Big Data and Analytics in Government

Deepak Mohapatra
Sr Consultant, Gov’t Big Data & Analytics
dmohapatra@us.ibm.com
Big Data Myths

- Big Data is primarily about *large datasets*
- We will have to *replace all older systems*
- Big Data is only *Hadoop*
- Older transactional data *does not matter anymore*
- Data warehouses are a thing of the past
- Big Data is only for *internet savvy* companies
- We do not have the need, budget or skills
Why Big Data Now?

The power of ALL data coming together

"Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."

with the power of technology & improved IT economics
Big Data & Analytics add value at the point of impact

- Systems of Record
  - Infrastructure
  - Security Intelligence
  - Enterprise Applications

- Systems of Engagement
  - Mobile Commerce
  - Call Center
  - Social Business

Extend and Integrate

© 2014 International Business Machines Corporation
Industry Movement from Traditional Environments to Cloud
One Step or An Evolution

Many clients are already on the way to cloud with consolidation and virtualization efforts

- **Consolidate**
  - Physical Infrastructure

- **Virtualize**
  - Increase Utilization

- **Standardize**
  - Operational Efficiency

- **Automate**
  - Flexible delivery & Self Service

- **Shared Resources**
  - Common workload profiles

- **Cloud**
  - Dynamic provisioning for workloads

Ready the Infrastructure
Keeping up with changing global and industry regulations

- **Russia:**
  - Computerization & Protection of Information / Participation in Int'l Info Exchange

- **Japan:**
  - Guidelines for the Protection of Computer Processed Personal Data

- **Korea:**
  - 3 Acts for Financial Data Privacy

- **Taiwan:**
  - Computer-Processed Personal Data Protection Law

- **Singapore:**
  - Monetary Authority of Singapore Act

- **Vietnam:**
  - Banking Law

- **New Zealand:**
  - Privacy Act

- **Philippines:**
  - Secrecy of Bank Deposit Act

- **Australia:**
  - Federal Privacy Amendment Bill

- **United Kingdom:**
  - Data Protection Act

- **Germany:**
  - Federal Data Protection Act & State Laws

- **Poland:**
  - Polish Constitution

- **China:**
  - Commercial Banking Law

- **France:**
  - Protection of Personal Data

- **China (Hong Kong):**
  - Privacy Ordinance

- **Hong Kong:**
  - Privacy Ordinance

- **Canada:**
  - Personal Information Protection & Electronics Document Act

- **USA:**
  - Federal, Financial, & Healthcare Industry Regulations & State Laws

- **Mexico:**
  - E-Commerce Law

- **Brazil:**

- **Argentina:**
  - Protection of Personal Data

- **Chile:**
  - Protection of Personal Data

- **Colombia:**
  - Political Constitution – Article 15

- **South Africa:**
  - Promotion of Access to Information Act

- **Mexico:**
  - Bank Secrecy Regulation 8

- **Indonesia:**
  - SEC Board of India Act

- **India:**
  - Protection of Privacy Law

- **Pakistan:**
  - Banking Companies Ordinance

- **Israel:**
  - Protection of Privacy Law

- **EU:**
  - Protection Directive

- **Switzerland:**
  - Federal Law on Data Protection

- **EU:**
  - Protection Directive

- **Russia:**
  - Computerization & Protection of Information / Participation in Int'l Info Exchange

- **United Kingdom:**
  - Data Protection Act

- **Germany:**
  - Federal Data Protection Act & State Laws

- **Poland:**
  - Polish Constitution

- **China:**
  - Commercial Banking Law

- **France:**
  - Protection of Personal Data

- **China (Hong Kong):**
  - Privacy Ordinance

- **Hong Kong:**
  - Privacy Ordinance

- **Canada:**
  - Personal Information Protection & Electronics Document Act

- **USA:**
  - Federal, Financial, & Healthcare Industry Regulations & State Laws

- **Mexico:**
  - E-Commerce Law

- **Brazil:**

- **Argentina:**
  - Protection of Personal Data

- **Chile:**
  - Protection of Personal Data

- **Colombia:**
  - Political Constitution – Article 15

- **South Africa:**
  - Promotion of Access to Information Act

- **Mexico:**
  - Bank Secrecy Regulation 8

- **Indonesia:**
  - SEC Board of India Act

- **India:**
  - Protection of Privacy Law

- **Pakistan:**
  - Banking Companies Ordinance
The Big Data Paradox:
The Complexity of Volume and Variety Can Lower Confidence

- Am I confident in the data used for this analysis?
- Is this data good enough to use?
- Which data is correct?
- Am I protecting sensitive big data
- Do I understand the risk of using this data ‘as is’?

Chief Marketing Officer

Chief Data Officer
“New information types and sources—driven by volume, variety and velocity—require CIOs to consider new technology, processes and personnel, such as information managers and data scientists.

Consider creating chief data officer or chief digital officer roles, or both, if your organization has a large number of information-focused roles.”

Customer Analytics are Driving Big Data Initiatives

- Digital connections giving customers greater voice and higher expectations
- Integrating data for complete customer picture
- Understanding behavior patterns and preferences provides new ways to engage

Big data objectives

- Customer-centric outcomes (49%)
- Operational optimization (18%)
- Risk / financial management (15%)
- Other functional objectives (14%)
- New business model (4%)
- Employee collaboration (4%)

Top functional objectives identified by organizations with active big data pilots or implementations. Responses have been weighted and aggregated.
Initial Big Data efforts are focused on gaining insights from existing and new sources

- **Untapped stores of internal data**
  - Size and scope of some internal data, such as detailed transactions and operational log data, have become too large and varied to manage within traditional systems

- **Focus on customer insights**
  - Customers expect information provided to an organization will then be “known” during future interactions
  - Combining disparate internal sources with advanced analytics creates insights into customer behavior and preferences

Respondents were asked which data sources are currently being collected and analyzed as part of active big data efforts within their organization.
Big Data requires a solid, flexible information management foundation that’s scalable and extensible.

<table>
<thead>
<tr>
<th>Big data infrastructure</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information integration</td>
<td>65%</td>
</tr>
<tr>
<td>Scalable storage infrastructure</td>
<td>64%</td>
</tr>
<tr>
<td>High-capacity warehouse</td>
<td>59%</td>
</tr>
<tr>
<td>Security and governance</td>
<td>58%</td>
</tr>
<tr>
<td>Scripting and development tools</td>
<td>54%</td>
</tr>
<tr>
<td>Columnar databases</td>
<td>51%</td>
</tr>
<tr>
<td>Complex event processing</td>
<td>45%</td>
</tr>
<tr>
<td>Workload optimization</td>
<td>45%</td>
</tr>
<tr>
<td>Analytic accelerators</td>
<td>44%</td>
</tr>
<tr>
<td>Hadoop/MapReduce</td>
<td>42%</td>
</tr>
<tr>
<td>NoSQL engines</td>
<td>42%</td>
</tr>
<tr>
<td>Stream computing</td>
<td>38%</td>
</tr>
</tbody>
</table>

Respondents with active big data efforts were asked which platform components were either currently in pilot or installed within their organization.
The 5 Key Use Cases

Big Data Exploration
Find, visualize, understand all big data to improve decision making

Enhanced 360° View of the Customer
Extend existing customer views by incorporating additional internal and external information sources

Security/Intelligence Extension
Lower risk, detect fraud and monitor cyber security in real-time

Operations Analysis
Analyze a variety of machine data for improved business results

Data Warehouse Augmentation
Integrate big data and data warehouse capabilities to increase operational efficiency
A New Architecture is required

IBM Big Data & Analytics Platform

- Information Ingestion & Operational Information Zone
- Real-time Analytics Zone
- Exploration, Landing & Archive Zone
- Enterprise Warehouse, Data Mart & Analytic Appliance Zone
- Information Governance Zone

IBM Big Data & Analytics Infrastructure

Systems, Security, Storage

New/Enhanced Applications

- Cognitive
- Decision Management
- Predictive Analytics & Modeling
- Reporting, Analysis, Content Analytics
- Exploration & Discovery
Emerging Key Usage Patterns

**Top Level Solution Segments**

- Social Program Effectiveness
- Budget, Finance & Revenue Mgmt
- Threat Prediction & Prevention
- Crime Prediction & Prevention

**Enabling Solution Components**

- Fraud Analytics
- Enhanced Single View Citizen
- Health Analytics
- Social Media Analytics
- Video Analytics
- Cyber Security
## Transformational Social Program Effectiveness

<table>
<thead>
<tr>
<th>Providing caseworkers with timely, relevant information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reduce Caseload Work</strong></td>
</tr>
<tr>
<td><strong>Alameda County Social Services Agency</strong></td>
</tr>
<tr>
<td>- Gives managers and caseworkers a deep, real-time understanding of case and program status, enabling them to find the best assistance programs for each situation</td>
</tr>
<tr>
<td>- Reduced caseload work by 90% or more</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifying clients at risk and improving program success rates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Increase program success rates</strong></td>
</tr>
<tr>
<td><strong>U.K. Youth Services Organization</strong></td>
</tr>
<tr>
<td>- Used a sophisticated analytics solution to gain insight into hidden predictors to determine whether a young person is more likely to stay in or drop out of school, occupational training or employment</td>
</tr>
<tr>
<td>- 250% improvement in accuracy of identification of at-risk youths</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifying and reducing fraudulent claims</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reduce Fraudulent Claims</strong></td>
</tr>
<tr>
<td><strong>U.S. State Health Agency</strong></td>
</tr>
<tr>
<td>- Identified fraud among millions of claims, by finding obscure connections among doctors, pharmacists, lab and medical supply companies</td>
</tr>
<tr>
<td>- Identified more than US $200m in questionable claims resulting in 22 criminal convictions and US $49m in recovered funds</td>
</tr>
</tbody>
</table>

© 2014 International Business Machines Corporation
Social Program Effectiveness
Driving Tangible Business Value

Customer Analytics & Smarter Campaigns

- Measure, Monitor, Analyze Performance & Outcomes
  - # of Clients Becoming Self-Sufficient
  - Employee Productivity
  - 12-15%
  - 18-20%

- Improve Efficiencies by Streamlining Case Mgmt
  - Citizen Satisfaction
  - Reduction in Operational Cost Per Case
  - 15-10%
  - 15-20%

- Fight Fraud, Abuse & Error
  - Reduction in # of Successful Appeals
  - Money Collected Per Case
  - 12-18%
  - 10-15%

- Gain a Holistic View of Client
  - Reduction in # of Multiple IDs Issued
  - Reduction in # of Untouched Cases
  - 15-13%
  - 15-20%

Sample Values

- Client Successful Outcomes
  - 12-15%
  - 15-%

15-20% # of Foster Children Who Emancipate Successfully
12-15% Reduction in Number of Repeat Clients
12-15% Reduction in # of Multiple IDs Issued
15-20% Reduction in # of Successful Appeals
12-18% Reduction in # of Untouched Cases
12-15% Money Collected Per Case
15-20% Reduction in # of Untouched Cases
IBM Smarter Care – Takes a Holistic approach
Leveraging More Data About an Individual Enables a Coordinated, Sustained and Multi-Disciplinary Approach to Care

Lifestyle
Choices have direct impact on an individual’s mental and physical wellness

Social
Demographic determinants such as where one is born, grows, lives, works and ages have direct impact on overall health, mental health and well-being

Clinical
Factors such as specific medical symptoms, history, medications, diagnoses, etc. are indicators of an individual’s health

Big Data plays a BIG role
Transformation Social Program Effectiveness
Common Business Requirements & Entry Points

1. **Data Explorer**
   - Data Explorer to assemble all claims and employment information to build a complete picture for adjudication

2. **BigInsights**
   - Land all data – transactional, structured, unstructured, machine, logs and correlate them for customer service

3. **PureData for Analytics**
   - Speed up delivery of analytics using social and health data models and made for purpose appliances

4. **Streams**
   - Ingest and assess claims in real time for more responsive program delivery, reduced “Pay and chase”

5. **MDM**
   - To build a single view of beneficiary and provider, and master relationships between them

© 2014 International Business Machines Corporation
**Transformational Budget, Finance & Revenue Management**

**Minimizing revenue drain from improper payments**

**Reduce Improper Payments**

*New York State Dept. of Taxation and Finance*

- Tax audit and compliance system closes loopholes by focusing audit resources on only those returns that seem most unusual
- Decreased revenue drain due to improper payments of $200 Million annually, $1.2B since inception

**Automating processes lower labor costs**

**Automate Processes**

*Dutch Tax and Customs Administration*

- Content Analytics is used to extract data from documents, including sales acts, codes, addresses, amount paid, seller and more to feed Content Manager hub and other systems
- Automation helps reduce labor costs by 60 percent and increases the amount and types of data that can be extracted

**Predicting who is likely to pay their taxes**

**Increase revenue with effective cross-sell & up-sell**

*European Tax Collection Agency*

- Uses predictive modeling that gathers information from tax assessments, train ticketing systems, TV licenses, police records and more to predict whether a person will reliably pay taxes
  - 18% reduction in collections workload
  - 70% risk prediction accuracy
Threat and Crime Prediction & Prevention is transforming government in many ways.

How do I connect data about related crimes and criminals across jurisdictions?
- Share local, state, and national law enforcement information
  - US Federal Bureau of Investigation
- Federally-hosted N-DEx national repository of criminal justice records
- 18,000 law enforcement agency participants
- Each agency controls sharing based on legal, jurisdictional, & privacy requirements
- N-DEx proactively notifies specific users if certain relationships are discovered

How can I predict where crime is going to happen?
- Recognize crime trends as they are happening
  - Memphis Police Department
- Predictive model uses numerous variables to predict locations of criminal activity
  - 30% reduction in serious crime overall
  - 15% reduction in violent crime
  - Gives precinct commanders ability to change tactics and redirect patrol resources to catch more criminals in the act

How can I identify and minimize risk and target illegal activity in customs?
- Identify anomalies in customs information
  - European Customs and Border Control Agency
  - Predictive modeling to detect predictable and unpredictable anomalies in customs and border control information
  - EUR 85.5 million in confiscated goods
  - EUR 300 million in seizures of illegal drugs
A Intelligent Command Center solution combines key capabilities

- Multi Source Data Ingestion & Analysis
- Advanced Situational Awareness
- Stakeholder Collaboration

Leverage information to make better decisions
Anticipate problems to resolve them proactively
Coordinate resources and processes to operate effectively
Intelligent Command Center, an end to end solution

Runtime Operations
Applying the features and functions contribute to effectively manage an environment

Intelligent Operations Center

- Performance Monitoring (KPIs)
- Alert Management
- Enhanced User interface
- Mobile Operations
- Investigation Filtering
- Intelligent Response
- Stakeholder Collaboration
- Integrated situational awareness
- Adaptive Business Intelligence charting
Together with big data, analytics is a major force multiplier

<table>
<thead>
<tr>
<th>Video Analytics</th>
<th>Advanced Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictive Analytics</td>
<td>Identity Analytics</td>
</tr>
<tr>
<td>Content Analytics</td>
<td>Contextual Analytics</td>
</tr>
<tr>
<td>Social Network Analytics</td>
<td>Information Sharing and Resource Coordination</td>
</tr>
</tbody>
</table>
Safe and Secure Border and Trade Environment: How do you build a trusted experience? Where do you want to begin?

**Business outcomes**
- Resources more efficiently deployed to address risks
- Travel security increased and clearing times improved through analytic based screening
- Cargo clearance capacity increased as much as 90% with automated declaration processes

**IBM Solution Areas**
- Advanced Passenger Analysis
- eCustoms Framework
- Single Window
- Information Management
- Analytics Signature Solutions

**Transformational stages**
- Manage data
- Comprehensive information and insight
- Automated clearance processes
- Analyze patterns
- Integrated information and analytics
- Public/Private sector collaboration
- Optimize outcomes
Customs 2015
New tools and capabilities to enhance Trade

1. **Big Data Analytics** for risk management. Find the needle in the haystack. Detect deviations from trade patterns.
   Improved risk assessment results in optimized control, fast customs clearance and hence trade facilitation.

2. **Information Management**. Extensive information sharing between supply chain partners and Customs.
   Data harmonization (WCO datamodel), information sharing and information management services.
   Integration serves customs risk management.

3. **Social media** for Customs employees to allow information sharing (ageing problem) and mobile work (Customs inspections) and to satisfy tech-savvy generation.
Big Data in Customs

Variety of sources for Economic Operator data: Customs systems, other government systems, commercial databases (e.g. D&B, Chamber of Commerce). Cargo data combined with people data. Vessel data, air traffic data, commodity data, sensor data, X-ray images, unstructured documents, social media...

Large amounts of historic data of customs declarations and manifests (in Customs DWH); Large number of commercial/logistics documents with unstructured data

Customs declarations and air cargo manifest require near-real time risk assessment.
2014 Predictions & Investment Guidance
Big Data & Analytics

1. More analytics, fewer gut feelings - "infuse analytics into everything that employees touch"
2. Get serious about big data privacy and security – “privilege and burden”
3. Invest in a Big Data Platform – “all data, all analytics, and all domains”
4. Welcome, Chief Data Officer – “champion of data”
5. Outside data is as important as inside data – “meta insights”
Eras of computing

- Programmable Systems Era
- Tabulating Systems Era
- Cognitive Systems Era

cognitive: of or pertaining to the mental processes of perception, memory, judgment, learning, and reasoning.
## Big data best practices

<table>
<thead>
<tr>
<th>Best Practices</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strategy</strong></td>
</tr>
<tr>
<td>• Start with a use case for big data and build a business case</td>
</tr>
<tr>
<td>• Adopt a data-driven mind set in day-to-day operations</td>
</tr>
<tr>
<td>• Build on existing infrastructure investments</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>People and Process</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Create a data science culture by fostering data experimentation</td>
</tr>
<tr>
<td>• Enable people to go hands-on with a self-service approach to data and analytics</td>
</tr>
<tr>
<td>• Maintain governance, security and privacy - dispose of data you don’t need</td>
</tr>
<tr>
<td>• Right interface for each person depending on skill set</td>
</tr>
<tr>
<td>• Ensure the stack allows collaboration between different types of users</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Technology</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Seek out reusability</td>
</tr>
<tr>
<td>• Embrace and think beyond Hadoop</td>
</tr>
<tr>
<td>• Optimize workload performance and costs</td>
</tr>
<tr>
<td>• Continually re-evaluate what is big data or not</td>
</tr>
<tr>
<td>• Accumulate context, mine and visualize information for answers</td>
</tr>
<tr>
<td>• Use tools that go across all big data sources, rather than tools for each data source</td>
</tr>
</tbody>
</table>
Recommendations for getting started

▪ **Assess which Use Case would you most benefit from?**
  ▪ What part of the organization would benefit from expanding the data set and analytics to provide more complete answers?
  ▪ What part of the organization is not using analytics today, but would benefit from analytics for their user community or to fuel their processes using new information sources?
  ▪ What information do I collect today, or what analytics do I perform, that would be highly valuable as an information set to others?
  ▪ What area has the greatest risk?

▪ **Assess existing skills. You may need to:**
  ▪ Evolve your existing analytics and information capabilities
  ▪ Raise your organization competency
  ▪ Invest in integration, governance, privacy, and security
  ▪ Get ready to address performance, scalability, simplicity and cost
THINK BIG