14th SOA for E-Government Conference October 2 2012

Service-Oriented Architecture For e-Government Conference
Announcement
Welcome
Registration
Logistics
Media Coverage
Tweets
Agenda
Speaker Biographies
  Jeanne Vasterling
  Ajay Budhraja
  Dave Mayo
  Nitin Naik
  Edward Ost
  Jason Bloomberg
  Wolf Tombe
  Brand Niemann
  Eric Little
  Victor Pollara
  Kate Goodier
  George Strawn
  Gadi Ben-Yehuda
  Steve Reinhardt
  Mark Guiton
  Tom Rindflesch
Blog By Brand Niemann
  Welcome & Introduction
  Keynote
  Establishing a Service Factory
  Case Study

http://semanticommunity.info/Federal_SOA/14th_SOA_for_E-Government_Conference_October_2_2012

Updated: Wed, 23 Sep 2015 06:05:39 GMT
Powered by mindtouch
Open Architecture
Overview of Exhibitors – SOA Tools Slides
Afternoon Keynote
SOA Pilots
Panel Discussion
Wiki Blog
Big Data Fall Forum 2012
Gartner Exclaims "uRiKA!"
YarcData’s uRiKA Shows Big Data Is More Than Hadoop and Data Warehouses

VIEW SUMMARY
Analysis
The Nature of Most Enterprise Data
Most Big Data Situations Are Addressed by MapReduce or Other Traditional Technologies
Graph Problems in Big Data Are Different
Figure 1. Impacts and Top Recommendations for Running Graph-Style Discovery Problems
Impacts and Recommendations
IT organizations faced with previously infeasible graph-style discovery problems may succeed using a focused solution like uRiKA
Why Graph-Oriented Problems Behave Pathologically on Traditional Systems
A Design Accommodating Graph Problem Peculiarities Works Where Classical Approaches Fail
Pick the Right Tools Based on the Nature of Your Big Data Problem
To address all their data requirements, IT organizations may be forced to duplicate data between systems such as uRiKA and transactional systems

Service-Oriented Architecture For e-Government Conference

Source: http://gov.aol.com/2012/09/05/service-oriented-architecture-for-e-government-conference/

McLean, VA 22102-7539
Organized by MITRE & Federal Government SOA Community of Practice
At MITRE Auditorium
14th Service-Oriented Architecture (SOA) for e-Government Conference
“Achieving Service Re-use”

Tuesday, October 2, 2012
8am – 4:30pm
MITRE and the Federal Government SOA Community of Practice invite you to attend the upcoming SOA Conference to share best practices and innovations in SOA and Cloud initiatives across government, industry and academia.

Who Should Attend?
The conference is open to federal government employees, members of industry and academia, and MITRE personnel.

Why Attend?
Re-use is commonly viewed as one of the main benefits of using Service-Oriented Architecture. However, many organizations that adopt SOA struggle to establish methods for identifying and provisioning reusable enterprise services. Hear case studies and interact with speakers and panelists as they describe best practices for service reuse in the federal government. Join the conversation as we explore the concept and benefits of a Service Factory to improve re-use. Learn about the Cloud and its role in Big Data.

Conference Registration and Information
Registration and participation is free. However, space is limited to the first 200 registrants and seats fill up fast.

For more information including registration, agenda and logistics please go to: https://register.mitre.org/soa/

Conference location:
MITRE’s McLean campus Main Auditorium
7515 Colshire Drive, McLean, VA 22102-7539

Agenda:
http://semanticommunity.info/Federal_SOA/14th_SOA_for_E-Government_Conference_October_2_2012
Announcement

14th Service-Oriented Architecture for e-Government Conference

“Achieving Service Re-use”

October 2, 2012
MITRE 1 Auditorium
McLean, VA

Hosted by The MITRE Corporation & Federal Government (SOA/CoP)
Keynote speakers

Ajay Budhraja
Chief Technology Officer (CTO)
U.S. Department of Justice, IR

Ajay Budhraja has over 23 years in Information Technology with experience in areas such as Executive leadership, management, strategic planning, enterprise architecture, system architecture, software engineering, training, methodologies, networks, databases, etc.

Ajay has provided Senior Executive leadership for nationwide and global programs and has implemented integrated Enterprise Information Technology solutions. He has a Masters in Engineering (Computer Science), Masters in Management and Bachelors in Engineering. He is a Project Management Professional certified by the PMI and is also CICM, CSM, ECM (AIIM) Master, SOA, RUP, SEI-CMMI, ITIL-F, Security + certified. Ajay is currently the Chief Technology Officer (CTO), US Department of Justice component led large scale projects for big organizations and has extensive IT experience related to telecom, business, manufacturing, airlines, finance and government. He has delivered internet based technology solutions and strategies for e-business platforms, portals, mobile e-business, collaboration and content management. He has worked extensively in the areas of application development, infrastructure development, networks, security and has contributed significantly in the areas of Enterprise and Business Transformation, Strategic Planning, Change Management, Technology innovation, Performance management, Agile management and development, Service Oriented Architecture, Cloud. He is the Co-Chair for the Federal SOA COP and has served as President DOL-APAC, AEA-DC, Co-Chair Executive Forum Federal Executive Institute SES Program. As Adjunct Faculty, he has taught courses for several universities. He has received many awards, authored articles and presented papers at worldwide conferences.

Wolf Tombe
CTO, Customs and Border Patrol, Department of Homeland Security (DHS)
Co-Chair of DHS SOA Working Group
As the Chief Technology Officer of US Customs & Border Protection, Wolf Tombe is responsible for the proactive formation of cross-cutting integrated technology strategies, architectures and solutions across CBP.

He is also responsible for establishing the Agency’s Strategic Technology Direction and associated Information Technology Transformation initiatives, which includes the CBP Enterprise Technical Architecture, CBP Technology Roadmap, Common Infrastructure architectures and high level designs, Technology Lifecycle implementation and Customer Transformation Services. Mr. Tombe Chairs both the SOA and Application Working Group and the Technology Review Committee where he provides leadership on technology planning, architecture, standards and adoption of industry best practices for technology management. As CTO, Mr. Tombe is the liaison for CBP with industry and other government agencies on technology issues and matters. In this capacity, Mr. Tombe founded the first US Government "Federal CTO Forum" in December of 2008. Recognizing both the need and tremendous benefits to be gained by sharing technology best practices and innovations between federal agencies the forum now has regular participation from twenty-six federal civilian and defense agencies.

Mr. Tombe joined US Customs and Border Protection in 2003 and supports CBP with more than 25 years of IT management experience in the Federal Government Sector serving in a variety of Agency’s on both the West and East Coast.

Mr. Tombe holds a Masters Certificate in Project Management from George Washington University and is a Certified Project Management Professional (PMP) with the Project Management Institute. Mr. Tombe is a recipient of multiple awards in the areas of Cloud Computing, SOA, Technology Innovation and an accomplished public speaker having presented at numerous public forums on issues pertaining to technology implementation and management in the US Federal Government.

Welcome

Reuse is commonly regarded as one of the main benefits of using Service Oriented Architecture, however many organizations adopting SOA have stumbled in establishing methods for identifying and provisioning reusable enterprise services.
This conference explores proven industry best practices for service reuse in the Federal government. Join the conversation as we explore the concept and benefits of a Service Factory, discuss the use of semantic information integration and the Cloud for managing Big Data through case studies, panels and interactive presentations.

View Conference Agenda [PDF] See Below

We look forward to seeing you at this event.

13th Conference Proceedings

Registration

This event is free and open to government personnel and contractors. Registration for all attendees is required and is limited to 200 people attending in person. If you are interested in attending please register below. Members of the media should contact Karina Wright at khw@mitre.org.

Doors to the MITRE facilities will open at 7:30am. A picture ID is needed for registration and badging. If you have any questions about the event, please contact Christine Custis (ccustis@mitre.org, 301-537-8979).

All participants must complete and submit the registration form. Conference exhibitors must also complete an Exhibitor Registration form to register their company.

Note to attendees: Media will be invited to this event and select presenters/presentations will be videotaped by MITRE Media Services. The entire conference proceedings may be audiotaped for MITRE use. Videotaped presentations will be made public with the permission of the presenters.

Logistics

Location: MITRE-1 Auditorium, 7515 Colshire Drive, McLean, VA 22102. The conference will take place as in previous years, on the MITRE campus located just inside the Beltway adjacent to Dolly Madison/Chain Bridge Road.

MITRE Shuttle Bus: A complimentary MITRE shuttle (white van) is available to and from the conference from the West Falls Church metro station. (See schedule.) Service available from the West Falls Church Metro Station starting at 6:40 a.m., continuing every 20 minutes, and last bus from MITRE at 5:40 p.m. Drop off at MITRE 1 and 2 (come to the back of MITRE-1 for entrance).

Accomodations: A list of hotels in the Tyson's Corner area is available for download.

Media Coverage

Members of the media should contact Karina Wright at khw@mitre.org.

Note to registrants: Media will be invited to this event and select presenters/presentations will be videotaped by MITRE Media Services. The entire conference proceedings will be audiotaped for MITRE use. Videotaped presentations will be made public with the permission of the presenters.

Tweets

https://twitter.com/search/?q=%2314thSOAGov&src=hash

Brand Niemann
Brand Niemann @bniemannsr

#14thSOAGov Excellent Panel and Conclusion with All Presentations, Tweets, and Blog Notes Posted in Preparation for the 15th SOA & Big Data

Expand
- Reply
- Delete
- Favorite

4m
Brand Niemann
Brand Niemann @bniemannsr

#14thSOAGov Gadi Ben-Yehuda IBM Watson & helping new HHS CTO with big data problems to make data more useful and usable by data scientists

Expand
- Reply
- Delete
- Favorite

15h
Brand Niemann
Brand Niemann @bniemannsr

#14thSOAGov Panel Discussion Big Data & Government Enterprise Kate Goodier, Gadi Ben-Yehuda, Eric Little, Victor Pollara, Mark Guiton

Expand
- Reply
- Delete
- Favorite
Victor Pollara See how far can go with Cray XMT2 as back-end for SOA Rest data services & representing data in graph form

Victor Pollara providing a tutorial on Semantic Medline for more detail see: http://semanticommunity.info/A_NITRD_Dashboard/Semantic_Medline

It would be interesting to compare Cray Graph and IBM Watson platform for Semantic Medline capabilities

Victor Pollara, Noblis Using Semantic Medline on the New Cray Graph Computer for Medical Research
History - Brought NLM (Tom Rindflesch) together with Noblis (Victor Pollara) which recently acquired the new Cray Graph Computer

Eric Little Utilizing Semantics for Integration of Data Across Multiple Sources - Many companies are doing it manually

Combining structured and unstructured data reveals clinical insights -- key component of translational medicine #14thSOAGOV

Eric Little, Orbis Technologies Simplifying Semantics to Query Relevant Information - SPARQL under the hood
Problems of health data integration: structural heterogeneity, semantic heterogeneity, inconsistency & redundancy #14thSOAGOV

Problems of data integration: #14thSOAGOV

Semantic Information Integration in Healthcare at #14thSOAGOV

#14thSOAGov Brand Niemann introduces pilots using slides: http://semanticommunity.info/@api/deki/files/19354/BrandNiemann10022012.pptx
Jeffrey Hall DHS SOA Working Group & DHS Agile Integrated Project Team (IPT) for "software intensive IT projects"

Custom Border Protection seeing value of SOA as migrate apps built on SOA to the cloud #14thSOAGov

Jeffrey Hall SOA Infrastructure is based in open architecture using federated ESB model across the department-SOA like agile

Jeffrey Hall CBP oldest & largest DHS components & processes millions of transactions & billions of data searches each day
Jeffrey Hall All slide notes http://semanticommunity.info/Federal_SOA/14th_SOA_for_E-Government_Conference_October_2_2012#Afternoon_Keynote

David Webber Oracle - I think Open Data Exchange Will Help NIEM and the ISE-PM With Its Problems http://semanticommunity.info/Information_Sharing_Environment#Story ...

Michelle Davis Redhat just acquired FuseSource: Integration + BPM = Business Agility

Michelle Davis Redhat/FuseSource SOA is not Dead it is Open Source integration & messaging - leading vendor - build your own ESB

#14thSOAGov Jason Bloomberg Now REST for Web-based enterprise distributed computing (WOA)-building enterprise SOA using the Web architecture

#14thSOAGov Jason Bloomberg Vendors said buy an ESB and hook it up - misleading - Client asks Where is my SOA? Really get the services right

#14thSOAGov Jason Bloomberg, President, ZapThink A Dovel Technologies Company talking about SOA is not Dead - wrote the book with that title

#14thSOAGov Question on governance with Open Source Software: Ed Ost answers that data scientists need to help with that by focusing on data
Jason Bloomberg, President, ZapThink thinks Edward Ost presentation Design for Re-use with Data Services on SOA he has heard!


Edward Ost now Design for Re-use with Data Services Data is the most important part See slide notes

Nitin Naik Challenges Ahead! Funding, Technical, Service Development Lifecycle not incorporated in existing ELC, Organizational
Brand Niemann @bniemannsr

#14thSOAGov Nitin Naik New Service: Taxpayer Identity Service which is needed across government Now doing Solution Architecture not just EA!

Expand
  • Reply
  • Delete
  • Favorite

4m

Brand Niemann
Brand Niemann @bniemannsr

#14thSOAGov Nitin Naik The Road to Adoption of Common Services at the IRS - difficult but worth it - systems to collect $2.5-2.8 Trillion

Expand
  • Reply
  • Delete
  • Favorite

29m

Brand Niemann
Brand Niemann @bniemannsr

#14thSOAGov Dave Mayo Service Factory Amazing talk! Now questions in person and on the phone

Expand
  • Reply
  • Delete
  • Favorite

30m

Brand Niemann
Brand Niemann @bniemannsr

#14thSOAGov Dave Mayo Service Factory Service Tooling for CMS/FFE Factory now targets MarkLogic formerly Oracle
Dave Mayo Service Factory Slide Notes

### Expand
- Reply
- Delete
- Favorite

Dave Mayo Microsoft pushed the Software Factory originally - the software - but Everware-CBDI definition is broader

### Expand
- Reply
- Delete
- Favorite

Dave Mayo Service Factory - An organizational construct with a set of tools, models, etc. to address the problem to be solved

### Expand
- Reply
- Delete
- Favorite

Dave Mayo provides State of Government: SOA has become a standard practice. But reuse is very limited.
Brand Niemann
Brand Niemann @bniemannsr

#14thSOAGov Ajay Budhraja provides Five Point Plan: Blueprint First

Brand Niemann
Brand Niemann @bniemannsr

#14thSOAGov Jeanne Vastering announces MITRE has a new Center for CMS Health

Brand Niemann
Brand Niemann @bniemannsr

#14thSOAGov Remote attendance Dialin: 703-983-6338 Passcode: 762762
Agenda

7:30 - 8:30 AM Conference Check In / Exhibitor Showcase

8:30 – 8:35 AM Welcome & Introduction – Jeanne Vasterling, Enterprise Modernization & Transformation Practices, MITRE Slides

8:35 – 9:05 AM Keynote: Government Trends and Innovation with Service Reuse, Cloud and Big Data – Ajay Budhraja, Chief Technology Officer (CTO), U.S. Department of Justice, IR Forbes ZdNet FedScoop Slides

9:05 – 9:45 AM Establishing a Service Factory – Dave Mayo, President and CEO of Everware-CBDI Slides

9:45 – 10:15 AM Case Study: The Road to Adoption of Common Services at the IRS – Nitin Naik, Director of Enterprise Architecture, IRS Slides

10:15 – 10:30 AM Break

10:30 – 11:30 AM Open Architecture: Design for Re-use with Data Services – Edward Ost, Application Integration Technical Director, Talend Slides

Extra Presentation: SOA: Not Dead Yet, Jason Bloomberg, President, ZapThink, A Dovel Technologies Company Slides

11:30 – 12:30 PM Overview of Exhibitors – SOA Tools Slides

1 Dovel Slides 2 Everware-CBDI Slides 3 Semantic Community Slides 4 Redhat / FuseSource Slides 5 IBM Slides 6 Oracle Slides 7 Software AG Slides 8 Talend Slides
12:30 – 1:30 PM Lunch, Networking and Exhibitor Showcase, Invitation-only Media Q&A

1:30 - 2:15 PM Afternoon Keynote: Delivering Mission Agility Through Agile SOA Governance – Wolf Tombe (Jeffrey Hall), Chief Technology Officer (CTO), at U.S. Customs and Border Protection Slides


Semantic Information Integration within the Healthcare Sector – Eric Little, Orbis Technologies Slides

Using Semantic Medline on the New Cray Graph Computer for Medical Research – Victor Pollara, Noblis Slides

3:15 – 3:30 PM Break

3:30 – 4:15 PM Panel Discussion: Big Data and the Government Enterprise Slides
Kate Goodier (Moderator, IC), and Gadi Ben-Yehuda (IBM) for Dr. George Strawn (OSTP/NITRD/NCO), Dr. Eric Little (Orbis Technologies), Dr. Victor Pollara (Noblis), Mark Guiton (Cray) for Steve Reinhardt (Cray), and Dr. Tom Rindflesch (NLM)

4:15 – 4:30 PM Wiki Blog – Brand Niemann, Federal SOA CoP

Speaker Biographies

Jeanne Vasterling

Jeanne Vasterling is an Acting Department Head leading the practice of Enterprise Business Transformation for MITRE’s Center of Connected Government. With over twenty years of experience in academia, private industry, information technology and management she works closely with the Government to help them address their mission challenges. Jeanne is actively engaged in the Process Design, Enterprise and Service Oriented Architectures, as well as cross agency collaboration, supporting many civil agency projects including the Affordable Health Care Act and the National Export Initiative.
Prior to joining MITRE in 2003, Jeanne has enjoyed a wide range of experiences in serving both the private and public sectors. In her roles she has led both field and corporate functions in the United States and abroad. Her experiences have included working for University of Missouri, Sprint Corporation, Global One and France Telecom in a variety of roles including designer of integrated business architectures between international telecommunications companies via collaborative intercompany collaboration.

Jeanne earned her Bachelor of Arts in English and Philosophy from Southeast Missouri State University. Jeanne’s work in academia led to research in the area of composition and she has served as a business advisor to academia in area of communications. Jeanne is an active volunteer in the Reston Virginia community and for the Washington Animal Rescue League.

Ajay Budhraja

Ajay Budhraja has over 23 years in Information Technology with experience in areas such as Executive leadership, management, strategic planning, enterprise architecture, system architecture, software engineering, training, methodologies, networks, databases etc. Ajay has provided Senior Executive leadership for nationwide and global programs and has implemented integrated Enterprise Information Technology solutions. He has a Masters in Engineering (Computer Science), Masters in Management and Bachelors in Engineering. He is a Project Management Professional certified by the PMI and is also CICM, CSM, ECM (AII) Master, SOA, RUP, SEI-CMMI, ITIL-F, Security + certified. Ajay is currently the Chief Technology Officer (CTO), US Department of Justice component led large scale projects for big organizations and has extensive IT experience related to telecom, business, manufacturing, airlines, finance and government. He has delivered internet based technology solutions and strategies for e-business platforms, portals, mobile e-business, collaboration and content management. He has worked extensively in the areas of application development, infrastructure development, networks, security and has contributed significantly in the areas of Enterprise and Business Transformation, Strategic Planning, Change Management, Technology innovation, Performance management, Agile management and development, Service Oriented Architecture, Cloud. He is the Co-Chair for the Federal SOA COP and has served as President DOL-APAC, AEA-DC, Co-Chair Executive Forum Federal Executive Institute SES Program. As Adjunct Faculty, he has taught courses for several universities. He has received many awards, authored articles and presented papers at worldwide conferences.
Dave Mayo

Dave Mayo – An IT practitioner for 25 years and an enterprise architect for over 15 years, Mr. Mayo is the President of Everware-CBDI, a firm dedicated to enabling and implementing service oriented architecture (SOA) for federal and commercial organizations. He has been a senior advisor to the Department of Homeland Security EA program as well as large commercial organizations. Mr. Mayo has been Vice Chair of the IAC/EA-SIG and Chairs the Services Committee. He led the government/industry team for the Practical Guide to Federal Service Oriented Architecture for the Federal CIO Council. For his steadfast efforts to improve government effectiveness and efficiency through service-based EA, he received a Federal 100 award from Federal Computer Week in 2009. His background includes economics, strategic planning, information engineering, and business process reengineering. He holds a MA in Economics from the University of British Columbia.

Nitin Naik

Dr. Nitin Naik is an information systems leader with over 25 years of experience in applying technology to business problems. He is currently the director of Enterprise Architecture (EA) and chief architect for the Affordable Care Act (ACA) program. As the EA director, Dr. Naik is managing the development of infrastructure strategy and enterprise transition plan, documenting and refining architectures of the more than 550 systems of the IRS, and establishing IRS technology standards. He is also responsible for collaborating with the ACA Program Management Office to establish the program requirements, solution architecture, tax administration system integration and IT roadmap for implementation.
Dr. Naik joined the IRS in 2007 as the IRS technical director, to provide leadership in areas of IT infrastructure, Web services and new technologies suitable for the IRS modernization effort. Prior to joining the IRS, Dr. Naik served at NASA as associate chief technology officer, where he managed the enterprise architecture program, Web presence consisting of 7,500 websites, IT infrastructure strategy, and information management program. Before federal service, he served as president of the Center for Educational Technologies at Wheeling Jesuit University, W.V., leading research and development efforts in applying cutting-edge IT for science research, simulation-based training and Internet commerce through competitive NASA, NSF, U.S. Department of Education and private sector funding. In addition, he established one of the premier computer-video local area networks using asynchronous transfer mode and real-time video streaming technology.

Dr. Naik has a Ph.D. in Computer Science, M.S. degrees in Electrical Engineering and Computer Science from Louisiana State University.

Edward Ost

Edward Ost is Technical Director at Talend, an Apache open source integration platform. Mr. Ost works with federal customers to build solution architectures using Talend’s Apache technology as well as Hadoop technology from Talend partners such as Horton Works and Cloudera. Previous to Talend Mr. Ost worked as a consultant for the Federal Aviation Administration on the System Wide Information Management (SWIM) program where he helped develop the service container concept and the FAA’s federated adoption approach. As part of the SWIM program Mr. Ost chaired the Architecture Working Group in facilitating adoption of SOA best practices by the SWIM Implementing Programs (SIP). He has helped support numerous aviation systems including Traffic Flow Modernization System, ACEP, ERAM, Terminal Data Distribution System, WMSCR, ITWS, and CIWS. Before joining the FAA team, Mr. Ost worked as a consultant for Lockheed Martin for eight years on federal IT systems. He is active in the Apache community where he is an evangelist for open source SOA and Cloud adoption.
Jason Bloomberg is President of ZapThink, a Dovel Technologies Company. He is a global thought leader in the areas of Cloud Computing, Enterprise Architecture, and Service-Oriented Architecture. He created the Licensed ZapThink Architect (LZA) SOA course and associated credential, and runs the LZA course as well as his Cloud Computing for Architects course around the world. He is a frequent conference speaker and prolific writer, including as a regular columnist on US Government IT for CIO Magazine and blogger for DevX.

Mr. Bloomberg is one of the original Managing Partners of ZapThink LLC, the leading SOA advisory and analysis firm, which was acquired by Dovel Technologies in August 2011. His book, Service Orient or Be Doomed! How Service Orientation Will Change Your Business (John Wiley & Sons, 2006, coauthored with Ron Schmelzer), is recognized as the leading business book on Service Orientation. His newest book, The Agile Architecture Revolution: How Cloud Computing, REST-based SOA, and Mobile Computing are Changing Enterprise IT (John Wiley & Sons), is due in the spring of 2013.

Mr. Bloomberg has a diverse background in eBusiness technology management and industry analysis, including serving as a senior analyst in IDC’s eBusiness Advisory group, as well as holding eBusiness management positions at USWeb/CKS (later marchFIRST) and WaveBend Solutions (now Hitachi Consulting). He also co-authored the books XML and Web Services Unleashed (SAMS Publishing, 2002), and Web Page Scripting Techniques (Hayden Books, 1996).
As the Chief Technology Officer of US Customs & Border Protection, Wolf Tombe is responsible for the proactive formation of cross-cutting integrated technology strategies, architectures and solutions across CBP. He is also responsible for establishing the Agency’s Strategic Technology Direction and associated Information Technology Transformation initiatives, which includes the CBP Enterprise Technical Architecture, CBP Technology Roadmap, Common Infrastructure architectures and high level designs, Technology Lifecycle implementation and Customer Transformation Services. Mr. Tombe Chairs both the SOA and Application Working Group and the Technology Review Committee where he provides leadership on technology planning, architecture, standards and adoption of industry best practices for technology management. As CTO, Mr. Tombe is the liaison for CBP with industry and other government agencies on technology issues and matters. In this capacity, Mr. Tombe founded the first US Government “Federal CTO Forum” in December of 2008. Recognizing both the need and tremendous benefits to be gained by sharing technology best practices and innovations between federal agencies the forum now has regular participation from twenty-six federal civilian and defense agencies.

Mr. Tombe joined US Customs and Border Protection in 2003 and supports CBP with more than 25 years of IT management experience in the Federal Government Sector serving in a variety of Agency’s on both the West and East Coast.

Mr. Tombe holds a Masters Certificate in Project Management from George Washington University and is a Certified Project Management Professional (PMP) with the Project Management Institute. Mr. Tombe is a recipient of multiple awards in the areas of Cloud Computing, SOA, Technology Innovation and an accomplished public speaker having presented at numerous public forums on issues pertaining to technology implementation and management in the US Federal Government.
Brand Niemann

Brand Niemann, former Senior Enterprise Architect and Data Scientist with the US EPA, completed 30 years of federal service in 2010. Since then he has worked as a data scientist for a number of organizations, produced data science products for a large number of data sets, and published data stories for Federal Computer Week, Semantic Community and AOL Government. Brand founded the Federal SOA CoP for the Federal CIO Council in 2006.

Brand recently completed 30 years of Federal service. He worked on assignment for the Federal CIO Council during 2002-2007 on a series of assignments: Founding Chair of the Web Services Working Group, Semantic Interoperability Community of Practice, and Federal SOA Community of Practice, and as Executive Secretariat of the Best Practices Committee. He has written an online book “A New Enterprise Information Architecture and Data Management Strategy for the U.S. EPA and the Federal Government”, published a paper entitled “Put My EPA Desktop in the Cloud to Support the Open Government Directive” and Data.gov/semantic (in response to Vivek Kundra's call), and implemented A Gov 2.0 Platform for Open Government in a Data Science Library (in response to Aneesh Chopra's call). Recently he has used two tools in the Amazon Cloud (Mindtouch and Spotfire) to extract, transform, and load a number of EPA and Federal databases to produce more transparent, open, and collaborative business analytics applications. See [http://semanticommunity.info/](http://semanticommunity.info/).

Eric Little

Eric Little is currently Director of Information Management at Orbis Technologies, Inc., in Orlando, FL. He received a Ph.D. in Philosophy and Cognitive Science in 2002 from the University at Buffalo, State University of New York. He later received a Post-Doctoral Fellow in the University at Buffalo's Department of Industrial Engineering developing ontologies for multisource information fusion applications (2002-04). Dr. Little then worked for several years as Assistant Professor of Doctoral Studies in Health Policy & Education and Director of the Center for Ontology and Interdisciplinary Studies at D'Youville College, Buffalo, NY (2004-2009). He left academia in 2009 to work as Chief Knowledge Engineer at the Computer Task Group (CTG) before joining Orbis.

His areas of specialization are: ontology, knowledge management, cognitive science, philosophy of mind/neuroscience, phenomenology and organizational theory. Dr. Little has designed and helped to implement formal ontologies for use in various applied domains including: biomedicine, medical device manufacturing, medical fraud, waste and abuse detection, pharmaceuticals, medical management, threat prediction/mitigation, disaster management, national defense/intelligence, steel production and petrochemicals. He has published in the areas of philosophy, cognitive science, ontology, information fusion, and human factors engineering. He has delivered lectures on ontology, philosophy, biomedicine, and cognitive science at numerous locations in Germany, Canada, Italy, United Kingdom and throughout the U.S. His research has been funded by The U.S. Air Force Office of Scientific Research (AFOSR), Development and Research for the Defense of Canada (DRDC)-Valcartier, Lockheed-Martin Corp., MIT-Lincoln Laboratories, The National Institute of Standards (NIST), the National Center for Ontology Research (NCOR), The U.S. Army Research Labs (ARL), the Boeing Corporation, British Petroleum, and the Computer Task Group (CTG). He is currently a co-chair of the Central Florida Semantic Web Meet-Up Group, as well as a member of the National Center for Ontology Research (NCOR), the National Center for Multisource Information Fusion (NCMIF), the federal government’s Geospatial Ontology Community, and numerous other international semantic communities. He has served on the scientific/review committees for various journals and international conferences.

Victor Pollara

Dr. Pollara is a Senior Principal Scientist at Noblis’ in the Health Innovation mission area. He applies several decades of experience in theoretical computer science, bioinformatics, knowledge extraction from text, and algorithm design to...
develop computational solutions for complex, data-driven problems. His current work is focused on applying formal modeling and semantic technologies to large, heterogeneous data sets and experimenting with Noblis’ Cray XMT2 as a multi-billion triplestore server.

Before joining Noblis, Dr. Pollara served as the lead bioinformatics scientist for the bioinformatics services company, Viaken, Inc. He was the project leader of a major bioinformatics infrastructure project for Corning, Inc., which included, software consulting, design of an enterprise-spanning database schema, and high-performance bioinformatics computing. Dr. Pollara worked as a bioinformatics scientist at the M.I.T./Whitehead Institute’s Center for Genomic Research. For the Human Genome Project, he focused on the computational challenge of cataloguing all repeated sequences in the human genome. He also designed, programmed, and refined a novel suite of software for SNP (single nucleotide polymorphism) discovery, used for production at Whitehead/M.I.T. for the SNP consortium project.

He holds a doctorate from the Technical University of Braunschweig, Germany, where his research involved the mathematical semantics of parallel constructs in programming languages. He has taught courses in formal languages, algorithms, complexity theory, compiler construction, and assembly language.

Kate Goodier

Ms. Goodier is a senior engineering consultant for the STRATIS division of L-3 Communications. She has more than 20 years of experience in the technical program management and systems development team leadership for both industry and the intelligence community. In addition to technical program management and management support, she has extensive systems engineering and integration experience with in large ACAT I programs. She maintains sponsored accounts in the Joint requirements Oversight Council (JROC) and other knowledge-bases. Ms. Goodier was fifth employee hired at the Center for Information Protection for Dept. of Treasury, FBI, and CIA. She was recognized by the Federal Enterprise Architecture (FEA) Program Management Office (PMO) as an expert in system Data Engineering and developed the Data Reference Model (DRM) version 1.5 Data Description guidance for the FEA. She is a member of the Scientific Committee for the Semantic Technologies in Intelligence, Defense and Security community.
George Strawn

Director, National Coordination Office, Networking and Information Technology Research and Development Program and Co-Chair, NITRD Subcommittee on Networking and Information Technology Research and Development, National Science and Technology Council Committee on Technology

Dr. George O. Strawn is the Director of the National Coordination Office (NCO) for the Federal government’s multiagency Networking and Information Technology Research and Development (NITRD) Program. He also serves as the Co-Chair of the NITRD Subcommittee of the National Science and Technology Council. The NCO reports to the Office of Science and Technology Policy (OSTP) within the Executive Office of the President.

Dr. Strawn is on assignment to the NCO from the National Science Foundation (NSF), where he most recently served as Chief Information Officer (CIO). As the CIO for NSF, he guided the agency in the development and design of innovative information technology, working to enable the NSF staff and the international community of scientists, engineers, and educators to improve business practices and pursue new methods of scientific communication, collaboration, and decision-making.

Prior to his appointment as NSF CIO, Dr. Strawn served as the executive officer of the NSF Directorate for Computer and Information Science and Engineering (CISE) and as Acting Assistant Director for CISE. Previously, Dr. Strawn had served as the Director of the CISE Division of Advanced Networking Infrastructure and Research, where he led NSF’s efforts in the Presidential Next Generation Internet Initiative. During his years at NSF, Dr. Strawn was an active participant in activities of the interagency IT R&D program that is now called NITRD.

Prior to coming to NSF, Dr. Strawn was a Computer Science faculty member at Iowa State University (ISU) for a number of years. He also served there as Director of the ISU Computation Center and Chair of the ISU Computer Science Department. Under his leadership, ISU became a charter member of MIDNET, a regional NSFNET network; he led the creation of a thousand-workstation academic system based on an extension of the MIT Athena system; and the ISU Computer Science department was accredited by the then-new Computer Science Accreditation Board.

Dr. Strawn received his Ph.D. in Mathematics from Iowa State University and his BA Magna Cum Laude in Mathematics and Physics from Cornell College.
Gadi Ben-Yehuda is the Director of Innovation and Social Media for The Center for the Business of Government.

Mr. Ben-Yehuda has worked on the Web since 1994, when he received an email from Maya Angelou through his first Web site. He has an MFA in poetry from American University, has taught writing at Howard University, and has worked in Washington, DC, for nonprofits, lobbying organizations, Fleishman-Hillard Global Communications, and Al Gore’s presidential campaign.

Prior to his current position, Gadi was a Web Strategist for the District of Columbia's Office of the chief Technology Officer (OCTO). Additionally, Gadi has taught creative, expository, and Web writing for more than 10 years to university students, private-sector professionals, and soldiers, including Marines at the Barracks at 8th and I in Washington, DC. (The lattermost by far the most disciplined.)

You can follow Gadi on Twitter, read his columns on Huffington Post, and see his posts on GovLoop, and read his blog entries on the IBM Center for the Business of Government site.
Steve Reinhardt is the chief solution architect at Yarc Data (a Cray company), blending semantic-graph technologies such as RDF and SPARQL with analysis of the structure of graphs to solve customer problems. He has led or co-led development of parallel systems and runtimes from Cray Research’s T3D and T3E distributed memory systems to SGI’s Altix and Altix UV globally-addressable systems to Interactive Supercomputing’s Star-P parallel implementation of the M language of MATLAB to the graph analytic Knowledge Discovery Toolbox for distributed-memory. Currently his primary focus is on bringing the power of complex large-graph analysis, with underlying parallel execution, to subject-matter experts who are not graph-analysis experts.

Graphs of vertices and edges naturally represent many phenomena in today’s world, from social networks to the Internet to metabolic networks. Analysis of such networks to understand key attributes of the interactions has led to the development of the RDF data format and SPARQL query language standards within the World Wide Web Consortium. YarcData, a Cray company, has developed uRiKA, a large-scale graph-analytic appliance based on the RDF and SPARQL standards, which delivers differentiated performance for SPARQL queries on very large data. This talk introduces RDF and SPARQL and demonstrates solving problems with them.
Mark Guiton serves as Director, Government Relations, responsible for working with federal executive and legislative branch officials on a variety of program, policy and procurement issues as it relates to advanced computing. Prior to joining Cray, Mr. Guiton served as legislative director in the U.S. Congress from 1999 to 2003 with a focus on appropriations and technology matters. From 1995 to 1998, he served as a technology policy advisor working closely with the House Government Management, Information and Technology subcommittee from 1995 to 1998. Before working in Congress, he was a computer programmer/analyst for Shared Medical Systems Corporation (now Siemens). Mr. Guiton received a B.S. in computer science with a concentration in electrical engineering from the University of Scranton, Pennsylvania.
Welcome & Introduction

Jeanne Vasterling, Enterprise Modernization & Transformation Practices, MITRE Slides

Welcome to the 14th conference. Acknowledge Planning Committee - stand for applause. Gabe Galvin has been promoted to the director of the MITRE programs for the VA and hopes to join us.

MITRE is a FFR&D Center, not-for-profit, that provides strategic views of government programs, science and engineering, etc., for solving complex problems, CMS Health (NEW), Homeland Security, Enterprise Modernization, Aviation, etc. MY NOTE: CMS Health is Big Data that our Data Science Teams is working with and is the Pilots presented this afternoon!

How to build services that MITRE clients can use and reuse? Speakers will address that need.

Introduce Ajay Budhraja, Chief Technology Officer (CTO), U.S. Department of Justice, IR. Co-chair of Federal SOA CoP

Keynote

Government Trends and Innovation with Service Reuse, Cloud and Big Data – Ajay Budhraja, Chief Technology Officer (CTO), U.S. Department of Justice, IR Forbes ZdNet FedScoop Slides

This is driven by a $200 Billion market. Cloud reduces procurement time and complexity.

Move from acquisition to a data-driven government. Move to hybrid cloud solutions that are agile and dynamic.

Big Data processing and analysis drives real-time decision making (business intelligence analytics) that can deliver data-as-a service for government scientific and business applications.
Federal Digital Government needs SOA, Cloud, and Big Data

AJ's Five Point Plan: Blueprint First

Question: Data Services as tables versus aggregations - which is the most valuable.

**MY NOTE:** Data.gov has about 500,000 "data sets" but only about 1500 high quality data tables. Berman Press publishes over 100,000 government data tables backed by a staff of statistical experts. Government data can be made valuable by statisticians and data scientists. The pilots show the value of both raw and aggregate data services.

---

**Establishing a Service Factory**

Dave Mayo, President and CEO of Everware-CBDI  Slides

State of Government: SOA has become a standard practice. But reuse is very limited. Very few domain service architectures can be scaled up, A project level service promoted to the enterprise is not working - lack of planning, etc.

Sun Tzu: Strategy without tactics is the slowest path to victory. Tactics without strategy is the noise before defeat.

Thus do this as a service factory - inputs and outputs.

Outline
Corporate Profile
Architecture – Engineering Model
What is a Service Factory?
Twin-Track Development
Product Line Engineering
Governance
Service Factory Methods & Tooling
  Service Architecture & Specification
  Model Driven Development
  Agile Methods

Corporate Profile - **Small Business with a Big Footprint**, Knowledge Base, Leader in Applet SOA
Architecture - Engineering Model - Business Mission, Enterprise Architect, etc.
What is a Service Factory? - An organizational construct with a set of tools, models, etc. to address the problem to be solved
Microsoft pushed the Software Factory originally - the software - but Everware-CBDI definition is broader
Solution Provisioning - Provides for the SWIM Lanes
Service Factory Context - Needs Capability Models

Twin-Track Development
Product Line Engineering
Governance
Service Factory Methods & Tooling
  Service Architecture & Specification
  Model Driven Development
  Agile Methods

Slides Notes:
Slide 1
State of SOA: 1. Much progress. 2. BAU – standard strategy. 3. Limited success in consumption/reuse. 4. Few instances of a service architecture. 5. Many single operation services (too fine grained). Leads to inability to locate & effectively consume services. 6. Service Specifications missing. Very few organizations are looking at reuse from the consumer’s perspective: Process (SDLC)/motivation; ambiguity; lack of spec with behavior (black box); risk/dependency; SLA/funding; etc. Promotion of local/tactical services to enterprise/strategic use has been problematic. Development of Service Architectures has focused on arch of the service itself, not the arch of the collaborating set of services within a domain.

Slide 2

This is a fairly advanced topic in SOA. I’m going to present it at an overview level.

BLUF: SF is a mechanism for achieving shared service objectives. It applies discipline, governance, standard processes, automation, etc. to provide services for consumption.

Establishing a service factory is a TACTIC, but it is important to place that tactic within the context of a strategy. So, I am going to spend some time on topics that are outside of the service factory. They provide direction to the factory.

Strategy: do the right things (eg, build the right services)

Tactic: do things right (efficient process)

Sun Tzu: Strategy without tactics is the slowest path to victory. Tactics without strategy is the noise before defeat.

Slide 3

We’re a small company with a big footprint.

KB comes with eLearning: SOA Fundamentals, …

Architecture Services

Enterprise and Solution Architecture

Service Oriented Architecture

Portfolio Management and Planning

SOA Enablement Services

Roadmap, Organization and Governance

Reference Architecture

SOA Education & Certification

Application Modernization & Development Services
Most engineering disciplines adopt this approach – why doesn’t IT?

House analogy. Customer is the owner who wants the house built. Architect translates a set of needs/desires into a balanced design. Engineer applies constraints (eg, the required size of the load bearing beam in the center of the house); creates the detailed blueprint to hand off to the developer (construction contractor) to implement the blueprint.

Not just top down – eg arch investigates other models that have worked to solve similar problems – harvest patterns.

Progressive refinement. Contract based – each role can count on what it gets from higher roles.

A factory takes things in (raw materials, subassemblies, resources), adds value, and produces something – using a standard process and automation tools.

Extended product life: Family based product is likely to be more generic, componentized AND configurable

Today IT is a custom code shop. We BUILD TO ORDER whatever is request with little attention paid to reusing components
Tomorrow we want to be an ASSEMBLE TO ORDER shop – use the enterprise services we have at hand and assemble the "legos" in new ways for the next request.

Need a Design Authority working in conjunction with a PMO to determine what needs to be built and to manage the process.

In the Provisioning, Implementation & Assembly area we have introduced the Legacy Application Reengineering discipline and the parallel Legacy to Service Reengineering discipline to support AM as illustrated in Figure 5.

The focus of these disciplines is to perform the transformation or reengineering of the current assets to meet new Solution Component or Service Implementation requirements.

Prod Lines emphasize building common assets with the ability to customize aspects of them (variability).

Identify things that are the same, but different.

Identify patterns of commonality à identify patterns of variation à predict variations à abstract/design for variability

Connect two points. Line. Now consider the two points are a Consumer/customer with set of requirements and a supplier with set of capabilities. Now there are two more consumers with requirements. Could create two new sets of capabilities. Or could analyze the commonality and discover that some capabilities are common and some are distinct. If you can configure the service with an articulation/configuration point you can meet all requirements with a common configurable service.

Entropy is rampant. Without governance, anything goes. Without controls, things move from structured state to unstructured.

Gov needs to deal with what is provided to the SF (inputs, like architecture) and what the SF does with it (process).
You can’t solve a strategic problem with a tactical solution. I.e., you can’t achieve enterprise service sharing by putting a few services in a registry and hope they get shared.

The term MDA and MDD are often interchanged – we see MDD as more holistic than MDA since MDA is concerned with the definition of a PIM and its conversion to PSM and then to an implementation. However we see MDD as more encompassing and the potential for the models to be used to produce the implementation as well as other needed artifacts for testing and documentation.

While not a new concept, MDD is not widely practiced and unfortunately the agile mantra of ‘working software above all else’ seems to drive teams to ‘code first’ – however in MDD not only are the models the code but they also accelerate the production of working software, facilitate communication about the software, enhance the quality of the software and even facilitate the maintenance and refactoring efforts that often go along with agile development. Oh and they allow you to run an agile practice that actually produces documentation as a natural side effect of the coding since much of the coding is done in model form.

Case Study

The Road to Adoption of Common Services at the IRS – Nitin Naik, Director of Enterprise Architecture, IRS

Long struggle at IRS to do SOA. Glad to be here today ot of the weather to talk about this subject.

IRS is a large, highly complex environment and plays a critical role in the US Government.

IRS has the equivalent of a Y2K every year – All IT that supports filing season must be delivered before next tax season

50 state revenue organizations besides the Federal IRS.
Overview of the IRS Enterprise - mammoth! Systems to collect $2.5-2.8 Trillion, about 100,000 employees, about 800 centers, 3 computing centers, lots of business partners, legacy data, etc.

New Service: Taxpayer Identity Service

Now doing Solution Architecture not just Enterprise Architecture!

Challenges Ahead!

Funding Challenge: The existing Project Based funding model is non-conducive for Shared or Common Services adoption as it leaves little incentive for them to look at the bigger picture, and produce Composable Service Components for future reuse.

Technical Challenges: SOA is a non-trivial, highly technical design paradigm which is not yet widely accepted at IRS. SOA adoption at IRS requires substantial field-guidance; helping the projects make the right architectural decisions, on the spot, at the right time.

Service Development Lifecycle is not incorporated into the existing ELC: Building Common Services requires additional steps and roles beyond what is covered under our standard ELC. The new steps like, Service Inventory Analysis, Service Modeling, Service-Oriented Design, Service Usage Monitoring, Service Discovery and Service Versioning needs to be incorporated in our existing ELC.

Organizational Challenge: Who will be the custodian of a Shared Service once it’s completed and delivered?

Open Architecture

Design for Re-use with Data Services – Edward Ost, Application Integration Technical Director, Talend Slides

Agenda
Identify re-use challenges
Apply SOA principles to layered architecture
Relate process, data, and application integration perspectives
Identify best practice for re-use at each layer
Relate governance to scopes of re-use

Best Practices: Design for Service Re-Use, Design for Process Re-use, Architect for System Re-Use, and Govern for Enterprise Re-Use

Slide Notes
Slide 1

Achieving service re-use in a scalable, high performance manner across multiple lines of business is a difficult challenge that requires balancing necessary variation with standardization for enterprise efficiency. Stakeholders needs from multiple organizations must be considered for managed re-use to be consistent with enterprise policy. Modular SOA
infrastructure can apply technologies such as BPM and ESB using a layered, responsibility driven design focused on information specifications to achieve an Open Architecture. The result is flexible re-use capable of supporting rapid development and deployment necessary for realizing SaaS benefits.

Slide 2

Bazaar is the factory of the Cathedral

Slide 3

Slide 4

Slide 5

Slide 6

Talend conducted a worldwide survey of integration professionals. 236 valid responses were received from around the world.

Slide 7

Slide 8

Talend:

Multiple domains and contexts

Separation of Concerns, Integration of Capability

Slide 6

Usage of Services by processes will vary, ESB provides flexibility via mediation and routing to accommodate different contexts to maximize re-use

BPM orchestrates Business Activities

Slide 7

Slide 8

Slide 9

Slide 10

Slide 11

Use ESB to loosely couple Services into Business Activity in an event driven manner using Choreography to accommodate variations in context

http://semanticommunity.info/Federal_SOA/14th_SOA_for_E-Government_Conference_October_2_2012
Updated: Wed, 23 Sep 2015 06:05:39 GMT
Powered by mindtouch™

40
• BPM delegates routing and connectivity to ESB
• Events decouple services and processes
• Routing slips decouple events and routing
• ESB splits, correlates, routes, and aggregates messages
• Data Services provide transforms and enrichment

Slide 12

Talend:

Wrap external interfaces when necessary

Refactor incrementally

Slide 13

Variation of processes can be managed with subprocesses

As number and complexity of process interactions change, the need for event driven process interfaces grows

Refactor incrementally to event driven process interfaces

Integrate ESB to BPM style Message events to isolate change

In this example all variation is encapsulate within the Request Fund Transfer, this is not always the case

When there is too much variation to encpasulate in a sub-process, encapsulate routing in a routing slip that allows variation across organizations while enforcing data constraints at targeted events

Slide 14

Hospital check-in: ER versus regular visit

Slide 15

Decouple process functional flow from asynchronous coordination

Decouple transport from routing logic

Decouple data format from message interface

Slide 16

Slide 17

Slide 18

Slide 19
MDM is a data-centric enterprise service that transforms data into trusted information

Not all data is mastered

Not all data can be mastered in system of record for some interactions consistent with latency constraints

Decouple policy enforcement from data flow

Allow local and enterprise policies to co-exist

These trends track together

Historic development in industry

Business Value

Complexity of solution

Scope of domains

Scope of organizational ecosystem
There is a fundamental tension between the technology supply chain and the solution supply chain. Both business and complexity conspire to make re-use more difficult. Not just standards but modularity are needed.

There is a fundamental tension between the preferred top-down governance and individual programs that needs to be balanced by the Community of Practice.

Are we building the product right for all stakeholders?

How do we collaborate with other information supply chain partners? Frequent test

How do we collaborate with multiple service consumers? Frequent test

Have we build the right product? Frequent Release

Multiple organizations

Multiple technology domains: OS, platform, application server, database, security, gui,

Multiple business domains: sales, marketing, finance, logistics

Many technology marketplaces within the bazaar
Overview of Exhibitors – SOA Tools Slides

1 Dovel Slides
Jason Bloomberg

2 Everware-CBDI Slides
John Butler

3 Semantic Community Slides
Brand Niemann

4 Redhat / FuseSource Slides
Michelle Davis

5 IBM Slides
Thomas Hall

6 Oracle Slides
David Webber

7 Software AG Slides
Manmohan Gupta
8 Talend Slides (none)

Edward Ost

**Afternoon Keynote**

Delivering Mission Agility Through Agile SOA Governance – Wolf Tombe (Jeffrey Hall), Chief Technology Officer (CTO), at U.S. Customs and Border Protection [Slides](#)

CBP is the oldest and the largest of DHS components

CBP processes millions of transactions and facilitates billions of data searches each day

Slide 1

Slide 2

Each day CBP is committed to improve border security, increase efficiencies, facilitate flow of legal trade and travel through our nation’s borders and ports of entry.

Some statistics on typical day include

- 6.5+ Petabytes of Raw Storage, 650 terabytes of operational data
- 40-50 billion data requests/transactions daily just on CBP
- 2,000 servers (mainframe, Unix, and Windows) at the National Data Center (NDC) plus 1200 servers in the field
- Over 15 million messages & Billions of records processed daily

Slide 3

To handle its day to day responsibilities CBP must maintain

- 24x7x365 systems
- with zero down time
- handle demand surge
- provide accurate, timely and verifiable data to users in seconds

Slide 4

Other statistics include

- Apprehensions which is a key indicator of illegal immigration decreased to 340252
- Enrolled nearly 290000 new travellers

Slide 5

- Secure America's borders by deploying largest law enforcement work force to protect at and between ports of entry
- Facilitate trade while enforcing U.S. trade laws that protect the economy, the health and the safety of the American people
- Facilitate travel at land, air and sea ports while assuring that individuals who have ties to terrorism or a criminal background are barred from entry
- Enforce hundreds of U.S. regulations, including immigration and drug laws
SOA Pilots

Federation of SOA and Semantic Medline introduced by Brand Niemann, Federal SOA CoP, Slides

Provides context for pilots, introduces agenda and presenters and panelists, and seed questions. See Slides

Semantic Information Integration within the Healthcare Sector – Eric Little, Orbis Technologies Slides

Simplifying Semantics to Query Relevant Information - SPARQL under the hood

Utilizing Semantics for Integration of Data Across Multiple Sources - Many companies are doing it manually

Summary

Enterprise-wide applications require unique providers of several integrated technologies
  • Requires several items: SOA, cloud computing, semantic technology, information fusion, cyber security, etc.

Need to provide real differentiators around cloud and semantic technology
  • Differentiators can help drive rapid development that brings about real change in your environment
  • Need to empower users by removing the complexities from the technology approach
  • Can improve security and auditing through the use of technology itself.

Using Semantic Medline on the New Cray Graph Computer for Medical Research – Victor Pollara, Noblis Slides

Brought NLM (Tom Rindflesch) together with Noblis (Victor Pollara) which recently acquired the new Cray Graph Computer

Conclusion

We believe that the XMT2 shows potential as a platform for providing semantic services on large semantic data sets

Over the next 12 months we will build a variety of services and test their utility and responsiveness

The internal graph representation is extensible in ways that could simultaneously support the logical queries of SPARQL and analytical methods that use other graph properties

This would enable us to tackle a much wider class of real world graph problems and we will build a catalogue of these problems and describe how this computational resource can be part of their solution

Panel Discussion

Big Data and the Government Enterprise Slides
Kate Goodier (Moderator, IC), and Gadi Ben-Yehuda (IBM) for Dr. George Strawn (OSTP/NITRD/NCO), Dr. Eric Little (Orbis Technologies), Dr. Victor Pollara (Noblis), Mark Guiton (Cray) for Steve Reinhardt (Cray), and Dr. Tom Rindflesch (NLM)

Gadi Ben-Yehuda IBM talking about Watson & helping new HHS CTO with big data problems to make data more useful and usable by data scientists at Code-a-Thons

Question: What about multiple languages?

Answer: Recorded Future
Wiki Blog
Brand Niemann, Federal SOA CoP

See the Above and Below

Most advanced semantic platform is Be Informed

Most difficult problem is unstructured and structured data

In memory processing is cheaper and faster than Hadoop

Need thousands of new data scientists to deal with data

---

Big Data Fall Forum 2012

Date: October 4, 2012
Event start time: 1:00 PM
Location: Grand Hyatt Washington

MY NOTE: Cancelled until further notice
Register Now for the October 4 ACT-IAC Big Data Forum!

Join ACT-IAC and colleagues in working sessions to discuss solutions and ideas for why so many agencies are talking about the value of Big Data, yet so few are actually building systems to handle the realities and opportunities of Big Data. Our half-day, innovative, professional development format will provide a forum for dialogue-driven workshops, designed specifically around the questions and issues being addressed by government and industry leaders.

Date: October 4, 2012
Location: Grand Hyatt Washington

Agenda
Registration: 12:30 PM – 1:00 PM
Welcoming Remarks: 1:10 PM – 1:25 PM
Opening (Keynote or Panel): 1:30 PM – 2:15 PM
Workshop: 2:20 PM – 3:05 PM
Break 3:15 PM – 3:30 PM
Workshop: 3:30 PM – 4:15 PM
Closing (Keynote or Panel): 4:15 PM – 5:00 PM
Networking Reception: 5:00 PM – 6:00 PM

Overview:
The Big Data Forum can serve as a key opportunity to answer some of the questions outlined here:
Why don’t more government agencies implement systems to handle Big Data?
Are government agencies pulling back because of budget issues or because there is no mandate to spend money on Big Data problems?
Is there a common mission that lends itself to having a Big Data system?
What are the mission-critical issues driving early adopters?

Keynote speaker: Dr. George Strawn, Director, National Coordination Office, Networking and Information Technology Research and Development Program See June 21, 2012: Big Data and The Government Enterprise Slides

MY NOTE: BIG DATA PART II SAVE THE DATE! December 12

http://semanticommunity.info/Federal_SOA/14th_SOA_for_E-Government_Conference_October_2_2012
Updated: Wed, 23 Sep 2015 06:05:39 GMT
Powered by Mindtouch™
After the keynote address attendees will participate in working sessions addressing the topics listed below. Each table will be hosted by a government moderator.

Workshop sessions: our workshops are drawn from the key topics addressed by the Federal Big Data Committee – each participant will select two breakout topics. Participants will engage in a 45-minute, interactive, “not for attribution” dialogues in each workshop, facilitated by a government leader who is actively engaged in the challenges and opportunities of Big Data. Topics include:

**Topic 1 - Large Scale Records Management in Health Care**
**Topic 2 - Cybersecurity**
**Topic 3 - Addressing Large Scale Fraud, Waste & Abuse in Federal Procurement and Entitlements Programs**
**Topic 4 - Conducting Extreme Scale Semantic Data Analysis**
**Topic 5 - Social Media Data Management**

(Additional information on the workshop sessions can be found below.)

During the registration process please select one topic from session 1 and one topic from session 2.

Sponsorship opportunities are available – please contact Judy Fry at jfry@actgov.org for more information.

Additional information on the workshop topics:

**Large Scale Records Management in Health Care** - The inaugural class of Presidential Innovation Fellows is focused on six key initiatives, and none have more potential impact to the economic and medical prosperity of the country than the Open Data Team. The Open Data Initiatives program aims to “liberate” government and voluntarily-contributed corporate data to fuel entrepreneurship creating jobs and improve the lives of Americans. Decades ago, the National Oceanic and Atmospheric Administration began making weather data available for free electronic download by anyone and these data were used to create weather newscasts, websites, mobile applications, insurance, and much more. More recently, the Health Data Initiative, launched by the Institute of Medicine and the U.S. Department of Health and Human Services in 2010, has opened growing amounts of health-related knowledge and information in computer-readable form from the vaults of the government and publicized the availability of these data to entrepreneurs and innovators. Hundreds of companies and NGOs have utilized these data to develop new products and services that are helping millions of Americans and creating jobs of the future in the process. Managing this growing ecosystem of Big Data will be critical to ensuring citizen privacy and finding creative ways to utilize cloud computing infrastructure to ensure people are not overwhelmed with costs to process this overwhelming volume of data.

**Cybersecurity** - President Obama has declared that the “cyber threat is one of the most serious economic and national security challenges we face as a nation” and that “America's economic prosperity in the 21st century will depend on cybersecurity.” However, the overwhelming growth and global adoption of the internet has resulted in an exponential increase in the number of attack surfaces and vectors. The trends towards TIC, Continuous Monitoring, and increased sharing of information across public-private partnerships means security professionals are going to need to find new ways of securely transmitting, storing, and processing this information to ensure critical awareness is not lost in the flood of information.

**Addressing Large Scale Fraud, Waste & Abuse in Federal Procurement and Entitlements Programs** - The White House reported that the Federal Government makes more than $2 trillion in payments to individuals and a variety of other entities each year. The American Recovery and Reinvestment Act of 2009 (ARRA) provided an opportunity to provide an unprecedented level of transparency into government spending through Recovery.gov and utilizing enhanced techniques for identifying fraud, waste and abuse. Even in an era of declining budgets, payments must be made and the enhanced focus on reducing improper payments, such as those identified here: [http://www.whitehouse.gov/omb/financial_fia_improper](http://www.whitehouse.gov/omb/financial_fia_improper), provides Big Data practitioners a fertile field to practice their craft and save the government money.

**Conducting Extreme Scale Semantic Data Analysis** - A semantic data model describes the meanings of information elements and that study is applied to enable agencies to take advantage of the full spectrum of information at their...
disposal, whether unstructured or structured. At a large scale, solutions utilize semantic analysis to valuable effect to mine the wealth of information and provide insights.

**Social Media Data Management** - Agencies such as the Department of State have shown the value of leveraging social media to engage citizens directly and HowTo.gov was established in part to assist agencies with building social media strategies in conjunction with engagement for Open Government activities. Agencies must ensure the right metrics and methods are applied to tracking engagement and leveraging social media tools.

---

**Gartner Exclaims "uRiKA!"**

**Source:** [http://www.datanami.com/datanami/2012-09-26/gartner_exclaims_urika_.html?featured=top](http://www.datanami.com/datanami/2012-09-26/gartner_exclaims_urika_.html?featured=top)

Ian Armas Foster

Hadoop is synonymous with big data, but perhaps, according to Carl Claunch of Gartner, it should not be. Instead, he suggests an in-memory system like YarcData’s uRiKA might better suited to handle big data graph problems.

According to Claunch, graph problems represent the epitome of big data analysis. The issue is that [graph problems are incredibly unpredictable](http://semanticommunity.info/Federal_SOA/14th_SOA_for_E-Government_Conference_October_2_2012) by nature. In principle, graph problems can be parallelized just like any other.

For example, if one were modeling global wind patterns, one could set aside a node for each particular cube of the globe. The nodes would then be made to interact with each other based on which cubes were next to which other cubes in the model.

Unfortunately, [says Claunch](http://semanticommunity.info/Federal_SOA/14th_SOA_for_E-Government_Conference_October_2_2012), this approach is problematic for several reasons. One of those reasons is that the nodes that take the most time do not necessarily correlate with the more complicated or interesting regions of the graph. “The region of the graph in which the search spends most time,” Claunch wrote "could concentrate in unknown spots or
spread across the full graph. A DBMS designed for known relationships and anticipated requests runs badly if the relationships actually discovered are different, and if requests are continually adapted to what is learned."

Essentially, in order for a Hadoop-like parallelization effort of a graph problem to be effective, it has to be known which relationships it should be picking out. But the whole point of graph problems is to recognize points of interest that were unknown. "When the relationships among data are mysterious, and the nature of the inquiries unknown, no meaningful scheme for partitioning the data is possible."

A practical application of this involves analyzing people’s interactions and actions across a wide network. Taken by itself, no particular action or interaction is suspicious. Added together, however, they may indicate a potential terrorist cell.

For obvious reasons, the US government is interested in big data analysis, particularly Hadoop, to solve the above problem. However, uRiKa may be more efficient. According to Claunch, uRiKa possesses three technologies that help it rise above the challenges presented by graph problems.

"YarcData's Threadstorm chip shows no slowdown under the characteristic zigs and zags of graph-oriented processing. Second, the data is held in-memory in very large system memory configurations, slashing the rate of file accesses. Finally, a global shared memory architecture provides every server in the uRiKa system access to all data."

As noted before, efficient graph processing requires handling unexpected jumps from certain regions of the graph to others. This calls for intense parallelization, "The Threadstorm processor runs 128 threads simultaneously, so that individual threads may wait a long time for memory access from RAM, but enough threads are active so at least one completes an instruction in each cycle."

Running 128 threads simultaneously is clearly an advantage. According to Claunch, other chips only have a few out of 128 active during a given cycle, making the Threadstorm chip a true, well, storm of threads.

But there is no one to say that chip cannot be made available to other systems. So why does that level of parallelization work when other systems, whose essential purpose is to partition and parallelize, come up short? It has a great deal to do with the third technology Claunch listed, the global shared memory architecture. Here, the data is not actually partitioned but shared.

"Employing a single systemwide memory space means data does not need to be partitioned," Claunch wrote "as it must be on MapReduce-based systems like Hadoop. Any thread can dart to any location, following its path through the graph, since all threads can see all data elements. This greatly reduces the imbalances in time that plague graph-oriented processing on Hadoop clusters."

Conceptually, it is easy to understand why a model that can freely interact with itself, where regions are not limited by their proximity, would be ideal. Frequently, however, solutions whose implications are easy to conceptualize are difficult to actually achieve. However, through parallelized threads that are not subject to the limitations of partitioning, Claunch notes that YarcData may have actually achieved it.

Finally, the in-memory portion of uRiKa hypothetically solves the inefficiencies caused by constantly re-accessing the data after shutdowns or referencing far away caches.
“The performance of almost all modern processors is dependent on locality of reference, to exploit very fast but expensive cache memories. When a series of requests to memory are near to one another, the cache may return all but the first of the requests. The first request is slow, as RAM memory is glacially slow in comparison with cache memory.”

The main argument being made here is that traditional divide and conquer methods in computer science are insufficient in solving vast modeling problems. The notion that data can forego partitioning and go straight into the model is a nice, but incredibly difficult to achieve. Claunch is perhaps implying that this kind of innovation is hard to come by, as people are more content hammering away at a certain process to make it faster instead of coming up with a whole new process.

Whatever the case, it is hardly important. What is important is that, for Claunch, uRiKA is an important first step in solving difficult to model graph problems efficiently.

YarcData’s uRiKA Shows Big Data Is More Than Hadoop and Data Warehouses

11 September 2012 ID:G00232737
Analyst(s): Carl Claunch

VIEW SUMMARY

The hype about big data is mostly on Hadoop or data warehouses, but big data involves a much wider and varied set of needs, practices and technologies. We offer recommendations for IT organizations seeking a solution to "graph" problems, including use of the uRiKA graph appliance.

Analysis

To listen to the hype, you might think big data is only about Hadoop, but Gartner deems big data to comprise a wider, more varied set of needs, practices and technologies. The uRiKA graph appliance from YarcData, a new company spinoff of Cray, illustrates our point and is used here to highlight just one of the classes of big data problems that are poorly addressed by traditional systems.

The Nature of Most Enterprise Data

IT groups understood the interrelationships of most of their data — orders connect to customer information, product data, inventory facts and financial figures. IT experts use this knowledge to enable efficient access. Technical staff use a variety of mechanisms in database management systems (DBMSs) and analysis software to optimize access to the data. Indexes speed up access to information expected to be needed. Techniques like aggregation records give high performance for commonly requested information. IT groups can apply such optimization methods to both operational systems and data warehouse (DW) or business intelligence (BI) systems, because they:

- Know the nature of the questions that will be asked about data
- Understand how data elements are related and connected
- Know the parts of the data where they expect most activity (for example, current-year orders versus historical data)
Experts use prior information about data access patterns to define the database, apply indexes or aggregations, spread data across storage devices, and apply the right tools to meet service-level expectations. It is obvious that product data is related to sales data, but building temperatures usually aren't related to sales or product data. Access to financial data clusters close to the current date, and the frequency of requests about past years drops steeply. Customer service representatives are likely to look at payment or charges for one customer. Metadata about access patterns, combined with predicted volumes and response time targets, is used to configure data and systems.

**Most Big Data Situations Are Addressed by MapReduce or Other Traditional Technologies**

Users now face volumes of data too large to work well or fit in traditional DBMSs or DW systems. To deal with the challenge, users looked to techniques invented to host global-scale public Web properties — sites like Google, Yahoo, Facebook and others — whose extreme volume of data must be partitioned and distributed across many servers. The design of most applications on these sites uses a MapReduce model\(^1\) in which the application will:

- Identify (map) which of a very large pool of servers should participate in running each transaction or inquiry
- Pass requests to and receive responses from potentially many servers for each transaction
- Combine responses (reduce) to produce the final result of the inquiry or transaction

MapReduce is the core of Hadoop\(^2\), an open-source project at the Apache Software Foundation. Hadoop includes an execution environment, development tools, file systems, DBMS software, BI tools and other capabilities to build large distributed systems. The effectiveness of the MapReduce model depends on the application mapping a request to just the servers with relevant data — if too many servers are involved, overall performance and capacity are impaired, the servers having the desired data are missed, and the outcome is incorrect. Success depends on the same metadata about access patterns used so effectively in traditional DBMSs and DW systems.

Knowing the related information, likely frequency, relationships and other information, technical staff can allot the huge volume to servers with a partitioning scheme optimized for performance, capacity and expected transaction types. Instead of the data on overloaded servers being serially dug through, properly distributed data can be searched in parallel by many servers, and a response produced in less time. IT groups can easily accommodate information that is too voluminous for traditional systems using a Hadoop-style system if they have solid metadata about access patterns.

**Graph Problems in Big Data Are Different**

Organizations have great difficulty in achieving good, consistent performance with a class of information searches called "graph problems." Information is represented by vertexes on a graph. These nodes are connected by edges, representing some relationship between the nodes. Any two nodes are connected by paths — a series of edges beginning at one vertex and ending at the other — that may pass many intermediate nodes. If one imagines many cities (vertexes) joined by roads, representing edges, many challenges of graph processing become clear. There may be many routes connecting two cities — some are longer or less direct than others. The task of discovering all paths and which are "best" can be difficult, as the number of cities and roads swells. Graph problems can represent many types of relationships as edges connecting vertexes.

In many graph applications, the user wants to discover patterns and connections between a myriad of facts, measure "distances" and paths between nodes, and choose the edge to traverse based on what has been learned up to that
moment. The course of the search may leap large distances, when a relationship is discovered between distant nodes. The region of the graph in which the search spends most time could concentrate in unknown spots or spread across the full graph. A DBMS designed for known relationships and anticipated requests runs badly if the relationships actually discovered are different, and if requests are continually adapted to what is learned.

Examples of graph problems:

- Spotting unrealized connections between actions and people — none suspicious in itself — that represent a coordinated threat to public safety, enabling plots to be blocked
- Finding patterns and correlations in treatments and results to enable health organizations to personalize treatment, improving outcomes for patients and institutions
- Learning unrecognized factors that shift market demand, drive swings in investment prices and alter portfolio risks, fast enough to take corrective action

**Figure 1. Impacts and Top Recommendations for Running Graph-Style Discovery Problems**

<table>
<thead>
<tr>
<th>Impacts</th>
<th>Top Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organizations faced with previously infeasible graph-style discovery problems may succeed using uRiKA.</td>
<td>• Select candidates where processing is graph-oriented, the data is large-scale, and discovery of relationships is the focus.</td>
</tr>
<tr>
<td>To address all their data requirements, IT organizations may be forced to duplicate data between systems such as uRiKA and transactional systems.</td>
<td>• Test the appropriateness of specialized systems and achievability of performance targets.</td>
</tr>
<tr>
<td>• Carefully define the volume and performance requirements for processing data.</td>
<td>• Use the fewest platform types possible.</td>
</tr>
</tbody>
</table>

Source: Gartner (September 2012)

**Impacts and Recommendations**

IT organizations faced with previously infeasible graph-style discovery problems may succeed using a focused solution like uRiKA

**Why Graph-Oriented Problems Behave Pathologically on Traditional Systems**

"Big data" is a broad term, but many of these are graph problems, discovering connections by roaming across a melange of data sources. When the purpose of the system is discovery of relationships, not extracting information from already known interrelations, achieving satisfactory performance is difficult. The key to managing MapReduce workloads is the scheme by which the data is partitioned and assigned to specific servers in the cluster. Every inquiry or other transaction must be mapped to the servers that are relevant to the task, leveraging the scheme by which the data was placed. When the relationships among data are mysterious, and the nature of the inquiries unknown, no meaningful scheme for partitioning the data is possible. The inquiries themselves are likely to have to run on all the servers, and when the trail from one bit of data to a related one takes several hops, each to a different portion of the data housed in another server, the time spent in each server can vary dramatically. The most overloaded server then determines the
response time to the original inquiry or transaction. As a result, it is challenging to pool multiple inquiries together in a way that spreads out the processing, if the pattern of processing is not predictable in advance.

Other challenges arise from graph-type processing due to irregular and unpredictable leaps along a route. The performance of almost all modern processors is dependent on locality of reference, to exploit very fast but expensive cache memories. When a series of requests to memory are near to one another, the cache may return all but the first of the requests. The first request is slow, as RAM memory is glacially slow in comparison with cache memory. Similarly, when a program asks again for data it recently accessed, that data is likely to still be in the cache, available with little delay. Thus, locality of reference in time and in area allows a large majority of memory requests to occur at cache speed, masking the impact of much slower RAM chips. Caches in disk drives, storage systems and server memory achieve a similar effective speedup over the access times of disk drives.

When graph problems are processed, the irregular pattern means a lessened locality of reference in time. The large leaps in unpredictable directions along the graph mean lessened locality of reference in area. Both cause the effective performance of the processor and the storage systems to decline, sometimes substantially, as requests are now more likely to require the full delay of RAM or disk access times, since they are not in the caches.

A Design Accommodating Graph Problem Peculiarities Works Where Classical Approaches Fail

YarcData has designed uRiKA with three technologies to minimize or eliminate the costs of the irregular, unpredictable leaps in graph processing. A unique approach to processor design, YarcData's Threadstorm chip, shows no slowdown under the characteristic zigs and zags of graph-oriented processing. Second, the data is held in-memory in very large system memory configurations, slashing the rate of file accesses. Finally, a global shared memory architecture provides every server in the uRiKA system access to all data.

The Threadstorm processor runs 128 threads simultaneously, so that individual threads may wait a long time for memory access from RAM, but enough threads are active so at least one completes an instruction in each cycle. All 128 are active simultaneously, unlike other chips where only a few threads are active in one cycle; the threads all move at a moderate rate that is insensitive to locality of reference.

Employing a single systemwide memory space means data does not need to be partitioned, as it must be on MapReduce-based systems like Hadoop. Any thread can dart to any location, following its path through the graph, since all threads can see all data elements. This greatly reduces the imbalances in time that plague graph-oriented processing on Hadoop clusters.

Pick the Right Tools Based on the Nature of Your Big Data Problem

Thus, uRiKA is a system designed for the characteristics of graph problems. This example underscores one of the ways that big data can be more than just huge quantities, and that IT organizations may need unfamiliar or novel technologies when they face unique big data situations. uRiKA is not the general solution to all big data challenges, nor is it the only technology that might work adequately with a graph-oriented application, but it does prove that Hadoop-style systems are not the universal tool for big data.

Recommendations:
• Survey opportunities across the business for leveraging discoveries from graph-oriented processing into meaningful business advantages.
• Select candidates to place on uRiKA where processing is graph-oriented, the scale of the data is large, and discovery of relationships is a core focus of the work.
• Validate the appropriateness of specialized systems and the achievability of performance targets with proof-of-concept and pilot tests.

**To address all their data requirements, IT organizations may be forced to duplicate data between systems such as uRiKA and transactional systems**

Data sources may be used for multiple purposes, and the system that best addresses each purpose is different. This can cause organizations to have to duplicate data across these islands; this can multiply the costs of storing big data volumes in one location. Some portion of the data that is searched to discover new relationships may be best placed on a system optimized for graph-oriented processing, such as uRiKA, but that data may also be used in operational, transactional systems whose needs are far better addressed by traditional DBMSs and analytics software. Other parts of the enterprise's data may be cost-effective to host and search only on large Hadoop-style clusters, searching for information using well-understood relationships, yet this may also be a source for the discovery tasks that are graph-oriented in nature.

The same information might be transactionally processed on operational systems, with a copy placed in data warehouses to extract BI with the mature and powerful tools available for those purposes, another copy pumped into a Hadoop-style cluster for very large scale inquiries, and yet a third copy streamed into a graph-processing-optimized system. However, the inflated costs that come from all that duplication, the increased complexities of managing multiple technology islands and the downsides of establishing isolated islands are serious disincentives.

For many, the optimization from discrete approaches may not be worth the ramped-up costs and other impacts. IT departments may select a few technology types to handle all requirements —accepting performance or processing rate limitations to avoid the costs of too much diversity. IT departments that can meet their raw scale or performance levels only by adopting the platform right for the task will have to bear the increased costs that ensue.

**Recommendations:**

• Carefully define the volume and performance requirements for all types of processing required against data.
• Calculate the impacts of duplication, complication and isolation for each potential additional technology platform to be implemented.
• When the requirements cannot be met without diversification, build plans around the optimized platform type.
• When the requirements can be achieved on a compromise employing fewer machine types, as long as the economics make sense, use the smallest number of platform types possible.