

# Big Data

Thought leaders discuss government efforts to harness their data assets.

## Inside

- 3 Big Data Boosts Baseball!
- 4 The Big Data Issue Is Big Data
- 6 Big Data Workshop At NIST
- 7 Making The Big Data Journey
- 8 108 Terabytes In 60 Seconds
- 10 The New Data Wave Has Arrived
- 12 Connecting Cross Domain
- 14 The Art Of The Algorithm
- 16 Imagine The Innovation
- 18 From Milliseconds To Microseconds





# SECURING YOUR BIG DATA

## CYBER SECURITY AT THE SPEED OF INNOVATION

- Rapidly index, correlate and store large volumes of network traffic.
- Perform internet searches across stored data sets.
- Conduct data forensics, to include both temporal and link analysis-style data analytics, to determine root-cause and correlated data.
- Assist with Sarbanes-Oxley compliance and other corporate requirements by having all network information stored, indexed and rapidly retrievable.

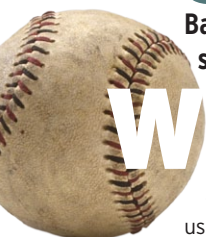
TO CONTACT A MERLIN  
SALES REPRESENTATIVE:  
[sales@merlin-intl.com](mailto:sales@merlin-intl.com)  
T 1.877.430.3021  
[www.merlin-intl.com](http://www.merlin-intl.com)



## It Occurs To Me...

By Jeff Erlichman, Editor,  
*On The FrontLines*

# Big Data Boosts Baseball!



**Baseball is benefiting from the power Big Data analytics provides. Now government is taking bold steps to do the same.**

**W**hen I was growing up the important stats were batting average, home runs, RBIs and a pitcher's ERA. It wasn't that complicated to know who was good. Managers (and fans) trusted their eyes. Grizzled scouts used their gut feelings.

Not today. It's WHIP and OPS. In Baseball, Big Data sabre metrics share the stage equally with trained eyes and grizzled guts. No self-respecting general manager would dare make any multi-million dollar player decision based purely on what they see or their gut feeling.

Today, every pitch and every play provides raw data that is sliced, diced and analyzed so defenses know exactly where to play; and batters know the pitcher he's facing throws a first pitch fast ball 90% of the time.

That's Big Data analytics doing its job in Baseball. The power of predictive analytics is there for all to see and benefit from. (As a Washington Nationals fan, I'm particularly glad.)

Now government is looking to benefit from the power of Big Data. That's why from census data to geospatial data to health care data to financial data to intelligence data, government is taking proactive steps to harness the power of Big Data the way Baseball already has.

The Obama Administration's first pitch is the March 2012 Big Data Initiative which is investing \$200 million on research (More on page 6).

Recent MeriTalk research on Big Data clearly demonstrates we are just in the 1<sup>st</sup> inning. Step 1 is Big Data education. Many don't even know the data they have is Big Data; and if they do, they don't know what to do with it; or are just learning how to monetize their data assets.

## The Three — Make That Four — “V”s



Dr. Michael Stonebraker  
MIT

I attended the June 2012 Big Data Workshop, sponsored by NIST (National Institutes of Standards and Technology). Thought leaders from the Big Data community gathered to share the current status, ongoing programs and future plans.

Renowned MIT database expert Dr. Michael Stonebraker said Big Data has three characteristics. He called them the three

“V”s — Big Volume, Big Velocity and Big Variety.

He said organizations are faced with data volumes growing faster than they can analyze it. “They are doing a good job with ‘small analytics’, but have problems with ‘big analytics’.” At the

same time, this information is pouring in with great velocity, “a fire hose of data” said Stonebraker.

Finally, challenges of dealing with a variety of data formats from a variety of data sources are not being met by traditional products according to Stonebraker.

“I would add a fourth V - Verification,” Kirk Borne, Professor of Astrophysics and Computational Science at George Mason University told Workshop attendees.

“Verifying inference based models from data and looking for algorithms that transform data into knowledge is what we are after.”

Borne spoke about Big Data in terms of dealing in “tonnabytes” whether it be Gigabyte, Petabyte or Zettabyte or Exabyte.

“We are looking for ways to extract relationships and patterns. How do we discover what the representations of the Big Data should be?”

With all enterprises being inundated with data, the knowledge discovery potential is enormous said Borne. Harnessing this is especially important for training and degree programs for the next generation of workers.

“Now is the time to implement data oriented information methodologies into the enterprise,” he urged, “and to address the four Big Data challenges from our “tonnabytes” data collections — volume, variety, velocity and verification.”

In the following pages, you can read how thought leaders from government, academia and industry are on their way to doing just that. ■



Kirk Borne  
Professor of Astrophysics  
and Computational Science  
at George Mason University

## Introducing the OTFL SME One-On-One

In this issue of *On The FrontLines*, we are launching a new series of articles under the title: OTFL SME One-On-One.

Each OTFL SME One-On-One is based on an interview with a Subject Matter Expert from a leading private sector provider.

Each SME provides practical information and “their best advice” on what government can do to further its goals and drive innovation.

Think of it as if you are having lunch at a conference, you are sitting next to an SME and here is a chance to learn from their experience.

We also would be interested in your feedback. Email me at [jefferlichman@onthefrontlines.net](mailto:jefferlichman@onthefrontlines.net).

— Jeff Erlichman

# The Big Data Issue Is Big Data

Can you guess how many grains of sand (bytes) are on this beach? Then multiply that number by 8 to get the number of bits. With numbers like these it's easy to see why the big data issue with Big Data is big data.

**I**f a byte of data was a grain of sand, a Petabyte (PB) of data would be all the sand in a beach a hundred yards wide, one foot deep and a mile long," explained Nick Combs of EMC Isilon during a recent Federal Executive Forum.

That is a 100 yard wide, 1 foot deep, 1 mile long beach.

An Exabyte of data is all the sand in a beach the entire length of the coast of the United States said Combs. "Then if you think about networks to support it you can multiply that times 8 because there are 8 bits in a byte."

## AMPing Up

"The total data created in 2010 was 1 Zettabyte. That's 1,000,000 Petabytes," said Dr. Michael Franklin of UC Berkeley's AMPLab during the NIST Big Data Workshop. Amazingly, or perhaps not, Big Data volume is growing more than 60% per year. Big is clearly getting bigger.



Michael Franklin  
UC Berkeley's AMPLab

AMPLab is a five-year collaborative effort at UC Berkeley; funded in part by the National Science Foundation as part of the Big Data Initiative.

AMP stands for "Algorithms, Machines, and People". Franklin said AMPLab envisions a world where massive data, cloud computing, communication and people resources can be continually, flexibly and dynamically be brought to bear on a range of hard problems.

Franklin defined Big Data as data that is expensive to manage and hard to extract value from.

"You don't need to be a big company to have a Big Data problem," Franklin declared, "You may just have inadequate tools to analyze the data you have."

Meeting these challenges will require an entirely new approach that transcends and reshapes disciplinary boundaries Franklin said.

**4 Big Data**

## At NASA: Big Questions, Big Answers

Government technologists like NASA Goddard CIO Adrian Gardner are the ones having looking for new approaches to deal with the practicality of Big Data's four "V"s.

During a recent interview with *On The FrontLines*, I asked Gardner to imagine he was attending a conference and sitting next to a Big Data supplier during lunch. What questions would he ask?

"My first question would be 'do you have a past performance in areas of the sizes of the kinds of data that we process on a daily and monthly and annual basis?'" For Gardner it is all about scale: "Can they scale to the limits of our Big Data problem, because we have an unusual problem?"



Adrian Gardner  
CIO  
NASA Goddard

"It is not necessarily having knowledge of exactly our Big Data problem," Gardner said, "but having some past performance in the area of managing Big Data at our current and future volumes."

He said the second thing he would ask about is "the notion of culture and being familiar with our scientific culture and what experience do they have in that area? It is an environment in which we collaborate across geographical boundaries and you have to get that. That's important."

The last thing Gardner would ask for are examples of real world proof of concepts. "Are there really proofs of concept we could look at and experience and then be able to draw parallels to some of the challenges and opportunities that we have here?"

## At NSA: The 64 Zettabyte Question

Neal Ziring is the technical director for the National Security Agency's Information Assurance Directorate (IAD), serving as technical advisor to the IAD Director and senior leadership. He is

responsible for setting the technical direction across the Information Assurance space.

From his vantage point, Ziring told the NIST Workshop audience about the clear advantages of having the tools with the speed to extract knowledge and actionable information from mountains of raw data. For Ziring the 64 Zettabyte question is: What challenges must be overcome?

He said the big challenge for Big Data is providing the information his consumers want. "We need to use Big Data to find more and better knowledge in order to make mission sense out of complex adversary behaviors."

He enumerated the challenges.

1. Fusing data from multiple sources and domains in order to discover not obvious relationships and bring data together that hadn't been together before. He also said we "still need humans to set up the parameters. This is essential for mission agility."

2. Spanning time scales because data comes in in intervals over time and there is a latency factor that makes calculating analytics difficult. He said "much of this needs to be automated and we need more cohesive approaches to the problem."

3. Creating actionable knowledge from data that is constricted in certain ways. "We have data but can't use it subject to conditions. We need better access controls and conditions that are computation friendly and make it easier for the analyst. The Intel community is trying to standardize some of the simplest aspects of this. This would benefit from standardization."

He said a side challenge is the "data pedigree, being able to track where the data comes from. "Can you verify that, especially under questioning from Congress?"

4. Enabling analytic science. "We need better tools for exploratory analysis and to support analytic multi-tenancy. We are doing that in the cloud by keeping everybody separate, that's not good for Big Data. We need to share resources; and people have different policies."

5. Improvements in indexing and algorithms for long term Big Data. "While Big Data is new, it is not going away. We need incremental algorithms to deal with things that come up in the short term."

Ziring also noted that "however big "Big Data" installations get, they will never get big enough. We need multi-site analytic computational capabilities. You just can't move a trillion objects or buy bigger pipes. We need new algorithms."

Ziring urged the Big Data community gathered at NIST to fulfill the enormous potential benefit for all, calling for government, industry and academia to come together to make progress on these challenges.

"We focus on national security, but all can benefit," said Ziring. "We need to solve some of these challenges for everyone to get full benefits from Big Data." ■



Neal Ziring  
National Security Agency's  
Information Assurance  
Directorate (IAD)

## Big On Big Data!

Government leaders are big on Big Data. Strategies for implementing Big Data solutions was one topic discussed during the recent Federal Executive Forum on Big Data.

View all these videos and more at the Trezza Media Group YouTube Channel at <http://www.youtube.com/user/Trezza-mediagroup>.

### US Army

#### Alan Hansen

Director, Intelligence Systems  
& Processing Division,  
US Army RDECOM



"We have multiple data models being used across the battlefield... we realize there is no one size fits all data model...and we are looking to serve information in the adaptive manner our users need..."

### National Geospatial-Intelligence Agency (NGA)

#### Dr. Robert Norris

Chief Architect



"A major Intelligence community project is IC efficiency...we are coming up with two clouds where we can relate information that had been in stovepipes..."

### National Oceanic and Atmospheric Administration (NOAA)

#### Zach Goldstein

Deputy CIO



"Strategies begin with planning...how do we handle telecommunications and processing...we have a plan in place to make sure we have investments in telecom to make sure we can do weather forecasts..."

### NASA

#### Adrian Gardner

CIO, NASA Goddard



"We have gravitated to a community of practice, where we have formally pulled together scientists and engineers to talk about Big Data...also we are talking about standards to address interoperability..."



## Fed Big Data Initiative Invests \$200 Million In New R&D

The Obama Administration sees big value in harnessing Big Data. It is investing \$200 million in its 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 the Nation's most pressing challenges," said Dr. John P. Holdren, Assistant to the President and Director of the White House Office of Science and Technology Policy (OSTP) when announcing the initiative.

Six departments and agencies have committed more than \$200 million "to greatly improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data," he said.

Participating are:

- **National Science Foundation and the National Institutes of Health:** Core Techniques and Technologies for Advancing Big Data Science & Engineering research focused on managing, analyzing, visualizing, and extracting useful information from large and diverse data sets.

NIH is particularly interested in imaging, molecular, cellular, electrophysiological, chemical, behavioral, epidemiological, clinical, and other data sets related to health and disease.

Learn more: [http://www.nsf.gov/news/news\\_summ.jsp?cntn\\_id=123607](http://www.nsf.gov/news/news_summ.jsp?cntn_id=123607)

- **Department of Defense:** Data to Decisions — The Department is investing approximately \$250 million annually (with \$60 million available for new research projects) across the Military Departments in a series of programs.

Learn more: [www.DefenseInnovationMarketplace.mil](http://www.DefenseInnovationMarketplace.mil) and <http://www.darpa.mil/NewsEvents/Releases/2012/03/29.aspx>

- **National Institutes of Health:** 1000 Genomes Project Data Available on Cloud — The National Institutes of Health is announcing that the world's largest set of data on human genetic variation — produced by the international 1000 Genomes Project — is now freely available on the Amazon Web Services (AWS) cloud.

Learn more: <http://www.nih.gov/news/health/mar2012/nh-gri-29.htm>

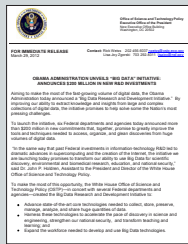
- **Department of Energy:** Scientific Discovery Through Advanced Computing — The Department of Energy will provide \$25 million in funding to establish the Scalable Data Management, Analysis and Visualization (SDAV) Institute in the Energy Department's Lawrence Berkeley National Laboratory.

Learn more: <http://science.energy.gov/news/>

- **US Geological Survey:** Big Data for Earth System Science — Big Data projects will improve understanding of issues such as species response to climate change, earthquake recurrence rates, and the next generation of ecological indicators.

Learn more: <http://powellcenter.usgs.gov> ■

6 Big Data



# NIST Hosts 300 At Workshop

The Big Data community gathered at the National Institutes of Standards (NIST) June 13-14 to discuss how to best benefit from the Obama Administration's Big Data Initiative \$200 million investment in America's future.

Presented by NIST and the National Science Foundation (NSF) Center for Hybrid Multicore Productivity Research (CHMPR), the Workshop brought together over 300 participants from academia, industry and other federal agencies.

**Dr. Ashit Talukder, Ph.D.,** Chief of NIST's Information Access Division told OTFL in an interview after the Workshop it "was a tremendous success."

"The audience liked the fact that the challenges and solutions on the multiple aspects of Big Data were presented and discussed at the Workshop," said Talukder.

Participants came to discuss state-of-the-art core technologies needed to collect, store, preserve, manage, analyze, and share Big Data that could benefit from standardization. Also to discuss potential measurements to make sure Big Data is delivering on what it is supposed to do.

"The panel discussions were very well received," said Talukder. "The presentation of various use-cases in Big Data and discussions about potential gaps and challenges for Big Data going forward raised various questions and helped the audience gain insights into the potential next advances in the field of Big Data."



Dr. Ashit Talukder, Ph.D.,  
Chief of NIST's  
Information Access  
Division

## Next Steps

In the aftermath of the Workshop, "NIST is considering various ways that we may be able to play a role relative to big data issues," explained Talukder.

"As is the case in many areas that NIST works we are considering ways we might work with industry, academia, and other agencies to foster the development of measurements, standards, guidance, and tools that could enable a uniform Big Data landscape."

# Making The Big Data Journey

Big Data research from MeriTalk reveals government is just at the beginning of its Big Data journey.

**IT** professionals report, on average, it will take their agency at least 3 years to fully take advantage of Big Data according to a May 2012 MeriTalk research study titled the **Big Data Gap**.



The research aims to shed light on how agencies are positioned to use Big Data to support their agencies. According to Meritalk, "the study highlights the Big Data opportunity and outlines the gap between Big Data possibility and Big Data reality for most agencies.

Respondents were 151 federal IT professionals; 50% civilian, 50% DOD; and 100% are familiar with their agency's data capture, management, and analysis capabilities.

## Work To Be Done

The first good news from the survey was 91% of civilian and 69% of DOD respondents said their agency is discussing Big Data. The bad news is the number is not 100%. Wake up people!

There is more good news on engagement: 60% of civilian respondents said they are learning about it; the number is 42% on the Defense side.

The not so good news is that only 6% of civilian respondents said they have the infrastructure/processes in place to successfully leverage Big Data; it's even less — only 3% for DOD.

When asked about the top three advantages of successfully managing Big Data, 59% pointed to improving overall agency efficiency, 51% to improving speed/accuracy of decisions and 30% the ability to forecast.

"Yet most of that promising progress is locked away in unused or inaccessible data," asserts the research.

"Just 60% of IT professionals say their agency is analyzing the data they collect and less than half (40%) are using their data to make strategic decisions.

Whether it is an opportunity or a challenge, data continues to grow: 87% of IT professionals say their stored data has grown in the last two years; 96% expect their data to grow in the next two years (by an average of 64%).

## Big Data Gap

This all adds up to the "Big Data Gap". The research says ownership of data is an issue. The amount of unstructured data that must be dealt with is growing and computing power and storage aren't yet up to handling Big Data.

"Agencies estimate they have just 49% of the data storage/access; 46% of the computational power; and 44% of the personnel they need to leverage big data and drive mission results," according to the MeriTalk research.

When asked what steps an agency can take to optimize data:

- **39%** — Invest in IT infrastructure to optimize data storage
- **33%** — Train IT professionals to manage/analyze big data
- **31%** — Improve the security of stored data
- **28%** — Educate senior management on big data issues
- **25%** — Solicit contractor support for big data initiatives
- **25%** — Invest in IT systems/solutions to optimize data processing

Click here to get your copy of the research. ■

## What You Can Do

The research offers federal technologists four recommendations for Big Data success:

- Tackle the management issue first: determine which department owns big data
- Start preparing for new solutions and approaches to unstructured data
- Look down the road and make decisions today that work with using big data tomorrow
- Expand your data management plan; more data is coming.

He said NIST will be promoting:

- Standards, interoperability of big data systems and database storage, security, networking.
- Common APIs for interoperable big data access.
- Metrics, evaluations and reference datasets for big data algorithms, analytics, search and retrieval.
- Guidelines, best practices and standards for Usability, Human Factors, Visualization and Personalization in big data.
- Reference architectures for big data systems, platforms and middleware, software and algorithms.
- Software-based measurements — metrology to determine

accuracy/uncertainty of applying processing and analysis steps to particular domains.

- Data models, information models, domain models.
- Security and privacy guidelines for big data systems.
- Metrics for measuring uncertainty and data quality.
- Virtual measurements, modeling and simulations of Big Data.

Click here to learn more about individual sessions. ■

**Paul G. Rutherford**  
Chief Technology Officer  
EMC/Isilon

# 108 Terabytes In 60 Seconds

With EMC/Isilon's high performing single pool of storage, you can handle your Big Data needs, while simplifying management and protecting your data.



It could be a couple or a couple of hundred Terabytes; or maybe a Petabyte; soon it could be a Zettabyte. It just doesn't matter.

"Someone has to deal with this Big Data," said **Paul Rutherford**, CTO at EMC/Isilon in a recent one-on-one with OTFL.

"Big Data is one of those things easily defined as being an amount of data or an expected amount of data that far exceeds your current capabilities as an IT professional."

"That's the way I like to define it," he said. "It's a rapidly changing, unexpected or unanticipated amount of data that is greater than the infrastructure currently supports."

Rutherford described himself as a product CTO for a company that develops products, one who spends a lot of time in the field talking technology.

While Big Data can provide unprecedented value and allow for faster analysis and decisionmaking, it poses new storage challenges for federal IT managers according to Rutherford.

"I go into federal government agencies and talk to them about their problems," he said. "I help people understand this technology relative to the problems they face and how life will change if they adopt this new technology."

He also mentions IT pros will get their weekends back.

"With EMC's high performing single pool of storage, they can handle their Big Data needs, while simplifying management and protecting data," Rutherford explained.

## In The Office Or Remote

Rutherford said federal storage requirements change depending on the setting. In federal offices everything is consolidating.

"They are building new data centers; space, power and cooling is a really big deal; and lights out management is a requirement. So they are all being driven that way."

It's different for those in remote offices, embassies and military in the field. The level of expertise is not high. Nor can they afford it.

"They don't have these great IT technicians out there. They need storage that will allow them to meet the mission, but not

require them to have a couple of experts sitting out there. So this is huge for them and they are catching on quickly and moving towards 'how do we get this stuff in place?'".

## 108 Terabytes In 60 Seconds

EMC/Isilon provides scalability and a system that transparently will scale according to Rutherford.

"Scale is a multifaceted word in this case," he said. "It could mean capacity; it can mean performance; it can mean the manageability of the system; and it can mean the availability of the system. They are all aspects of scalability."

Rutherford said they demonstrate how easy it is to add storage regularly.

"We can show customers a live demo of us doing that while we are reading and writing to a system. We can add 108 terabytes to a system in 60 seconds with a single button push."

The process not only adds storage, but rebalances all the data across these additional disk drives, so that every disk drive in the system still has available space. "That's really important in the scalability of performance," Rutherford said.

"You always want every disk drive in your system to be active for both writes and reads. And that's what we make happen for them, and they never touched a button, it just happens for them."

A further benefit according to Rutherford is IT groups no longer have to have dedicated staff to expand storage and re-allocate it on a daily basis. Existing staff can be used in areas where they can provide more value.

## Future View

For Rutherford, two or three years down the road are just around the corner. He said from an engineering standpoint, EMC/Isilon is already working on the architectural changes required to keep up with this growth.

And when talking federal IT professionals, Rutherford urges them to not look backwards on their storage decisions, but to look forward as new technologies are being developed.

"You are not alone. Look forward. Don't let somebody keep going back to the incumbent vendors and saying this is the way we should do things, but always look for something new." ■





# BIG DATA IT'S ON ISILON

**GAME-CHANGING  
TECHNOLOGIES  
FOR ISR**

- STORE & ANALYZE YOUR DATA
- PROTECT YOUR DATA
- SECURE YOUR NETWORK

[www.emc.com/federal](http://www.emc.com/federal)

**EMC<sup>2</sup>**

**Dave Ryan**

Chief Technology Officer/VP Technology & Innovation  
Navy Air Force Systems  
General Dynamics Information Technology

# The New Data Wave Has Arrived

General Dynamics Information Technology urges organizations to focus on the cultural and demographic changes of the workforce that Big Data will drive.



Whether it is mission data or data created socially — the new data wave has arrived. With it comes a new focus on how to increase and leverage its value.

"Big Data is not a new technology. It has the potential to solve significant problems for all types of organizations," **Dave Ryan** of General Dynamics Information Technology

told OTFL in a recent interview.

"Clearly understanding the challenges and opportunities that Big Data will offer is the first step," said Ryan. "This understanding provides a platform to evaluate and identify the solution that will best support the unique business objectives of an individual organization."

## Focus On Culture

"Data scientists and the practice of data management will come into greater focus as organizations try to figure out how to increase the value and usage of existing data, while the frontier of new data continues to expand on a daily basis," he said.

Ryan said from a technology standpoint it is critical to understand the difference between structured and unstructured data and determine how each type of data factors into Big Data solutions.

"While structured data can often be handled by more traditional relational database systems, unstructured data — which does not have a pre-defined data model and does not fit well into relational tables — poses a new challenge and a new opportunity."

If harnessed appropriately, unstructured data — like social media content, video files, health records and smart meter data — can cut out multiple steps of the analysis process. This enables organizations to quickly analyze massive collections of records without first forcing data to be modeled, cleansed and loaded he explained.

"Culturally, organizations have to adapt to this new, "data-savvy" business strategy to achieve a competitive advantage," Ryan went on.

"Big Data advocates an open innovation approach through data sharing that requires organizations to break their traditional methods for organizing and analyzing data. In addition, analyst skillsets must be addressed to ensure the correct adop-

tion of these new analytical tools."

Big Data is manageable but only when it is strategically evaluated and implemented to best serve the needs of the given domain or organization explained Ryan.

"Once the solution is identified and implemented, organizations can leverage Big Data to greatly reduce the processing time needed to move from raw data to actionable information, and as a result, can lead to many new data analysis capabilities and smarter business decisions."

## Tools For Big Data Problem Solving

As a system integrator with decades of experience in data management, General Dynamics IT builds, integrates and maintains a number of solutions in the Big Data space said Ryan.

"One such solution integrates an open-source platform with domain-based applications that provide the fundamental tools for a big data effort," he said.

"This approach consists of the open-source Apache Hadoop distribution, an open-source data management solution called the Data and Processing Syndicate (DPS), and in-house analytics, geospatial and visualization capabilities."

Ryan said the DPS has been implemented successfully in the Intelligence Community, and is an open-source platform in the public domain that is designed to be simple, fast and applicable to any domain.

He noted that General Dynamics IT stands behind the integration and maintenance of the solution; the wider developer and user audience also has access to all the tools and technologies at a source level. "This approach fosters a no 'lock in' risk for the government manager and optimizes cost effectiveness in any budget environment."

## Looking Ahead

Ryan sees no limits when it comes to ways to use Big Data.

"But to capture its full potential organizations need to make sure they have the right technology, talent, workflows and policies related to privacy and security in place."

Big data represents a whole new source of competitive advantages, "making information transparent and usable at a much higher frequency, improving productivity and using advanced analytics to improve decision making," said Ryan.

"Government managers need to ensure that the infrastructure, solution and processes are in place," Ryan counseled, "so that their organizations will see the value Big Data can offer." ■



## Turning Big Data into Actionable Information

Today's data driven environment requires a sophisticated framework and advanced analytics to turn massive amounts of structured and unstructured data into actionable information. General Dynamics Information Technology looks beyond basic big data approaches and provides solutions tailored to the unique needs of defense, intelligence and federal civilian organizations. We are dedicated to solving big data integration problems in any domain.

**GENERAL DYNAMICS**  
Information Technology

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



**Ed Hammersla**

Chief Operating Officer

Raytheon TCS (Trusted Computer Solutions)

# Connecting Cross Domain

**Raytheon TCS provides software so secure the government trusts it to move and share classified information from one classified network to another.**



NSA networks are top secret. The CIA, DOD, DHS and the FBI also have top secret networks. Normally these top secret networks are not allowed to connect electronically. That is understandable. Maintaining security is paramount.

But what happens when real-world scenarios make it necessary for these top secret networks to connect so information can be moved and shared? How is the transfer of Big Data accomplished? How are policies enforced?

"When they want to connect and share information they use "cross domain" software explained **Ed Hammersla**, COO of Raytheon TCS in a recent interview with OTFL.

Cross domain is software and hardware mechanisms that are approved to move sensitive information from one network to another, Hammersla said. "It automates the process of moving information back and forth, and implements certain procedures and strategies."

Hammersla said Raytheon TCS's "core business is software that is so secure that the intelligence community uses it to allow them to share classified information from one classified network to another."

These accredited products enable them to automate security, save time and resources, and reduce risk while supporting critical objectives and goals he said.

## Control Issue

"We think that as Big Data gets bigger as an issue, the control of that data, making sure it is being sent to the right people and not the wrong people becomes an incredibly important piece of the puzzle," he said.

Hammersla described how when data is moved, cross domain software controls the placement of data at rest by very secure containers.

"Then when data does move from one user group to another, it's done in an automated fashion. Even though the user may just point and click, behind the scenes the software is checking to see if that user is allowed to send that data."

Hammersla said networks range from the open Internet, to sensitive but unclassified (SBU), to secret and top secret.

"You've got lots of different top secret networks and the different layers that maintain their security by limiting the number of users and not connecting to other networks, he said. "The problem with that is 'how do you share?'"

This gets to the heart of how intelligence agencies share information.

"Cross domain solutions came out of the 9/11 Commission Report as an approved way for the intelligence community to share information with each other," Hammersla said.

"Expand that to a controlled environment for containing and sharing Big Data," he said. "If you want a proven solution already in place, in use and approved for the highest sensitivity levels of information then you can look to cross domain solutions for that."

"We are all about controlling Big Data, both the movement of it, the acquisition of it and the security it of when it sits at rest," Hammersla said. "What's more, at this point we've done enough of the implementation that we are pretty good at it."

## Security and Policy Drivers

As data volumes increase, the need to analyze it properly and control the flow and movement of that information becomes even more important said Hammersla.

"Organizations ought to be thinking about built-in security because security has to be built-in to any Big Data environment if you are going to have any confidence in it," he said.

"The kind of security that we are talking about is built in down at the OS and the bare metal levels. That's why we can do the things we can do in the Intel community, said Hammersla. "Organizations have a much higher degree of confidence when they are sharing information in a controlled environment they can trust."

Much of Big Data is, of course, about analytics. To do that you have to put things in the right categories so you can analyze it.

While Raytheon doesn't sell analytic software, "but what we do sell is a framework where you can categorize the software into compartments and then run your analytics against it," said Hammersla.

"You can get better results once you've categorized it and secured it." ■

# CROSS DOMAIN SOLUTIONS

## RAYTHEON TRUSTED COMPUTER SOLUTIONS: AT THE FOREFRONT OF CYBERSECURITY.

Raytheon Trusted Computer Solutions has been the global leader in cross domain solutions for the past two decades. Delivering secure data sharing technologies for military, intelligence, and civilian agencies is what we do. With the most accredited enterprise-wide deployments in the United States, 5 Eyes nations, and NATO member countries, we are the leader in delivering the right data, to the right people, at the right time. When protecting the world's most sensitive data, choose Raytheon Trusted Computer Solutions.



INNOVATION IN ALL DOMAINS

[www.raytheon.com/rctcs](http://www.raytheon.com/rctcs)

Follow us on:



# Raytheon

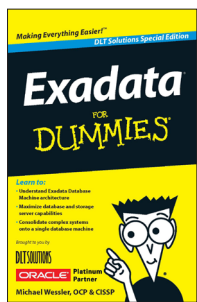
*Customer Success Is Our Mission*

**Steve Stratton**

Senior Vice President of  
the Enterprise Applications Division  
DLT Solutions

# The Art Of The Algorithm

**DLT/Oracle Big Data solutions help customers ask the right questions and connect the dots.**



Ask the right questions, and your chance improves you'll get the right answers.

"I believe that in analytics it is about knowing what to ask for; and how to connect the dots, said **Steve Stratton**, SVP at DLT Solutions. "We can ingest a lot of data, we can put it in main memory and we can look at it, but what are the right questions to ask?"

Stratton said DLT/Oracle is involved with customers across the government in projects addressing Big Data solutions capabilities investigations and implementations.

"We are working with several Law Enforcement agencies to bring their solutions online," he said in a recent interview with OTFL. "The Oracle Sales Team and DLT Solutions team is making use of the products available from Oracle Corporation including Endeca, Big Data Appliance and Exalytics."

## Force Multiplier

"Big Data and the analytics that go along with it can be a force multiplier," said Stratton. For example if you have four analysts delving into a problem, now you can now have two analysts do the work of four.

The other two analysts are now free to work on another project because the automation in these systems makes them more efficient and gets them looking at the right data instead of lots of data Stratton added.

"The problem is sometimes they are so enabled with data and information that you can't swallow it all, you can't ingest it. So I would say that it's more like a force multiplier. We can make the work of four be the work of two. And we can actually get to outcomes and results quicker by applying the right algorithms to these petabytes of data that we are getting stuffed down our throats."

The "art of the algorithm" is knowing the right questions to ask said Stratton.

To demonstrate Stratton described how questions the Secret Service might ask someone could be a subset of what a police department would ask on a larger scale for a metropolitan area.

"The Secret Service is looking at what do I need to know about certain people; where a law enforcement agency is look-

ing at a lot of different questions about protecting a metropolitan area."

Big Data analytics makes it possible for different law enforcement agencies to access the same data and ask the different questions for their unique situation and desired outcome.

"The 'art' is understanding how to bring these tools to life to the mission that you are supporting whether that is in a commercial environment or the government environment," Stratton noted.

## Understanding Big Data Tools

Stratton spends a lot of time consulting with customers, helping them understand how new Big Data tools fit into their day-to-day mission and showing them what they could accomplish.

"Law enforcement agencies have really smart intelligence people, but they've never had to have the analytic talent on board, or the ability to bring on a complex system, like when we are talking Big Data or data analytics in particular," Stratton said.

"We are helping them understand what these kinds of tools can do for you. How can they save you manpower and money and increase your effectiveness, whatever your job might be."

Stratton believes that Big Data can transform an organization.

"The organization needs to really understand it and see it there's transformation in that process," he said. "That is where you start to see the value very quickly of some of these tools and the analytic power that it provides."

Stratton also believes that transformation takes time.

"I'm an old school IT guy. I believe in proofs of concept with stakeholders; using much smaller data sets, starting off with a premise that says 'this is the outcome I'm looking for' and prove these tools can get me there," said Stratton.

He urged agencies to think basics and fundamentals first starting with "we believe that data is important. How can we unlock that data we already have and get the most benefit out of it?"

Stratton offered a word of caution to CIOs who are considering truly open source implementations. "Do they have the luxury of having the time and money to put people on using truly open source tools and solutions? Will they get the outcome when more COTS tools might be a better, faster solution?" ■



$$\frac{\text{Data} \\ \text{Cost} \\ \text{Time} \\ - \\ \text{Space}}{\text{EXadata}}$$

Solving for 'X' can be complicated when you have competing factors such as decreasing budgets and increasing data loads. When it comes to Oracle and DLT Solutions, X equals Exadata but it also equals 10 times faster access to 10 times more data.

Agencies are challenged to use, store and analyze huge volumes of data – how can you unlock that data to get the most benefit? Oracle's Exadata is your answer. Designed for large database deployments, Exadata enables analytics alongside your existing data to find new insights and capitalize on hidden relationships.

To learn more about how Exadata enables you to connect the dots visit [www.dlt.com/BigData](http://www.dlt.com/BigData).



#DoMore  
WithLess

DLT SOLUTIONS

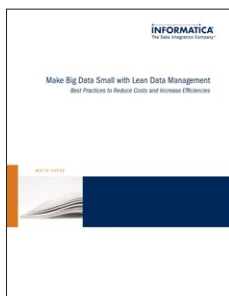
ORACLE® Platinum  
Partner

**John Haddad**

Director of Enterprise Data Integration  
Informatica Corporation

# Imagine The Innovation

Informatica® 9.5, the newest release of the Informatica Platform, enables you to maximize your return on Big Data.



Moving air traffic more efficiently; making better weather and climate predictions; and finding fraudulent claims in Medicare and Medicaid.

Those are just three of many current Big Data applications in government. More are surely on the horizon.

"I think Big Data will open up possibilities we haven't even imagined,"

John Haddad, Director of Enterprise Data Integration at Informatica Corporation told OTFL in a recent interview.

Haddad thinks there is tremendous growth opportunity for Big Data applications that leverage emerging Big Data technologies such as Hadoop and innovations related to social, mobile, and cloud.

"We are no longer restricted by time and cost as to the amount of data and type of data we can use to improve government services and the lives of our citizens," he said.

Haddad stressed that in order to build these applications, organizations need a holistic data management strategy in their vision for a Big Data future.

"Hadoop can complement an existing information management strategy by enabling IT organizations to cost-effectively process massive amounts of data with lower cost commodity hardware," he explained.

The new reality is that data is stored, processed, consumed, and managed across a myriad of systems, platforms, applications, and devices Haddad continued.

"Without a holistic data management platform, the risk is that Hadoop becomes yet another silo of information. Data management processes can now span a hybrid IT infrastructure with some workloads executed on Hadoop, others on traditional RDBMS based platforms, both on-premise and in the Cloud," Haddad said.

This challenge requires data management tools that can manage the end-to-end data pipeline across multiple systems and platforms.

"My advice to government managers is to choose a holistic data management platform that can orchestrate and optimize the performance of the end-to-end Big Data processing pipeline independent of the underlying storage and processing technology."

## Big Data For The Rest Of Us

Haddad said Informatica® 9.5, the newest release of the

Informatica Platform, enables you to maximize return on Big Data — "whether you're strategizing over how to deal with large-scale increases in data volume, variety, and velocity; how to derive greater business value from more traditional data sets; or how to combine new and conventional data for business advantage."

He said Informatica 9.5 "enables customers to not merely cope, but to thrive in the era of Big Data by increasing the value of data, decreasing its costs and fully realizing the promise of Big Data."

Customers get universal data replication, ultra messaging, and data integration that delivers transaction data to Hadoop analytical processing to lower processing costs; Social MDM that integrates social customer profiles with existing customer data to increase customer retention and share of wallet; Hadoop-native parsers and transforms that harness the power of Hadoop with optimized transforms to accelerate business value delivered; and a data integration platform for Hadoop that lets you use reuse existing assets and skillsets to make Hadoop enterprise-ready.

These are solutions that are currently successfully used in the private sector most notably by US Xpress and eHarmony.

## Big Data Challenges

Some of the big challenges relate to volume, variety, and velocity said Haddad, specifically:

- How to access all types of data consistently across projects and services.
- How to store and process massive amounts of data cost-effectively.
- How to integrate both real-time data and historical data up to years in the past to create better models and predictors.
- How to extract out just what is relevant and deliver it through different applications and channels like mobile devices in a timely fashion.
- How to govern massive amounts of data from a variety of data sources and data types in terms of quality, security, and transparency.
- How to get various stakeholders collaborate and then how to manage this process in an agile and responsive fashion.

As these issues are solved Haddad predicted "as more data becomes available to more people across agencies and the public, we'll see even more innovation." ■

# BIG DATA?

## NO BIG DEAL

You don't need to be overwhelmed  
by the downpour of data.

With Informatica, you can embrace it all,  
and turn Big Data into big advantage.

### Gain Big Advantage In a Dynamic Big Data Environment

Informatica understands that it's critical to effectively process extreme volumes of data to support missions, serve citizens and provide the right information to the right individuals. Informatica helps government agencies to digest, transform, manage, and analyze their data — regardless of how much, what kind, or where it is.

Learn how Informatica can help  
your agency with Big Data.



.....  
**INFORMATICA®**

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



**Gary Nakamura**  
General Manager, Terracotta  
Software AG

# From Milliseconds To Microseconds

Terracotta's BigMemory solution allows you to scale up *and* out, simultaneously increasing efficiency, density, and capacity.



Big Data is pushing us to think in new ways about how to manage data and, most importantly, the value we can extract from data when we can manage it differently.

"Having a broad set of skills and expertise in various Big Data techniques will help agencies leverage Big Data to achieve their missions in the most efficient and cost effective manner," Gary Nakamura, Manager at Terracotta a Software AG company told OTFL in a recent interview.

"Our customers are at the forefront of the Big Data trend," explained Nakamura. "As leaders in financial services, telecommunications, media and other fields, they are experiencing high growth in the volumes of data that they need to access and process. They also are looking for solutions which allow them to work with Big Data at much faster speeds."

Nakamura explained how Terracotta's flagship product — BigMemory — provides the ability to persist, distribute, search, monitor and manage Big Data in memory.

"This means that, instead of waiting tens of milliseconds to access their data from a database, they can achieve microsecond access to terabytes of data in-memory," he said. "We engineered BigMemory for large-scale, high transaction, mission-critical environments that have massive amounts of operational data that they want to access in real time."

Nakamura described how customers are challenged at first by the cultural mind shift that BigMemory requires. They are used to having to grow their data centers and their staffs to manage increasing volumes of data.

"The BigMemory solution allows them to scale up and out, simultaneously increasing efficiency, density, and capacity," Nakamura noted. "This allows them to maximize the hardware investments they've made and reduce future hardware purchases."

## Meeting Big Data Demands

The number of our customers focused on Big Data projects has grown sharply in the past few years said Nakamura.

"Our financial services customers first adopted BigMemory

to address the demands of managing their trade verifications and also detecting fraud. In these situations, the service levels on terabytes of data required microsecond response times."

Nakamura described how one large financial institution they serve was struggling with a 45 minute response time on fraud detection using their older relational and grid techniques.

"In under two months, we implemented the BigMemory solution and reduced their fraud detection to less than 4 seconds for the same process, he said. "That ability to speed up their processing was transformational, resulting in hundreds of millions of dollars in prevented fraud as well as happy customers and merchants who experienced the benefits during the busy holiday shopping season."

Fraud detection is just one use case for BigMemory according to Nakamura. "There are other entities that have a similar need for fast processing on Big Data," he noted. "For example, the military collects data and security information at a very significant pace. They may not be looking for fraud, but they want to detect other things specific to their data where detection could be a life or death situation where they have to make an immediate critical determination."

## Again The Four "V"s

Nakamura advises government managers to think about the four V's of Big Data: volume, velocity, variety, and value.

"Make sure you have a solution that addresses each of those areas well and realize that there are solutions, like our BigMemory solution, that don't force you to make tradeoffs between those important aspects of Big Data."

Nakamura said one big challenge he faces is actually convincing prospective customers that the results we claim are real. "The data access speeds and scale we achieve on very large data sets are literally unbelievable to them."

Other solutions they are familiar with can manage large data sets slowly or small data sets quickly but not terabytes of data at blazing fast speeds he said.

"Once we demonstrate BigMemory, however, we find that our prospects quickly start thinking bigger about the many ways we can help them experience transformational value using BigMemory to handle their unique Big Data needs." ■

# BigMemory

## In-Memory Data Management for the Government Enterprise

BigMemory stores “big” amounts of data in machine memory for ultra-fast access in the most efficient and cost effective manner. It snaps into federal enterprise applications to deliver high-speed performance at any scale. BigMemory completely changes what your agency can do—whether that’s doubling order throughput with real-time processing, cutting risk analysis from 45 minutes to less than 4 seconds, or streaming full motion video to mobile devices. What could your agency do with high-volume, real-time data access?



Big Data

+



In-Memory

=



BigMemory

# CYBER SECURITY EXCHANGE



**Cyber Security Exchange** is a public-private community focused on collaborating and sharing best practices.

JOIN THE CONVERSATION TODAY AT  
[www.meritalk.com/CSX](http://www.meritalk.com/CSX)