

# CONFERENCE ABSTRACT SUMMARY

It is my pleasure to invite you to participate in this year's premier single day conference on data warehousing, the **BIG DATA, DATA WAREHOUSING AND ENTERPRISE ARCHITECTURE CONFERENCE 2017** featuring **Bill Inmon**. This event is the 25th annual event sponsored by DAMA National Capital Region and Data Management Forum. Conference Day takes place **Wednesday, March 15, 2017, at the Crystal City Hilton Hotel**, 2399 Jefferson Davis Highway at Reagan Washington National Airport, minutes from the **Washington Mall, DC Metro, and Crystal City Underground Shopping Complex and Olde Town Alexandria**. Art Chantker, President of the Potomac Forum, will be moderating this event. A program summary can be downloaded in PDF format by going to website below and clicking on "Download Program Summary" (highlighted in yellow) <http://www.dmforum.org/>

This event is described by past attendees as one of this year's most anticipated special events on BIG DATA, Data Warehousing, Business Analytics and Enterprise Architecture. Bill Inmon, "The Father of Data Warehousing", will provide a Pre-Conference One-Day Seminar **(Tuesday, March 14) on "Architecting for Business Intelligence and Data Warehousing: Integrating the Structured and Unstructured Data World"** as well as provide the opening keynote to Conference Day -- **"The Role of Unstructured Data Within the BIG DATA Arena"**. **Mike Hermida**, Past Lead Architect on the Space Shuttle and SpaceLab Programs, will provide a Pre-Conference One-Day Seminar **(Monday, March 13) on "BIG DATA WAREHOUSING -- Data Modeling, Architecture and Design"** with a Special Section on Identifying, Evaluating and Selecting BIG DATA Software and Tools.

## Theme and Program

This year's theme is:

### **Award-Winning BIG DATA Programs – Best Practices in Technologies and Techniques**

The conference speakers are among the world's best known experts in the BIG DATA, data warehouse, enterprise architecture and information management arena, and represent the leading organizations in the field (Program Enclosed). The past few keynotes have included such luminaries as **John Zachman, Ralph Kimball and David Marco**. This year we are honored to have assembled the foremost speakers ever included on this subject.

## Conference Overview

Since the dawn of time up until 2003 mankind has created 2 exabytes of data, but today at least that much is created each and every day. This means that traditional databases and tools cannot process the huge volumes of data, nor can they discern and separate the "high value" from the "low value" data. The world of data management has evolved rapidly just over the last three years, into an ecosystem that is architected with traditional data, Big Data, real-time data, unstructured data, semantic interfaces, complex algorithms, new infra-structures, cloud computing, massively parallel computing, machine learning, data streaming, and many more technologies, all to serve the analytical needs of customers world-wide.

In the Federal Sector, the amazing speed with which this technology has progressed has led President Obama in March of 2012 to launch the Presidential Initiative on BIG DATA Research and Development, to capitalize of the fast growing data in today's world. By investing in over \$200 Million dollars seed money as an initial investment, federal agencies hope to extract more information and better analysis in order to make better decisions that drive growth and efficiency. The goals of this effort are:

- To advance the state-of-the-art core technologies needed to collect, store, manage, analyze and share huge quantities of data
- Harness these technologies to accelerate the pace of discovery in Science and Engineering enabling the United States to become more competitive
  - Expand the workforce needed to develop and use BIG DATA technologies
  - Place America back in the forefront of Science and Technology, while pioneering break-through innovations in every part of our life from the study of exoplanets and galaxy formations, imaging and genome mapping, weather and climate analysis, to national security encryption, customer sentiment and customer service, intelligent manufacturing, machine learning and energy development.

Coupled with analytics, BIG DATA can pinpoint waste in government spending, prioritize resources and help provide better service to citizens. There is also huge potential for benefits in the Defense and Intelligence area, Health Care and Finance arenas, Science and Technology. This conference will open a window to activities in both the private and public sectors regarding existing "use cases" and applications, current "real world" projects underway, as well as to future goals, challenges and expectations of what can be accomplished in tomorrow's enterprise architecture and ecosystem of technologies utilizing the capabilities of BIG DATA.

**Bill Inmon, featured keynote, President and Chief Technology Officer,**

Inmon Consulting, "The Father of Data Warehousing", speaking on

**THE ROLE OF UNSTRUCTURED DATA IN THE BIG DATA WORLD.**

This year's featured keynote, Bill Inmon, President and Chief Technology Officer, Inmon Consulting, has been best known as "The Father of Data Warehousing". In addition, he has been a leading authority on the "World of Unstructured Data", and How to Integrate with the traditional information and databases. In addition to authoring more than 50 books and 650 articles, Bill holds 7 software patents, and has been a monthly columnist with the Business Intelligence Network. As creator of the Government Information Factory and the Corporate Information Factory, Bill was named by Computerworld as one of the ten most influential people in the first 40 years of the computer profession. Bill will overview what he envisions to be the "Next Generation of Data Warehousing" and show how this technology is redefining itself. Bill will illustrate how breakthroughs in BIG DATA, Data Warehousing Methods, Tools and Approach are shaping Enterprise Architecture within the major corporations, governments and businesses on a world-wide basis, while citing from practical applications developed throughout the industry. Bill has been developing a blueprint for future data warehousing design and development. Within this framework, Bill describes how some important new architectural features are intertwined – the life cycle of data, the need to connect unstructured data with structured data, the need to tightly intertwine metadata with the warehouse itself, the need for an enterprise metadata repository – all these features are vitally important for the next generation of data warehousing. Perhaps the most important is the capability to discern what kinds of information must be kept or archived for BIG DATA Analytics, and what kinds of data can be stored for short term or discarded.

The Focus of the Event will be on some of the most famous organizations and their Success Stories in BIG DATA Analytics, Data Warehousing and Business Intelligence Data Transmission, Data Mining, Data Exploration and Discovery. Featured User Experience Speakers include the following:

**Peter Lyster, Deputy Director, NCO, NITRD, Executive Office of the President  
on THE PRESIDENTIAL INITIATIVE ON BIG DATA RESEARCH AND DEVELOPMENT**

will be our featured BIG DATA in Government Presentation.

The Obama Administration in March 2012 announced a "Big Data Research and Development Initiative." By improving our ability to extract knowledge and insights from large and complex collections of digital data, the initiative promises to help solve some of the Nation's most pressing challenges.

To launch the initiative, six Federal departments and agencies announced more than \$200 million in new commitments that, together, promise to greatly improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data, and this expenditure has only been upgraded in recent months with the announcement of added funding for this critical program."In the same way that past Federal investments in information-technology R&D led to dramatic advances in supercomputing and the creation of the Internet, the initiative we are launching today promises to transform our ability to use Big Data for scientific discovery, environmental and biomedical research, education, and national security," said Dr. John P. Holdren, Assistant to the President and Director of the White House Office of Science and Technology Policy (OSTP).

In order to make the most of this effort, The White House Office of Science and Technology in concert with six federal agencies launched this program to make the United States more competitive in the world marketplace by exploring and exploiting the benefits of BIG DATA. The National Science Foundation (NSF) and National Institutes of Health (NIH) took a lead role in conjunction with the Defense Advanced Research Projects Agency (DARPA). Chaitan Baru, who has done enormous research at University of California San Diego Supercomputing Center, has accepted a chair as Computer Information Science and Engineering (CISE) at NSF. He will overview the background and purpose of the BIG DATA INITIATIVE, identify the major agencies and projects underway throughout the federal government, and describe the challenges and accomplishments. He will discuss efforts in Educational Research, Biomedical Research, Defense Intelligence, and Scientific Discovery. Projects such as UC Berkeley Expeditions in Computing, DARPA's XDATA, NIH 1000 Human Genomes Mapping Project (Amazon Web Service and Cloud Computing), Dept of Energy's Effort to create the Scalable Data Management Analysis and Visualization (SDAV) Institute, led by the Lawrence Berkeley National Laboratory, USGS's Earth System Science and the ambitious DoD Warfarer Project to improve situational awareness to help warfighters and analysts and provide increased support to operations in defense of the United States. "Earthcube" – a system that will allow geoscientists to access, analyze and share information about our planet across an integrated network and Big Data database, and XDATA, a project which intends to invest approximately \$25 million annually for four years to develop computational techniques and software tools for analyzing large volumes of data, both semi-structured (e.g., tabular, relational, categorical, meta-data) and unstructured (e.g., text documents, message traffic). The BIG DATA Initiative is truly an ambitious project that will set the tone for BIG DATA Analytics well into the future.

**For the featured Government Executive, we are honored to have Dr. John Sprague, Deputy Chief Technology Officer, NASA Office of the CIO, NASA Headquarters in Washington, D.C. as our speaker. With control over the technology projects and budgeting effort for all programs inside NASA, John Sprague will speak on:**

**NASA'S STORY OF BIG DATA INNOVATIONS AND APPLICATIONS:  
USING BIG DATA ANALYTICS, TECHNOLOGY AND ARCHITECTURE**

In the time it has taken you to read this sentence, NASA has gathered about 1.73 gigabytes of data from nearly 100 currently active missions, and the collection rate is growing exponentially. Handling, Storing, and Managing this data is a massive challenge, and NASA is preparing to use optical laser communication technology that will transmit and stream more than 24 terabytes a day in support of their missions, that is 2.5 times the entire Library of Congress, each and every day!

John will discuss the challenges and opportunities of dealing with an array of Big Data Management issues focusing on Large Volume Databases and Data Warehousing, the integration and data sharing of structured and unstructured data, the telemetry of data communications over long distances, and the Success Stories so critical to the nation's un-manned space flight program which he oversees. Data Mining and Business Intelligence have become key instruments for managing information at NASA projects, and across the organization at NASA. Working closely with Dr. Scott Bernard as part of the Federal Enterprise Architecture, John Sprague has been instrumental in overseeing the development of HADOOP and other key technologies and advanced projects working with DARPA.

BIG DATA has played a more prominent role in activities at NASA than ever before, and John will discuss key projects that figure into these technologies such as Hubble, Earth Observatory, Solar and Deep Space Programs. Missions such as Climate Change Data Repository will dramatically impact weather forecasting, Australian Square Kilometer Array Pathfinder (ASKAP) will unlock mysteries of our universe in the search for forms of life, Mission Data Processing and Control System (MPCS) supported the Curiosity Rover on Mars interfacing with NASA's Deep-Space Network to relay data in real-time, which once had taken days to accomplish. NASA's Pleiades Supercomputer is used to help analyze data from Kepler spacecraft leading to the discovery of new Earth-sized planets (exoplanets) in the Milky Way Galaxy. NASA is also involved in "non-Space" like Airline Safety in order to help commercial airlines improve maintenance procedures and prevent equipment failures while in flight, as well as map the terrain for missing airliners as well as for weather and geological research on energy findings

The Open Government Plan outlines numerous ways to drive innovations in technology around big data, starting with data.nasa.gov -- a directory where NASA makes internal data available externally for inter-agency and third party use. Covering everything from Surveying Polar Ice and National Climate Assessment to using TIKA to detect and extract MetaData from unstructured text, John will discuss the future plans and strategies that have made NASA one of the world's most respected organizations for technical accomplishment.

**Krish Krishnan, one of the world's most knowledgeable experts on BIG DATA TECHNOLOGIES, will present on DATA WAREHOUSING IN THE AGE OF BIG DATA.**

Krish Krishnan is a recognized expert worldwide in the strategy, architecture and implementation of high performance data warehousing solutions and unstructured Data. A sought after visionary data warehouse thought leader and practitioner, he is ranked as one of the top strategy and architecture consultants in the world in this subject. Krish is also an independent analyst, and a speaker at various conferences around the world on Big Data and teaches at TDWI on this subject. Krish along with other experts is helping drive the industry maturity on the next generation of data warehousing, focusing on Big Data, Semantic Technologies, Crowdsourcing, Analytics, and Platform Engineering. A very popular lecturer worldwide, Krish also writes a blog on the IBM HUB for Data Management Forum and has just completed a series of articles on the Internet of Things (IOT). Author of "Building the Unstructured Data Warehouse" with Bill Inmon ("The father of data warehousing"), and more recently "Data Warehousing in the Age of Big Data" (Morgan Kaufman), Krish has recently completed "Social Data Analytics: Collaboration for the Enterprise" with Shawn Rogers.

The biggest phenomena that has captured the attention of the modern computing industry since the "internet" is Big Data. Technology platforms that have emerged along with Big Data are providing the capability to process information in multiple formats and varying structures without the constraints of traditional databases. Data can be processed in extremely high volumes, generated at various velocities, and in varying degrees of ambiguity, more than ever before. New platforms like Hadoop and NoSql significantly lower the cost and increase the scalability and flexibility. Typical forms of data that are stored in Big Data databases span everything from Financial Reporting and Performance data, Compliance such as in the Healthcare and Government sectors, to Contract and Textual Data, Weather and Geospatial Data, Machine-Generated Sensor Data, data generated by Social Media like Linked In, Face book and Yahoo, and Customer Sentiment information such as required by Sales and marketing. The Variety, Velocity and Volumes of data needed to satisfy business needs keeps growing by "leaps and bounds".

Krish Krishnan will not only define what big data is, how it's used, and typical applications and industries that turn to Big Data Technologies to solve their problems, he will cover some of the most popular big data tools. databases, reporting and analytics software. Krish will review products like Pig, Hive, Map Reduce, YARN, HDFS, HBase, HCatalog, Zookeeper and Sqoop. He will discuss why Hadoop has so many versions in the marketplace, and why it has been a "go to" database in the big data market space. Typical Use Cases for click stream databases, sensor-driven data (e.g. airline engines), contract analysis (e.g.doctor notes). GPS (e.g.onstar), and Securities Tracking and Surveillance, are just a few examples. He will show how critical insights can be discovered from the proper analysis of Big Data, and how it will "add value" to an organization, in terms of both increasing earnings, reducing cost, and adding better control.

**Juan Gorricho, WALT DISNEY WORLD STORY: USING ENTERPRISE WAREHOUSING AND ANALYTICS TO ACHIEVE BUSINESS STRATEGIES**

Juan will discuss some of the most successful business programs and strategies to increase market share and customer loyalty, improve business performance and expansion. As part of a tremendous strategy called Destination Disney, Walt Disney World will be used as a test bed for one of Americas most ambitious programs for business use of IT convergence, the combination of global positioning satellites, smart sensors, wireless technology, mobile devices, including one that looks like Mickey Mouse himself (Pal Mickey) to reinvent the customer experience, influence visitor behavior and ease crowding throughout the parks.

Data warehousing and business intelligence (DW/BI) has a relatively short history at Disney Parks and Resorts. One of the key things to understand is the structure of the company overall and the role IT and lines of business have played in developing DW/BI. The purpose of this presentation is to share the evolution of DW/BI at Disney P&R, some of the best practices and things we have learned over time, and a vision of how the future might look in terms of where we see DW/BI going.

He will provide an overview of some of the DW/BI/analytics initiatives that have been put in place thus far in the company, including revenue management and customer relationship management. This will serve to show and hopefully open a discussion with other companies around the role of IT and the business in DW/BI initiatives. Particularly, to understand what prompts the business to continue to pursue shadow IT DW/BI initiatives on their own. According to Juan Gorricho, the role of IT and Data Warehousing is changing. it is not simply an organization that deploys technology, but one that now integrates technology from a lot of different angles to improve the customer experience. Data Warehousing technologies used with high intensity business intelligence and analytics is at the heart of this effort.

Juan F. Gorricho has lead the delivery of business intelligence and analytical solutions within the Information Technology division of Walt Disney Theme Parks & Resorts. More recently, Juan successfully led the design and implementation of an \$8 MM business intelligence and data warehousing solution for the Walt Disney Theme Parks and Resorts Merchandise line of business as part of a \$40+ MM supply chain and inventory management system implementation. This included end to end design of the solution, delivery of operational, tactical and strategic reports, as well as the foundation to support advanced analytics business processes, key for the merchandise line of business. Juan now is VP and Chief Analytics Officer for the Partners Federal Credit Union Division of Walt Disney Parks and Resorts. Juan's educational background includes a BSc in Industrial Engineering from Universidad de los Andes in Bogota, Colombia, and a MBA from the Darden Graduate School of Business Administration at the University of Virginia. Juan is married and lives in Orlando, Florida.

**THE CHIEF DATA OFFICER (CDO) AGENDA: INTEGRATING BIG DATA WITH THE EXISTING DATA ARCHITECTURE will be presented by Dr. Peter Aiken, Founding Director of Data Blueprint, Professor at Virginia Commonwealth University and Past President of DAMA International**

Dr. Peter Aiken, an Award-Winning Presenter who has consistently won Best Presenter at DAMA International, will discuss the planning, requirements and monetization of BIG DATA Projects, in order to show how a Big Data Plan should be assembled and monitored to ensure project success. He will examine a series of Business Cases for Big Data Projects, and outline how the project payback can be demonstrated in programs that identify results of Big Data Analytic insights. The business value and metrics of a proposed project must be properly analyzed and developed in order to show to management that there will be a Big Payback for expenditures on the tools, databases and data collection efforts involved. Peter will discuss the role of the Chief Data Officer (CDO) in bringing in the Big Data Technologies that can add value for an organization's business goals and strategies. Peter will cover the myriad of "Use Cases" that can be undertaken using BIG DATA tools and techniques. He will also define which projects are candidates for Big Data databases and reporting and analytical tools. Big Data is very simply a collection of data sets so large and complex that your legacy IT systems cannot handle them, store their data or process efficiently. When organizations get to the point where their volume, velocity, variety, or veracity of data exceed storage capacity, there are some big challenges to be addressed. You know you have a big data challenge when your traditional data management systems and analysis tools are overwhelmed and it becomes too difficult to process the information. Today's Big Data discussion is often centered around how to target advertisements or customize a user experience, which makes sense given the growth of the marketplace is so closely tied to how we interact with customers, and is more and more dependent on use of mobile devices that connect to the work through sensors.

Big Data is focused on meeting mission requirements for actionable information to perform business functions in a more efficient or effective way. Peter will cover the architecture, technology and business cases that are typical in a large organization, and show how a CDO can truly add monetary value to an organization. Most organizations have difficulty integrating their current information pool with the Big Data pool. This presentation will focus on BIG DATA Use Cases and projects that utilize BIG DATA and how data management best practices have supported those efforts.

**Pre-Conference Day - Bill Inmon Seminar (Tuesday, March 14, 2017):  
Architecting For Business Intelligence and Data Warehousing --- Integrating the Structured and Unstructured Data World**

Bill Inmon provides an enlightening view into the latest technologies and methodologies for capturing and processing in the world of unstructured data, one of the major components of a BIG DATA architecture . This seminar is the definitive tutorial guide that will overview how to integrate both structured and unstructured information that permeates throughout every organization. It will show in a simple straight forward fashion how to tap into the rich sources of data that need to be managed in order to truly capture your enterprise data assets, from one of the World's Leading Information Management Experts. In order to architect your organization to optimize your data warehousing and business intelligence capabilities, you need to capture and organize all your enterprise information. A wealth of invaluable information exists in unstructured textual form, but organizations have found it difficult or impossible to access and utilize it. This is changing rapidly: new approaches finally make it possible to glean useful knowledge from virtually any collection of unstructured data.

William H. Inmon — the father of data warehousing -- introduces the next data revolution: unstructured data management. Bill Inmon will cover all you need to know to make unstructured data work for your organization. You'll learn how to bring it into your existing structured data environment, leverage existing analytical infrastructure, and implement textual analytic processing technologies to solve new problems and uncover new opportunities. Bill Inmon will introduce breakthrough techniques covered in no other seminar, including the powerful role of textual integration, new ways to integrate textual data into data warehouses, and new SQL techniques for reading and analyzing text, whether in emails or spreadsheets, WORD documents or Communities of Interest (COI) web data. Bill will also present real-world case studies that are indispensable to every organization trying to make sense of a large body of unstructured text: Program Managers, executives, project leaders, data warehouse Managers and staff, architects, consultants, database designers, data modelers, DBAs, researchers, and end users alike will all benefit from this one day comprehensive overview on the subject.

Break-throughs in BIG DATA technologies and techniques for managing and bridging the world of unstructured, semi-structured and structured databases can pay huge dividends for organizations, and Bill will share success stories from efforts from across the globe describing how leading companies and governments can gain an edge in the marketplace.

**Pre-Conference Day - Mike Hermida Seminar (Monday, March 13, 2017):**

**Big Data Warehousing Architecture, Modeling and Design --- A One-Day Tutorial on the optimal design techniques and architectural frameworks for building BIG DATA WAREHOUSES with new section on BIG DATA Architecture, Software Tools Evaluation and Typical Use Cases.**

Michael Hermida, very popular seminar leader, provides a fascinating overview of this topic. Data Warehousing has taken hold throughout the business world and is providing great insight into business performance, and timely information to allow decision makers to enhance this performance. There are similarities between designing databases for operational systems and data warehouses, but there are also significant differences. Organizations have skilled technology personnel that have been modeling data and building databases for years in support of the business. The break-throughs in BIG DATA Databases and Computing have only increased the potential for challenges ahead. How can the current skills of these technologists be leveraged to produce an effective data warehouse? What are the differences between data modeling for a production environment vs. a data warehousing environment or BIG DATA environment? How should historical data be maintained in the data warehouse? What role do ER models and dimensional models play in data warehouse development? Are snowflake schemas appropriate? What database design options facilitate end user access?

What is the role of BIG DATA in Data Warehousing and Data Management? This tutorial highlights all aspects of Data Warehousing and Data Marts from Architecture, Design, Data Modeling to Warehouse Management Roles and Responsibilities, Data Accessing and Reporting. There is also coverage on the New Software Tools and Databases utilized in the BIG DATA WORLD. The Methodology for Selecting, Analyzing and Recommending Big Data Databases, ETL Tools, Reporting Tools, Analytical Tools and MetaData Extractors is discussed with usage of the Gartner Magic Quadrant and Forrester Wave Techniques. This one day seminar will address these questions and others related to designing databases for your data warehouse and assembling the proper Big Data Architecture that is "right" for your organization. A Data Warehouse Glossary will be presented by Mike Hermida to overview the Terms and Definitions pertinent to any Data Warehousing Environment.

Mike Hermida held various technical and management positions during a sixteen-year career with IBM. He designed and developed operating systems and access methods for the nation's en-route **air traffic control system** and ground-based **Space Shuttle programs**, and was a lead architect for the **Spacelab's command and control system**. He was also a technical member of the field test team for the rollout of **IBM's DB2 database management system**. Within IBM he managed a staff responsible for the development of tools and utilities to enhance the DB2 development environment, and also managed a **Data Resource Management (DRM)** department of data modelers, DBAs, and performance analysts supporting relational and other database systems.

On leaving IBM, Mike, with a partner, built and subsequently sold a successful consulting and education firm specializing in relational database and applications design and development. Aside from being the firm's Chief Technology Officer, he was personally responsible for providing technical and management consulting services to key corporate clients. He started the **Knauer DB2 User Group in New York City** that grew to over one thousand members, and developed and implemented a set of DB2 value added services to support a national DB2 practice.

## **OVERVIEW**

The BIG DATA and Enterprise Architecture Conference is produced and managed by the Data Management Forum in cooperation with DAMA-National Capital Region. This event will attract Data Warehouse Directors, Program Managers, IT Directors, Directors of MIS and Data Warehousing, Executive Management, CIO's, CEO's, CTO's, Chief Financial Officers, Chief Architects, Data Management Experts, Data Governance and Repository Managers, VP of Architecture, CIO Council, VP's of Database and Planning Directors, Assistant Deputy Directors, Data Scientist, Chief Data Officers, Data Custodians, Data Designer or Data Modeler and any Application Developer, Business Analyst, Web Developer or Systems Analyst interested in acquiring an overview of Big Data, Analytics, Data Warehousing or related technologies and techniques.

### **How to Register**

Please go to <http://www.DMForum.org/> and Click on the button "(TO REGISTER)" or Email [registration@dmforum.org](mailto:registration@dmforum.org) or Call 516-221-5560.