Panel:
Key Challenges for Future Big-Data to Knowledge (BD2K) Technologies

Panelists:

Moderator: Vijay Raghavan

1) Ricardo Baeza-Yates, VP, Yahoo!Labs
2) Anil Goel, VP and Chief Architect, SAP Hana Data Platform
3) H.V. Jagadish, Professor, University of Michigan, USA
4) Laura Pullum, Research Scientist, Oak Ridge National Laboratory
5) Ranjan Sinha, Head of Data Science eBay
6) Kristin Tolle, Director Microsoft, USA

Bios of Moderator and Panelists

Moderator:

Vijay Raghavan is the Alfred and Helen Lamson/ BoRSF Endowed Professor in Computer Science at the *Center for Advanced Computer Studies* and the Director of the NSF-sponsored *Industry/ University Cooperative Research Center for Visual and Decision Informatics*. His research interests are in data mining, information retrieval, machine learning and Internet computing. He has published over 250 peer-reviewed research papers- appearing in top-level journals and proceedings- that cumulatively accord him an *h-index* of 32, based on citations. He has served as major advisor for 24 doctoral students. He has also directed industry-sponsored research, on projects pertaining to Neuro-imaging based dementia detection and literature-based biomedical hypotheses generation, respectively. Dr. Raghavan serves as a member of the Executive Committee of the IEEE Technical Committee on Intelligent Informatics (IEEE-TCII), the Web Intelligence Consortium (WIC) Technical Committee and the Web Intelligence and Intelligent Agent Technology Conferences' Steering Committee. He is one of the Editors-in-Chief of the Web Intelligence journal and an Associate Editor of the ACM Transactions on Internet Technology. He is an ACM Distinguished Scientist and served as an ACM Distinguished Lecturer from 1993 – 2006. In addition, he served as a member of the Advisory Committee of the *NSF Computer and Information Science and Engineering* directorate *(CISE-AC)* during 2008 – 2010.

Panelists:

Ricardo Baeza-Yates is VP of Research at Yahoo Labs, Sunnyvale, USA. He is a Fellow of both ACM and IEEE. His research interests are Web Search and Data Mining as well as scalability in general. He received his Ph.D. from the University of Waterloo, Canada. He is co-author of the best-selling textbook, Modern Information Retrieval, published in 1999 by Addison-Wesley with a second enlarged edition in 2011 that won the ASIS&T 2012 book of the year award. He can be reached at rbaeza@acm.org.

Anil K. Goel is a Chief Architect at SAP where he works with the globally distributed HANA Platform and Database engineering team to drive forward looking architectures, vision, strategy and execution for all SAP data management products and technologies. He oversees data platform related co-innovation projects with hardware and software partners as well as collaborative research and internship programs with many universities in North America and Europe. His interests include database system architecture, in-memory and large scale distributed computing, self-management of software systems and cost modelling. Anil earned a PhD in CS from University of Waterloo. He also holds M.Tech in CS from the Indian Institute of Technology, Delhi, and B.E. (Electronics and Communications Engineering) from the University of Delhi.

H. V. Jagadish (Jag) is a computer scientist in the field of database systems research. He is the Bernard A. Galler Collegiate Professor of Electrical Engineering and Computer Science at the University of Michigan at Ann Arbor and Distinguished Scientist at the Michigan Institute for Data Science. Prior to joining the Michigan faculty, he spent over a decade at AT&T Bell Laboratories as a research scientist where he would eventually become head of the Database division. Jagadish earned his bachelor's degree from the Indian Institute of Technology, Delhi and a doctorate in Electrical Engineering from Stanford University in 1985. He was elected fellow of the Association for Computing Machinery in 2003 and trustee of the VLDB Endowment in 2004. He was the founding editor of the Proceedings of the VLDB Endowment (PVLDB) in 2008.

Laura Pullum is a research scientist in the Computer Science and Engineering Division at Oak Ridge National Laboratory (ORNL), a Department of Energy laboratory. She has led numerous research projects in industry, non-profit organizations, academia and government laboratories. Laura's research has focused on the dependability of software-intensive systems that incorporate state-of-the-art technology. Laura is the author of *Software Fault Tolerance: Techniques and Implementation* (Artech House) and lead author of *Guidance for the Verification and Validation of Neural Networks* (IEEE Computer Society Press/Wiley). She has authored numerous book chapters and peer-reviewed papers; holds a patent, serves on technical advisory boards and international science review panels, and serves on the standards working group for the *IEEE Standard for System Verification and Validation*. She has served as a reviewer for numerous international journals, conferences and books; served on the organizing or program committees for international conferences; and is a senior member of the IEEE Computer Society.

Laura's research interests include software dependability and intelligent systems. She currently conducts research in the evaluation, verification and validation of predictive analytics and machine learning systems, as well as using "big data" and learning to provide insights into disease dynamics. She holds a BS in Mathematics; Master's degrees in Operations Research, Business Administration, and Geology; and a doctorate in Systems Engineering and Operations Research.

Ranjan Sinha is Head of Data Science Engineering & Technology for Customer Analytics and Personalization at eBay, San Jose. Earlier, as lead scientist at eBay, he led several business-impacting projects in recommendations and personalization, which have significantly enhanced consumers' shopping experiences. He has also contributed in domains such as infrastructure availability, security, and identity linking.Prior to joining eBay, Ranjan was a research academic and chief investigator at the University of Melbourne. He earned his PhD in computer science from RMIT University and has published over 30-refereed works, including in top-tier venues such as ACM JEA, ACM SIGMOD, Bioinformatics journal, IEEE Big Data, and VLDB journal. He was amongst the Top 12 Asia-Pacific Young Inventors and appeared in the article on *Cutting-Edge Crusaders* in the WSJ. He was awarded the Sort Benchmark medals for both JouleSort and PennySort in 2009. He presented a tutorial on *E-commerce Personalization at Scale* at the 2014 ACM CIKM conference and has been recently interviewed by KDnuggets. He is also a co-organizer of the *Bay Area Search Meetup* consisting of over 2,000 members.

Ranjan's current interests include scaling data science solutions, introducing engineering practices in data science pipelines, developing real-time predictive analytic solutions, and in the application of semantic relationships between words in large text corpus.

Kristin M. Tolle, Ph.D. is the Director of the Data Science Initiative in Microsoft Research Labs. Since joining Microsoft in 2000, Dr. Tolle has acquired numerous patents and worked for several product teams including the Natural Language Group, Visual Studio, and the Microsoft Excel Team. Since joining Microsoft Research's outreach program in 2006, she has initiated and managed several major initiatives from Biomedical computing and environmental science to more traditional computer and information science programs around natural user interactions and data curation. She was also directed the development of the Microsoft Translator Hub and the Environmental Science Services Toolkit, among other successful research development projects. Dr. Tolle is a co-editor with Tony Hey and Stewart Tansley, as well as one of the authors of one of the earliest books on data science, The Fourth Paradigm: Data Intensive Scientific Discovery. She has two major areas in which she focuses, reducing time to scientific discovery by leveraging cloud computing and machine learning services, like Microsoft Azure and Azure ML and enabling the education of the next generation of data scientists.