Big Data in Healthcare
Roundtable Discussion

Juergen A. Klenk, PhD
NAHDO 27th Annual Conference
October 2012 | New Orleans, LA

This document is Booz Allen proprietary and is intended solely for the information of the audience to which it is delivered
Booz Allen serves clients across many markets, and Big Data is rapidly becoming a key component for all services provided.

### Capabilities¹

- **Strategy and Organization**
  - ~3,000 staff
  - Strategy and change management
  - Human capital, learning and communications
  - Organization and process improvement

- **Analytics**
  - ~6,000 staff
  - Business analytics
  - Mission and performance analytics
  - Intelligence and operations analytics
  - Advanced analytics

- **Technology**
  - ~8,000 staff
  - Cyber technologies
  - Systems engineering and integration
  - Strategic technology and innovation
  - Systems development

- **Engineering and Operations**
  - ~6,000 staff
  - Acquisition and program management
  - Engineering and science
  - Enterprise integration
  - Supply chain and logistics

### Senior Advisors with Operating Experience
Reachback to 25,000+ Experts Across Booz Allen

---

¹ Does not include ~2500 infrastructure staff
The world’s collection of data is growing at an exponential rate

- According to an IDC Digital Universe study, the total digital data created and replicated in 2011 was 1.8 billion terabytes, growing nine times in just five years.

- How much is that?
  - IDC says that to store this information you would need 57.5 billion of 32-gigabyte iPads – enough to erect a 61-foot-high wall running from Miami to Anchorage.


Organizations are drowning in data and are making it a priority to leverage Big Data to support their mission and improve their decision making.
Health HITECH and Meaningful Use: Legislation that promotes Big Data in Health

The Health Information Technology for Economic and Clinical Health (HITECH) Act promotes the adoption and meaningful use of health information technology, and addresses the privacy and security concerns associated with the electronic transmission of health information.
Big Data and Advanced Analytics are inextricably linked, and the “Cloud” is becoming the premier paradigm that is used as a synonym for both.

1. Analytic Capabilities
   - **Data Input**
     - Data Infrastructure Design
     - Requirements Gathering
   - **Data Analysis**
     - Data Mining
     - Statistical Analysis
     - Predictive Modeling
     - Discrete Event Simulation
   - **Data Output**
     - Reporting, Dashboard Design
     - Data Visualization

2. Outputs
   - **Advanced Analytics**
     - Advanced analytic techniques on limited amounts of data/samples
     - Simulations
     - Basic Reporting Dashboards
     - Spreadsheets
     - Forecasts
   - **Cloud Analytics**
     - Scalable, complex analytics to mine large amounts of data
     - Advanced Data Calculations
     - Business Intelligence Systems
     - Data Visualization Exploration
     - Cloud Computing
     - Basic analytic techniques on large amounts of data

3. Problem Space
   - **Payment & Delivery Reform** (e.g., Quality of Care)
   - **Financial Integrity** (e.g., Bad Actors, Fraud/Waste/Abuse)
   - **Scientific Computing** (e.g., Biomedical/Transl. Research)
   - **Surveillance** (e.g., Bio-Threats, Food/Drug Safety)
   - **Real-time Decision Support** (e.g., Personalized Medicine)

Vectors indicate current market trends.
A closer look at some key Big Data sources in Health

<table>
<thead>
<tr>
<th>Domain or Problem Space</th>
<th>Example</th>
<th>Data Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payment &amp; Delivery Reform</td>
<td>Quality of Care</td>
<td>Electronic Health Records, Claims Data, Demographic Data</td>
</tr>
<tr>
<td>Financial Integrity</td>
<td>Fraud, Waste, Abuse</td>
<td>Claims Data, Financial Network Data, Social Media Data</td>
</tr>
<tr>
<td>Scientific Computing</td>
<td>Biomedical Research</td>
<td>Genomic Data, Clinical Data (studies, trials), Scientific Literature</td>
</tr>
<tr>
<td>Surveillance</td>
<td>Bio-Threats</td>
<td>Public Health Data, Social Media Data, Market Data</td>
</tr>
<tr>
<td>Real-time Decision Support</td>
<td>Personalized Medicine</td>
<td>Personalized Med Record, Real-time measurements</td>
</tr>
</tbody>
</table>
A configurable Cloud Based Reference Architecture is required to scalably manage and analyze Big Data.
Core Principles of a Cloud Based Reference Architecture

**In Situ Processing**
Most of the Analytics is done locally where the data are stored to avoid delays due to large data transfers. “Send the question to the data.”

**Use Commodity Hardware**
Hardware must be cheap to scale horizontally and is expected to fail. The architecture supports both scalability and fault tolerance.

**Throw Away Nothing**
Near linear scalable hardware and software systems allow much more data to be stored, which enable re-processing of historical data with new algorithms.

**NoSQL**
All data can be ingested ‘as-is’, without the need to first develop data models, avoiding expensive and slow ETL processes.

**Data Tagging**
All data can be richly tagged and indexed, including management of complex ontologies, for rapid querying.
The Reference Architecture can be implemented using a variety of COTS and Open Source products.

<table>
<thead>
<tr>
<th>Cloud Category</th>
<th>Sample Vendor Landscape</th>
<th>Summary Vendor Perspective(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web-based Service</td>
<td><img src="image" alt="Vendor Landscape" /></td>
<td>- Development of web-based APIs for accessing information / data is becoming more prevalent especially with the advent of more advanced PDAs/phones</td>
</tr>
<tr>
<td>Software as a Service</td>
<td><img src="image" alt="Vendor Landscape" /></td>
<td>- With more advanced Rich Internet Application (RIA) technology, more and more hosted software applications will continue to be offered as services</td>
</tr>
<tr>
<td>Application Component as a Service</td>
<td><img src="image" alt="Vendor Landscape" /></td>
<td>- Core application functionality supporting content management and business process automation is the most prolific</td>
</tr>
<tr>
<td>Software Platforms as a Service</td>
<td><img src="image" alt="Vendor Landscape" /></td>
<td>- Software platforms as a service will move more into production-grade hosting environments in addition to services for developers and testers</td>
</tr>
<tr>
<td>Virtual Infrastructure as a Service</td>
<td><img src="image" alt="Vendor Landscape" /></td>
<td>- Platform vendors will continue to offer transparent, aggregated, scalable, on-demand access to idle hardware resources scattered throughout networks</td>
</tr>
<tr>
<td>Physical Infrastructure as a Service</td>
<td><img src="image" alt="Vendor Landscape" /></td>
<td>- Providing enterprise-grade computing, storage, network infrastructure, and IT operations services will continue to advance and become a mature, competitive market</td>
</tr>
</tbody>
</table>
Big Data Value Proposition: Solving hard and pressing problems

- Sepsis, Severe Sepsis, and Septic Shock (S4) is a national healthcare priority
- Mortality rate: 30-50%
- Significant prevalence: > 750,000 cases diagnosed annually
- High burden of cost: ~ $16.7 billion nationally per year
- A 2007 ruling by the CMS limited payment to hospitals for certain preventable hospital-acquired infections. In 2009, CMS added sepsis to the list of conditions covered by this ruling
- Average total costs per S4 patient is estimated at $13,900 in the general ward, and roughly $29,900 in the ICU
- In addition to preventing and reducing the severity and mortality of S4, evidence-based approaches to the treatment of S4 can help to reduce the treatment costs absorbed by hospitals and other providers

Example: In 2010, Booz Allen started a “Meaningful Use” project to study Sepsis using EHR data

- Using over 27,000 individual patient Electronic Health Records (EHRs) containing both structured and unstructured data spanning across 4 hospitals a 2-year period, we addressed two important research areas:
  - **Compliance Analysis**: Evaluate and measure effectiveness of hospital compliance with SSC guidelines for addressing Severe Sepsis and Septic Shock (S4) by analyzing patient EHR records
  - **Early Detection Analysis**: Mine EHR records for potential clinical indicators that could lead to early detection of S4

![Graph showing Compliance vs. Mortality with Hospital data points and trend line](image)

![Risk Score distribution with Vital Measurements](image)
Conclusions/Discussion

- Significant advancements across all Health domains can be expected from Big Data
  - Improved outcomes, cost savings
  - Reduction of fraud, waste, and abuse
  - Advanced research
  - Improved safety from bio-threats
  - Patient centered care

- Affordable platforms are required to manage, store, analyze, and visualize Big Data
  - Configurable Cloud Reference Architecture

- Analytic techniques must be developed that can leverage Big Data to solve hard problems
  - Within 9 months mortality was reduced from 28% to 14.5% (severe sepsis) and from 47% to 18.5% (septic shock)
  - Integrate ‘continuous monitoring’ system into a Big Data Analytics environment to stream and analyze large quantities of patient data
THANK YOU

Juergen A. Klenk, PhD
8283 Greensboro Drive
McLean, VA 22102
klenk_juergen@bah.com
(703) 377-7205