The Sixth International Conference on Big Data Analytics

(BDA 2018)

18-21 December, 2018, NIT Warangal, India

(One workshop, Three keynotes, 18 research papers, One panel, Six invited talks, five tutorials)

BDA 2018 is organized by Department of Computer Science and Engineering, National Institute of Technology (NIT), Warangal, Telangana State, India.

Chief Patrons	• Ajay Prakash Sawhney, Secretary, Meity, GOI, India • N.V. Ramana Rao, Director, NIT Warangal, India
Honorary Chairs	• S.K. Gupta, IIT Delhi, India • V. Rajanna, Vice President & Regional Head, Tata Consultancy Services, Hyderabad, India
General Chair	• D.V.L.N. Somayajulu, NIT Warangal, India
Steering Committee Chair	• P. Krishna Reddy, IIIT Hyderabad, India
Program Committee Chairs	• Anirban Mondal, Ashoka University, India • Himanshu Gupta, IBM Research, India • Jaideep Srivastava, University of Minnesota, USA
Organizing Chair	• R. B. V. Subramanyam, NIT Warangal, India
Steering Committee	• S.K. Gupta, IIT Delhi, India • Srinath Srinivasa, IIIT Bangalore, India • Krithi Ramamritham, IIT Bombay, India • Sanjay Kumar Madria, Missouri University of Science and Technology, USA • Masaru Kitsuregawa, University of Tokyo, Japan • Raj K Bhatnagar, University of Cincinnati, USA • Vasudha Bhatnagar, University of Delhi, India • Mukesh Mohania, IBM Research, Australia • HV Jagadish, University of Michigan, USA • Ramesh Kumar Agrawal, Jawaharlal Nehru University, India • Divyakant Agrawal, University of California at Santa Barbara, USA • Arun Agarwal, University of Hyderabad, India • Subhash Bhalla, The University of Aizu, Japan • Jaideep Srivastava, University of Minnesota, USA • Anirban Mondal, Ashoka University, India • Sharma Chakravarthy, The University of Texas at Arlington, USA
Finance Chair	• S. Ravi Chandra, NIT Warangal, India
Sponsorship Chair	• D. V. L. N. Somayajulu, NIT Warangal, India
Publication Chair	• P. Krishna Reddy, IIIT Hyderabad, India
Workshop Chairs	• Sanjay Chaudhary, Ahmedabad University, India

	• Punam Bedi, University of Delhi, India • Subhash Bhalla, The University of Aizu, Japan
Tutorial Chairs	• Vikram Goyal, IIIT Delhi, India • Sanjay Kumar Madria, Missouri University of Science and Technology, USA
Publicity Chairs	• Shelly Sachdeva, National Institute of Technology Delhi, India • Vasudha Bhatnagar, University of Delhi, India
Panel Chairs	• Naresh Manwani, IIIT Hyderabad, India • Sharma Chakravarthy, The University of Texas at Arlington, USA
Web Site Chair	• T. Ramakrishnudu, NIT Warangal, India
Program Committee	• Alok Singh, University of Hyderabad, India • Arkady Zaslavsky, CSIRO, Australia • Arnab Basu, IIM Bangalore, India • Asoke Talukder, Precision Genomics, India • Atul Singh, Fidelity Investments, India • Bharat Bhargava, Purdue University, USA • Danish Contractor, IBM Research, India • Dhaval Patel, IBM Research, USA • Dhruba Bhattacharyya, Tezpur University, India • Girish Agrawal, Jindal Global University, India • Hoang Tam Vo, IBM Research, Australia • Ladjel Bellatreche, ENSMA, France • Lili Jiang, Umea University, Sweden • Lukas Pichl, International Christian University, Japan • Naresh Manwani, IIIT Hyderabad, India • Niloy Ganguly, IIT Kharagpur, India • Philippe Fournier-Viger, Harbin Institute of Technology, China • Pradeep Kumar, IIM Lucknow, India • Prasad Pathak, FLAME University, India • Prem Jayaraman, Swinburne University of Technology, Australia • R.K. Agrawal, JNU, India • Raja Sengupta, McGill University, Canada • Rajmohan C., IBM Research, India • Rakesh Pimplikar, IBM Research, India • Samant Saurabh, Shiv Nadar University, India • Samiulla Shaikh, IBM Research, India • Santhanagopalan Rajagopalan, IIIT Bangalore, India • Shelly Sachdeva, NIT Delhi, India • Soumyava Das, Teradata-Aster, USA • Srinath Srinivasa, IIIT, Bangalore, India • Uday Kiran, University of Tokyo, Japan • Vadlamani Ravi, IDRBT Hyderabad, India • Vasudha Bhatnagar, University of Delhi, India • Vijil Chenthamarakshan, IBM Research, USA • Vikram Goyal, IIIT Delhi, India • Rema Ananthnarayanan, IBM Research, India • Sonia Khetarpaul, Shiv Nadar University, India • Nitin Gupta, IBM Research, India • Manju Bharadwaj, University of Delhi, India • Satish Narayana Srirama, University of Tartu, Estonia • Deepak Vijaykeerthy, IBM Research, India • Manu Awasthi, Ashoka University, India • K. Shashi Prabh, ABB Research, India • Shikha Mehta, Jaypee Institute of Information Technology, India • Subhash Bhalla, University of Aizu, Japan • Ravi Kothari, Ashoka University, India

	• Akhil Kumar, Penn State University, USA • Aswin Kannan, IBM Research, India • Harshit Kumar, IBM Research, India • Sanjay Madria, Missouri S&T, USA • Avinash Sharma, IIIT Hyderabad, India • Manu Sood, Himachal Pradesh University, India • Anand Gupta, NSIT Delhi, India
Organizing committee	Faculty members of Department of Computer Science and Engineering, NIT Warangal, Telangana State, India: B. B. Amberker, S. G. Sanjeevi, K. Ramesh, Ch. Sudhakar, S. Ravi Chandra, Raju Bhukya, R. Padmavathy, K. V. Kadambari, U. S. N. Raju, P. V. Subba Reddy, Rashmi Ranjan Rout
External Reviewers	• Sharanjit Kaur, Acharya Narendra Dev College, University of Delhi, India • Alo Peets, University of Tartu, Estonia • Pelle Jakovits, University of Tartu, Estonia • Mohan Liyanage, University of Tartu, Estonia • Jakob Mass, University of Tartu, Estonia • Parul Agarwal, Jaypee Institute of Information Technology, Noida, India • Anuradha Gupta, Jaypee Institute of Information Technology, Noida, India
Conference Services offered by Institutions	Department of Computer Science and Engineering, NIT, Warangal; International Institute of Information Technology, Bangalore; University of Aizu, Japan; Indraprastha Institute of Information Technology, Delhi; School of Computer and Information Sciences, University of Hyderabad, Hyderabad; Department of Computer Science and Engineering, Indian Institute of Technology, Delhi; Department of Computer Science, University of Delhi, India; International Institute of Information Technology, Hyderabad.
Sponsors	Platinum Sponsors: National Institute of Technology Warangal TEQIP-III Project of NIT Warangal E&ICT Academy, NIT Warangal NUPA, New Delhi Silver Sponsors: Ministry of Electronics & Information Technology, Govt of India, New Delhi. Bronze Sponsors: Event Sponsors: Conference Bags Sponsors: Papadams Blue Restaurant, Hyderabad & Hanamkonda Aabmatica Technologies Pvt Ltd, New Delhi. Gudlavalleru Engineering College, Gudlavalleru.

I. 18 December 2018: Workshop (Venue: NIT Warangal)

18 December 2018 (Tuesday)
Workshop Program
International Workshop on Blockchain Technology,
National Institute of Technology Warangal (NITW), Telangana State, India, December 18, 2018
Website: https://easychair.org/cfp/IWBT18

09:30-09:45	Registration
09:45-10:00	Inauguration
10:00-10:45	Inaugural Talk: Blockchain 101 and Applications by Mukesh Mohania, IBM Australia
10:45-11:00	TEA BREAK
11:00-12:00	Paper 1 "A Comparative Study of Permissioned Blockchain Platforms", by Ravi Kanth Kotha IDRBT, Hyderabad, Narendra Kumar N IDRBT, Hyderabad and Ramakrishnudu T. NIT, Warangal, India Paper 2 "Secure digital service payments using zero knowledge proof in distributed network", by Harikrishnan M and Lakshmy K V. TIFAC-CORE in Cyber Security, Amrita Vishwa Vidyapeetham, India Paper 3 "Redactable Blockchain using Enhanced Chameleon Hash Function", by Ashritha Kondapally, Sindhu M and Lakshmy K V. TIFAC-CORE in Cyber Security, Amrita Vishwa Vidyapeetham, India
12:00-13:30	Panel Discussion Panel Topic: Challenges and Opportunities in Blockchain Technologies Moderator: Sanjay Chaudhary, Ahmedabad University / Vikram Sorathia, Kensemble Panelist 1: Sitaram Chamarty, Principal Consultant, TCS Panelist 2: Sriram Lakshmanan, Big Data Expert, Optum Panelist 3: Srinivas Naik, Cybersecurity Expert, InfoMagnum IT Services
13:30-14:30	LUNCH BREAK
14:30-16:00	Invited Talk 1 "IEE Standards Communities - Means to Access and Grow Global Markets; A focus on IEEE Blockchain Standards", by Sri Chandrasekaran, Program Manager, IEEE Standards Association Invited Talk 2 "Open Innovation, ICT (Information and Communications Technology) and Entrepreneurship", By Sriram Birudavolu, SVP Information Sciences, T-Hub Invited Talk 3 "Enterprise Blockchains - Challenges in Industry and Opportunities for Innovation", By Sridhar Vedhanabatla Sr Security Architect, Member, DSCI
16:00-16:15	TEA BREAK
16:15-17:15	Paper 4 "Blockchain for the Internet of Vehicles", by Ramaguru R and Sindhu M. TIFAC-CORE in Cyber Security, Amrita Vishwa Vidyapeetham, India Paper 5 "Run Time Verification and Vulnerability Testing of Smart Contracts", by Misha Abraham TIFAC-CORE in Cyber Security, Amrita Vishwa Vidyapeetham, India and Jevitha K. P. Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India. Paper 6 "MMPL(Medicine Multi-Participant Ledger) : A blockchain based Solution to tackle the problem of illegal use and manufacturing of medicine", by Raj Dhamsaniya, Dipkumar Patel and Harshkumar Patel School of Engineering and Applied Science, Ahmedabad University, Ahmedabad, India
17:15-17:30	Closing Remarks

II. 19-21 December 2018: Main Conference Program (Venue: NIT Warangal)

(Three keynotes, 18 research papers, six invited talks, five tutorials, One Panel)

19 December 2018 (Wednesday)		
08.30-09.00	Registration	
09.00-09.30	Inauguration	
09.30-10.30	Keynote 1	
10.30-1100	TEA BREAK	
11.00-12.30	Invited Talk 1, Invited Talk 2, Paper 1	
12.30-13.30	LUNCH BREAK	
13.30-03.00	Paper 2, Paper 3, Paper 4	Tutorial 1
15.00-15.30	TEA BREAK	
15.30-17.15	Paper 5, Paper 6, Paper 7, Paper 8	Tutorial 2
18.00-19.30	Cultural program	
19.30-21.00	Banquet Dinner	
20 December 2018 (Thursday)		
09.00-10.00	Keynote 2	
10.00 -10.30	TEA BREAK	
10.30-12.30	Invited Talk 3, Invited Talk 4, Paper 9, Paper 10, Paper 11	Tutorial 3
12.30-13.30	LUNCH	
13.30-15.15	Paper 12, Paper 13, Paper 14, Paper 15	Tutorial 4
15.15-15.30	TEA BREAK	
15.30-17.00	Panel: Can DBMS technologies play a significant role in Big Data Analytics? Moderator: Prof. Sharma Chakravarthy, The University of Texas at Arlington, USA,	
17.00-18.00	Tour of NIT Warangal	Steering committee meeting (invited members only)
18.00-19.30	Cultural program	
21 December 2018 (Friday)		
09.00 - 10.00	Keynote 3	
10.00 -10.15	TEA BREAK	
10.15-12.30	Invited 5, Invited 6, Paper 16, Paper 17, Paper 18	Tutorial 5
12.30-13.30	LUNCH BREAK	
13.30-17.00	Local Tour	

Keynote 1
Title: “Blockchain-powered Big Data Analytics Platform” by Dr. Mukesh Mohania, IBM Australia

Abstract: As crypto-currencies and other business blockchain applications are becoming mainstream, the amount of transactional data as well as business contracts and documents captured within various ledgers are getting bigger and bigger. Further, blockchains provide enterprises and consumers with greater confidence in the integrity of the data that have been captured. This gives rise to the new level of analytics that marries the advantages of both blockchain and big data technologies to provide trusted analysis on validated and quality big data. Particularly, blockchain-based big data is a perfect source for subsequent analytics because the big data maintained on the blockchain is both secure (i.e., tamper-proof and cannot be forged) and valuable (i.e., validated and abundant). In addition, data integration and advanced analysis across on-chain and off-chain data present enterprises with even more complete business insights. In this paper, we first discuss a blockchain-based business application for micro-insurance and AI marketplaces, which render blockchain-generated big data scenarios. Then, we describe the design of a blockchain-powered big data analytics platform as well as our initial steps being taken along the development of this platform.

Biography: Mukesh Mohania was an IBM Distinguished Engineer in IBM Research - India, and is currently working in IBM Research - Australia, in the areas of Blockchain, and Cognitive Data and Analytics. He is also an Adjunct Professor at University of South Australia, Australian National University, and University of Melbourne. He has worked extensively in the areas of Information Management and Autonomic Computing. His work in these areas has led to the development of new products and also influenced several existing IBM products. He has received several awards within IBM, such as "Best of IBM", "Excellence in People Management", "Outstanding Innovation Award", "Technical Accomplishment Award", "Leadership By Doin", and many more. He has published more than 120 Research papers in International Conferences and Journals and also filed more than 80 patents in these or related areas, and more than 50 have already been granted. He is an IBM Master Inventor and a member of IBM Academy of Technology. He has held several visible positions in professional activities, like VLDB 2016 Conference Organizing Chair, ACM India Vice-President. He is an ACM Distinguished Scientist and currently chairing ACM Distinguished Service Award Committee in 2017-2018.

Keynote 2
Title: “Fault Tolerant Data Stream Processing in Cooperation with OLTP Engine”,
by Prof. Yoshiharu Ishikawa, Nagoya University, Japan.

Abstract: In recent years, with the increase of big data and the spread of IoT technology and the continual evolution of hardware technology, the demand for data stream processing is further increased. Meanwhile, in the field of database systems, a new demand for HTAP (hybrid transactional and analytical processing) that integrates the functions of on-line transaction processing (OLTP) and on-line analytical processing (OLAP) is emerging. Based on this background, our group started a new project to develop data stream processing technologies in the HTAP environment in cooperation with other research groups in Japan. Our main focus is to develop new data stream processing methodologies such as fault tolerance in cooperation with the OLAP engine. In this paper, we describe the background, the objectives and the issues of the research.

Biography: Yoshiharu Ishikawa received the B.Eng., M.Eng., and Dr.Eng. degrees in computer science, all from University of Tsukuba in 1989, 1991, and 1994, respectively. He joined Nara Institute of Science and Technology (NAIST) as an assistant professor in April 1994. In April 1999, he moved to University of Tsukuba and worked as an assistant professor and an associate professor. From April 2006, he is a full professor in Nagoya University. Currently he belongs to Graduate School of Informatics, Nagoya University. His research interests include databases, especially in spatial and spatio-temporal databases, data stream management, mobile databases, and scientific databases. He was a visiting researcher of University of Maryland and Carnegie Mellon University from 1998 to 1999. He is a member of ACM, IEEE CS, IEICE, IPSJ, and DBSJ. He is now working as a general chair for VLDB 2020 in Tokyo.

Keynote 3
Title: “Demystifying Blockchains: Decentralized and Fault- tolerant Storage for the Future of Big Data?”, by Prof. Divyakant Agrawal, University of California at Santa Barbara, USA

Abstract: Bitcoin is a successful and interesting example of a global scale peer-to-peer cryptocurrency that integrates many techniques and protocols from cryptography, distributed systems, and databases. The main underlying data structure is blockchain, a scalable fully replicated structure that is shared among all participants and guarantees a consistent view of all user transactions by all participants in the cryptocurrency system. The novel aspect of Blockchain is that historical data about currency transactions is maintained in the absence of any central authority. This property of Blockchain has given rise to the possibility that the future applications will transition from centralized databases to a fully decentralized storage based on blockchains. In this talk, we start by developing an understanding of the basic protocols used in blockchain, and elaborate on its main advantages and limitations. To overcome these limitations, we provide the necessary distributed systems background in managing large scale fully replicated ledgers, using Byzantine Agreement protocols to solve the consensus problem. Finally, we expound on some of the most recent efforts to design scalable and efficient blockchains.

Biography: Divyakant Agrawal is a Professor of Computer Science at the University of California at Santa Barbara. His research interests are in the areas of databases, distributed systems, cloud computing, and big data infrastructures and analysis. He is the Fellow of the ACM, the IEEE, and the AAAS. He serves as the Editor-in-Chief of Journal of Distributed and Parallel Databases and serves on the Editorial boards of ACM Transactions of Spatial Algorithms and Systems and ACM Books. He has published 400+ articles on databases and distributed systems and has supervised 35+ PhD students during his tenure at the University of California at Santa Barbara.

INVITED TALKS

Invited Talk 1
Title: “Analysis of Narcolepsy Based on Single-Channel EEG Signals”, by Prof. Yanchun Zhang, Victoria University, Australia

Abstract: A normal person spends about a third of his life in sleep. Healthy sleep is vital to people's normal lives. Sleep analysis can be used to diagnose certain physiological and neurological diseases such as apnea, insomnia and narcolepsy. This paper will introduce the sleep stage and the corresponding electroencephalogram (EEG) characteristics at each stage. We used the deep convolutional neural network (CNN) to classify original EEG data with narcolepsy. We use perturbations based on frequency to generate adversarial examples to analyze the characteristics of narcolepsy in different sleep stages. We find that perturbations at specific frequencies affect the classification results of deep learning.

Biography: Yanchun Zhang is a Professor and Director of Centre for Applied Informatics at Victoria University since 2004. Dr Zhang obtained a PhD degree in Computer Science from The University of Queensland in 1991. His research interests include databases, data mining, web services and e-health. He has published over 300 research papers in international journals and conference proceedings including ACM Transactions on Computer and Human Interaction (TOCHI), IEEE Transactions on Knowledge and Data Engineering (TKDE), VLDBJ, SIGMOD and ICDE conferences, and a dozen of books and journal special issues in the related areas. Dr. Zhang is a founding editor and editor-in-chief of World Wide Web Journal (Springer) and Health Information Science and Systems Journal (Springer), and also the founding editor of Web Information Systems Engineering Book Series and Health Information Science Book Series. He is Chairman of International Web information Systems Engineering Society (WISE). He was a member of Australian Research Council's College of Experts (2008-2010), and also serves as expert panel member at various international funding agencies such as National Natural Science Fund of China (NSFC), “National 1000 Talents Program” of China, the Royal Society of New Zealand Marsden Fund and National Natural Science Fund of China (NSFC). He is one of the National "Thousand Talents Program" Experts in China since 2010 (currently with Fudan University).

Invited Talk 2
"Design of the Cogno Web Observatory for Characterizing Online Social Cognition", Prof. Srinath Srinivasa, IIIT Bangalore, India

Abstract: It is important to occasionally remember that the World Wide Web (WWW) is the largest information network the world has ever seen. Just about every sphere of human activity has been altered in some way, due to the web. Our understanding of the web has been evolving over the past few decades ever since it was born. In its early days, the web was seen just as an unstructured hypertext document collection. However, over time, we have come to model the web as a global, participatory, socio-cognitive "space". One of the consequences of modeling the web as a space rather than as a tool, is the emergence of the concept of Web observatories. These are application programs that are meant to observe and curate data about online phenomena. In this talk, we present the underlying model behind our Web observatory called Cogno, that is meant to observe online social cognition. Social cognition refers to the way social discourses lead to the formation of collective worldviews. Social media is modeled as a "marketplace of opinions" where different opinions come together to form "narratives" that not only drive the discourse, but may also bring some form of returns to the opinion holders. The problem of characterizing social cognition is defined as breaking down a social discourse into its constituent narratives, and for each narrative, its key opinions, and the key people driving the narrative.

Biography: Srinath Srinivasa heads the Web Science lab and is the Dean (R&D) at IIIT Bangalore, India. Srinath holds a Ph.D (magna cum laude) from the Berlin Brandenburg Graduate School for Distributed Information Systems (GkVI) Germany, an M.S. (by Research) from IIT-Madras and B.E. in Computer Science and Engineering from The National Institute of Engineering (NIE) Mysore. He works in the area of Web Science — that models of the impact of the web on humanity. Technology for educational outreach and social empowerment has been a primary motivation driving his research. He has participated in several initiatives for technology enhanced education including the VTU Edusat program, The National Programme for Technology Enhanced Learning (NPTEL) and an educational outreach program in collaboration with Upgrad. He is a member of various technical and organizational committees for international conferences like International Conference on Weblogs and Social Media (ICWSM), ACM Hypertext, COMAD/CoDS, ODBASE, etc. He is also a life member of the Computer Society of India (CSI). As part of academic community outreach, Srinath has served on the Board of Studies of Goa University and as a member of the Academic Council of the National Institute of Engineering, Mysore. He has served as a technical reviewer for various journals like the VLDB journal, IEEE Transactions on Knowledge and Data Engineering, and IEEE Transactions on Cloud Computing. He is also the recipient of various national and international grants for his research activities.

Invited Talk 3

"Humble Data Management to Big Data Analytics/Science: A Retrospective Stroll", Prof. Sharma Chakravarthy, The University of Texas at Arlington, USA,

Abstract: We are on the cusp of analyzing a variety of data being collected in every walk of life in diverse ways and holistically as well as developing a science (Big Data Science) to benefit humanity at large in the best possible way. This warrants developing and/or using new approaches -- technological, scientific, and systems -- in addition to building upon and integrating with the ones that have been developed so far. With this ambitious goal, there is also the risk of these advancements being misused or abused as we have seen so many times with respect to new technologies. In this presentation, we take the audience on a retrospective stroll on the approaches that have come about for managing and analyzing data over the last 40+ years. Since the advent of Database Management Systems (or DBMSs) and especially the Relational DBMSs (or RDBMSs), data management and analysis have seen several significant strides. Today, data has become an important tool (or even a weapon) in society and its role and importance is unprecedented. The goal of this paper is to provide the au dience an understanding of data management and analysis approaches with respect to where we have come from, motivations for developing them, and what this journey has been about in a short span of 40+ years. We sincerely hope this presentation provides a historical as well as a pedagogical perspective for those who are new to the field and provides a useful perspective that they can relate to and appreciate for those who have been working and contributing to the field.

Biography: Prof. Chakravarthy is an ACM Distinguished Scientist. He is also an IEEE Senior Member. He organized (General Co-Chair) the 13 th international Conference on Distributed Event-Based Systems (DEBS) in 2013 at UT Arlington. He spent the summers of 2013, 2014, and 2017 at the Rome Air Force Research Laboratory (AFRL-Rome) working on, respectively, continuous query processing over fault-tolerant networks and applying stream processing framework to video stream analysis. He is a co-author of the book "Stream Data Processing: A Quality of Service Perspective" published by springer in 2009. Sharma Chakravarthy is Professor of Computer Science and Engineering Department at The University of Texas at Arlington, Texas since 2000. He established the Information Technology Laboratory (IT Lab) at UT Arlington in Jan 2000. Sharma Chakravarthy has also established the NSF funded, Distributed and Parallel Computing Cluster at UT Arlington. He is the recipient of the university-level "Creative Outstanding Researcher" award for 2003 and the department level senior outstanding researcher award in 2002. He is well known for his work on stream data processing, semantic query optimization, multiple query optimization, active databases (HiPAC project at CCA and Sentinel project at the University of Florida, Gainesville), and more recently scalability issues in graph mining, social network analysis, and graph analysis of multilayered networks. His group at UTA is currently adapting map/reduce and other paradigms for scaling graph mining algorithms to very large graphs and for answering graph queries. He has applied machine learning techniques to rank answers, identify general- and topic-based experts in a Question-Answer (or Q-A) social network. His work on InfoSift - a classification system for text, email, and web - has used graph mining techniques. His current research includes fusion using multi-layered networks, stream data processing for video analysis, scaling graph mining algorithms for analyzing social networks, active and real-time databases, distributed and heterogeneous databases, query optimization, and multi-media databases. He has published over 200 papers/book chapters in refereed international journals and conference proceedings. He has given tutorial on a number of database topics, such as graph mining, active, real- time, distributed, object-oriented, and heterogeneous databases in North America, Europe, and Asia. He is listed in Who's Who Among South Asian Americans and Who's Who Among America's Teachers. Prior to joining UTA, he was with the University of Florida, Gainesville for 10 years. Prior to that, he worked as a Computer Scientist at the Computer Corporation of America (CCA) for 3 years and as a Member, Technical Staff at Xerox Advanced Information Technology, Cambridge, MA for a year. Sharma Chakrvarthy received the B.E. degree in Electrical Engineering from the Indian Institute of Science (IISc), Bangalore and M.Tech from IIT Bombay, India. He worked at TIFR (Tata Institute of Fundamental Research), Bombay, India for a few years. He received M.S. and Ph.D degrees from the University of Maryland in College park in 1981 and 1985, respectively.

Invited Talk 4

"Polystore Data Management Systems for Managing Scientific Data-sets in Big Data Archives", Prof. Subhash Bhalla, The University of Aizu, Japan

Abstract: Large scale scientific data sets are often analyzed for the purpose of supporting workflow and querying. User need to query over different data sources. These systems manage intermediate results. Most prototypes are complex and have an ad hoc design. These require extensive modifications in case of growth of data and change of scale, in terms of data or number of users. New data sources may arise to further complicate the ad hoc design. The polystore data management approach provides 'data independence' for changes in data profile, including addition of cloud data resources. The users are often provided a quasi- relational query language. In many cases, the polystore systems support distinct tasks that are user defined workflow activity, in addition to providing a common view of data resources.

Biography: Dr. Subhash Bhalla teaches at University of Aizu since 1993. His area of study is Distributed Information Systems. He completed studies up to PhD from IIT Delhi. He started teaching at School of Computer and Systems Sciences at JNU in 1986. After that, he briefly worked at Sloan School of Management at MIT during 1987-88. His current interests include- standards in Electronic Health Records Databases, data modeling in Big data archives in Time-domain Astronomy and Polystore database systems.

Invited Talk 5

"Fusion of Game Theory and Big Data for AI Applications", Dr. Praveen Paruchuri, IIIT Hyderabad, India.

Abstract: With the increasing reach of the Internet, more and more people and their devices are coming online which has resulted in the fact that, a significant amount of our time and a significant number of tasks are getting performed online. As the world moves faster towards more automation and as concepts such as IoT catch up, a lot more (data generation) devices are getting added online without needing the involvement of human agents. The result of all this is that there will be lots (and lots) of information generated in a variety of contexts, in a variety of formats at a variety of rates. Big data analytics therefore becomes (and is already) a vital topic to gain insights or understand the trends encoded in the large datasets. For example, the worldwide Big Data market revenues for software and services are projected to increase from 42 $Billion$ USD in 2018 to 103 $Billion$ in 2027. However, in the real-world it may not be enough to just perform analysis, b ut many times there may be a need to operationalize the insights to obtain strategic advantages. Game theory being a mathematical tool to analyze strategic interactions between rational decision-makers, in this paper, we study the usage of Game Theory to obtain strategic advantages in different settings involving usage of large amounts of data. The goal is to provide an overview of the use of game theory in different applications that rely extensively on big data. In particular, we present case studies of four different Artificial Intelligence (AI) applications namely Information Markets, Security systems, Trading agents and Internet Advertising and present details for how game theory helps to tackle them. Each of these applications has been studied in detail in the game theory literature, and different algorithms and techniques have been developed to address the different challenges posed by them.

Biography: Dr. Praveen Paruchuri is an Associate Professor at IIIT Hyderabad in the Machine Learning Lab. He obtained his Ph.D. from the University of Southern California in 2007 and his post doctorate from the Carnegie Mellon University. His research interests span Applied Artificial Intelligence and Machine Learning, Multi-agent Systems and Game theory. Dr. Paruchuri's Ph.D. thesis is well-known for initiating the development of a deployed game theoretic resource allocation application called the ARMOR system that currently performs efficient scheduling of security resources at the Los Angeles International airport. This work laid the foundation for enhancement of ARMOR for deployment at many locations of international importance such as the New York, LA, Boston ports and others. Dr. Paruchuri's research in the area of security scheduling has received wide publicity and his work has been described in various international news media such as Newsweek, LA Times, Times of India, Lenta.Ru etc. He is the first author of a book, has 40+ technical publications including two finalists for best paper awards and has two patents based on his work. He was officially nominated by USC in 2008 for the TR-35 award and was profiled as an innovator by the USC Stevens Institute for Innovation.

Invited Talk 6
"Deep Neural Network Based Image Captioning", Prof. Ravi Kothari, Ashoka University, India

Abstract: Generating a concise natural language description of an image enables a number of applications including fast keyword based search of large image collections. Primarily inspired by deep learning, recent times have witnessed a substantially increased focus on machine based image caption generation. In this paper, we provide a brief review of deep learning based image caption generation, the datasets and metrics used to evaluate the captioning algorithms. We conclude the paper with some discussion on promising directions for future research.

Biography: Ravi Kothari started his professional career as an Assistant Professor in the Department of Electrical and Computer Engineering of the University of Cincinnati, OH, USA where he later became a tenured Associate Professor and Director of the Artificial Neural Systems Laboratory. After about 10 years in the academia, he joined IBM Research and held several positions over the years including that of Chief Scientist of IBM Research, India and the Global Chief Architect of the IBM-Airtel outsourcing account where he introduced the first-ever stream based processing of telecom data (Airtel is one of the world's largest full service telecom providers and the Chief Architect is responsible for the IT strategy, design and realization across more than 20 countries). He also became IBM's first Distinguished Engineer from India. After about 15 years in IBM, he joined Accenture for a short stint as a Fellow, Technical Labs prior to returning to academia as Professor of Computer Science at Ashoka University.
Dr. Kothari's expertise lies in Machine Learning, Pattern Recognition, AI, Data Mining, Big Data and other data-driven technologies. His present research focuses on multiple aspects of Artificial Intelligence including the exploration of "creative" machines.

Dr. Kothari has served as an Associate Editor of the IEEE Transactions on Neural Networks, IEEE Transactions on Knowledge and Data Engineering, Pattern Analysis and Applications (Springer) as well as on the program committees of various conferences. He was an IEEE Distinguished Visitor (2003-2005 and 2006-2009) and was a member of the IBM Academy of Technology and the IBM Industry Academy. He was a recipient of the 2008 Gerstner Award (IBM's highest team award), the Best of IBM award (IBM's highest individual award) and is a fellow of the Indian National Academy of Engineering.

I. RESEARCH PAPERS

Data Systems and Frameworks

Paper 1	K.T.Sridhar, M.A.Sakkeer, Shiju Andrews and Jimson Johnson. MPP SQL Query Optimization with RTCG. Abstract. Analytics database dbX is a cloud agnostic, MPP SQL product with both DSM and NSM stores. One of the techniques for better micro optimization of SQL query processing is runtime code generation and JIT compilation. We propose a RTCG model that is both query aware and hardware conscious extending analytics SQL query processing to a high degree of intra-query parallelism. Our approach to RTCG, at system level targets to maximize benefits from modern hardware, and at use level focuses on typical, industry type SQL, somewhat different from standard benchmarks. We describe the model, highlighting its novel aspects, techniques implemented and product engineering decisions in dbX. To evaluate the efficacy of the RTCG model, we perform experiments on desktop and cloud clusters, with standard and synthetic benchmarks, on data that is more commensurate in size with industry applications.
Paper 2	Purnima Shah and Sanjay Chaudhary. Big Data Analytics Framework for Spatial Data Abstract. In the world of mobile and Internet, large volume of data is generated with spatial components. Modern users demand fast, scalable and cost-effective solutions to perform relevant analytics on massively distributed data including spatial data. Traditional spatial data management systems are becoming less efficient to meet the current users demand due to poor scalability, limited computational power and storage. The potential approach is to develop data intensive spatial applications on parallel distributed architectures deployed on commodity clusters. The paper presents an open-source big data analytics framework to load, store, process and perform ad-hoc query processing on spatial and nonspatial data at scale. The system is built on top of Spark framework with a new input data source NoSQL database i.e. Cassandra. It is implemented by performing analytics operations like filtration, aggregation, exact match, proximity and K nearest neighbor search. It also provides an application architecture to accelerate ad-hoc query processing by diverting user queries to the suitable framework either Cassandra or Spark via a common web based REST interface. The framework is evaluated by analyzing the performance of the system in terms of latency against variable size of data.
Paper 3	Sanket Mishra, Mohit Jain, B Siva Nagasasank and Chittaranjan Hota. An ingestion based analytics framework for Complex Event Processing Engine in Internet of Things Abstract. Internet of Things (IoT) is the new paradigm that connects the physical world with the virtual world. The interconnection is generated by the optimal deployment of sensors which continuously generate data and streams it to a data store. The concept drift and data drift are integral characteristics of IoT data. Due to this nature, there is a need to process data from various sources and decipher patterns in them. This process of detecting complex patterns in data is called Complex Event Processing which provides near real-time analytics for various IoT applications. Current CEP deployments have a inherent capability to react to events instantaneously. This leaves room to develop CEPs which are proactive in nature which can take the help of various machine learning (ML) models to work together with CEP. In this paper, the usage of Complex Event Processing (CEP) engine is exhibited that allows the inference of new scenarios out of incoming traffic data. This conversion of historical data into actionable knowledge is undertaken by a Long Short Term Memory (LSTM) model so as to detect the occurrence of an event well before time. The experimental results suggest the rich abilities of Deep Learning to predict events proactively with minimal error. This allows to deal with uncertainties and steps for significant improvement can be made in advance.
Paper 4	Vaibhav Pandey and Poonam Saini, An Energy-Efficient Greedy MapReduce Scheduler for Heterogeneous Hadoop YARN Cluster

	Abstract: Energy efficiency of a MapReduce system has become an essential part of infrastructure management in the field of big data analytics. Here, Hadoop scheduler plays a vital role in order to ensure the energy efficiency of the system. A handful of MapReduce scheduling algorithms have been proposed in the literature for slot-based Hadoop system (*i.e.*, Hadoop 0.*x* and Hadoop 1.*x*) to minimize the overall energy consumption. However, YARN-based Hadoop schedulers have not been discussed much in the literature. In this paper, we design a scheduling model for Hadoop YARN architecture and formulate the energy efficient scheduling problem as an Integer Program. To solve the problem, we propose a *Greedy* scheduler which selects the best job with minimum energy consumption in each iteration. We evaluate the performance of the proposed algorithm against the FAIR and Capacity schedulers and find out that our greedy scheduler shows better results for both CPU- and I/O intensive workloads.
Financial Data Analysis and Data Streams	
Paper 5	Rao Casturi and Rajshekhar Sunderraman. Distributed Mortgage Factor Aggregation Calculation Framework on Cloud Computing Environment Abstract. Even though the recent technological innovations in cloud computing, distributed data base architecture grew to a point where they can now address Big Data process, still most of the companies are struggling to implement their solutions on cloud computing environment. This is mainly due to lack of proper case studies and application frameworks available on a public research domain. Most of the implementations are vendor provided, high cost, consulting type and long drawn projects which consume valuable Business and IT resources for an organization. In this paper we propose a simple to implement (parallel data load and any aggregation calculation framework), easy to maintain and flexible architecture framework which can be adapted as a tool for small to mid-size investment organization in implementing a Distributed Cloud Computing Architecture. Our framework is an extension of traditional Distributed Database Design with the horizontal partition of the relations to parallelize the computation on Azure SQL instance and materialize the aggregated results with SQL Views for users, Business Intelligence (BI) Reporting, Data Mining and Knowledge Discovery applications. The solution is implemented on Azure SQL Cloud Computing platform to build the financial calculation framework.
Paper 6	Narinder Punn and Sonali Agarwal. Testing Concept Drift Detection Technique on Data Stream Abstract: Data mutates dynamically, and these transmutations are so diverse that it affects the quality and reliability of the model. Concept Drift is the quandary of such dynamic cognitions and modifications in the data stream which leads to change in the behaviour of the model. The problem of concept drift affects the prognostication quality of the software and thus reduces its precision. In most of the drift detection methods, it is followed that there are given labels for the incipient data sample which however is not practically possible. In this paper, the performance and accuracy of the proposed concept drift detection technique for the classification of streaming data with undefined labels will be tested. Testing is followed with the creation of the centroid classification model by utilizing some training examples with defined labels and test its precision with the test set and then compare the accuracy of the prediction model with and without the proposed concept drift detection technique.
Paper 7	Sourabh Yadav and Nonita Sharma. Homogenous Ensemble of time-series Models for Indian Stock Market Abstract. In the present era, Stock Market has become the storyteller of all the financial activity of any country. Therefore, stock market has become the place of high risks, but even then it is attracting the mass because of its high return value. Stock market tells about the economy of any country and has become one of the biggest investment place for the general public. In this manuscript, we present the various forecasting approaches and linear regression algorithm to successfully predict the Bombay Stock Exchange(BSE) SENSEX value with high accuracy. Depending upon the analysis performed, it can be said successfully that Linear Regression in combination with different mathematical functions prepares the best model. This model gives the best output with BSE

	SENSEX values and Gross Domestic Product(GDP) values as it shows the least p-value as 5.382e-10 when compared with other model's p-values.
Paper 8	Pulkit Mehndiratta, Mohit Gupta, Ayushi Asthana and Nishant Joshi. Improving Time Series Forecasting Using Mathematical and Deep Learning Models Abstract. With the increase in the number of internet users, there is a deluge of traffic over the web, handling Internet traffic with much more optimized and efficient approach is the need of the hour. In this work, we have tried to forecast Internet traffic on TCP/IP network using web traffic data of Wikipedia articles, provided by Kaggle1. We work on the stationarity of time series and use mathematical concepts of log transformation, differencing and decomposition in order to make the time series stationary. Our research presents an approach for forecasting web traffic for these articles using different statistical time series models such as Auto-Regressive (AR) model, Moving Average (MA) model, Auto-Regressive Integrated Moving Average (ARIMA) Model and a deep learning model - Long Short-Term Memory (LSTM). This research work opens the possibility of efficient traffic handling thus, leading to improved performance for an organization as well as better experience for the users on the internet.
Web and Social Media Data	
Paper 9	Shriyansh Agrawal, Lalit Sanagavarapu and Y. Raghu Reddy. Automated Credibility Assessment of Web Page Based on Genre Abstract. With more than a billion web sites, volume and variety of content available for consumption is huge. However, credibility, an important quality characteristic of web pages is questionable in many cases and tends to be non-uniform. Credibility can increase or reduce the importance of web page leading to potential gain or loss of user base. Credibility without factoring genre of content (for example, Help, Article, Discussion, etc.) can lead to incorrect assessment. Depending on the genre, the importance of features such as web page date time modified, grammar, image to text ratio, in and out links, and other web page features differ. We propose a genre credibility assessment based on web page surface features and their importance in a genre. Further, we built a *WEBCred* framework to assess *GCS* (Genre based Credibility Score) with flexibility to add/modify genres, its features and their importance. We validated our approach on 10,429 'Information Security' related web pages; the assessed score correlated 35% with crowd sourced Web Of Trust (WOT) score and 39% with Alexa ranking.
Paper 10	Sonu Gupta, Shelly Sachdeva, Prateek Dewan and Ponnurangam Kumaraguru CbI: Improving Credibility of User-Generated Content on Facebook Abstract. Online Social Networks (OSNs) have become a popular platform to share information with each other. Fake news often spread rapidly in OSNs especially during news-making events, e.g. Earthquake in Chile (2010) and Hurricane Sandy in the USA (2012). A potential solution is to use machine learning techniques to assess the credibility of a post automatically, i.e. whether a person would consider the post believable or trustworthy. In this paper, we provide a fine-grained definition of credibility. We call a post to be credible if it is *accurate, clear, and timely*. Hence, we propose a system which calculates the *Accuracy, Clarity, and Timeliness* (A-C-T) of a Facebook post which in turn are used to rank the post for its credibility. We experiment with 1,056 posts created by 107 *pages* that claim to belong to news-category. We use a set of 152 features to train classification models each for A-C-T using supervised algorithms. We use the best performing features and models to develop a RESTful API and a Chrome browser extension to rank posts for its credibility in real-time. The random forest algorithm performed the best and achieved ROC AUC of 0.916, 0.875, and 0.851 for A-C-T respectively.
Paper 11	Shivani Batra, Shelly Sachdeva, Aayushi Bansal and Suyash Bansal. Modeling Sparse and Evolving Data Abstract. Existing relational database management system (RDBMS) excels in providing transactional support. However, RDBMS performance declines when sparse and evolving data needs to be stored. Modeling of highly evolving and sparse data is a major issue that needs attention to provide faster and competent technology solutions. This research work is focused on providing a solution to handle sparseness and frequent evolution of data with adherence for the transactional support. Recently, authors propose an extension of binary table approach to overcome the lacking aspects. The proposed approach is termed as Multi Table Entity Attribute Value

	(MTEAV) model. To make users completely unaware about the underlying modeling approach, MTEAV is augmented with a translation layer. It translates conventional SQL query (as per the relational model) to a new SQL query (as per MTEAV structure) to provide the user friendly environment. In this research, authors extend the functionality of the translation layer to provide support for data definition (creating, reading, updating and deleting schema). Authors have experimented MTEAV for analyzing the effect of sparseness on the performance of MTEAV. Results achieved clearly indicate that the MTEAV performance increases with increase in sparseness.
Paper 12	Keshab Nath. A Parallel Approach to Detect Communities in Evolving Networks Abstract. To understand the dynamics, functional and topological aspect of the real-world networks it is necessary to segregate the network into sub-networks, where each member of a sub-network possess analogous characteristics. Numerous number of community finding approaches are proposed in the last few decades to overcome the issues associated with community detection. Although, most of the conventional approaches rely on the premises that networks are static in nature and there won't be any alternation over time. Moreover, all these approaches are single machine approach and hence exhibits poor scalability. In this work, we propose a new incremental parallel community detection method, PcDEN (Parallel Community Detection approach in Evolving Networks). Our proposed method can detect communities in dynamic distributed networks. We define a new Affinity score based on intracommunity strength between nodes and their neighbors. We also derive a new model to perform community merging, based on common high degree nodes present in both the communities. We tested our algorithm onvarious real-world networks for our experimentation. Results show that, PcDEN produce satisfactory output with respect to various assessment indices.
Predictive Analytics in Healthcare and Agricultural Domains	
Paper 13	Sakthi Ganesh and Asoke Talukder. Robotic Formal Methods, Artificial Intelligence, Big-data Analytics, and Knowledge Engineering in Medical Care to Reduce the Burden of Disease Abstract: Medical errors and overtreatment combined with growing noncommunicable disease population are responsible for increase in the *burden of disease* and *health disparity*. To control this burden and disparity, automation with zero defects must be introduced in evidence based medicine. In *safety critical* systems, zero defects are achieved through formal methods. A formal model is tested (proved) and the target system is generated through automation with the removal of error prone programming or construction phase. Inspired by similar ideas, we created *DocDx*, a novel formal method driven medical care framework without any programming phase involved. We convert *clinical pathways* into a *multipartite directed weighted graph* (MDWG) that embeds the medical intelligence. The autonomous interpreters in the server presents *natural language generator* (NLG) pathophysiology questions a doctor would normally ask a patient to understand the signs and symptoms of a disease. The biological terms and human understandable unstructured text entered in DocDx client is made machine understandable through AI NLP engine and translated into biomedical ontology concepts. A new medical condition or presentation of disease in DocDx will need a new clinical pathway translated into a MDWG without the need for any programming or application development process either at the client or at the server end.
Paper 14	Rashmi Priya and Dharavath Ramesh. Adaboost. RT based Soil N-P-K Prediction Model for Soil and Crop Specific Data: A Predictive Modelling Approach Abstract. In relation to the evaluation of the soil breeding status of a region or realm, the soil characteristics are an important aspect in terms of agricultural production. Nitrogen, phosphorus, potassium, and sulfur are important elements of soil that regulate its fertility and yield of crops. Due to the low efficiency of other inputs or due to the use of unbalanced and inadequate fertilizer, the reaction of chemical fertilizer nutrients (production) efficiency in recent years has reduced considerably under intensive agriculture. Stability in crop productivity cannot be extended without the judicial use of macro and micro nutrients to overcome existing deficiencies. The information on the availability of macro nutrients in the study area is low. Therefore, the current study has been done to know the condition of soil nutrients. Use of advanced agricultural technology can help in predicting soil nutrient content and can help farmers to decide the amount of fertilizers to use on a particular land. The proposed study focuses on the accurate prediction of N-P-K content in the

	given land by utilizing the predication method using Adaboost.RT method. A comparison is also made in between the nutrient utilized using traditional methods and the proposed method. Experimental results show that the proposed stream outperforms with other existing methodologies.
Machine Learning and Pattern Mining	
Paper 15	Mohit Agarwal. PRISMO Priority Based Spam detection using multi objective optimization Abstract. The rapid growth of social networking sites such as Twitter, Facebook, Google+, MySpace, Snapchat, Instagram, etc., along with its local invariants such as Weibo, Hyves, etc., has made them infiltrated with a large amount of spamming activities. Based on the features, an account or content can be classified as spam or benign. The presence of some irrelevant features decreases the performance of the classifier, understandability of dataset, and the time requirement for training and classification increases. Therefore, Feature subset selection is an essential phase in the process of machine learning mechanism. The objective of feature subset selection is to choose a subset of size 's' (s<n) from the total set of 'n' features that results in the least classification error. The feature subset selection problem can be represented as a problem of optimization in which the objective is to choose the near-optimal subset of features. Based on the literature survey, it is found that the classifier will offer its best performance if the data with high dimension is reduced such that it includes only appropriate features having lesser redundancy. The contribution of this paper comprises feature subset and its cost optimization simultaneously. The fundamental aspect PRISMO is to generate a primary feature subset through various optimization algorithms for the initialization stage. Further, the subset has been generated using the initial feature set based on their priority using basic rules of conjunction and disjunction. To evaluate the overall efficiency of PRISMO, various experiments were carried out using different dataset. The obtained result shows that the proposed model effectively reduces the cardinality of features without any bias to a specific dataset and any degradation to the classifier accurateness.
Paper 16	Himaja D, Maruthi Padmaja T and Radha Krishna P. Oversample Based Large Scale Support Vector Machine for Online class Imbalance Abstract. Dealing with online class imbalance from evolving stream is a critical issue than the conventional class imbalance problem. Usually, the class imbalance problem occurs when one class of data severely outnumbers the other classes of data, thus leads to skewed class boundaries. In the case of online class imbalance problem, the degree of class imbalance changes over time and the present state of imbalance is not known a prior to the learner. To address such problem, in this paper, we present an Oversampling based Online Large Scale Support Vector Machine (OOLASVM) algorithm which is a hybrid of active sample selection and over sampling of Support Vectors and thereby both oversampling and under sampling coexists while learning the new boundary. Further, OOLASVM maintains the balanced boundary throughout the learning process. Results on simulated and real world datasets demonstrate that proposed OOLASVM yields better performance than existing approaches such as Generalized Oversampling based Online Imbalanced Learners and Over Online Bagging.
Paper 17	Ashish Patel, Satyendra Singh Chouhan and Rajdeep Niyogi. Using Crowd Sourced Data for Music Mood Classification Abstract. Music has been part of human lives since ancient times. We have hundreds of millions of songs representing different cultures, mood and genres. These songs are readily accessible using Internet and streaming services. However, the discovery of the right music piece to listen is hard and an automated assistance to find the right song among the millions is always desired. There have been several attempts to classify music on the basis of their genres but their efforts have not been much fruitful because of lack of good and large datasets. Moreover, identifying the set of features to represent the music in a summarized way is also a challenging task. In this work, we present an automated music mood classification approach that uses crowd-sourced platforms to label the songs. It eliminates the subjectivity of one's perception of mood on a song. We have confined our work to two classes of mood: happy and sad. The proposed approach is tested with three machine learning models: artificial neural networks (ANN), Decision Tree (DT) and Support Vector Machine (SVM). The experimental results show that ANN performs better than DT and SVM.
Paper 18	Amrit Pal and Manish Kumar. Applying Big Data Intelligence for Real Time Machine Fault

	Prediction Abstract: Continuous use of mechanical systems requires precise maintenance. Automatic monitoring of such systems generates a large amount of data which require intelligent mining methods for processing and information extraction. The problem is to predict the faults generated with ball bearing which severely degrade operating conditions of machinery. We develop a distributed fault prediction model based on big data intelligence that extracts nine essential features from ball bearing dataset through distributed random forest. We also perform a rigorous simulation analysis of the proposed approach and the results ensure the accuracy/correctness of the method. Different types of fault classes are considered for prediction purpose and classification is done in a supervised distributed environment.
	# IV.Tutorial Talks
Tutorial 1	Title: "AI Models and Trust", Himanshu Gupta, IBM Research, India; DiptiKalyan Saha, IBM Research, India;Vijay Arya, IBM Research, India. Abstract: In this tutorial, we will discuss - how we can trust AI models? Put simply, we trust things which behave as we expect them to. We will discuss four key requirements - bias, lineage, explainability and robustness, which go a long way towards trusting an AI model and understanding their behavior. We will discuss few AI trust and governance use-cases and how these use-cases inter-relate with bias, explainability, lineage and robustness requirements. We will provide a brief overview of the research efforts in these four domains and outline some research problems in this space. At IBM, we have been working on building a trusted AI platform and we will also discuss some insights from our experience. Biography: Himanshu Gupta is a senior researcher at AI Engineering department of IBM Research - India. He received his MS and BTech in Computer Science from IIT Delhi and IIT Kanpur respectively. His research interests include data management, data mining, information integration, Spark/map-reduce based processing etc. His current focus in on building scalable lineage services on IBM watson data platform. Diptikalyan Saha is a senior researcher and manager in AI Engineering department of IBM Research - India. He received his Ph.D. degree in Computer Science from the State University of New York at Stony Brook and his BE from Jadavpur university. His research interests include Artificial Intelligence, NLP, Knowledge representation, Program Analysis, Security, Software Debugging, Testing, Verification, and Programming Languages. His current focus is on building dependable AI systems using bias and adversarial testing, debugging, and verification. Vijay Arya is a senior researcher at AI Engineering department of IBM Research - India. He received his Ph.D. degree in Computer Science from INRIA, France. His research interests include security, machine learning and data management. His current focus is on developing explainable ML models.
Tutorial 2	Title: "Post facto of Cambridge Analytica- The Dawn of Social Computing", Amitava Das, Mahindra Ecole Centrale, Hyderabad, India; Tanmoy Chakraborty, IIIT Delhi, India Abstract: The unexpected win of Trump at least has proven one fact that social computing --- personality, values, ethnicity, gender, political view analysis from social media content have immense power and it can change the shape of the world! Facebook and the data analytics company called Cambridge Analytica are at the center of a dispute over the harvesting and use of personal data. The question comes - what to learn from the history? Social Science is on the verge of a paradigm shift that allows us to ask and answer questions that were unthinkable a few decades ago. We can now collect data about human behavior on a scale never before possible and with

	tremendous granularity and precision, but as always said - great power comes with great responsibility. Therefore, Social Computing has been emerging as an independent research problem. Social Computing, mostly deals with a data-driven understanding of complex social networks, which often requires knowledge about graph analysis, data mining, natural language processing, and machine learning. Therefore, this tutorial is formed around five major topics that all fall under the emerging field of computational social science: Psychology and Sociology, Natural Language Processing, Complex Network, Big Data, and Machine Learning and how to use them together to answer few buzzing questions of our times. Biography Amitava Das is an Associate Professor in the department of Computer Science & Engineering at Mahindra Ecole Centrale, Hyderabad. Earlier he worked for IIIT Sri City. In his research career he has spent significant time in USA, Europe and Japan and also worked for Samsung Research India. His research area is sentiment analysis, NLP, Social Computing, conversational AI, and Deep Learning. Currently he is consulting for AI-NLP with several Indian IT companies including Wipro. Tanmoy Chakraborty is an Assistant Professor and a Ramanujan Fellow at the Dept. of Computer Science & Engg., IIIT Delhi, India. Prior to this, he was a postdoctoral researcher in University of Maryland, College Park, USA. He completed his Ph.D as a Google India PhD fellow at IIT Kharagpur, India in 2015. His primary research interests include social network analysis, Data Mining, and Natural Language Processing. He has received several awards including the Google India Faculty Award, Early Career Research Award, DAAD Faculty fellowship, Best reviewer award in WWW'18, best PhD thesis award by Xerox Research, IBM Research and Indian National Academy of Engineering (INAE). He has been serving as a PC member of several conferences including WWW, WSDM, NAACL, AAAI, IJCAI, PAKDD.
Tutorial 3	Title: “Malware Detection using Machine Learning and Deep Learning”, Hemant Rathore, BITS, Pilani, Goa Campus, India; Swati Agarwal, BITS, Pilani, Goa Campus, India; Sanjay K. Sahay, BITS, Pilani, Goa Campus, India; and Mohit Sewak, BITS, Pilani, Goa Campus, India. Abstract: Research shows that over the last decade, malware have been growing exponentially, causing substantial financial losses to various organizations. Different anti-malware companies have been proposing solutions to defend attacks from these malware. The velocity, volume, and the complexity of malware are posing new challenges to the anti-malware community. Current state-of-the-art research shows that recently, researchers and anti-virus organizations started applying machine learning and deep learning methods for malware analysis and detection. We have used opcode frequency as a feature vector and applied unsupervised learning in addition to supervised learning for malware classification. The focus of this tutorial is to present our work on detecting malware with (1) various machine learning algorithms and (2) deep learning models. Our results show that the Random Forest outperforms Deep Neural Network with opcode frequency as a feature. Also in feature reduction, Deep Auto-Encoders are overkill for the dataset, and elementary function like Variance Threshold perform better than others. In addition to the proposed methodologies, we will also discuss the additional issues and the unique challenges in the domain, open research problems, limitations, and future directions. Biography Hemant Rathore is an Assistant Professor Department of CS and IS at BITS, Pilani, Goa Campus, India. He is currently pursuing his PhD in Malware Analysis and Detection. He graduated with an M.E. degree from BITS Pilani in 2013. He worked in the area of Computer Security for 3 years at Symantec, India. His research interests are in the area of Data Mining, Malware Analysis, Network Security, Cryptography, Machine Learning, and Operating Systems.

	Swati Agarwal is a Visiting Assistant Professor in Computer Science Department at BITS Pilani-Goa, India. Her research interests are in the area of Social Computing, Natural Language Processing, Security Informatics, and Text mining and Analytics. She has a PhD in Computer Science (Social Media Analytics and Security Informatics) from IIIT-Delhi in 2017. She has been serving as a TPC member of various conferences and workshops including PAKDD, ADMA, NAACL, ACL, COLING and many more. Sanjay K. Sahay is an Associate Professor in the Department of CS and IS at BITS Pilani, Goa Campus, India. He is also a Visiting Associate of IUCAA, Pune. His research interests are in the area of Network Security, Information Security, Data Science, and Gravitational Waves. Under his supervision three PhD has been graduated, one has submitted the PhD thesis and currently supervising two students. Mohit Sewak is a senior Research Scholar at BITS Pilani, Goa Campus. He has 14 years of rich Industry experience in the space of Machine Learning and Cognitive Computing. Mohit has so far authored two books, and he has been the lead/solo inventor of 12+ patents 4 (Granted) and 8+ (Applied) with the USPTO.
Tutorial 4	Title: "Spatial Co-location Pattern Mining", Venkata M. V. Gunturi, IIT, Ropar, Punjab, India. Abstract: Widespread use of spatial computing technologies has lead to increasing interest in mining interesting and non-trivial patterns from spatial data. Over the years several works have made progress towards this end by exploring different aspects of the problem of finding patterns from data which is embedded in a geographic space. This talk would focus on one particular pattern family called the spatial co-location patterns which have gained widespread attention due to their potential uses. The talk would cover well known algorithms for mining spatial co-location patterns from large dataset in a time-efficient manner. We would also cover some of the new directions of research being explored in this area. Biography Dr. Venkata Gunturi is an Assistant Professor at the Indian Institute of Technology Ropar. He obtained his PhD from the Dept of Computer Science and Engineering at the University of Minnesota, Minneapolis, USA. His research interests include spatial and spatio-temporal databases, spatial data mining, navigation algorithms on spatial networks. He is a recipient of the Early Career Award from DST, SERB
Tutorial 5	Title: "Emerging Technologies and Opportunities for Innovation in Financial Data Analytics: A Perspective", Anirban Mondal, Ashoka University, Sonipat, Haryana, India; Atul Singh, Fidelity Management and Research, Bengaluru, India. Abstract: Several key transformations in the macro-environment coupled with recent advances in technology have opened up tremendous opportunities for innovation in the financial services industry. We discuss the implications and ramifications of these macro-environmental trends for data science research. Moreover, we describe novel and innovative IT-enabled applications, use-cases and techniques in retail financial services as well as in financial investment services. Furthermore, this tutorial identifies the research challenges that need to be addressed for realizing the full potential of innovation in financial services. Examples of such research challenges include context-aware analytics over uncertain and imprecise data, data reasoning and semantics, cognitive and behavioural analytics, design of user-friendly interfaces for improved expressiveness in querying financial service providers, personalization based on fine-grained user preferences and financial Big Data processing on Cloud- based infrastructure. Additionally, we discuss new and exciting opportunities for innovation in financial services by leveraging the new and emerging financial technologies as well as Big Data technologies. Biography Anirban Mondal is an Associate Professor of Computer Science at Ashoka University. His broad

	research interest is in big data analytics, urban informatics, financial analytics, mobile crowdsourcing, large-scale distributed systems and database indexing. He has an established reputation, key presence and high visibility in the international research community. He has numerous publications in key conferences/journals and has also been actively involved as a PC Chair/Co-chair, PC member, journal reviewer as well as keynote/tutorial speaker at reputed international conferences/workshops. He has served as an ACM India Eminent Speaker and has also been awarded the prestigious JSPS (Japanese Society for Promotion of Sciences) Fellowship. He has research collaborations with prestigious Universities in Japan, Singapore, USA, Australia and India. Prior to this, he has worked in organizations such as University of Tokyo, Xerox Research Lab and IIIT Delhi. Based on his industry experience in designing practical research applications in urban informatics and financial analytics, he has multiple USPTO granted patents as well as several filed patents. He holds a PhD in Computer Science from the National University of Singapore, an MBA from the University of Massachusetts Amherst (UMass) and a BTech from the Indian Institute of Technology (IIT) Kharagpur. His technological expertise coupled with his business capabilities as well as his ability to create a big vision and execute it to completion in diverse multi-cultural settings make him an exciting innovator. Atul Singh is a data scientist at a reputed financial firm. His research interests include Natural Language Processing (NLP), geo-spatial analytics and reinforcement learning with a focus on finance. He has over sixteen years of software industry work experience in research, and innovation. Prior to his current employment, he has worked at Xerox Research Centre India and Robert Bosch Research Technology Centre India. He has nine granted US patents, eleven pending US patent applications, and several research publications in various international forums. He has given several seminars in the field of financial data analytics and big data. He has a PhD in Computer Science from Trinity College Ireland, and a B.Tech from Indian Institute of Technology Kanpur.