Auditing Big Data - Internal Audit Role

Sajay Rai, CPA, CISSP, CISM
sajayrai@securelyyoursllc.com
Auditing Big Data

- What is BIG data?
- Business Significance of Big Data
- Opportunities of Big Data in Insurance
- V’s of Big Data?
- Auditors’ Responsibilities
- Big Data Audit Constraints
- Big Data Audit Approach
- Risks / Audit Controls
What is Big data? - Definition

Google

- Extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions.

Graphic from bigdatasp.com
What is Big data? - Architecture

- Many sources of data
- Many communication links
- Many different hardware
- Many types of reports
- Many different users
Business Significance of Big data

- New product development and faster go-to-market
- Strategic advantage
- Increased revenue / market share
- Cost reductions
- Time reductions creating operational efficiency
- Smarter business decisions and ability to do better market demand products

EXAMPLES
- Amazon
- Netflix
- Walmart
Insurance Example:
Opportunities of Big data

- Better Customer engagement
- Far more information around risks for Marketing, underwriting and pricing
- Enriched context around claims and claimants
- Insights from analysis of trends
- Brand protection, marketing opportunities, fraud detection and sentiment analysis
- Predictive analytics in claims
INSURANCE EXAMPLE:
Predictive analytics in Claims

- Forecast Medical costs
- Predict likelihood of return to work
- Identify litigation indicators
- Identify early settlement indicators
- Detect potential fraud earlier in the process
- Review incoming claims to route to appropriate adjuster
- Optimize workload across internal and external adjustors
Original 3 v’s* of big data: Volume

- Originally coined by Doug Laney, 2001
- Graphic from www.rosebt.com
Original 3 v’s* of big data: Variety

- Originally coined by Doug Laney, 2001
- Graphic from 1.bp.blogspot.com
Original 3 v’s* of big data: VELOCITY

- Originally coined by Doug Laney, 2001
- Graphic from blog.lionbridge.com
Original 3 v’s* At A Glance

- Graphic from CaseWare IDEA, INC.
Audit Implications of the 3 “V”s

- **Volume of data** – Share amount of data
  - *Audit sample*
  - *Continuous monitoring*

- **Variety of data** – mixing many sources and types of data
  - *Access control for the different data sources and type*
  - *Data classification*

- **Velocity of data** – rate of data accumulation
  - *Is the required audit data retained?*
  - *Are historical data retained past policy?*
Additional 4 v’s* of big data: Veracity

Quality of data – data must be accurate and in context

- From Mark van Rijmenam’s “Why the 3 V’s are not sufficient to describe big data”.
- Graphic from www.iri.com
Additional 4 v’s* of big data: Variability

- Meaning of data.
- Big data is always changing.
- Variability is relevant in performing sentiment analyses.

*From Mark van Rijmenam’s “Why the 3 V’s are not sufficient to describe big data”.
*Graphic from vint.sogeti.com
Additional 4 v’s* of big data: Visualization

Visually depicting data – analytic results are hard to interpret. Graphs and pictures highlight additional insights.

- From Mark van Rijmenam’s “Why the 3 V’s are not sufficient to describe big data”.
- Images from [www.sas.com](http://www.sas.com), [www.businessweek.com](http://www.businessweek.com), [www.hpi.org.uk](http://www.hpi.org.uk)
Additional 4 v’s* of big data: Value

Data improving outcomes – when analytics are translated to action

* From Mark van Rijmenam’s “Why the 3 V’s are not sufficient to describe big data”.
Preparing For big data

- Alignment of strategic plans, resources and technology solutions to requirements
- Governance – unclear roles responsibilities and requirements
- Time to implement
- Inadequate training and support
- Single points of failure
- Inappropriate access to data
Auditors’ Responsibilities

● Verify that big data objectives align with organization’s goals
● Educate the audit committee
● Provide multiple or continuous auditing
● Become an advisor to the big data team
  ● Perform pre- and post implementation informal reviews
    ● Review processes
    ● Review technology
● Determine if big data should be leveraged for auditing
● Define measurement criteria for success and security
Big Data Audit Constraints

- System is not homogenous
- Data is both structured and unstructured
- Sources of data are many
  - Geographic locations
  - Internal and external
- Access points to data are many
- Specialized talents required
- Audit and security controls are not part of Big Data design

“Through 2016, fewer than 30% of Hadoop deployments will be secured and governed in accordance with the enterprise’s information governance standards.”
– Gartner
Big Data Audit Approach

- Business Goals
- Presentation
- Communications
- Infrastructure
- Storage & Operating Systems
- Audit Committee
- Security Policies, Standards, & Procedures
- IT Policies, Standards, & Procedures
- Data Sources
- Personnel
- Middle Tier
- Continuous Monitoring
- Continuous Auditing
Key RISK Areas: IT General Controls

- Change Management
  - Coordinated
    - Platforms
    - Systems
    - Vendors
  - Defined change and configuration management policy
  - Documented change and configuration management process
    - Approvals
    - Impact analysis
    - Testing
    - Back-out
  - Must address availability of big data during change
Key RISK Areas: IT General Controls

- Access Controls
  - Defined and approved access control policy
  - Control over access while stored in Big Data
  - Control over analytical assessments more critical
  - Access approval by data owners
  - Data classification and least privilege
  - Periodic access review by data owners coordinated by information security group
  - Separation of duties and role based access
  - Proper control and management of privileged access
  - Proper management of 3rd party vendors/partners
Key RISK Areas: IT General Controls

- System Development Life Cycle
  - Security check-point during development
  - Source vulnerability scan before introduction to production
  - Source code control during and after development
  - Integration with change management
  - Defined services and system acquisition policy and procedure
  - Defined policy and procedure for external information system services
  - Protection of data supply-chain
Key RISK Areas: IT General Controls

- IT and Security Operations
  - Documented backup policy and procedures
  - Documented incident and problem management policies and procedures
  - Audit log collection and monitoring
  - Periodic vulnerability scans and penetration tests
## Key RISK Areas: Program governance

### Risk: Failure or lack value

<table>
<thead>
<tr>
<th>Control Area</th>
<th>Test Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Alignment to business goals and funding</strong></td>
<td>• Defined objectives aligned to business goals</td>
</tr>
<tr>
<td></td>
<td>• Defined, implement, and monitor measurement metrics</td>
</tr>
<tr>
<td></td>
<td>• Defined customer-centric service level agreement (SLA)</td>
</tr>
<tr>
<td></td>
<td>• Proper business and technical requirements</td>
</tr>
<tr>
<td></td>
<td>• Properly executed proof-of-concept (POC) and pilot</td>
</tr>
<tr>
<td><strong>Roles and Responsibilities</strong></td>
<td>• Roles and responsibilities for internal personnel</td>
</tr>
<tr>
<td></td>
<td>• Roles and responsibilities for business partners/vendors</td>
</tr>
<tr>
<td><strong>3rd Party Vendors</strong></td>
<td>• Provision for security</td>
</tr>
<tr>
<td></td>
<td>• Defined SLA and vendor performance monitoring</td>
</tr>
<tr>
<td></td>
<td>• Vendor transition</td>
</tr>
<tr>
<td></td>
<td>• Vendor termination</td>
</tr>
<tr>
<td><strong>Data Governance</strong></td>
<td>• Data management policies and procedures (to be discussed)</td>
</tr>
<tr>
<td></td>
<td>• Definition of metadata</td>
</tr>
<tr>
<td></td>
<td>• Inventory of databases, tables, etc.</td>
</tr>
<tr>
<td></td>
<td>• Defined authoritative data sources</td>
</tr>
</tbody>
</table>

*Courtesy of IIA documents*
## Key RISK Areas: Personnel

### Key to a successful big data implementation

<table>
<thead>
<tr>
<th>Resources</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sponsor</td>
<td>Executive level person that can obtain necessary funding</td>
</tr>
<tr>
<td>Business/Data Owners</td>
<td>Functional business people that own the data and have been trained as to what will be required of them</td>
</tr>
<tr>
<td>CDO</td>
<td>Chief Data Officer who might be the executive responsible day-to-day management of big data</td>
</tr>
<tr>
<td>CISO</td>
<td>Chief Information Security Officer responsible for securing big data infrastructure and data without hindrance</td>
</tr>
<tr>
<td>Technical Data Analyst</td>
<td>Including developers, DBAs, system administrators, subject matter experts, etc.</td>
</tr>
<tr>
<td>Data Scientist</td>
<td>Analytics professional(s) with a good understanding of the business and the technical tools</td>
</tr>
<tr>
<td>Big Data Architect</td>
<td>Responsible for the design and scalability of big data solutions</td>
</tr>
</tbody>
</table>

*Courtesy of IIA documents*
Key RISK Areas: Policies & Procedures

Policies, Procedures, and Guidelines

- Ethical collection of data
- Use of sensitive data
  - Data classification
  - Access control
  - Provisioning
  - Access review
- Data ownership
  - Assigned business owners
  - External owners of data
- Awareness and Training
## Key RISK Areas: Technology

**Risk:** Inappropriate and ineffective technologies can result in degradation and/or availability issue

<table>
<thead>
<tr>
<th>Control Area</th>
<th>Test Criteria</th>
</tr>
</thead>
</table>
| Hardware                   | • Annual capacity planning  
                            | • Right-sized hardware  
                            | • Ability to scale based on need  
                            | • High availability and redundancy                                          |
| Network                    | • Appropriate bandwidth  
                            | • Network segregation  
                            | • Secure transmission  
                            | • Right-sized devices  
                            | • Proper, detailed, and up-to-date network documentation                     |
| Tools                      | • Appropriate end user tools  
                            | • User training                                                      |
| Notification Engine        | • Proper configuration of ETL/ELT tools for timely notification  
                            | • Verification of record counts and totals  
                            | • Job failures and problem management                                      |
| Upgrades and Patches       | • Procedure for upgrade and patches  
                            | • Analysis of upgrades and changes for impact                           |

_Courtesy of IIA documents_
Key RISK Areas: Availability

Risk: Inappropriate and ineffective technologies can result in degradation and/or availability issue

<table>
<thead>
<tr>
<th>Control Area</th>
<th>Test Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd Party Partners</td>
<td>• Service Level Agreement (SLA)</td>
</tr>
</tbody>
</table>
| Disaster Recovery (DR)      | • High availability and redundancy  
• Accelerated recovery procedures documented (Triage – Post Mortem)  
• Proper and adequate backups |
| Monitoring                  | • Defined and configured alerts and thresholds  
• Documented product support processes  
• Problem management processes and procedures  
• Escalation process |
| Storage and Retention       | • Data retention policy  
• Data retention schedule defined and implemented  
• Data destruction procedure |
Key RISK Areas: Performance

Risk: Under-performing technologies can result in degradation and/or availability issue

<table>
<thead>
<tr>
<th>Control Area</th>
<th>Test Criteria</th>
</tr>
</thead>
</table>
| Scalability               | • Plans for growth  
                            | • Networking and transmission devices  
                            | • Storage systems  
                            | • Data collection systems (e.g. databases)  
                            | • Procedure to maintain level of performance |
| Performance Testing       | • Testing of new and existing analytics  
                            | • Time results   |
| Solution Reassessment     | • Periodic reassessment of solutions  
                            | • Reassessment of data integrity |
Key RISK Areas: Information Security

*Risk: Unauthorized access to data, inappropriate modification, and non-compliance to regulations*

<table>
<thead>
<tr>
<th>Control Area</th>
<th>Test Criteria</th>
</tr>
</thead>
</table>
| Information Security | • Information security and cybersecurity program  
                        • Minimum security baseline and system hardening program  
                        • Access to system and auditing tools  
                        • 3rd party assessment  
                        • Patch management process and procedures |
| Data Security        | • Information classification and access control based on classification    
                        • Privileged access management  
                        • Periodic access review to data by data owners  
                        • 3rd party access to data properly managed |
| Data Privacy         | • Data inventory and classification  
                        • Sanitization of PII, PHI, and other sensitive data  
                        • Documented incident management process and procedures |
### Key RISK Areas: data Quality/Reporting

**Risk: Data quality effect on reporting and management decision making**

<table>
<thead>
<tr>
<th>Control Area</th>
<th>Test Criteria</th>
</tr>
</thead>
</table>
| Data Quality   | • Quality standards and procedures  
• Definition authoritative source of data  
• Data integrity in database design  
• Time-to-reception of data to big data  
• Time-to-delivery of data to analytics |
| Management     | • Procedure for the verification purchased data from 3rd parties  
• Data and computational integrity  
• Timely data availability for decision making |
| Reporting      | • Process and procedure for reviewing and approving output from analytics  
• Definition of internally and externally consumed data  
• Restrictions in modification of report data  
• Provision for ‘ad hoc’ reporting  
• Process for selecting reporting options (e.g. Tableau, Splunk, etc.) |

*Courtesy of IIA documents*
Conclusion

- Create an enterprise-wide and holistic view of big data implementation
- Educate the audit committee and executive management
- Implement a continuous audit of big data or several single audits
  - You may need a dedicated group of auditors
- Review Confidentiality, Integrity, and Availability of big data systems and data
- Verify that big data implementation has the right people
SEE, I TOLD YOU THAT BIG DATA WAS TOO SCARY

Workforce Innovation That Works™

KRONOS.com
QUESTIONS / DISCUSSIONS