



# **INFORMS** CONFERENCE THE **BUSINESS** OF **BIG** **DATA**

**June 22-24, 2014** **San Jose, California**  
San Jose Convention Center

Conference Sponsors	<b>2</b>
Conference Committee	<b>2</b>
Program Highlights	<b>3</b>
Conference Mobile App	<b>4</b>
Technology Workshops	<b>4</b>
Networking & Social Events	<b>5</b>
Guide to Exhibits	<b>6</b>
Poster Sessions	<b>8</b>
Sunday Events	<b>9</b>
Monday Events/Abstracts	<b>9</b>
Tuesday Events/Abstracts	<b>14</b>
Schedule at a Glance	<b>18</b>
Floor Plans of Meeting Space	<b>20</b>

**Online Proceedings**

View slides for most conference sessions:  
<http://meetings2.informs.org/bigdata2014/>

**Conference Mobile App**

Search your App Store (Android or Apple) for "INFORMS Mobile" to download

## Thanks to Our Sponsors



**Portfolio Pads and Signage**



**Conference Bag**

Booz | Allen | Hamilton  
100 YEARS

**Badge and Lanyard**  
**Online Proceedings**



**Welcome Reception**



UNIVERSITY OF  
SAN FRANCISCO

Master of Science  
in Analytics

**Monday Refreshment Break**



**Tuesday Refreshment Break**



## Conference Committee

**Margery H. Connor**  
Conference Co-Chair  
Senior Operations Researcher  
Advanced Analytics, Chevron

**Diego Klabjan**  
Conference Co-Chair  
Professor, Founding Director  
Master of Science in Analytics  
Northwestern University

**Paul Bachteal**  
Director, Americas Technology Practice,  
SAS

**Donald L. Buckshaw**  
Chief Operations Research Analyst  
Leidos

**Arnold Greenland**  
IBM Distinguished Engineer  
Office of the Public Sector CTO, IBM

**Mani Janakiram**  
Director, Supply Chain Intelligence  
& Analytics, Intel

**Brian Keller**  
Lead Associate, Booz Allen Hamilton

**Natalie Kortum**  
Director of Consumer Analytics  
Humana

**Veena Mendiratta**  
Consulting Member of Technical Staff  
Alcatel-Lucent

**Jeff Winters**  
Operations Research Division Manager  
UPS

# Program Highlights

## THOUGHT-PROVOKING KEYNOTE PRESENTATIONS

**Monday, 8:30am – 9:30am, LL20A**

### Putting Big Data To Work

Bill Franks, Chief Analytics Officer,  
Teradata Corporation



Bill provides insight on trends in the analytics & big data space and helping clients understand how Teradata and its analytic partners can support their efforts. In addition, Bill is a faculty member of the International Institute for Analytics and the author of the book *Taming The Big Data Tidal Wave* (John Wiley & Sons, Inc., April, 2012). He is also an active speaker and blogger. Bill's focus has always been to help translate complex analytics into terms that business users can understand and to then help an organization implement the results effectively within their processes. His work has spanned clients in a variety of industries for companies ranging in size from Fortune 100 companies to small non-profit organizations.

#### Abstract:

Big data is everywhere. You can't avoid being exposed to discussions around big data, and the analysis of it, on a regular basis. The downside of this attention is that there is a lot of hype and misinformation in the marketplace. Many companies are confused about how to get started, what actions to take, and what pitfalls to avoid. Based on content from the popular book *Taming The Big Data Tidal Wave*, this talk will provide an overview of important themes to understand regarding big data. The talk will address technological, organizational, and cultural points that must be considered. Most importantly, the talk will aim to provide attendees a solid direction to take their big data initiatives.

**Tuesday, 8:30am – 9:30am, LL20A**

### Moving from Big Data to Big Outcomes on the Journey to ROI

Michael Svilar, Managing Director, Accenture



As a managing director at Accenture Analytics, now part of Accenture Digital, Michael Svilar serves as the Delivery Lead and Capability Lead for Accenture Analytics. In these roles, Michael runs a group of more than 900 data scientists and analytics professionals globally. Michael oversees Accenture's network of Analytics Innovation Network, including our Advanced Analytics Innovation Centers in Athens, Barcelona, Dublin, Madrid and Singapore. Michael has been with Accenture for 14 years, and was a founding member of its Marketing Sciences practice. Most recently, Michael led the Marketing Analytics group in Accenture Interactive. For more than 30 years, Michael has run analytics projects across multiple industries including Retail, Communications, Financial Services, Automotive, Consumer Packaged Goods, and Electronics. His primary area of focus is applying analytics to marketing, merchandising and price promotions. He holds a patent in forecasting analytics, and has published on a variety of analytics topics. Michael has a BS in Economics from John Carroll University, an MA in Economics from Miami University, and a Ph.D. from the University of Maryland.

#### Abstract:

In this Internet of Everything world, systems, devices and physical objects are "talking" to one another. There are upwards of a trillion connected and instrumented things: cars, appliances, cameras, roadways, pipelines...even pharmaceuticals and livestock. The talk will focus on how organizations can drive positive business outcomes in the connected world using big data analytics.

## POSTER PRESENTATIONS

### WITH DESSERT

**Monday & Tuesday: 2:00pm – 3:00pm  
LL20 Foyer**

Each day, we're serving up a delectable afternoon dessert during a high quality poster session. These poster presentations offer you an inside look into a full range of big data analytics, including all applications areas, industries and methodologies. Presenters will be available during the afternoon poster sessions from 2:00pm-3:00pm to describe their work and answer questions. The posters will also be displayed during the entire day, so you can browse at your leisure. As a guide for identifying which posters you want to visit, watch for the color-coded signs that indicate general topic areas. See full poster listing on page 8.

## ONLINE PROCEEDINGS

### Preview Presentation Slides to Help Select Sessions to Attend

Our online proceedings, featuring presentation slides for almost every session in the conference, is a great resource for selecting the talks you want to attend. A quick check provides a detailed look at what speakers will be presenting, and helps ensure a good match with what you want to learn. The online proceedings will also be available to attendees after the conference. Go to the conference web page and click on "Conference Proceedings." If you have questions, please go to the INFORMS registration desk.

## Conference Mobile App



INFORMS is pleased to offer a free mobile app for smart phones and tablet users. Use the app to access conference information on the go, including the schedule of talks, speaker bios, maps showing meeting locations, exhibit details and much more. As this is a native app, you do not need an Internet data connection to use it during the conference.

When you enter the conference program in the app, you can either browse through the parallel sessions and special events or search for a specific speaker or keywords. As you locate the events, sessions and presentations you are interested in, click on the star button which bookmarks it into the My List itinerary. Use My List for a quick reference tool.

### Android Devices

To download, go to the Google Play Store and search for INFORMS Mobile.

### Apple Devices

To download on either iPhone or iPad, go to the App Store and search for INFORMS Mobile.

### Please Provide Your Input

We encourage you to use the survey feedback form available for every presentation on the program. You can respond to the survey as you browse the presentations. A full conference survey is on the conference screen of the main menu. The conference committee greatly appreciates your feedback on individual sessions and on the conference overall.

## Technology Workshops

**SUNDAY, JUNE 22**

**1:45pm – 3:15pm**

### FICO

LL21B

Turn Big Data into Better Decisions

### Forio

LL21C

Creating and Publishing Interactive Online Operations Research Applications

### ProbabilityManagement.org

LL21D

Unambiguous Uncertainty Stored as Big Data

**3:30pm – 5:00pm**

### Frontline Systems

LL21B

Analytics Made Easy: Data Mining, Simulation and Optimization for Excel and the Web

### SAS

LL21C

An Overview of Machine Learning with SAS® Enterprise Miner™

## INFORMS Future Meetings

**2014**

November 9-12

### INFORMS Annual Meeting

#### 2014 San Francisco

Hilton San Francisco & Parc 55 Wyndham  
San Francisco, California  
<http://meetings2.informs.org/sanfrancisco2014/>

**2015**

April 12-14

### INFORMS Conference on Business Analytics & Operations Research

Hyatt Huntington Beach  
Huntington Beach, California

June 14-17

### CORS-INFORMS International

Sheraton Montreal  
Montreal, Canada

November 1-4

### INFORMS Annual Meeting

#### 2015 Philadelphia

Pennsylvania Convention Center  
& Marriott Philadelphia Downtown

# Networking and Social Events

## SUNDAY

**5:30pm – 7:00pm** **LL20 Foyer**

### Welcome Reception and Exhibits

We welcome you to the conference with this first opportunity to visit the exhibits, network with colleagues and enjoy heavy hors d'oeuvres, beer, wine, soft drinks and bottled water.

## MONDAY

**7:30am – 8:30am** **LL20BCD**

### Full Buffet Breakfast

Connect with colleagues over breakfast.

**12:00pm – 1:00pm** **LL20BCD**

### Networking Lunch

Join your colleagues and meet someone new to discuss how you are using big data in your company. Seating is not assigned for this event.

**2:00pm – 3:00pm** **LL20 Foyer**

### Poster Session, Dessert & Exhibits

Enjoy dessert, visit the exhibits, and learn from poster presentations that showcase a full range of big data analytics, including many applications areas, industries and methodologies. As a guide for identifying which posters you want to visit, watch for the color-coded signs that indicate general topic areas. Prizes for the top 3 overall projects will be awarded at 2:50pm. The criteria are interesting use of big data analytics, effective display of results and business impact. See page 8 for the full list of posters.

**4:00pm – 5:30pm** **LL20BCD**

### Birds-of-a-Feather Discussion Groups

Stop by for beer, soda, snacks and good conversation around important issues. Topics will be selected by the attendees. We need your input! Look for the white board near registration to suggest topics or vote for a topic already posted. This is one of the most popular and highly-rated aspects of the conference – don't miss it.

**5:30pm – 7:00pm** **Marriott Hotel  
General Reception San Jose Ballroom**

Stop by this informal reception where you can connect with other attendees and unwind from a jam packed day. Members of the Conference Committee will be your hosts. Enjoy a hearty display of heavy hors d'oeuvres and food stations, beer, wine, soft drinks and bottled water.

## TUESDAY

**7:30am – 8:30am** **LL20BCD**

### Full Buffet Breakfast

Connect with colleagues over breakfast.

**12:00pm – 1:00pm** **LL20BCD**

### Networking Lunch

Join your colleagues and meet someone new to discuss what is keeping you up at night, as a big data expert or data scientist, and where do you see big data analytics going in the next 3 years. Seating is not assigned for this event.

**2:00pm – 3:00pm** **LL20 Foyer**

### Poster Session, Dessert & Exhibits

Enjoy dessert, visit the exhibits, and learn from poster presentations that showcase full range of big data analytics, including all applications areas, industries and methodologies. As a guide for identifying which posters you want to visit, watch for the color-coded signs that indicate general topic areas. Prizes for the top 3 overall projects will be awarded at 2:50pm. The criteria are interesting use of big data analytics, effective display of results and business impact. See page 8 for the full list of posters.

## For Your Career & Your Organization

### ANALYTICS CERTIFICATION – CAP®

Stop by the INFORMS table to reserve your copy of the official CAP Study Guide. This brand new resource provides an easy and efficient way to prepare for CAP. Be sure to pick up a copy of the CAP Candidate Handbook. This contains all the information you need to decide if CAP is for you. Quantities are limited.

### INFORMS ANALYTICS Maturity Model

How well does your organization use analytics? What can you do to progress from good to great? Attend this presentation of the new INFORMS analytics maturity model to see how you can score your organization, set target goals, and attain the enormous advantage of analytics leadership. For information about doing your first analytics maturity assessment, stop by the INFORMS booth for a helpful brochure or attend the following in depth presentation.

#### Monday, Track T5

**9:40am – 10:30am, LL21B**

### INFORMS CONTINUING EDUCATION

INFORMS' two-day, in-person courses provide real-world value in skills, tools and methods that can be implemented in your work. Essential Practice Skills for Analytics Professionals was offered just before the INFORMS Big Data Conference and a second course is being offered immediately following the conference. If you would like to attend the Data Exploration & Visualization course offered June 25-26 off site, please inquire at the registration desk. For the full 2014 schedule and more information, contact Thedra White at [continuinged@informs.org](mailto:continuinged@informs.org) or visit [www.informs.org/continuinged](http://www.informs.org/continuinged).

## Guide to Exhibits

### **Booz Allen Hamilton**

[www.boozallen.com](http://www.boozallen.com)

Booz Allen Hamilton is a leading provider of management consulting, technology, and engineering services to the US government in defense, intelligence, and civil markets, and to major corporations, institutions, and not-for-profit organizations. Booz Allen is headquartered in McLean, Virginia, employs approximately 23,000 people, and had revenue of \$5.76 billion for the 12 months ended March 31, 2013. In 2014, Booz Allen celebrates its 100th anniversary year. To learn more, visit [www.boozallen.com](http://www.boozallen.com). (NYSE: BAH)

### **Booth #109**

### **Contact Singapore**

[www.contactsgapore.sg](http://www.contactsgapore.sg)

Contact Singapore is an alliance of the Singapore Economic Development Board and Ministry of Manpower. We engage overseas Singaporeans and global talent to work, invest and live in Singapore. Contact Singapore actively links Singapore-based employers with professionals to support the growth of our key industries. We work with investors to realise their business and investment interests in Singapore. For information on working, investing and living in Singapore, please visit [www.contactsgapore.sg](http://www.contactsgapore.sg) or contact our worldwide offices.

### **Booth# 104**

### **Enrich Consulting**

[www.enrichconsulting.com](http://www.enrichconsulting.com)

Enrich was born out of a restless desire to help companies do more—create more game-changing technology, develop more life-saving drugs, find more opportunity. Our emphasis is on building self-sufficient business forecasting software for our clients, based on customized analytic solutions. The Enrich Analytics Platform (EAP) provides the enterprise computing power; our high-touch, one-on-one approach builds the custom solutions that empower clients to make strategic business decisions—and follow through on them.

### **Booth# 102**

### **Exponent®, Inc.**

[www.exponent.com](http://www.exponent.com)

Exponent is a leading engineering and scientific consulting firm providing solutions to complex technical problems. Our multi-disciplinary organization of engineers, scientists, physicians, and business consultants addresses complicated issues facing industry and government today. Exponent's staff has extensive experience in advanced analytics techniques and the application of these methods to a wide range of applied and research studies, including clinical investigations, epidemiologic studies, public health evaluations, health economics analyses, and other outcomes research. Exponent has a full-time staff of over 900 located in 25 international offices. Exponent is certified to ISO 9001 and is authorized by the GSA to provide professional engineering services.

### **Booth #116**

### **Forio**

[www.forio.com](http://www.forio.com)

High performance predictive analytics with the Forio analytics platform. Put interactive analyses within the hands of decision-makers with online analytic applications that let users change assumptions and instantly see results through their web browsers. With the Forio Analytics Platform, you can develop models in Julia (a new, high-performance technical computing language with capabilities similar to R and MATLAB) or other languages, import data and formulas from Excel, and enable users to share and compare scenarios online. The Forio Analytics Platform combines sophisticated analysis, universal accessibility via online data visualizations, and centralized model with secure access.

### **Booth #111**

### **FICO**

[www.fico.com/Xpress](http://www.fico.com/Xpress)

FICO is a leading analytics software company, helping businesses around the world make better decisions that drive higher levels of growth, profitability and customer satisfaction. FICO's analytic tools, which include FICO® Model Builder and FICO® Xpress Optimization Suite, provide functionality to turn Big Data into better decisions. FICO's analytics capabilities are now available on the FICO® Analytic Cloud and the FICO® Decision Management Platform, an easy, cost-effective way to evaluate, customize, deploy and scale state-of-the-art analytics and decision management solutions. Learn more at [www.fico.com/Xpress](http://www.fico.com/Xpress). FICO: Make every decision count™.

### **Booth #108 & 110**

### **Frontline Systems, Inc.**

[www.solver.com](http://www.solver.com)

See the new release of Analytic Solver Platform, introduced at this conference – the surprisingly easy, lightning-fast, surprisingly powerful integrated analytics software from Excel Solver developer Frontline Systems, the company that's democratizing analytics. See Solver SDK Platform and Solver Server, your fastest, most flexible path to Web deployment of your own analytics applications. Learn how little this costs, and how quickly you can create data visualizations, data mining and forecasting models, simulation and risk analysis models, and conventional and stochastic optimization models in Microsoft Excel or your favorite programming language, from C++, C# and Java to PHP and JavaScript. Take home a CD with full-power versions of our tools, and free trial licenses.

**INFORMS**[www.informs.org](http://www.informs.org)

INFORMS is offering complementary print issues of our member magazine *OR/MS Today* and our online publication *Analytics*, as well as our 13 scholarly journals including *Interfaces* and our online-only journal *Service Science*. Attendees can also purchase books, CDs, tutorials, and INFORMS t-shirts. Information is available on INFORMS Certification, Continuing Education, Subdivisions, Awards, Meetings, Publications, and Memberships. Come by the booth to speak to an INFORMS representative about our programs and products.

**Near Registration****Northwestern University Booth #115****Master of Science in Analytics**[www.analytics.northwestern.edu](http://www.analytics.northwestern.edu)

The Master of Science in Analytics program at Northwestern University is a full-time, 15-month professional master's degree that immerses students in a comprehensive and applied curriculum exploring the underlying data science, information technology and business of analytics. Supplemented by an internship placement and industry supplied projects, graduates will be exceptionally well equipped to harness and communicate the full value of data to the organizations they serve.

**ProbabilityManagement.org Booth #107**[www.probabilitymanagement.org](http://www.probabilitymanagement.org)

Non-profit ProbabilityManagement.org promotes standards for communicating uncertainties through arrays of big data known as SIPs (Stochastic Information Packets).

SIPs are:

- Actionable - may be used directly in Excel to perform thousands of Monte Carlo trials before your finger leaves the <Enter> key using only the Data Table command.
- Additive - may be aggregated across platforms across the enterprise.
- Auditable - are represented as unambiguous data with provenance.

Packages such as Crystal Ball, @Risk, Risk Solver, Matlab, R, and SAS may be leveraged through shared stochastic libraries using the open SIPmath™ standard. Visit SIPmath.org for videos and Excel demo models.

**Provalis Research****Booth # 113**[www.provalisresearch.com](http://www.provalisresearch.com)

Provalis Research is a leading developer of text analytics software with ground-breaking qualitative and quantitative analysis programs, such as QDA Miner, an innovative mixed-methods qualitative data analysis software; WordStat, a powerful add-on module for computer assisted content analysis and text mining; and Simstat, an easy yet powerful statistical software. One of the most distinctive features of these tools is their interoperability, allowing researchers to integrate numerical and textual data into a single project file and to seamlessly move back and forth between quantitative and qualitative data analysis, as well as to easily explore relationships between numerical and textual data. Visit [www.provalisresearch.com](http://www.provalisresearch.com).

**SAS**[www.sas.com](http://www.sas.com)

What if you could use big data to make faster, more precise decisions – while your competitors are still running analyses?

No matter the size or complexity of your data demands, high-performance analytics simply outperforms traditional computing methods. SAS provides several high-performance distributed processing options: in-memory, in-database and grid computing. Plus, you can seamlessly explore, visualize and model big data stored in open-source platforms like Hadoop. Visit [sas.com/bigdata](http://sas.com/bigdata) to learn how you can:

- Gain faster, more profitable insights into your business.
- Analyze and model data with domain-specific analytics.
- Deploy a scalable analytics infrastructure.
- Optimize the use of your IT resources.

**Booth #112 & 114**

**UCLA Anderson Master of Financial Engineering Program**  
[www.anderson.ucla.edu/mfe](http://www.anderson.ucla.edu/mfe)

The UCLA Anderson Master of Financial Engineering (MFE) Program equips graduates with skills required to succeed in careers in quantitative finance and data analytics. Companies in financial services and technology are increasingly looking for employees with data science skillset. Our one-year specialized degree meets that need by providing our students with a rigorous educational experience that blends mathematical and statistical modelling skills, computational expertise, and finance theory. Our program is based on the business school paradigm of merging theory and principle with up-to-the-minute business practice – so that our graduates can immerse themselves in a rewarding career immediately upon graduation.

# Poster Presentations with dessert

## MONDAY SESSION

2:00pm – 3:00pm, LL20 Foyer

### 1 - Just-in-Time Scheduling to Maximize Weighted Number of Jobs on Flow Shop Machines

Muminu Adamu, Senior Lecturer, University of Lagos

### 2 - Zero Inflated Transformation Hazard Modeling for Corporate Bankruptcy Prediction

Shaonian Tian, Assistant Professor, San Jose University

### 3 - How Do Business Models Evolve through Big-data Analytics

Christoph Breidbach, Postdoctoral Scientist, University of California, Merced

### 4 - Scalable, Real-time Big Data Analytics for Connected Cars

Dirk Van den Poel, Professor of Business Analytics/Big Data, Ghent University

### 5 - Sensitivity Analysis of the VPIN Metric

Kesheng Wu, Staff Computer Scientist, Lawrence Berkeley National Laboratory

### 6 - Salient Features of TCS' GNDM™ (Global Network Delivery Model) Framework

Natarajan Vijayarangan, Senior Scientist, Tata Consultancy Services Ltd (TCS)

### 7 - Exploration vs. Exploitation in the Information Filtering Problem

Xiaoting Zhao, PhD Candidate, Cornell University

### 8 - Twitter Analysis Captures Antiretroviral Drug Toxicity and Community

Cosme Adrover, Postdoctoral Scholar in Data Science, Pennsylvania State University

### 9 - Detection of Manipulated Movie Reviewers and its WOM Implications: Experience in a Movie Porta Site in Korea

Jaehyeon Ju, Doctoral Student, KAIST

### 10 - Organization, Segmentation, and Hierarchies for Large Number of Time Series

Michele Trovero, Research Statistician, SAS Institute Inc.

### 11 - Large Scale Hierarchical Text Classification

Zheng Zhao, SAS Institute, Inc.

### 12 - Experiments in Deep Learning

Patrick Hall, SAS Institute, Inc.

### 13 - Establishing Functionally Similar Sites Based on Direct Impressions

Stephen Burning, University of Illinois Mathematics Department

## TUESDAY SESSION

2:00pm – 3:00pm, LL20 Foyer

### 1 - Business Analytics with Forio Epicenter using Julia: Delivery Route Optimization with a Mobile Interface Example

Brian Piper, Data Scientist, Forio

### 2 - An Application of Principle Component Analysis in Clinical Trials – Choosing the 'Best' Sites

Chao Deng, Sr. Operational Analytics Specialist, Quintiles

### 3 - Bridging the Semantic Gap in Multimedia Retrieval with Topic Extraction From User Reviews

Eunjeong Park, Seoul National University

### 4 - Generating Ideas from Hidden Customer Needs in Social Media for New Product Development Process

Taehoon Ko, Seoul National University

### 5 - Operationalizing Big Data for the Tactical Edge

Arkady Godin, Research Associate, Naval Postgraduate School

### 6 - Research on Big-Data to Increase Wheat Yield and Efficiency of "Granary of Bohai Bay"

Chupeng Xie, Professor, Shandong Agricultural University

### 7 - Large Scale Optimization for Stochastic Economic Dispatch

Harsha Gangammanavar, Visiting Assistant Professor, University of Southern California

### 8 - Analyzing Historical Exchange Rate

Ali Alhamdan, Graduate Student, Marywood University

### 9 - Post-Marketing Safety Monitoring of Self-Reported Symptom-Treatment-Outcome Measurement in Web-based Media Sources

Michael Wallis, SAS Institute, Inc.

### 10 - Sales Prediction for New Product with Kaggle Data

Shuanglong Wang, PhD Candidate, University of Illinois Urbana Champaign

### 11 - A New Approach to Time Series Data Preparation in SAS®

Kenneth P. Sanford, PhD, SAS Institute, Inc.

### 12 - Using Big Data and Text Analytics to Manage Attrition

Vinh Nguyen, Toyota Financial Services

## Sunday June 22

**12:00pm – 7:00pm**  
Registration

### 1:45pm – 5:00pm Technology Workshops

See page 4 for complete information on sessions and rooms. Workshops held during these times:  
1:45pm – 3:15pm  
3:30pm – 5:00pm

**LL20 Foyer**

## Monday June 23

**7:00am – 5:00pm**  
Registration

### 7:30am – 8:30am Continental Breakfast

**9:30am – 4:00pm**  
Exhibits Open

### 8:30am – 9:30am Plenary Presentation

#### Putting Big Data to Work

Bill Franks, Chief Analytics Officer  
Teradata Corporation

Big data is everywhere. You can't avoid being exposed to discussions around big data, and the analysis of it, on a regular basis. The downside of this attention is that there is a lot of hype and misinformation in the marketplace. Many companies are confused about how to get started, what actions to take, and what pitfalls to avoid. Based on content from the popular book Taming The Big Data Tidal Wave, this talk will provide an overview of important themes to understand regarding big data. The talk will address technological, organizational, and cultural points that must be considered. Most importantly, the talk will aim to provide attendees a solid direction to take their big data initiatives.

**9:40am – 10:30am**  
**Track 1: Case Studies**

### Optimal Fusion for Predictive Road-way Traffic Speeds

Toby Tennent, HERE, a Nokia Business

The Estimation of Short Term Future Traffic Conditions (15 mins - 12 hours ahead) can be Accomplished using a Multitude of Methods. In this case-study, we discuss our development of a principled algorithmic approach to fusing different estimates of future traffic speed conditions for an 'optimal estimate'. Data sparseness, noise and competing model inputs were issues to overcome. In our case, though use of a Minimum Variance Unbiased Estimator (MVUE) approach to data-fusion, we were able to see a 25%+ improvement over our legacy approach. Lessons from this case-study can be applied to many applications involving current and predictive state-modelling."

**LL20 Foyer**

**LL20BCD**

**LL20 Foyer**

**LL20A**

### Track 2: Case Studies

**LL21E**

#### Big Data in Practice - Scientific Information as a Business Asset – Driving Productivity at Merck Research Labs Through Novel Approaches to Scientific Information Management

John Erik Koch, Director, Informatics, Merck & Co.

BioPharma companies often struggle to manage scientific information – study results, analyses and historical records are lost due to poor information management practices and failure to steward information in a way that can be leveraged for future purposes. We will share examples of our strategy, execution and progress to date for improving information management through a set of innovative capabilities focused on information Search, Access and Analytics.

### Tracks 3: Big Data 101

**LL21D**

#### Getting Started with Hadoop and Big Data

Brian Keller, Data Scientist  
Booz Allen Hamilton

Getting started with big data analytics can be a daunting task because of the complexities of the technologies and the fast evolution of the software ecosystem. This talk will expose analytics professionals to established open source big data technologies and provide them with practical methods to prototype and develop big data analytics without requiring deep knowledge of software development or distributed computing. Come learn the secrets that hard core software developers don't want you to know! Keller will cover:

- An overview of Hadoop and Map Reduce,
- Alternatives to writing Java code to develop analytics on Hadoop,
- The big data analytic development life-cycle; how to go from data to actions,
- A worked example using the principals discussed,
- How you can get started with big data technologies today.

<b>Track 4: Emerging Trends</b>	<b>LL21C</b>	<b>11:10am – 12:00pm</b>	<b>Tracks</b>	<b>Tracks 3: Big Data 101</b>	<b>LL21D</b>
<b>Big Data Combining Large and Semantically Disparate Datasets to Come to Meaningful Conclusions</b>		<b>Track 1: Case Studies</b>	<b>LL21F</b>	<b>Hadoop Beyond Java: Pig, Hive &amp; Other Friends</b>	
Chris Hallenbeck, Global Vice President, SAP		<b>Big Data in Official Statistics</b>		Diego Klabjan, Professor Northwestern University	
While the three Vs of big data (volume, velocity, and variety) are largely focused on the capturing and storage of the data, we have learned a lot over the last few years about the characteristics and challenges of Big Data. The real challenge of Big Data is combining large and semantically disparate datasets to come to meaningful conclusions. Even in 2014, it's hard enough to 'join' text to tabular data! How can this data be mined to derive a model that can be scored? What about semi-structured, Machine or Graph data? The answer to the problem is all about semantics. We have SQL for tabular data and SPARQL for RDFs, but how can these meaningfully be combined and exposed to solve larger problems? How do we expose the resulting data to developers in a meaningful way? What is the best way to present the data to the end-users for the consumption and application of the data?		Cavan Capps, Big Data Lead U.S. Census Bureau		Director, Master of Science in Analytics	
<b>Track 5: Software Tutorials</b>	<b>LL21B</b>	<b>Track 2: Case Studies</b>	<b>LL21E</b>	It takes two to tango, and Hadoop and MapReduce are an excellent example of this idiom. In order to maximize performance of MapReduce, low level java implementations must be used which requires teams of data scientists. Scripting engine Pig offers a nice balance between ease of use and performance. Similarly, Hive offers a more SQL-like syntax for data stored in files within Hadoop. This talk will highlight the basic principles of these two tools. It will focus on their pros and cons and simple examples will be discussed. If you have to introduce these technologies to your organization, employ the necessary talent, or are behind the selection decision making process, this is a must attend presentation. A substantial portion of the talk will revolve around the question of selecting the right tool for the task at hand.	
<b>INFORMS Analytics Maturity Model</b>		<b>Improving Efficiency of Google's Infrastructure Using Big Data Tools</b>			
<b>How Mature Are You? Preview the Beta Version of the New INFORMS Analytics Maturity Model</b>		Behdad Masih, PhD, Quantitative Analyst Google			
Aaron Burciaga, Senior Manager Operations Analytics, Accenture Analytics Co-author; Nagaraj Reddi, Director of Information Technology, INFORMS		In this talk, we will review how Google's big data tools are being used internally to improve the efficiency of Google's infrastructure. Every minute, hundreds of thousands of servers across Google's fleet report performance metrics of hardware components such as cpu, disk, memory and network bandwidth usage and these metrics are saved in different formats, in different databases, across multiple datacenters. But as consumers of the data, Google's data analysts can easily merge and aggregate these metrics and perform data mining techniques to identify and address inefficiency of the resource deployment. As the first case study, we will show how we have used Google's application programming interfaces to merge performance metrics and forecast the required networking capacities at server, cluster and datacenter levels as a function of other hardware resources such as cpu and disk. During the second case study, we share our experience setting up a bandwidth monitoring dashboard used to inform flash deployment decisions.			
<b>10:30am – 11:00am</b>	<b>LL20 Foyer</b>				
<b>Refreshment Break with Exhibits</b>					

<b>Track 4: Emerging Trends</b>	<b>LL21C</b>	<b>Track 5: Software Tutorials</b>	<b>LL21B</b>	<b>1:10pm – 2:00pm</b>	<b>Tracks</b>
<b>Taming Big Data with Berkeley Data Analytics Stack (BDAS)</b>		<b>FORIO- Publishing Interactive Data Analytics on the Web</b>		<b>Track 1: Case Studies</b>	<b>LL21F</b>
Ion Stoica, Professor, UC Berkeley CEO, Databricks, CTO, Conviva		Michael Bean, Forio		<b>Conception to Deployment of Big Data Projects at a Large Financial Institute</b>	
One of the most interesting developments over the past decade is the rapid increase in data; we are now deluged by data from on-line services (PBs per day), scientific instruments (PBs per minute), gene sequencing (250GB per person) and many other sources. Researchers and practitioners collect this massive data with one goal in mind: extract “value” through sophisticated exploratory analysis, and use it as the basis to make decisions as varied as personalized treatment and ad targeting. Unfortunately, today’s data analytics tools are slow in answering even simple queries, as they typically require to sift through huge amounts of data stored on disk, and are even less suitable for complex computations, such as machine learning algorithms. These limitations leave the potential of extracting value of big data unfulfilled. To address this challenge, we are developing BDAS, an open source data analytics stack that provides interactive response times for complex computations on massive data. To achieve this goal, BDAS supports efficient, large-scale in-memory data processing, and allows users and applications to trade between query accuracy, time, and cost. In this talk, I’ll present the architecture, challenges, early results, and our experience with developing BDAS. Some BDAS components have already been released: Mesos, a platform for cluster resource management has been deployed by Twitter on 6,000+ servers, while Spark, an in-memory cluster computing frameworks, is already being used by tens of companies and research institutions.		Forio’s web platform makes your model available to hundreds of people within your organization through the browser. We will start with an introduction to the platform and sample interactive online models. Then we’ll divide the workshop into two parts. In the first part we will teach you how to get your analysis on a server so it can be shared. In the second part we’ll focus on creating a user interface for your model.	Mike Aguilina, Executive Director JP Morgan Intelligent Solutions	What does it take to directly impact top line revenue with Big Data projects? What are the typical hurdles to overcome as you move from a research project to integrate directly into an existing business? We review the lifecycle (conception to deployment) of Big Data projects at a large financial institute. The talk will cover topics including project concept and design, guidelines to Big Data project development, a high level overview of a ‘Big Database’, and end project deployment. The focus will be on the specific business use case, the analytic methodology used, and the measurable outcome. In addition, there will be a review of lessons learned and typical challenges faced working as the “Big Data Team” in a large business. We will also briefly cover regulatory hurdles framing Big Data solutions (specific to financial institutions).	
<b>Track 6: Software Tutorials</b>	<b>LL20A</b>			<b>Track 2: Case Studies</b>	<b>LL21E</b>
		<b>SAS - High-performance Data Mining and Big Data Analytics</b>		<b>Big Data in Health</b>	
		Jaren Dean, Senior Director of Research and Development for Advanced Analytics SAS		Juergen A. Klenk, PhD, Principal Scientist Exponent®, Inc.	
		In this era of big data, organizations strive to make the best use of information. High-performance data mining and big data analytics drive better, faster decisions based on vast amounts of data. During this tutorial, learn about the evolution of hardware platforms; predictive modeling methods, including statistical and machine learning; and the benefits of segmentation. Through case studies and straightforward examples, you’ll walk away with new ideas on how to implement these methods and technologies.		Erik Brynjolfsson of MIT’s Sloan School of Management found a remarkable correlation between an organization’s performance and its data driven decision making (DDD) ability, thus demonstrating the actual value of data. Fueled by findings like this and other promises, hospital systems, health insurance providers, pharmaceutical and biotech firms, and medical devices manufacturers alike now have high expectations to generate value from Big Data. Indeed, we are nearing the “Peak of Inflated Expectations” in Gartner’s Hype Cycle, and organizations must quickly learn to develop and implement Big Data strategies. In this talk, we will review key components of a Big Data strategy, and demonstrate how it can be successfully implemented to benefit your organization, whether you are interested in improvement of the quality of your healthcare delivery, the safety and benefits of your drugs and devices, the effective management of your operations, or the control of cost.	
<b>12:00pm – 1:00pm</b>	<b>LL20BCD</b>				
		<b>Networking Lunch</b>			
		Join your colleagues and meet someone new to discuss how you are using big data in your company. Seating is not assigned for this event.			

<b>Tracks 3: Big Data 101</b> <b>An Introduction to NOSQL Databases and their Analytic Uses</b> Aaron Cordova, CTO & Co-Founder Koverse	<b>LL21D</b>	<b>Track 5: Software Tutorials</b> <b>ProbabilityManagement.org-The SIPmath™ Modeler Tools</b> Sam L. Savage, Executive Director ProbabilityManagement.org	<b>LL21B</b>	<b>3:00pm – 3:50pm</b> <b>Track 1: Case Studies</b> <b>Smart Machine Learning to Lend to Small Businesses</b> Pinar Donmez, Chief Data Scientist Kabbage, Inc.	<b>Tracks LL21F</b>
NOSQL technology has the potential to provide powerful data storage and analytic capabilities, but the space is just emerging, crowded with many approaches and difficult to navigate. This talk will expose analytics professionals to some of the core principles that make NOSQL unique and dive into the specifics of several of the more established efforts. Included will be sample use cases and lessons learned. Specifically this talk will cover:		The Open SIPmath Standard from non-profit ProbabilityManagement.org allows uncertainties to be communicated as big data for driving interactive simulation in native Excel and other environments without add-in software. We will demonstrate the SIPmath Modeler Tools, which facilitate the generation of such models in Excel, and also how to import and export results from @RISK, Crystal Ball, Risk Solver and Matlab to leverage those packages. Visit the SIPmath page of SIPmath.org for videos and example files.		The way traditional financial institutions underwrite loans and manage risk has been changing incredibly with the help of emerging technologies. Consumer/business loans and other areas of credit scoring have been disrupted by interdisciplinary technologies combining risk management with machine learning and advanced data analytics. The talk will cover how we turn data into decisions for underwriting and dynamic risk management at Kabbage Inc.	
<ul style="list-style-type: none"> <li>• An overview of NOSQL principals and history</li> <li>• An overview of core analytic use cases for NOSQL</li> <li>• An overview of leading NOSQL solutions and analytic use cases</li> <li>• Ideas and tools for getting started with NOSQL technology.</li> </ul>					
<b>Track 4: Emerging Trends</b> <b>Watson in the Emerging Era of Cognitive Computing</b> Rob High Jr., IBM Corporation	<b>LL21C</b>	<b>Track 6: Software Tutorials</b> <b>Booz Allen Hamilton - Machine Learning on Streaming Data with Storm and MOA</b> Paul Yacci, Data Scientist Booz Allen Hamilton	<b>LL20A</b>	<b>Track 2: Case Studies</b> <b>Developing a Lean Data Science Strategy</b> Steven Hillion, Chief Product Officer Alpine Data Labs	<b>LL21E</b>
In this talk I will describe the role of cognitive computing in addressing some of the worlds most important business and social problems. I will discuss how Watson leverages human-readable information at massive scale to meet those needs. In addition, I will provide an update on how IBM has applied that capability to assist clinical decision support for Oncology, and for engaging clients in Insurance and Financial Services institutions.		Analysis of streaming data enables real-time actions in applications such as network defense and fraud detection. Storm provides an open source distributed stream processing framework designed to scale to the demands of big data. Storm can be extended to include machine learning by integrating the open source library MOA, which provides machine learning algorithms capable of online learning. This tutorial will introduce concepts to build a scalable, streaming machine learning platform using open source software.		When developing a data science project strategy, the scope, tools, and team are crucial. In Lean Data Science projects, these considerations require special attention prior to kicking off the engagement. We will present possible paths to success, by focusing on topics including: <ul style="list-style-type: none"> <li>• Setting “bite-sized” project goals</li> <li>• Using tools that provide a display of process and information</li> <li>• Utilizing a cross-functional team</li> </ul> It is important to map out the path to a minimally viable product such that critical points of failure are evaluated early when developing a process.	
		<b>2:00pm – 3:00pm</b> <b>Poster Session with Dessert Break</b> Enjoy dessert, visit the exhibits, and learn from poster presentations that showcase a full range of big data analytics, including many applications areas, industries and methodologies. As a guide for identifying which posters you want to visit, watch for the color-coded signs that indicate general topic areas. Prizes for the top 3 overall projects will be awarded at 2:50pm. The criteria are interesting use of big data analytics, effective display of results and business impact. See page 8 for the full list of posters.	<b>LL20</b>	<b>Foyer</b>	

**Tracks 3: Big Data 101****What is Data Science, and Is a "Data Scientist" a Myth?**

Steve Mills, Senior Associate  
Booz Allen Hamilton

Data Scientist was named the sexiest job of the 21st Century. Given the rapid increase in the term "Data Science" over the last year, many organizations are left with figuring out who a data scientist is and what they do, leading many to think that it is a "purple unicorn". Many organizations believe they can simply turn an existing analytics team into a "data science team" or that they can hire a "data scientist". In addition, many organizations are evaluating Big Data systems and are challenged with providing value from these deployments. Through client examples, this talk will discuss the terminology around OR and data science; discuss the next generation of data science technology, Big Data, and subsequent skills; showcase data science organizational design, deployment, and tips for building the team; and present case studies on how businesses and governments used data science to gain competitive advantage and to better serve citizens.

**Track 4: Emerging Trends****LL21C  
Data Products - Changing Lives, Disrupting Markets, Hard to Build**

Shawn Scully, Graph Lab/UC Berkeley

What the heck is a data product and why should you care? Applications that drive online product recommendations, predict machine failures, forecast airfare, social match-make, identify fraud, and predict churn are all examples of data products. Under the hood, these products consume all that Big Data and apply machine learning and statistics magic to drive real-time actions. Data products are quickly becoming the most valuable part of the Big Data pipeline both because many of them drive a company's most important customer experiences and because these can directly affect the bottom line. If you don't have data products driving new business at your company, you should be asking why. In spite of this value, teams aren't developing these products fast

**LL21D**

enough and those just getting started have a lot of trouble knowing where to begin. At GraphLab we've been working hard to make it easy for both developers and data scientists to build, deploy, and manage their data products. How many times have you heard that things don't scale or have had to completely rewrite your applications just to deploy a proof of concept? R, Mahout, sci-kit learn, Hadoop, relational, key-value, and graph databases; Pig, Hive, Flume, and array of visualization and other tools.... Is there a better way? We think so. Whether you are a data scientist trying to do more, or a developer or business owner trying to get started, this talk is for you. I'll define data products and how they are different from traditional data analytics. I'll give some examples of the cool applications you can build, show some code and systems behind the scenes, and show how data scientists can be more productive, developers can get started faster, and businesses can define new strategic opportunities.

**Track 5: Software Tutorials****LL21B  
Provalis Research  
Presentation of Provalis Research**

**Text Analytics Tools**  
Adam Bendriss Alami, Senior Marketing Manager, Provalis Research

Provalis Research will showcase its integrated collection of text analytics software. QDA Miner is an easy to use qualitative and mixed methods software that meets the needs of researchers performing qualitative data analysis and would like to code more quickly and more consistently larger amounts of documents. It offers high level computer assistance for qualitative coding with innovative text search tools that help users speed up the coding process as well as advanced statistical and visualization tools. Users with even bigger text data, can also take advantage of WordStat. This add-on module to QDA Miner can be used to analyze huge amounts of unstructured information, quickly extract themes, find trends over time, and automatically identify patterns and references to specific concepts using categorization dictionaries.

**Track 6: Software Tutorials****LL20A  
Contact Singapore  
Data Analytics Career Opportunities in Singapore**

Kelly Tan, Area Director  
Contact Singapore

Interested to find out more about Big Data trends and career opportunities in Singapore? Join us at the session by Contact Singapore where we will cover topics ranging from the latest Big Data developments in Singapore to career opportunities in data analytics and general information on working and living in Singapore. Come find out more!

**4:00pm – 5:30pm****LL20BCD****Birds-of-a-Feather Discussion Groups**

Stop by for beer, soda, snacks and good conversation around important issues. Topics will be selected by the attendees. We need your input! Look for the white board near registration to suggest topics or vote for a topic already posted. This will be the most popular and highly-rated aspects of the conference—don't miss it!

**5:30pm – 7:00pm      Marriott San Jose General Reception      Ballroom**

Stop by this informal reception where you can connect with other attendees and unwind from a jam packed day. Members of the Conference Committee will be your hosts. Enjoy a hearty display of heavy hors d'oeuvres and food stations, beer, wine, soft drinks and bottled water.

# Tuesday June 24

**7:00am – 3:00pm**  
**Registration**

**LL20 Foyer**

**7:30am – 8:30am**  
**Continental Breakfast**

**LL20BCD**

**9:30am – 3:00pm**  
**Exhibits Open**

**LL20 Foyer**

**8:30am – 9:30am**  
**Plenary Presentations**

**LL20A**

## Moving from Big Data to Big Outcomes on the Journey to ROI

Michael Svilar, Managing Director  
Advanced Analytics-Accenture Analytics,  
Accenture

In this Internet of Everything world, systems, devices and physical objects are “talking” to one another. There are upwards of a trillion connected and instrumented things: cars, appliances, cameras, roadways, pipelines...even pharmaceuticals and livestock. The talk will focus on how organizations can drive positive business outcomes in the connected world using big data analytics.

**9:40am – 10:30am**  
**Track 7: Case Studies**

**Tracks**  
**LL21F**

## Optimizing Media Purchasing Through Big Data

Alan Papir, Software Engineer  
Analytics Media Group (AMG)

Analytics Media Group (AMG) is using lessons learned from the 2012 Obama presidential campaign to bring new, data-driven insights into the world of media buying. Using various modeling and data mining techniques in conjunction with large and rich datasets (such as billions of set top box records), AMG discovers who is most likely to “convert” to a product or candidate at the person-level. AMG then takes these desirable targets and uses a trove of set top box data to produce a near-optimal solution to problem of purchasing the most valuable placements given a limited budget (a multi-objective variation of the knapsack problem). This presentation will cover some of AMG’s techniques for identifying targets, strategies for efficiently storing and retrieving tens of billions of TV

viewing records, and heuristics for finding a near-optimal media buy plan. AMG has been featured on the cover of the *New York Times* magazine as well as in *Bloomberg*, *Politico*, the *Cook Political Report* and elsewhere.

## Track 8: Big Data 101

**LL21E**

### Panel Discussion: Analytics Talent

*Moderator:* Anne Robinson, Director, Chain Strategy & Analytics, Verizon Wireless, *Panelists:* Colin Kessinger, Theresa Kushner, Thomas Olavson , Aditya Rastogi

MIT claims that 67% of companies see having analytics capabilities as a driver for their competitive advantage. However, according to TDWI, 46% of companies listed inadequate staffing or skills as the top barrier for realizing value from their big data and analytics investments. What does it take to have a successful big data and analytics capability in an organization? How do you attract the quintessential data scientist? What are the executive sponsor and leadership qualities required to be successful? Listen to a panel of industry expert's answer these questions and more.

## Track 9: Emerging Trends

**LL21D**

### The Role of the Big Data Platform in a Smart System

Kaushik Kunal Das  
Senior Principal Data Scientist  
Pivotal, Inc.

There is a proliferation of data in today's world in terms of quantity, type and sampling speed. This is matched by the growth in tools to store and extract value from this data. But what is missing is an effective way to bring these together to form a smart system. I am going to define a smart system and describe how a big data platform which is based on open standards and compatible with many analytical and data processing tools is essential for building such a system. I am also going to describe a few specific use cases of such systems and the technology, methodology and algorithms needed to create them.

## Track 10: Software Tutorials

**LL21C**

### FICO

#### Analytics Tools for the Area of Big Data

Benjamin Baer, Senior Director,  
Product Marketing

Get a hands-on look at how to use new innovations in FICO® Model Builder that detect predictive signals from massive and unstructured data for extracting real, actionable insight. Learn how to turn these predictions into optimized decisions with the new robust modeling and optimization capabilities of FICO® Xpress Optimization Suite. Come see how Model Builder and Xpress can help you make better decisions fueled by Big Data.

## 10:30am – 11:00am

**LL20 Foyer**

### Refreshment Break with Exhibits

## 11:10am – 12:00pm

**Tracks**

**LL21F**

### Real-world Big Data Applications

Jane Uyvova, Teradata

Real-world Big Data Applications: Learn how customers are utilizing bid data in the enterprise to uncover valuable insights and operationalize them across large organizations and complex business processes. Topics include data discovery strategies and applications in telecommunications, financial services, life sciences and manufacturing industries. Use cases address solutions for digital marketing, fraud loss prevention, customer churn prediction, sales force enablement and more.

**Track 8: Big Data 101****Python, R, and SQL in MPP Databases**

Anton J. Mobley

Kaiser Permanente

Python, R, and SQL are some of a data scientist's favorite tools for exploring data. Massively Parallel Processing (MPP) databases provide one way to leverage these tools at scale. This talk introduces MPP databases and shows how to combine them with analytic tools to start developing models. Topics include:

- A brief introduction to how MPP databases work from the data scientist's perspective
- Which type of data science problems MPP databases are well suited to solve and why they are used even when HDFS is likely cheaper
- How to use Python and R natively in MPP databases
- Sample problems that leverage Python, R, and SQL in MPP

**Track 9: Emerging Trends****Big Data and Big Analytics – So Much more Gunpowder!**Paul Kent, Vice President  
SAS

Mathematical Computing is changing for the better – Modern Analytic Platforms such as Hadoop are embracing the Massively Parallel Cluster. Spread your data out over the nodes, and find ways to send the work to the data (instead of the other way around), and you'll soon be enjoying some awesome advances in computing firepower. This talk details the transition to this new style of computing and gives examples of software techniques that were un-reachable before but are now practical. Interactivity and visual data exploration are available across all your data and now you can access sophisticated programmatic model development that was previously computationally intractable! The talk should challenge you to consider the status quo at your organization – are you doing enough to adapt to the new wave of computing?

**LL21E****Track 10: Software Tutorials****Frontline Systems, Inc.****Analytic Solver Platform: Integrated Data Mining, Simulation and Optimization in Microsoft Excel**

Daniel H. Fylstra, Frontline Systems, Inc.

Analytic Solver Platform in Microsoft Excel has everything you need for data visualization, forecasting and data mining, Monte Carlo simulation and risk analysis, and conventional and stochastic optimization – where its solving power actually surpasses “enterprise” analytic software costing far more. See how you can use it to build your own analytic expertise and teach others, leveraging what you already know, build and solve industrial-scale models with the world's best Solvers, and effectively communicate business results.

**LL21D****12:00pm – 1:00pm LL20BCD**  
**Lunch and Roundtable Discussion**

Join your colleagues and meet someone new to discuss what is keeping you up at night, as a big data expert or data scientist, and where do you see big data analytics going in the next 3 years. Seating is not assigned for this event.

**1:10pm – 2:00pm****Tracks LL21F**  
**Track 7: Case Studies****Big Data Analytics Application to Jet Engine Diagnostics**

Link C. Jaw, Fellow, Intel Corporation

The “explosion of information” that we have witnessed is not just contributed by a large number of data sources, but also by the large amount of data originating from these sources. Together they have created the so-called “big data” environment characterized by variety, volume, and velocity. The challenge of information explosion is how to extract the right information at the right time from the big data environment which we live in. Extracting the right information from the data is an analytic process, and this process uses data platforms and analytic algorithms to spot trends and patterns so as to derive predictive indicators. These indicators are then used to make proper recommendations or to take timely actions. In this case study, an aircraft engine diagnostic problem is described and a solution is discussed.

Specifically, the discussion will include:

- Background of the aviation industry.
- Characteristics of machine diagnostic problem.
- Data elements.
- Algorithms.
- Results of analytics.

This problem is analogous to a large class of machine monitoring and control problems, hence, the solution approach discussed here is applicable to various industrial sectors.

**Track 8: Big Data 101****LL21E****Big Data Solutions on AWS Platform**

Vikram Garlapati, Manager Solutions Architect, Amazon Web Services

Scientists, developers, and many other technologists from many different industries are taking advantage of Amazon Web Services to meet the challenges of the increasing volume, variety, and velocity of digital information. Amazon Web Services offers an end-to-end portfolio of cloud computing resources to help you manage big data by reducing costs, gaining a competitive advantage, and increasing the speed of innovation. It feels like everything generates data today, from your customers on social networks to the instances running your web applications. AWS makes it easy to provision the storage, computation, and database services you need to turn that data into information for your business. AWS also has data transfer services which can move big data into and out of the cloud quickly such as AWS Direct Connect and our Import/Export service. In this presentation we will talk about

- AWS Direct Connect and our Import/Export service
- Redshift - Petabyte-scale data warehousing in minutes.
- Amazon Elastic MapReduce (EMR) provides - Easy analytics for everyone
- Amazon Kinesis - streaming data at massive scale

<b>Track 9: Emerging Trends</b>	<b>LL21D</b>	<b>2:00pm – 3:00pm</b>	<b>LL20 Foyer</b>	<b>Track 8: Big Data 101</b>	<b>LL21E</b>
<b>Can Machine Learning Really Do That?</b>		<b>Poster Session with Dessert Break</b>		<b>Babies, Brains, and Buses... and Why Stream Computing is the Right Approach to Real-time Predictions and Decision Making</b>	
Anthony Goldbloom, Founder and CEO Kaggle		Enjoy dessert, visit the exhibits, and learn from poster presentations that showcase a full range of big data analytics, including many applications areas, industries and methodologies. As a guide for identifying which posters you want to visit, watch for the color-coded signs that indicate general Prizes for the top 3 overall projects will be awarded at 2:50pm. The criteria are interesting use of big data analytics, effective display of results and business impact.	See page 8 for the full list of posters.	Kevin Foster, Big Data Solutions Architect IBM	Babies, brains, and buses. What do they have in common? Each of these have been the subject for stream computing projects that monitor data in real-time with decisions made quicker and in higher volume than is possible with any disk or even memory based store-then-query architectural approach. Real-time decision making can often require sub-second results, but can also just be within seconds or even minutes based on the needs of the project. In this talk we'll discuss why real-time is better approached from the perspective of just-in-time, and discuss examples of use cases across industries including their algorithms, data volumes, latency requirements, project methodology, and implementation architectures that include analytics with R, MATLAB, SPSS, and other technologies.
<b>Big data is one of today's hot technology trends. We're told that our ability to store petabytes of data and the profusion of sensors will transform modern business. One element of the big data story that's often overlooked is the development of machine learning techniques that allow us to do more with the data that we're storing. This talk will introduce data mining competitions as a way to solve challenging problems. In doing this, it will also cover some of what Kaggle is seeing at the cutting edge of machine learning, including some of the applications that are now possible with modern techniques.</b>		<b>3:00pm – 3:50pm</b>	<b>Tracks LL21F</b>	<b>Track 9: Emerging Trends</b>	<b>LL21D</b>
<b>Track 10: Software Tutorials</b>	<b>LL21C</b>	<b>Track 7: Case Studies</b>	<b>A Case Study on Grid Data Analytics</b>	Simon Zhang, Senior Director, Business Analytics, LinkedIn	
<b>Enrich Consulting</b>		Marina Thottan, Director Bell Laboratories, Alcatel-Lucent		Please visit the conference website for additional information.	
<b>Finding the Strategy Hidden in Your New Product Portfolio with the Enrich Analytics Platform</b>					
Dan Smith, VP of Consulting Richard Sonnenblick, CEO Enrich Consulting		Smart grid evolution involves tremendous growth in sensor deployment in both the distribution and the transmission grid, and an exponential growth in the availability of both structured and unstructured data collected from sensors and other external sources such as demographic and environmental data. To realize the return on investment in smart grid technology it is important to maximize the benefit of the substantial amount of data flowing into utility control centers. The wealth of data obtained can be leveraged to extend visibility into the grid from the traditional boundaries of the Supervisory Control and Data Acquisition (SCADA) network (i.e., substations and transformers) all the way down to the end user customer premises. Utilities have to develop a data analytics strategy for timely and precise analysis of the grid data to define novel business and operational applications that will greatly enhance grid operations. In this talk we will describe the design of an online grid data analytics system along with lessons learned.			
The Enrich Analytics Platform helps R&D companies value, communicate, and refine tens of billions of dollars in R&D investments annually. Come see for yourself how easy it can be to build new product forecasts that include critical uncertainties, create and compare what-if scenarios, and judge the strategic value of a portfolio of investments, using the Enrich Analytics Platform. We'll also introduce a free set of tools available on our web site to build visualizations specially developed to communicate the trade-offs between portfolio value and portfolio balance.					



**November 9-12, 2014**

Hilton San Francisco & Parc 55 Wyndham  
San Francisco, California

**Registration Now Open**

Thanks to our Sponsors:



Sunday, June 22

12:00pm - 7:00pm	<b>Registration - LL20 Foyer</b>
1:45pm - 5:00pm	<b>Technology Workshops</b>
5:30pm - 7:00pm	<b>Welcome Reception and Exhibits - LL20 Foyer</b>

Monday, June 23

7:00am - 5:00pm	<b>Registration - LL20 Foyer</b>
7:30am - 8:30am	<b>Breakfast - LL20BCD</b>
9:30am - 4:00pm	<b>Exhibits Open - LL20 Foyer</b>
8:30am - 9:30am	<b>Plenary - Bill Franks Teradata - LL20A</b>

TRACKS	T1 CASE STUDIES	T2 CASE STUDIES	T3 BIG DATA 101	T4 EMERGING TRENDS	T5 SOFTWARE TUTORIALS	T6 SOFTWARE TUTORIALS
ROOMS	LL21F	LL21E	LL21D	LL21C	LL21B	LL20A
9:40am - 10:30am	Optimal Fusion for Predictive Road-way Traffic Speeds <i>Toby Tennent</i> HERE	Big Data in Practice - Scientific Information as a Business Asset <i>John Erik Koch</i> Merck & Co., Inc.	Getting Started with Hadoop and Big Data <i>Brian Keller</i> Booz Allen Hamilton	Big Data Combining Large and Semantically Disparate Datasets to Come to Meaningful Conclusions <i>Chris Hallenbeck</i> SAP	<i>INFORMS</i> INFORMS Analytics Maturity Model Preview the Beta Version	

10:30am - 11:00am	Refreshment Break with Exhibits - LL20 Foyer					
11:10am-12:00pm	<b>Big Data in Official Statistics</b> <i>Cavan Capps</i> U.S. Census Bureau	<b>Improving Efficiency of Google's Infrastructure Using Big Data Tools</b> <i>Behdad Masih</i> Google	<b>Hadoop Beyond Java: Pig, Hive &amp; Other Friends</b> <i>Diego Klabjan</i> Northwestern University	<b>Taming Big Data with Berkeley Data Analytics Stack (BDAS)</b> <i>Ion Stoica</i> UC Berkeley/ Databricks/Conviva	<b>Forio Publishing Interactive Data Analytics on The Web</b>	<b>SAS High-performance Data Mining and Big Data Analytics</b>

12:00pm-1:00pm	Lunch and Roundtable Discussions - LL20 BCD					
1:10pm-2:00pm	<b>Conception to Deployment of Big Data Projects at a Large Financial Institute</b> <i>Mike Aguiling</i> JP Morgan Intelligent Solutions	<b>Big Data in Health</b> <i>Juergen A. Klerk</i> Exponent®, Inc.	<b>An Introduction to NOSQL databases and their Analytic Uses</b> <i>Aaron Cordova</i> Koverse, Inc.	<b>Watson in the Emerging Era of Cognitive Computing</b> <i>Rob High Jr.</i> IBM Corporation	<b>ProbabilityManagement.org</b> The SIPmath™ Modeler Tools	<b>Booz Allen Hamilton</b> Machine Learning on Streaming Data with Storm and MOA

2:00pm-3:00pm	Poster Session - Dessert Break with Exhibits and Posters - LL20 Foyer					
3:00pm-3:50pm	Smart Machine Learning to Lend to Small Businesses <i>Pinar Donmez</i> Kabbage Inc.	Developing a Lean Data Science Strategy <i>Steven Hillion</i>	What is Data Science, and is a "Data Scientist" a Myth? <i>Steven Mills</i>	Data Products - Changing Lives, Disrupting Markets, Hard to Build <i>Shawn Scully</i> GraphLab	<i>Provalis Research</i> Presentation of Provalis Research Text Analytics Tools	Contact Singapore Data Analytics Career Opportunities in Singapore

5:30pm - 7:00pm General Reception in Marriott Hotel - San Jose Ballroom

## Tuesday, June 24

7:30am - 3:00pm	Registration - LL20 Foyer
7:30am - 8:30am	Breakfast - LL20BCD
9:30am - 3:00pm	Exhibits Open - LL20 Foyer
8:30am - 9:30am	Plenary - Michael Svilar, Accenture - LL20A

TRACKS	T7 CASE STUDIES	T8 BIG DATA 101	T9 EMERGING TRENDS	T10 SOFTWARE TUTORIALS
ROOMS	LL21F	LL21E	LL21D	LL21C
9:40am - 10:30am	Optimizing Media Purchasing Through Big Data <i>Alan Papir</i> Analytics Media Group (AMG)	Panel Discussion: Analytics Talent <i>Anne Robinson</i> Verizon Wireless	The Role of the Big Data Platform in a Smart System <i>Kaushik Kunal Das</i> Pivotal, Inc.	<i>FICO</i> Analytics Tools for the Area of Big Data
10:30am - 11:00am	<b>Refreshment Break with Exhibits - LL20 Foyer</b>			
11:10am-12:00pm	Real-world Big Data Applications <i>Jane Uyova</i> Teradata	Python, R, and SQL in MPP Databases <i>Anton J. Mobley</i> Kaiser Permanente	Big Data and Big Analytics – So Much more Gunpowder! <i>Paul Kent</i> SAS	<i>Frontline Systems, Inc.</i> Analytic Solver Platform: Integrated Data Mining, Simulation and Optimization
12:00pm-1:00pm	<b>Lunch and Roundtable Discussions - LL20BCD</b>			
1:10pm-2:00pm	Big Data Analytics Application to Jet Engine Diagnostics <i>Link Jaw</i> Intel Corporation	Big Data Solutions on AWS Platform <i>Vikram Garlapati</i> Amazon Web Services	Can Machine Learning Really Do That? <i>Anthony Goldbloom</i> Kaggle	<i>Enrich Consulting</i> Finding the Strategy Hidden in Your New Product Portfolio with the Enrich Analytics Platform
2:00pm-3:00pm	<b>Poster Session - Dessert Break with Exhibits and Posters - LL20 Foyer</b>			
3:00pm-3:50pm	A Case Study on Grid Data Analytics <i>Marina Thottan</i> Bell Laboratories	Babies, Brains and Buses... and Why Stream Computing is the Right Approach to Real-Time Predictions and Decision Making <i>Kevin Foster</i> IBM	<i>Simon Zhang</i> LinkedIn	

# San Jose Convention Center

Lower Level

