This Symposium brought to you by

www.ttcus.com

Linkedin/ Group:
Technology Training Corporation

@Techtrain

Technology Training Corporation

www.ttcus.com
Big Data and Deep Analytics at NASA Langley Research Center

Manjula Ambur, Chief Knowledge Officer
Ed McLamey, Chief Technology Officer for IT
Office of Chief Information Officer
NASA Langley Research Center
September 2013
Outline

- NASA Langley Overview
- Big Data and Deep Analytics possibilities
- Langley Initiatives
  - IBM Watson Technologies and Content Analytics
- Fourth Paradigm of Scientific Discovery
  - Example Use Cases
- Big Data Strategy
- Comprehensive Digital Transformation
  - 20 year horizon
- Questions and Discussion
“To reach for new heights and reveal the unknown so that what we do and learn will benefit all humankind”
NASA Langley at a Glance (2012)

Founded in 1917
1st civil aeronautical research lab

~$853m PY2012 Budget
~$823m NASA Langley budget
~$30m External business

~3,600 Workforce
~1,900 Civil Servants
~1,700 Contractors (on/near-site)
(~250 students)

Langley’s Economic Impact (2011)
National economic output of ~ $2b and generates over 17,000 high-tech jobs
Virginia economic output of ~ $1b and generates over 9,000 high-tech jobs

Infrastructure/Facilities
788 acres, 169 Buildings
~$3.3b replacement value

Aeronautics $192m
Science $122m
Space Tech $84m
Human Exploration $47m
Education $12m

Cross-Agency Support Program & Construction/Environmental Compliance & Restoration
- Center Management & Operations
- Agency Management & Operations
- Construction/Environmental Compliance & Restoration
Mission: NASA Langley is a research, science, technology and development center that provides game-changing innovations to enable NASA to make significant contributions to the nation.

**Aerosciences**
Research for Flight in All Atmospheres

**Aerospace Systems Analysis**
Entry, Descent & Landing

**Characterization of all Atmospheres**
(Lasers & LIDAR)

**Aerospace Structural & Material Concepts**
Big Data and Deep Analytics Possibilities
Big Data and Big Analytics
Data Deluge

Scientific Data

Digital Media

VIDEO

BLOGS

MOBILE

IM

Human Sensors

Public

Social

Person

Source: Sajal Das, Keith Marzullo

Health Care

Source: Biomechanics, Department of Biomechanical Engineering, University of Wisconsin
ABC’s of Big Data

**Analytics** - To gain insight and take advantage of the digital universe, and turning it into information and intensive workloads at really high speeds.

**Bandwidth** - Performance for data-intensive workloads.

**Content** - Boundless secure scalable data storage to write it, can find it, and keep it forever – ‘Knowledge base’ or ‘Data Lake’.
Promising Possibilities

- Data intensive science appears to be revolutionary science; Data Analytics and big data are major opportunities.

- Big Data will continue to provide the basis for new services perhaps as important as the web, Google and Facebook.

- Big Data and Data Scientists are poised to be key areas of expertise for the next 20 years.

- Active multimillion dollar (~$600M) research in many areas.

- Dr. Watson in becoming reality with Well Bearing and Sloan Kettering cancer institute.

- Incredible force multiplier for science, engineering & research worldwide.

We need to leverage this for transforming NASA work.
IBM Watson Technologies- NASA Langley Journey......
NEW ERA OF COGNITIVE COMPUTING

- Discovery
- Probabilistic
- Big Data
- Natural language
- Intelligent options
Langley Watson Journey So Far

**Investigation: 2011**
- 2011: Center wide IBM Watson Seminar
- 2011: Center Workshop: Developed Use cases

**Prototype: 2012**
- Visit to Watson lab and discussions with IBM Experts; Selected Content Analytics as a starting point
- Prototype with IBM Content Analytics (ICA): 3 use cases; collaboration with mission organizations

**Pilot: 2013**
- Pilot with ICA: 4 use cases
- Workshop to explore the next steps: Focus on Knowledgeability and Innovation
  - Our Chief Technologist, Scientist and Engineer, subject matter experts with IBM experts in Watson and NLP technologies

**Capability: 2013**
- ICA content analytics being offered as part of our capabilities
- Knowledge Assistant Pilot being formulated in specific discipline with Q and A capability - beginnings of “NASA Watson”
Use Cases - Prototype

NTSB Accident Reports Analysis

Sonic Boom Search and Analysis

Time Series View
(Approx. 3500 Documents)

Bird Names Joined with Impact Synonyms
(Frequencies Spike in April and August) Why?

20 clusters of 1500 documents clustered based on machine learning algorithm.
Other Uses Cases

- **Finding Business Opportunities from FBO.gov**
  - Analyze data looking for opportunities (current, emerging and niche), trends, experts in chosen technology areas. Data Source: 120,000 XML current and archived records

- **Analysis of Flutter research**
  - Concept search and cluster analysis, classification and identifying entries with potential data from NASA records

- **Analysis of Langley Publications**
  - Concept search, pattern analysis and classification of publications in structures and materials areas from 14,350 publications/reports

- **Research Opportunities in Autonomy areas**
  - Subject search and analysis, trends, experts and opportunities (current, emerging and niche) in fields related to autonomy; 1,500 records from many different sources (NASD, AIAA, Engineering Village, etc...)
**Purpose:** Enable and Improve Center Knowledgeability

**Current Methodology:**
- A significant amount of time is spent mining for targeted knowledge, manually by SME.
- Data sits unexplored. Connections not made.
- Insights missed.
- A knowledge assistant would serve as a virtual colleague.

**Goals:**
- Keeping up with technical and competitive intelligence
- Making sense quickly: Find wheat in the chaff.
- Identify strategic business opportunities
- Enable cross-discipline innovation
- Identify and connect networks of experts.

**Value:**
- Help/Improve Center Knowledgeability - Market/Competitive/Technical Intelligence
- Identify key trends, emerging experts and expert networks; summaries, alerts, recommendations, non-obvious relationships and intuitive visualization of results
- Give users the ability to ask questions and get answers -- Deep Q&A
Data Intensive Scientific Discovery – Potential for NASA Langley
Emergence of a Fourth Research Paradigm
Data Intensive Scientific Discovery

- Thousand years ago - **Experimental Science**
  - Description of natural phenomena; Galileo

- Last few hundred years - **Theoretical Science**
  - Newton’s Laws, Maxwell’s Equations

- Last few decades - **Computational Science**
  - Simulation of complex phenomena

- Today - **Data Intensive Science**
  - Data from Instruments, simulations, sensors and satellites
  - Set of tools and technologies
    - For data mining, analysis and visualization
    - For collaboration and dissemination

Massive Climate Research Data - Opportunity

[Image of Earth with various research data points]
Non Destructive Evaluation
4-D Imagery Reconstruction and analysis of X-Ray CT Scans

**Purpose:** Increase scientific understanding of material properties and how pressure causes fracturing and material distortions

**Current Methodology:**
- The current process to find flaws or anomalies in the material is entirely visual assessment and requires SMEs
- Only a fraction of the data taken can be evaluated, roughly 10% are currently analyzed
- A significant amount of time is required for computational volume alignment when multiple data sets are acquired on same

**Goals:**
- Utilize Big Data mining capability with an algorithm capable of analyzing and detecting the anomalies in multiple types of images
- Identifying key features such as flaws or other out of specification indications
- Reconstruct the key features of the images within a single 4-D environment that is capable of being viewed from multiple angles

**Value:**
- Big data mining capabilities will allow new forms of analysis and the potential for insights that are currently impossible from completely manual analysis techniques
- All of the data collected will be used in the analysis
- Analysis time will be significantly reduced
- SMEs/Analysts will be able to redirect their efforts to other tasks, improving productivity
- Efforts could benefit other scientists performing data analysis
**Aeroelasticity**

*Assessing the Causes of Flutter Effects from Wind Tunnel Testing*

**Purpose:** Increase scientific understanding of variables that cause flutter effects on the wing of an aircraft

**Current Methodology:**
- Current methodology assesses flutter characteristics at prespecified points using Matlab
- Current methods do not assess flutter over time or one or several variables are changing
- Data is not currently treated as a time series and therefore only a fraction of the total data available can be assessed

**Goals:**
- Let the data help determine the variables that predict flutter effects
- Increased understanding of the interaction between pressure on various points on the wing of a plane and flutter effects
- Create better tools to visualize the properties that predict flutter

**Value:**
- Big data mining capabilities will allow new forms of analysis
- **Potential new insights from analyzing multiple wind tunnel data sets simultaneously**
- Allow all of the wind tunnel data to be used in the analysis
- Analysis time will be significantly reduced
- SMES/Analysts will be able to redirect their efforts to other tasks, improving productivity
Big data, Deep Analytics, and Machine Intelligence Strategy

**OBJECTIVE:** Enable NASA employees to utilize and apply these transformational technologies as force multipliers for scientific and engineering discoveries and systems innovation and optimization.
Researchers, Engineers and Project Teams have ‘Virtual Expert(s)/Colleague(s) at their disposal that can
- Answer specific questions
- Synthesize & makes sense of volumes of big data/information
- Processes modeling & simulation data in real time
- Provides predictions for new technologies and design configurations

Human cognition and machine cognition augmenting each other providing unimaginable capabilities
Approach

- Big Data Team with IT experts and end users (Researchers and Engineers)
- Understand the possible value to our disciplines/missions
  - Develop use cases for now and future
  - Collaborate with discipline experts
- Understand and investigate current technologies and tools and applicability to us. Include open source tools
- Identify and Learn about the research initiatives – DARPA; IARPA; NSF; NIST; Oak Ridge; ……
- Start seeking partnerships – NASA; Federal; Industry; Universities…. 
- Work simultaneously on prototypes/pilots and long term strategy as they are synergistic
Challenge Problems

- Provide automated findings, analysis, visualization, and insights generated from relevant data/information from massive, distributed data sources worldwide: 30 minutes

- Identify emerging technology trends from actively seeking and analyzing web and other content and provide reports with evidence on a regular basis (Every 3-6 months)

- Analyze and visualize massive volumes of static and streaming data with sub-millisecond response times to take action in real-time

- Provide precise and accurate answers to specific natural language questions in 1 minute
Comprehensive Digital Transformation Strategy
A Comprehensive Digital Transformation: The Future of R&D at Langley Research Center

**Vision:**
Strategically position Langley Research Center to maximize relevant, innovative and persistent contributions to NASA and the Nation

**Intent:**
• Provide a technical strategy and plan that complements the Facilities and Workforce plans

**Scope:**
• Simulation-based engineering & science (SBES) at the discipline-level and system-level
• High performance computing, data processing, information technology, machine intelligence…
• 20-30 year time horizon
Re-Imagine Work in an Emerging World Where...

- Work location largely ceases to matter
- People seamlessly innovate and invent via virtual collaboration teams, distributed worldwide and even in space
- Knowledge and wisdom created in teams rather than individual minds
- Science & Industry accelerate exponentially thru:
  - Pervasive sensors, Advanced modeling, simulation, analysis, visualization & computing
  - Manufacturing anything via any-material 3D printers
  - Advanced electronic tools continue to augment human capability
  - Miniaturized embedded mobile devices
  - Improved senses - augmented reality displays & more
  - Two-way direct brain interfaces – know/learn anything
  - Computers and humans are cognitive partners
  - Robots faster & better than humans at many physical & cognitive tasks
Trends for NASA Langley in 2050

- **Work from anywhere**
  - Telepresence, tele-operation, tele-exploration; 5-senses virtual reality
  - Benefits – *Now: best minds, morale, agility. Later: facilities savings, cleaner environment*

- **Global workforce**
  - Any nationality; any organization; any location; Social network of scientists
  - Shifting World Demographics – massive scientist population worldwide
  - Benefits – *best minds, less expensive minds, expanded international partnerships*

- **Blending of virtual and physical testing**
  - Rebalancing of virtual and physical
  - Rapid, automated, iterative looping between virtual and physical
  - Advanced computing and modeling & simulation capability
  - Benefits – *maximize research/science/engineering output* (some combination): better, faster, cheaper, safer

- **Leveraging Machine Intelligence**
  - Machine intelligence at human level or smarter
  - Rethink relationship among man and machines
  - Benefits – *maximize human capacity to innovate and think*

- **Human/Computer Augmentation & Symbiosis**
  - Cyborgs; augmented humans
  - Space travel by modified human/machine symbiots
  - Benefits – *maximize human capacity to innovate and explore*
The Question

- How can we (LaRC, NASA, USA, World) best prepare ourselves for the Digital Transformation?
Contacts

- Ms. Manjula Ambur
  manjula.y.ambur@nasa.gov

- Mr. Edward McLamey
  edward.l.mclamey@nasa.gov