Big Data Storage Platforms for Large Scale Data Storage

Week 5 : Introduction to Big Data Streaming Platforms for Fast Data

Week 6 : Introduction to Big Data Applications (Machine Learning)

Week 7 : Introduction of Big data Machine learning with Spark

Week 8 : Introduction to Big Data Applications (Graph Processing)