

Big Data Innovation

Nail the strategy. Nurture the culture. Access the goldmine.

September 25 & 26, 2014 Westin Copley Place I Boston, MA



Confirmed Speakers



Confirmed Speakers Include...

- CEO, WolframAlpha
- Chief Data Officer, MapQuest
- Chief Architect, Jet Propulsion Lab at NASA
- Head of Data Management, Shell
- Principal Architect, Lockheed Martin
- SVP, Global Data Strategy, Sony Pictures
- CMIO, Kaiser Permanente
- Data Scientist, Mozilla
- Director, Analytics, Consumer United
- Chief of Data Governance, **Humana**
- Chief Analytics Officer, Berg
- Director, Big Data, Caesars Entertainment
- Snr Manager, Information Architecture, **Best Buy**
- Senior Data Scientist, Nike Sport Research Lab
- Vice President, Data Gov, American Cancer Society
- Scientist, MIT
- Chief Data Scientist, Kabbage
- Senior Data Scientist, Staples Innovation Lab
- SVP, Business Information Officer, US Bank
- Principal Staff Engineer, LinkedIn
- Chief Economist, State of Utah
- Senior Director, Pharmacy IT, Walgreens
- Director, Analytics & Data Science, Facebook
- Chief Economist, U.S. Department of Commerce

- SVP, Data & Insight, **Dun & Bradsheet**
- Associate Development Engineer, ESPN
- Head Scientist: Big Data, **eBay**
- Data Science Informatics Leader, **GE**
- Director, Data Governance, Schneider Electric
- Head of Risk Analytics, **Santander**
- Chief Data Scientist, Chegg
- Director, IT, Caesars Entertainment
- Data Scientist, General Mills
- Senior Data Scientist, Jawbone
- Senior Data Scientist, Staples
- Director, Business Intelligence, Delhaize
- Software Engineer, Twitter
- Senior Director, Clinically Integrated Network, NYU
- Engineering Fellow, Schneider Electric
- Senior Database Engineer, **Etsy**
- Data Management Specialist, Halliburton
- Director, Big Data, Sears
- Director, Analytics, The Venetian I Palazzo
- Data Analytics Engineer, Staples
- Head of Information Management, Cigna
- Enterprise Data Architect, Shell
- Director, Streaming Science, Netflix
- Chief Model Risk Officer, Capital One

Who Will You Meet?

There is no question that IE. provides the gold standard events in the industry and will connect you with decision makers within the Big Data space. You will be meeting senior level executives from major corporations and innovative small to medium size companies.



Company Size Of Attendees

- 1000+ Employees
- 300-999 Employees
- 50-299 Employees
- Less than 49 Employees



Past Delegates Include

- Solutions Architect AOL
- Director, BI CNN
- Senior Vice President Samsung
- Data Architect Toyota
- Director, Analytics, Kaiser Permanente
- VP, Information Management MGM Resorts





About The Summit

The Big Data Innovation Summit brings together thought-leaders from the industry for an event acclaimed for its interactive sessions and high-level speakers.

As many organizations are now working with unmanageably large data sets, the importance of using and maintaining an analytics platform which can cope with this scale of information is essential. This presents both a challenge and opportunity as organizations must identify patterns and gain actionable results in order to gain a crucial advantage over competitors.

Illustrated intermittently with case studies, interactive panel sessions and deep-dive discussions, this summit

offers solutions and insight from the leaders operating in the Big Data space.

Big Data Innovation will help your business understand & utilize data-driven strategies and discover what disciplines will change because of the advent of data. With a vast amount of data now available, modern businesses are faced with the challenge of storage, management, analysis, visualization, security and disruptive tools & technologies.



Confirmed Speaker Information



Slava Akmaev, Chief Analytics Officer Berg

Slava Akmaev is Chief Analytics Officer at Berg. Dr. Akmaev oversees technology development in advanced analytics and application of innovative computational techniques such as Bayesian Networks in biotechnology. Slava leads the Berg Analytics division and directs the research informatics, healthcare analytics, and personalized medicine programs within Berg and its subsidiary companies. Prior, Dr. Akmaev led a multidisciplinary team of scientists and computational biologists at Genzyme Genetics. He launched multiple commercial diagnostic products in prenatal care, oncology and rare genetic diseases, managed patient clinical trials, and developed intellectual property in genetics and molecular biology. Akmaev has published numerous articles in computational biology, bioinformatics and genetics. Slava holds a Ph.D. in Applied Mathematics from the University of Colorado at Boulder.

Next Generation Medicine: A Real World Use Case on Commercialization of the Big Data Insight in Healthcare

Healthcare organizations need to look beyond genomics and learn how to utilize a myriad of data types available to the pharmaceutical companies in the form of highthroughput molecular data, the healthcare providers in the form of patient and physician records, the payers in the form of claims and pharmacy data, and, most importantly, data available to the patients in the form of self-reported outcomes and social media interactions. Berg has developed a Big Data platform that provides a clear view into the causal mechanisms of the disease, treatment, and population healthcare management. Artificial intelligence techniques such as Bayesian Networks have been successfully applied by Berg in precision medicine, patient stratification, and in the development of in-silico diagnostics by integration of real world Big Data in healthcare.





John Hogue Data Scientist General Mills

John Hogue is part of a new group at General Mills focused on making Consumer data more accessible and usable for marketing. This includes utilizing Big Data technologies and techniques to find new uses for data. John's focus so far has been on tackling the consumer packaged goods company. He is set to complete his Masters of Software Engineering with concentrations in Big Data and Data Management in spring of 2015.

variety and veracity problems of big data at a





Mark Chang Chief Data Scientist Chegg

Dr. Mark Wang is Chief Data Scientist at Chegg, a fastgrowing Silicon Valley startup. Since 2010, he has built and led its big data and advanced analytics team that uses data and analysis to finds stories buried in the data, then communicates insights and opportunities to the business. Trained as an MIT PhD particle physicist, he previously had a long career at the RAND Corporation "think tank." There, he received Gore's "Hammer Award" ΑI for work dramatically improving high performance military supply chains, and he held leadership positions of two federally funded research and development centers supporting the Department of Defense and the White House Office of Science and Technology Policy.

Big Data Analytics for Insights and Business Grow

This presentation will describe how advanced analytic techniques are used to find high impact insights, and to accelerate business growth. Chegg is an e-commerce that helps high school and college students startup save time, save money, and get smarter. Known originally for renting textbooks, Chegg has evolved to offer students a wide range of digital products and services. Chegg grew quickly as a venture-funded startup, and had an NYSE IPO in November 2013. Big data analytics are leveraged to generate insights into customer behavior and business opportunities, then guide roadmaps and operations.



Justin Dangel CEO

Consumer United

Justin Dangel created Consumer United in late 2007 and has cultivated its growth from just a few employees in a small office to over 400 employees across multiple locations. Consumer United was named one of the 'Top 10 Fastest Growing Companies in Insurance' by Inc. Magazine in 2012. Prior to Consumer United, Justin was Founder and CEO of Voter.com where he helped create the most popular independent political channel on the Internet with more than 3.5 million monthly users and 500,000 newsletter subscriptions. He led a team that developed strategic partnerships with Microsoft, NBC, CNN, the AFL-CIO, the U.S. Chamber of Commerce, the National Democratic and Republican parties and several others.

The Role of Predictive Analytics in **Transforming Customer Acquisition**

This session will tell the story of how Consumer United transformed customer acquisition through predictive analytics and machine learning. The talk will focus on specific steps taken to significantly boost marketing ROI and expand marketing reach through better use of data and customer intelligence. In this session attendees will learn the following:

- The role of predictive analytics in transforming customer acquisition

- Tips and tricks on building the team and infrastructure to support this transition

- A perspective on how marketing will evolve over the coming years, and what one company is doing to get ahead of the curve





JR Yasgur SVP, Global Data Strategy Sony Pictures Entertainment

JR Yasgur is the Senior Vice President, Global Data Strategy & Operations at Sony Pictures Entertainment (SPE). Ms. Yasgur is responsible for intellectual property and operational data management throughout the content lifecycle. Ms. Yasgur's oversight of data governance best practices enables emergent digital business models, operational efficiencies, and improved content discovery. Ms. Yasgur leads change management initiatives to optimize content exploitation, margin management, and digital workflows. Ms. Yasgur is involved with industrywide initiatives to define future standards and is the co-inventor of two patents for metadata management.

Maximizing the Organization: Turning Data Into a Competitive Advantage

To efficiently capitalize on emergent opportunities in a consumer-focused world, organizations must embrace a data-driven culture and understand how data management can contribute to strategic growth, product discovery, market penetration, and revenue generation. Consumers are demanding a personalized, and in the case of entertainment, immersive experience that some organizations are not yet fully positioned to support. Ms. Yasgur will share an entertainment studio executive's perspective on technology, process, and organizational challenges and opportunities that lie ahead.





Chris Mattmann Chief Architect Jet Propulsion Lab at NASA

Chris Mattmann has a wealth of experience in software design, and in the construction of large-scale dataintensive systems. His work has infected a broad set of communities, ranging from helping NASA unlock data from its next generation of earth science system satellites, to assisting graduate students at the University of Southern California (his Alma mater) in the study of software architecture, all the way to helping industry and open source as a member of the Apache Software Foundation. When he's not busy being busy, he's spending time with his lovely wife and son braving the mean streets of Southern California.

Handling the NASA Big Data Tsunami with Surfboards from the Apache Software Foundation

The National Aeronautics and Space Administration (NASA) and its Jet Propulsion Laboratory are involved in several international projects in Astronomy, Climate Science, Remote Sensing, Intelligence and other areas that are pushing the boundaries in volume, velocity, complexity and that are redefining architectures and requirements for data management, processing and dissemination. Whether it's the 700 TeraBytes (TB) of data per second that will be generated by the International Square Kilometre Array (SKA) next generation radio telescope; the petabytes of climate data being analyzed by the U.S National Climate Assessment; or the terabyte-per-flight snow hydrology remote sensing data being generated by JPL's Airborne Snow Observatory (ASO) being translated into water management decisions in the Western US and Alaska, the data tides are rising and you don't want to be caught without a life boat. Open Source platforms are readily available but it is admittedly difficult to cut through the hype and jargon to usable technologies and systems that can assist in these science big data challenges. In this talk, I will cover a representative sampling of some of the world's most used Big Data technologies available from the Apache Software Foundation, the stewards of de facto Big Data platforms like Apache Hadoop; emerging technologies including Apache Spark; and Apache Mesos; and scientific data technologies including Apache OODT, Apache Tika, and search engines like Apache Solr. Put together these platforms, address the aforementioned challenges and ride the data waves towards shore as I cover real use cases; results, and the future of where the Big data, from the ripples to the heavy currents!





Vipul Kashyap Senior Director NYU Langone Medical

Vipul Kashyap, PhD, is a seasoned professional with more than 20 years of experience in information management, analytics, and healthcare sectors holding executive positions at Pegasystems, NYU Langone Medical Center, Partners Healthcare System, Cigna Healthcare, National Library of Medicine, and has been an advisor to many healthcare analytics companies. At NYU Langone Medical Center, Dr. Kashyap established the program structure and IT strategy for Clinically Integrated Networks. This included identifying core clinical and analytical capabilities for care coordination and population health management, identifying the gaps in current EMR systems and population health tools, and designing pragmatic solutions across them. At Cigna Healthcare, he implemented a Predictive Modeling and Rule-based Repository in conjunction with a welldefined process for creation, development and deployment of predictive models and rule-based engines. At Partners Healthcare, he was the architect for the Knowledge Management Platform for providing Clinical Decision Support and developed a portal for the complete Knowledge Management Lifecycle, including the creation of authoring and reviewing workflows, life-cycles, versioning and change management policies.

Aligning the Healthcare Ecosystem

A holistic view of the broader healthcare ecosystem will be presented with a special emphasis on the role of data analytics and innovations across various stakeholders in the ecosystem. Interactions between the patient and various stakeholders in the various contexts are presented and different types of data, insights and knowledge created are identified. The notion of a "coordinated healthcare intervention" which could leverage all these insights is presented. This is followed by a presentation of two scenarios: (a) Improvement in Outcomes: A medication adherence scenario where all the stakeholders can collaborate with each other to improve outcomes

(b) Cost Reduction: A scenario illustrating shortage/overage of drugs/medical supplies where a collaborative approach between hospitals and pharmaceuticals can help optimize their respective supply chains and reduce wastage/cost in the system.





Atul Arora & Dr. Kate Webster Director, Data Governance & Lead Data Scientist Schneider Electric

Atul Arora has over 20 years of experience in the areas of Data Governance & Quality, Information Technology, Auditing and Process Quality. He has held various leadership positions within Manufacturing, Healthcare and Retail industry groups. Since joining Schneider Electric as Director, Data Governance, Atul has played a key role in establishing and formalizing their data governance program. Prior to this, Atul was in a similar role at a healthcare firm where he was instrumental in starting their data governance from a growth and risk perspective.

Dr. Kate Webster has over 20 years experience implementing and teaching multivariate predictive analytics within business and academic settings. Her focus of expertise is in applying data driven decision making through the development of predictive and segmentation models for customer acquisition, retention, and optimization. Priorities have included predicting stage based consumer behavior, persona development, increasing digital engagement, tracking demand generation performance, documenting and predicting customer journey through marketing automation, segmenting populations for profitability and enhancing client relationships across their lifecycle. She currently holds the position of Lead Data Scientist for North America Marketing within Schneider Electric.

Creating Agile Data Governance: Fostering Access and Growth

The word "Governance" can mean different things to different people and the applied definition may get even more ambiguous based on the culture and openness of associates to change within an organization. Governance over data/information gathered internally or externally is no different. In this session, we'll share our experiences in developing and sustaining a data governance program within organizations where transformations are ubiquitous and the quality of the customer's experience, insight into their buying journey and time to market define success. In addition to gaining acceptance, a dynamic and practical information governance framework enables agile business processes that foster adherence, increase go-to market velocity, and drive growth.





Mario Vinasco Data Scientist, Marketing Analytics Facebook

Mario Vinasco has over 18 years of progressive experience in data driven analytics with emphasis in database programming and predictive models creatively applied to eCommerce, advertising, customer acquisition/retention and marketing Mario specializes in developing and investment. applying leading edge business analytics to complex business problems Mario holds a Masters in engineering economics from Stanford University and currently works for facebook as data scientist; in this role he has been providing optimization recommendations to internal search and other critical IT operations. Prior roles included VP of business intelligence in digital textbook startup, people analytics manager at Google and eCommerce Sr manager at Symantec.

Predictive Analytics & Business Insights: Practical Uses in eCommerce Marketing

A/B testing is the best way to optimize the customer experience, select acquisition channels and allocate investments; however, customer segmentation, identification of risk and potential and close monitoring of testing performance are essential components of the strategy. Predictive models can be of great help, as these can identify likely customer shopping behaviors, or they can rank channels based on several factors. These can then be run through tests to determine winning customer experiences. Predictive models can be simple and easy to implement, as in macro investment allocation; they can also be elaborated as in deeper customer segmentation. HR analytics has also benefited from the use of these techniques; at Google, we helped to facilitate the promotion process of thousands of engineers. Marketing analytics to me is a field that leverages methods, algorithms and ideas from other scientific fields to optimize customer acquisition, retention and satisfaction; I will share with you, the reader, some experiences where I have helped in those areas.





Beena Ammanath Data Informatics Leader GE

Beena is the Data Science Informatics Leader at GE. She leads the data efforts to support data science at GE. She works across the GE businesses to drive advanced analytics development leveraging big data technologies. She is passionate about data and analytics to aid cross functional teams to derive data insights, aid teams in articulating questions they did not know they had and help view data in more effective ways. Beena has over 20 years' experience in the data arena with a number of international organizations including British Telecom, E*trade and Thomson Reuters. She holds a Masters in Computer Science and MBA in Finance.

Industrial Internet: Big Data and Analytics Driving Big Outcomes

GE is one of the world's leading provider of industrial equipment and machinery - moving, powering, and healing the world. GE provides services to industrial clients worldwide, minimizing unplanned downtime for assets and driving operational efficiencies. Data management and advanced analytics are core to GE's recent success in delivering superior software-based services to customers across aviation, power generation, oil & gas, healthcare, and transportation. The torrent of data generated from machines, networks, devices and data centers in industry verticals provide challenges and opportunities. The challenge is to make this machine data meaningful and actionable to deliver on opportunities around operational efficiencies. I will share real-world case studies demonstrating tangible operational benefits - ranging from fuel savings to improving productivity to reducing unscheduled maintenance to enhancing on-time performance - by tightly integrating machines, networked sensors, industrial-strength data, and software to enable intelligent insights and affect measurable outcomes





Robert Yerex Senior Data Scientist Nike Research Lab

Robert Yerex is a Senior Data Scientist at Nike where he develops predictive models that characterize the effect of interventions on athlete performance, protection, and perception.Robert's doctoral research at the Cornell resulted in the development of algorithms for predicting outcomes of genomic perturbations applied to large data. After studying economics at the Cornell Johnson School of Management, and receiving an MBA, Robert worked for several organizations developing predictive models in the social sciences. As Chief Economist at Kronos, Inc. Robert developed the Kronos Retail Labor Index, which predicts the future disposition of the US retail sector.

Data Science and Sports Medicine

In his 2003 bestseller, Moneyball, Michael Lewis describes how sophisticated analytics can lead to useful insights in sports. In the last decade, advances in sports "wearables" have made vast amounts of data available to coaches of professional and amateur teams alike. As Yogi Berra might say, "getting the data is half the battle, the other 90% is asking the right questions". A key question is "can I detect athlete injuries before they occur". The application of survival analysis commonly used in medicine to the activity data pre and post injury is starting to answer this question.





Connor Johnson Data Management Specialist Halliburton

Connor Johnson is a Data Manager at Halliburton where he performs various data related tasks from data collection and cleaning, to modeling and analysis. He has a master's in mathematics, and several years of experience writing code to interpret and exploit typically messy, unstructured data.

Machine Learning with Open Source Tools

Using Python one can leverage mature open source machine learning tools. First we will discuss how to pose a machine learning problem, next we will discuss two example cases. The first example will of use decision tree classifiers from scikitlearn to train data and predict geological formations from log curves. The second will consider using Tkinter and PyBrain for quickly generating training data and then building a neural network for prediction. The advantage of using open source tools is that they can be picked apart, modified, and recombined in ways that more user-friendly analysis software and dashboards cannot.





Glenn Finch Global Leader, Data & Technology IBM

Glenn Finch is the Global Leader for Technology and Data in the Strategy and Analytics Practice. Mr. Finch's emphasis on Business Analytics and Optimization is focused on the most Challenging and Transformative engagements which IBM undertakes. Those engagements have leveraged broader Customer Centricity and Customer Experience work across the Customer Acquisition, Customer Management, and Servicing processes. He has over 25 years of experience with such organizations as WellPoint, RIMAC, Banorte, Vodafone, H3G, Maybank, JPMC, HSBC, Bank of America, Capital One, American Express, Citibank, Wells Fargo, Barclays, LTSB, GE Capital, The Associates, State Farm, Norwich Union, First Interstate Bank, North American Mortgage, Household, and GM.

Your Chief Data Officer: An Agent of Change for an Era of Data

As the importance of data and fact-based decision making rises, organizations are recognizing the need for a new point of accountability – someone who manages enterprise data as a strategic corporate resource. Through this study, IBM will explore the emerging role of Chief Data Officer, uncovering why organizations feel compelled to establish such a role and the value they're deriving from it. The research will also outline advice from CDOs and their executive peers on key areas of focus for this new position.





Donnie Yancey Chief Data Officer MapQuest

As Chief Data Officer at MapQuest, Donnie Yancey is responsible for the overall data & analytical strategy along leading all audience development efforts and the beginning of a new business line - MapQuest DaaS (data-as-a-service), which focuses on providing location based data solutions for mapping & traffic, advertising, and research initiatives across multiple digital platforms. Prior to joining MapQuest, Mr. Yancey (an 11+ year veteran at AOL, Inc.) was Head of Business Metrics & Analytics under the CAO organization and held various positions at AOL within Finance (Head of Financial Strategy & Analysis; CFO for Advertising.com Group, now known as AOL Platforms), Global Strategic Planning & Analytics, and Marketing Analysis. Prior to AOL, Donnie was a Product Manager at Capital One with primary responsibilities in CRM/retention focused strategies, online account services, and new product development.

From Here to There and Everywhere Inbetween

As growth continues with geo-location targeting and visitation attribution through mobile devices, MapQuest looks to take advantage of one of the most unique consumer behaviors it obtains from its userbase – where are they planning to go (aka routes). Whether its going to the new store at the mall or to a relative's new house, we understand where the start, where the plan to go, and where they go before they get there when they are using. This creates unique and dynamic opportunities in the marketplace to reach our valuable MapQuest mobile audience.



Sign up for a DIAMOND PASS for access to the co-located **Data Visualization Summit**

featuring speakers from...









The Washington Post

For more information visit the summit website



Anirudh Todi Senior Software Engineer Twitter

At Twitter I work on the Data Platform team. My team is chartered with processing and understanding the vast body of data that is generated by the operation of the Twitter platform. Our technologies are used to provide insights both to platform engineering and to other teams throughout the company. We build a range of cutting-edge services that can process petabytes of data per month in real time for insights into the usage patterns of the Twitter platform, and to build highperformance infrastructure to deliver those insights to ourselves and our partners. I have previously worked at Facebook helping scale their HBase cluster to process the billions of messages that are sent every day. In college, I built "Politify" a startup that used financial data to model the US economy and simulate the impacts of political policies on American households. Additionally, I have experience working at a genomics startup and have taught a class on Open Source software while at UC Berkeley.

Tsar (the TimeSeries AggregatoR), - How to Count 50 Billion Daily Events in Real Time using Open Source Technologies

Twitter's 250 million users generate over 50 billion tweet views per day. Aggregating these events in real time - in a robust enough way to incorporate into our products - presents a massive scaling challenge. In this talk I'll introduce Tsar (the TimeSeries AggregatoR), a robust, flexible, and scalable service for real-time event aggregation designed to solve this problem and a range of similar ones. I'll discuss how we built Tsar from the ground up, almost entirely on open-source technologies (storm, summingbird, kafka, aurora, and others), and describe some of the challenges we faced in scaling it to process tens of billions of events per day.





Stephen Dillon Engineering Fellow Schneider Electric

Stephen Dillon is an Engineering Fellow with Schneider Electric focused on Big Data and database technologies. He is a practitioner and researcher in the fields of in-memory and distributed data. He has over 15 years of experience in database technologies and architecture and has been working with Big Data and Cloud technologies since 2010. Stephen holds a Masters Degree in Computer Information Systems from Boston University and numerous certifications including Microsoft, IBM, and VoltDB products. He is a native New Englander currently living in New Hampshire with his wife and two children and many adopted pets.

In-Memory Databases, Solving the Big Data Velocity Crunch

In-Memory Database Management Systems (IMDBS) have evolved into modern systems that allow us to solve the Velocity needs of Big Data. By removing the processing bottlenecks faced by traditional RDBMS, we are able to solve both legacy system performance issues as well as create new and innovative solutions to support the domains of Smart-City and the Internet of Things. We will define the characteristics of an IMDBS as well as explore how they support efficient and durable processing of data. We will also discuss RAM as the new disk, performance metrics, what the term "real-time data" actually means, and why distributed scale-out architecture is the better solution.





Vijay Gadepally Scientist MIT

Vijay Gadepally is a Member of the Technical Staff at the Massachusetts Institute of Technology (MIT) Lincoln Laboratory in the Computing and Analytics team. Vijay's research is in the area of big data hardware, software and analytics. Vijay holds a PhD in Electrical and Computer Engineering from The Ohio State University and a B.Tech degree from the Indian Institute of Technology, Kanpur. Vijay has also worked at Raytheon Company and Rensselaer Polytechnic Institute.

Addressing the Challenges of Big Data through Innovative Technologies

The ability to collect and analyze large amounts of data is a growing problem within the scientific community. The growing gap between data and users calls for innovative tools that address the challenges faced by big data volume, velocity and variety. The Massachusetts Institute of Technology's Lincoln Laboratory has taken a leading role in developing a set of tools to address these challenges. This talk will concentrate on how we utilize innovative technologies in our mission to apply advanced technology to problems of national security.





Stephen Dillon Engineering Fellow Schneider Electric

Scott is a 'Big Data' and Advanced Analytics expert with extensive experience in workforce, finance, and technology consulting spanning multiple industries. Prior to joining Right, Scott served as an analytics consultant for a global valuation, advisory, and liquidation services company as well as the Managing Director of Revenue Cycle and Technological Transformation Consulting for a Healthcare Research and Advisory firm. Scott leverages his data science expertise to build predictive models of individual, institutional, and market behavior while employing his extensive training in statistics, mathematics, and computer science to answer questions, test hypotheses and build solutions for clients.

The Human Age

ManpowerGroup has introduced the concept of "The Human Age"- when optimizing human potential will be the single most important determinant of future success and growth. As the world's largest workforce solutions company, ManpowerGroup is uniquely able to optimize talent acquisition by leveraging big data techniques, advanced statistical modeling, creativity, and an unparalleled understanding of the world of work to create and leverage predictive "job matching" algorithms which align candidate and employer preferences in an unprecedented way. Scott's presentation will discuss approaches to increase efficiency and profitability by improving talent matching and retention, enabling companies to thrive in the Human Age.



-

Dr. Milorad Krneta Fonuder, CEO & CTO Soft10

Dr. Krneta is an accomplished data scientist and entrepreneur. A former Governor of the Fields Institute of Mathematics who has won multiple awards in the field for his revolutionary analytics technologies; with over 30 years of research invested in self-learning predictive analytics technology. His first application in automatic predictive analytics software, called MWM, for Generation5 Mathematical Technologies, remains a leading influencer in the North American Marketing field. Today, at Soft10 - Dr. Krneta has developed his smartest, automated technology yet: a revolutionarily multi-model, self-learning prediction software called-Dr. MO the only software in the world that can fully analyze big data.

Big Data Predictive Modeler – Dr. Mo: Live Demo!

More than 30 years ago, Dr. Krneta had already experienced the Big Data Predictive Modeling Challenge. As the youngest university professor of statistics at the time, Dr. Krneta was engaged for a special UN project to predict behavior of economic system. Computational power was one restriction, but even more challenging was what statistical methodology to use. It soon became apparent that the only solution was in unobserved, self-learning, automatic predictive technology. Ten years later, Dr. Krneta developed a self-learning clustering system at his first company Generation5: probably the first of its kind. Today at Soft10, Dr. Krneta developed his best technology yet, a self-learning, automatic, multi-model predictive analytics technology with the capability to use unlimited database size, with any type of variables, called: Dr. Mo, Virtual Statistician. Dr. Mo's speed and accuracy power is so great that its predictions actually come with a Guarantee. Dr. Krneta will Live Demo Dr. Mo on a database with more than 100 independent variables with superb prediction accuracy - in less than 3 minutes.





Prasanna Gautam Software Engineer ESPN

Prasanna is a Software Engineer at the Data Platform and Architecture group at ESPN. He is building the next generation of data services at ESPN with a passionate team around him. He is keen on better understanding fault tolerance in distributed systems. He is always looking for better abstractions to make sense of these systems.

A Bridge to a Modern Data Workflow: Modeling Live TV Data for Second

This talk is about data modeling process used by ESPN to build personalized second screen products. It required transforming data designed for linear TV to be usable by a wide variety of devices. At ESPN, we built composable wrapper APIs around services that power TV. We validated our models in isolation and composed them as needed. In the end, we got a cohesive data model for rapidly prototyping services and products around live TV. I will walk you through the process we used to model and expose linear TV data for exploring second screen experiences.





Abe Gong Data Scientist Jawbone

Abe was the first data scientist at Jawbone, where he builds data systems to nudge people to form good habits and live healthier. Prior to Jawbone, Abe was the lead data scientist at Massive Health. He earned his PhD in Public Policy, Political Science, and Complex Systems at the University of Michigan. All told, Abe has worked as data scientist/statistical consultant in education, health, and public policy for over a decade. In previous lives, Abe has been a pollster, journalist, refugee, and amateur historian.

Data Science on Deadline

Training for data scientists focuses largely on nouns, such as algorithms, software, infrastructure, and data products. Noun-based training teaches tools and technical skills, but has little to say about time. Therefore, fundamental concepts such as scheduling, deadlines, efficiency, and return on investment. This presentation shares a verb-based framework for reasoning about time in data science workflows. From identifying the day-to-day activities that occupy most of data workers' time, to accelerating essential activities to eliminating wasted steps, data science on deadline teaches the cognitive reflexes that separate abstract thinkers from productive doers in the world of data.

JAWBONE



Nauman Sheikh Data Architect Exelon

Nauman Sheikh is a veteran of the data architecture profession who has built dozens of large scale operational, data warehouse and analytics systems over the last 18 years. He has worked in three continents solving business challenges in consumer credit, risk, fraud and direct marketing areas dealing with a variety of cultural, technological and legal challenges surrounding data and its use. He is a handson practitioner with skills ranging from data warehousing to predictive modeling to analytics driven business decisions and their audit and control frameworks. He is the author of a recently published book "Implementing Analytics: A Blueprint for Design, Development & Adoption". He has worked with companies like Experian, Fidelity Information Systems, Navistar and Exelon. He lives near Washington D.C in USA.

Creating Value in the Internet of Things Using Big Data Streaming Analytics

The current trends in Big Data Analytics suggest an approach where the idea seems to be large upfront capital investment, acquisition of big data technology, accumulation of all the possible data, hiring of data scientists and then hoping that something good comes out of it. There is a lot of potential for innovation but at a high cost and high risk of failure. As a hands-on practitioner, I will share a step-by-step approach to Big Data Analytics that fosters a culture of innovation through a bottom-up approach starting with line managers and business operations and yielding benefits of innovation driven by analytics for the entire organization. This detailed methodology relies on 4 objectives that lead to achieving this culture of innovation that every organization aspires to

- 1. Simplification
- 2. Commoditization
- 3. Democratization
- 4. Innovation

The session will help simplify analytics clearing the confusing terminology and jargon and will put analytics in its proper context with respect to existing technologies.





Dale Skeen Co-Founder & CTO Vitria

Dr. Dale Skeen co-founded Vitria in 1994. He has been designing and building stream processing systems for over 25 years. Dr. Skeen is a recognized serial entrepreneur having co-founded three technology start-ups. He is considered a pioneer and thoughtleader in stream processing, streaming analytics, complex-event processing, messaging, and business process management. Prior to co-founding Vitria, Dr. Skeen was the co-founder of TIBCO Software, where he served as the Chief Scientist. Dr. Skeen has held faculty positions at the University of California, Berkeley and Cornell University. He has a Ph.D. in Computer Science from the University of California, Berkeley.

Creating Value in the Internet of Things Using Big Data Streaming Analytics

Dr. Dale Skeen, Co-Founder & CTO of Vitria Technology, will discuss how customer-centric organizations are utilizing Big Data Streaming Analytics to create business value in their Internet of Things initiatives. From monetizing customer location information to predicting equipment and device failures to securing the Smart Grid, Dr. Skeen will discuss how some of the most innovative companies are leveraging Big Data Streaming Analytics to generate incremental revenue, enrich their customers' experience, and reduce operational risks and costs.





Courosh Mehanian & Karthik Kumara Senior Data Scientists Staples Innovation Lab

Courosh Mehanian received a Ph.D. in Physics at Cornell University. He has applied the mathematical and modeling skills of the physicist to solve problems in computer vision and machine learning. He has worked in the fields of automatic object recognition, semiconductor quality control, computer aided pathology, and data science. He is currently a Senior Data Scientist at Staples Innovation Lab.

Karthik Kumara has over 10 years experience using machine learning in production systems. He has a masters degree in Computer Science from the Indian Institute of Science and has worked in the fields of collaborative filtering, handwriting recognition, image de-blurring, and text mining. He is currently a Senior Data Scientist at Staples Innovation Lab, working on algorithms to maximize sales through price optimization.

Staples Innovation Lab: Data Modeling for Personalization based on User Behavior

A major goal of data modeling efforts at Staples Innovation Lab is to optimize outcomes for both the company and the consumer. One of the projects focuses on price personalization, which can be used to optimize various business metrics, such as purchases, total sales, or margin. A rich set of historical data on consumer website sessions is used to train machine learning models that encapsulate price elasticity. These models are used to determine a realtime price for a consumer based on user, product, and session attributes. In this presentation, we discuss the modeling approaches, challenges, and prospects for future work.





Anthony J. Scriffignano SVP, Data & Insight Dun & Bradsheet

With 30+ years experience in IT, Big-4 management consulting, and international business, Scriffignano leads D&B's discovery, curation, and synthesis of information in multiple languages, geographies, and contexts worldwide. He has extensive background in linguistics and advanced algorithms and is primary inventor on multiple patents and patents pending, primarily relating to identity and related process. Scriffignano has presented at various venues in the U.S., Europe, Latin America, and Asia as a keynote speaker, guest instructor, and forum panelist. He was profiled by InformationWeek in the series "Big Data. Big Decisions" and by BizCloud regarding Big Data problem formulation and data privacy.

Unlocking the Value of Big Data

Clearly, Big Data is central to the strategic thinking of today's innovators and business executives. Companies are scrambling to figure out the secret to transforming Big Data to Big Insight and that Insight into Action. As many companies struggle with the emerging technologies and nascent capabilities to discover and curate massive quantities of highly dynamic data, new problems are emerging in the form of how to ask meaningful questions that leverage the "V's" of large amounts of data (e.g. volume, variety, velocity, veracity). In the Business-to-Business space, these challenges are creating both significant opportunity and ominous new types of risk. In this session, D&B's SVP of Worldwide Data & Insight, Dr. Anthony Scriffignano, will discuss how companies are reacting to these changes and provide valuable insight into new ways of thinking in a world with overwhelming quantities of data.



Presentation details coming soon from...











Humana

Click here to view the online agenda

Interactive Sessions

There are a limited number of places available at the interactive workshops and roundtables listed below. Avoid missing out by registering in advance. Email <u>Heather</u> <u>James</u> with the name of the session(s) you wish to attend to secure your place.

Workshop: Hadoop for Dummies: Understanding the Foundations of Hadoop



Hadoop is becoming an established component in today's data centers, but there are still misconceptions on how it works, where it's best suited, and what direction it's going in. This session, led by the lead author of the Hadoop for Dummies book, explains the basics of Hadoop at an introductory level.

Workshop Leader:



Dirk deRoos World-Wide Technical Sales Leader IBM

Dirk deRoos, B.A., B.Comp Sci., is IBM's World-Wide Technical Sales Leader for IBM InfoSphere BigInsights. Dirk has spent the past three years helping customers build Big Data solutions featuring BigInsights and Apache Hadoop, along with other components in IBM's Big Data Platform. Dirk has co-authored thee books on this subject area: Hadoop for Dummies (Wiley, 2014), Harness the Power of Big Data (McGraw-Hill, 2012), and Understanding Big Data (McGraw-Hill, 2011). Prior to this Dirk worked in the IBM Toronto Software Development Lab in the DB2 Database development team. Dirk has earned two degrees from the University of New Brunswick in Canada: a Bachelor of Computer Science, and a Bachelor of Arts (Honours English). You can reach him on Twitter at @Dirk_deRoos.

Roundtable Discussion: Implementing MongoDB in Windows Azure using a Windows Operating System

MongoDB, the leading NoSQL database, has traditionally been implemented on Linux based operating systems. In this session we will explore my research and observations regarding the implementation of this technology on a cluster of virtual machines using a Windows Server 2012 OS and Microsoft's Azure Cloud offer. Discussion points include but are not limited to Azure services and storage accounts, virtual machine configuration, and MongoDB architecture. This session is intended to promote discussion, idea sharing and networking amongst those interested in similar deployments.

Roundtable Leader:



Stephen Dillon Engineering Fellow Schneider Electric

Stephen Dillon is an Engineering Fellow with Schneider Electric focused on Big Data and database technologies. He is a practitioner and researcher in the fields of in-memory and distributed data. He has over 15 years of experience in database technologies and architecture and has been working with Big Data and Cloud technologies since 2010. Stephen holds a Masters Degree in Computer Information Systems from Boston University and numerous certifications including Microsoft, IBM, and VoltDB products. He is a native New Englander currently living in New Hampshire with his wife and two children and many adopted pets.

Multi-Track Event

The summit will comprise of multiple tracks, covering the most current topics in Big Data today - click for more information on each track:

- •Big Data Analytics
- •Hadoop & NoSQL
- Data Governance
- •Big Data & the Cloud
- •Big Data & Marketing
- Chief Data Officer
- •<u>Data Science</u>
- •Data Visualization
- Interactive Sessions



Multiple Session Types

The Big Data Innovation Summit hosts a selection of session options, within the summit you can mix & match the sessions that are most relevant to your needs - content from keynote presentations will also be made available post-summit subject to the presenters permission.

- **Presentations** everything from use cases to the challenges & successes faced by those working in the Big Data space.
- **Panel Sessions** Interactive panels will allow you intimate access to the thought leaders at the top of their game, allowing extensive Q&A time.
- Round Table Discussions Sit with your peers and discuss the latest innovations, sharing ideas, insights & connecting with those close to you.
- Workshops Hands on demonstrations of the latest technologies helping to drive the space forward. Solutions & Services for your needs.
- Access to all Sponsor Zones Get the technologies that you need, everything from Hadoop Technologies to Business Intelligence needs, Big Data Innovation has it for you



"I got a lot from this and will be recommending to several of my colleagues" eBay, Big Data Innovation Summit, Boston 2013 Attendee

"It was very successful for our Big Data team - looking forward to it again in the future" IBM, Big Data Innovation Summit, Boston 2013 Gold Sponsor

The Information

Big Data Innovation Summit

September 25 & 26, 2013 Westin Copley Place Boston, Massachusetts **Online Reservations** +1 888 627 7216 and quote 'IE Group'

Registration Pricing

Early Bird Prices available until July 18 please see below



- 5 Silver Passes:
- 3 Gold Passes:
- 5 Gold Passes:
- 3 Diamond Passes:
- 5 Diamond Passes:

\$3000 (\$1000 per attendee) \$4500 (\$900 per attendee) \$3900 (\$1300 per attendee) \$6000 (\$1200 per attendee) \$4500 (\$1500 per attendee) \$7000 (\$1400 per attendee)

email pedro@theiegroup.com

* Team discounts are applicable at the point of registration only.

Ways to Register





+1 323 446 7673



<u>Click Here to Register</u>

Registration Form



Big Data Innovation Summit

September 25 & 26, 2014 | Westin Copley Place | Boston, MA

For registration or more information on the program, please call Pedro on +1 415 992 5658, or fax this registration form to +1 (323) 446 7673

1. Delegate Information...

NAME OF EACH ATTENDEE

TITLE OF EACH ATTENDEE			DEPARTMENT		
COMPANY			INDUSTRY		
ADDRESS			CITY		
STATE/PROVINCE		ZIP/POSTAL CODE	COUNTRY	COUNTRY	
EMAIL OF EACH ATTENDEE			BUSINESS PHONE NUMBER		
2. P	ass Types				
Early Ea Ea Ea Ea Ea Control Ea	Bird Pass Options u rly Bird Silver: \$1295 rly Bird Gold: \$1595 rly Bird Diamond: \$1795 Ilar Pass Options afte ver Pass: \$1495 Id Pass: \$1795 amond Pass: \$1995 Descriptions:	ntil July 18, 2014 Attendees Attendees Attendees r July 18, 2014 Attendees Attendees Attendees	Group Discount Pass (3 Silver Passes \$3000 (\$100 5 Silver Passes \$4500 (\$90 3 Gold Passes \$3900 (\$130 5 Gold Passes \$6000 (\$120 3 Diamond Passes \$4500 (\$ 5 Diamond Passes \$7000 (\$ For larger groups or special requests con +1 415 992 5658 or email <u>pedro@theiegr</u> Group passes only available when all par	Options 00 per attendee) 0 per attendee) 0 per attendee) 0 per attendee) 51500 per attendee) 51400 per attendee) tact Pedro Yiakoumi by calling toup.com ticipants register together.	
Silver P Gold Pc Diamor all cont 3. P	ass: Access to all sessions & r ass: Access to all sessions, net ad Pass: Access to all sessions ent on the Big Data & Analytic Cayment Optior	networking events working events & unlimited ac including co-located Data Vis cs channels via ieOnDemand 15	ccess to ieOnDemand sualization Summit, networking eve	ents, annual subscription to	
Check (Make checks payable to The Visa Mastercard		he Innovation Enterprise Ltd) American Express EXPIRATION DATE	Diners Club	Invoice meDiscover	

CARDHOLDERS NAME

BILLING ADDRESS

INDUSTRY

CARDHOLDER'S SIGNATURE

Prices are exclusive of VAT. Places are transferable without any charge to another Summit occurring within 12 months of the original purchase. Team discounts are applicable at the point of registration only. Any cancellations within a group registration will in turn incur an increase in registration fee for the remaining group participants. Cancellations before August 18, 2014 incur an administrative charge of 50%. If you cancel your registration after August 18, 2014 you will be charged the full fee. You must notify The Innovation Enterprise in writing of a cancellation, or you will be charged the full fee. The Innovation Enterprise reserve the right to make changes to the program without notice. NB: FULL PAYMENT MUST BE RECEIVED BEFORE THE EVENT.

Subject to Change

Schedule

Day One

September 25



Day Two

September 26



Big Data Innovation Platinum Sponsors





BeyondCORE



Sponsors



Gold Sponsor



Gold Sponsor



Gold Sponsor

RedPoint

Gold Sponsor





Silver Sponsor

Gold Sponsor

Sponsors

VITRIA

Silver Sponsor



Silver Sponsor





Accurate. Automatic. Predictions.

Silver Sponsor



Silver Sponsor



Exhibitor

Exhibitor





Sponsorship opportunities are designed to maximize your ROI.

Customizable packages ensure your company, brand, products and services are promoted to senior decision makers in todays leading industry organizations.

"The people we need to be talking to are here!" - 2012 Sponsor

Contact Giles your dedicated sponsorship representative today. Giles - <u>ggodwinbrown@theiegroup.com</u>