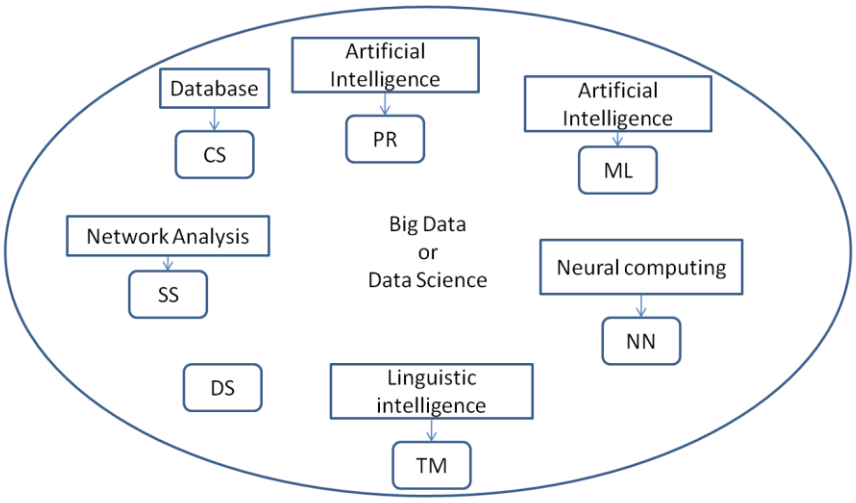
Briefing of Expert Talk at IIT-Indore on

"Advanced analytics tools: State of the art and Practice"

Speaker: Dr. Srinivas Padmanabhuni

ACM India President and Associate Vice President at Infosys Limited

The talk was on the importance of various domains expertise in Big data. Big Data is influenced by many domains like artificial intelligence, neural computing, machine learning, statistics, databases, network theory and linguistic intelligence. Big data analyst needs to be aware of above domains because no single domain alone can solve a Big data problem completely.



Need to know everything. Because no single field can solve a Big data problem completely

Further the fundamental of computing based on lambda calculus was also discussed. Lambda calculus is a function definition equivalent to Turing machine. Anonymous functions have arguments being passed to higher-order function or used to construct the result of a higher-order function that returns a function. Lambda calculus is used in Lips, Haskell, ML, Scala and Java8. A function executes in-memory and without destroying old calculation using factorial recursive functions example F(1) =1; F(N) = N*F(N-1).

Additionally, he motivating the audience for the various ACM awards like Turing award, Doctoral Dissertation Award. He discussed about few recent Turing awards winners. Matei Zaharia a PhD scholar of University of California at Berkeley created "Advanced Data Processing Architecture" won 2014 Doctoral Dissertation Award http://awards.acm.org/doctoral_dissertation/

Further discussion was on Google File System which is a system interface extensions designed to support distributed applications http://research.google.com/archive/gfs.html. He described Hadoop Distributed File System (HDFS) and Map Reduce. Map reduce is a batch-oriented processing of big data. Batch processing is a series of programs or jobs execution on a machine without any manual intervention. Here, hard-disk is involved which gives approx 1 millisecond of throughput. In contrast to batch processing, a real time processing example is Resilient Distributed Datasets (RDD) and an in-memory database is Spark (open source) and SAP Hana.

In the end he had explained few applications of big data analytics. For example fraud detection in banking system, in software and machine automation based on previous data. In future extensive automation on Big Data Analysis may create unemployment problems. Thus we need to go for “Augmented learning” (i.e. semi automation) where a machine needs help of human to make a decision.

Date: 17-August-2015

Indian Institute of Technology, Indore