Applying Data Analytics to IS Audit

Michael Hoesing, CISA, ACDA, CDP, CIA, CISSP, CMA, CPA, has more than 30 years of experience in the areas of information systems audit and assurance, information systems implementation, and financial audit. His experiences span a variety of industries during his years with public accounting firms. His last 18 years have focused on financial services with firms such as First Data Corp., First National Nebraska Inc., PricewaterhouseCoopers and American Express. Hoesing has been involved in both the external and internal audit processes and has served as a software trainer.

Efficiency in today's information systems (IS) audit process is critical to achieve the cost-effectiveness desired in the current economic environment. Applying automated data analytic techniques to audit functions can enable IS audit teams to do more with the same, or fewer, resources. Use of these techniques can assist the IS auditor in complying with ISACA® standards.[1] Effective 1 January 2009, internal auditors, including IS auditors, *must* (instead of *should* in the prior standard) "consider the use of computer-assisted technology-based audit tools and other data analysis techniques when conducting internal audits" according to the Institute of Internal Auditors' (IIA) professional practice standard section 1220.A2.[2]

Previously, the *ISACA Journal* has included articles discussing the use of data analytics. [3, 4]

This article wishes to expand on the prior articles by providing specific implementations of data analytics to the IS audit scope of the audit universe, moving beyond financial audit techniques. The following sections are an anecdotal collection of IS audit data analytic techniques used by the author and other IS auditors recently.[5]

PLANNING

Many times auditors equate data analytics with fieldwork. While fieldwork may be the most frequent area in which data analytics is deployed by auditors, what is often overlooked is the process of using data analytics as an effective planning tool.

When risk-assessing the audit universe to determine the annual audit plan, data analytics may be used to help evaluate risk components that drive the annual deployment of IS audit resources. Most organizations' IT shops collect, for tracking purposes, incident or help desk tickets across all IT areas. This database of issues, usually rated "high" to "low" in severity, can be loaded into the IS auditor's data analytics tools[6] and assessed by frequency and severity across multiple departments in the IT organization or across multiple software applications. This evaluation can be combined with other risk factors to help determine the overall IT department risk rating.

Related to efficiency, that same IT incident/help desk ticket database can be reused for planning at the individual engagement level. If, for example, an audit of the systems team is on this year's plan, the incidents may be grouped by platform (Windows, UNIX, LINUX, etc.) and evaluated for severity and frequency. That evaluation may drive more audit time during the engagement to platforms experiencing the most challenges.

IT organizations' development efforts, both planned and in process, are usually stored in a project management tool. IS auditors' analysis of this database can help:

- Drive the IS annual audit plan toward applications development teams experiencing the higher-risk changes
- As an integrated audit tool, inform financial and operational auditors of significant application changes that may affect the timing of a financial or operational audit

FIELDWORK

There is a wealth of examples showing how to apply data analytics to financial transactions. Following are some ideas of how to apply data analytics to IS audit test procedures.

Interpretation of Complex Data

Analysis of Active Directory Groups

To support analysis of access controls over storage repositories (shares) in which sensitive customer or company data are stored, many auditors review user membership in Active Directory (AD) groups for rights appropriate to the user's job duties. While there are commercial tools to analyze AD,[7] the assumption in this article is that the IS auditor has access to a data analytic tool but not to a commercial AD tool. Built into the AD server is a command that extracts data from AD into a comma-separated values (CSV) file type. The command *csvde –f outputfilename.csv* places the content of all the AD objects (users, computers, groups, etc.) and their settings into a CSV file that loads into most data analysis tools. The groups a user is a member of can then be counted and listed by separating the field "memberOf" based

on the semicolon delimiter, using parsing functions (e.g., SPLIT) built into most data analytic tools, as shown in **figure 1** .

Configuration Management

Operating system auditing conquers one of the more typical audit challenges, the lack of data. In fact, operating systems usually provide a wealth of excess data that need to be culled down in order to audit the relevant facts. For example, a VMware ESX 3.5 virtualization host can provide details of the current configuration and operational metrics by issuing the *esxcfg-info* command. Redirecting the output to a text file results in a file of approximately 10,000 records, mostly performance metrics (as shown in **figure 2**). Loading the data into an analytics tool requires testing each of the records for an identifier (e.g., level 2 is "\==+" at position 4) using the TEST or equivalent function, as shown in **figure 3**, then locating descriptions and data within that record.

Once the raw data are read into the analytic tool, filters and searches can be conducted to narrow the data displayed to relevant audit topics. Audit-relevant data for this host can be extracted and saved to a separate file. Repeating this process in the future or with additional hosts could enable building cumulative files to trend the configuration over time to evaluate remediation efforts or for comparison among multiple hosts.

Assessment Tool's XML Output

Many assessment tools provide data in an Extensible Markup Language (XML) output. While easy to read and navigate, the XML output tends to stack data top to bottom, rather than the more traditional file structure of data flowing from left to right. Also, many free assessment tools provide assessment data for one scan or one host and are not designed to collect data over multiple time periods or for multiple machines. One free tool that

Figure 1—Active Directory Analysis

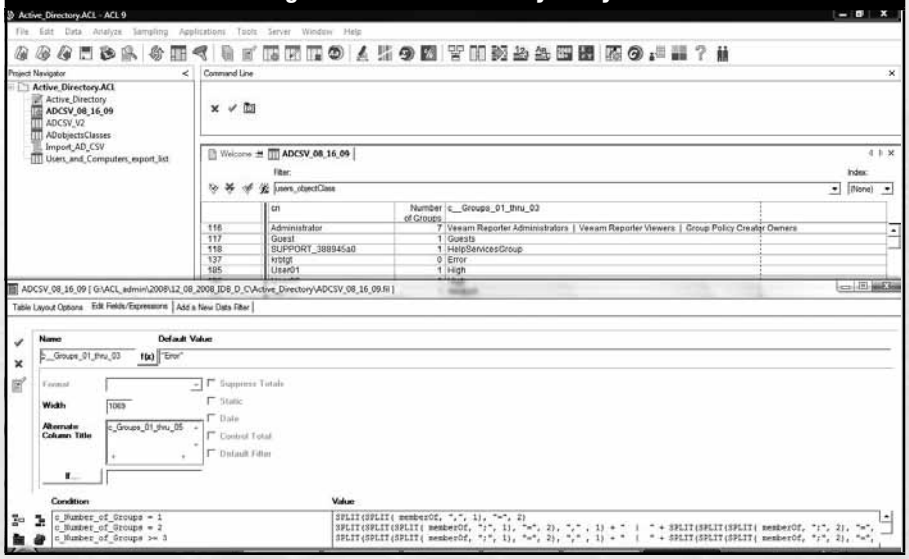


Figure 2—Configuration and Operational Metrics Output Text File in Notepad

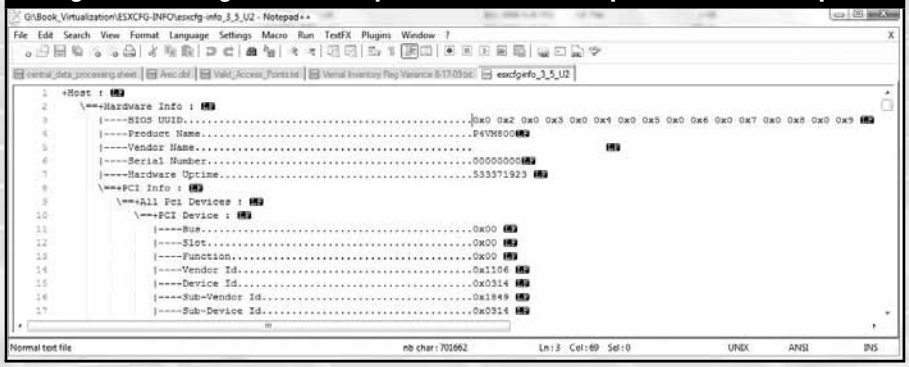
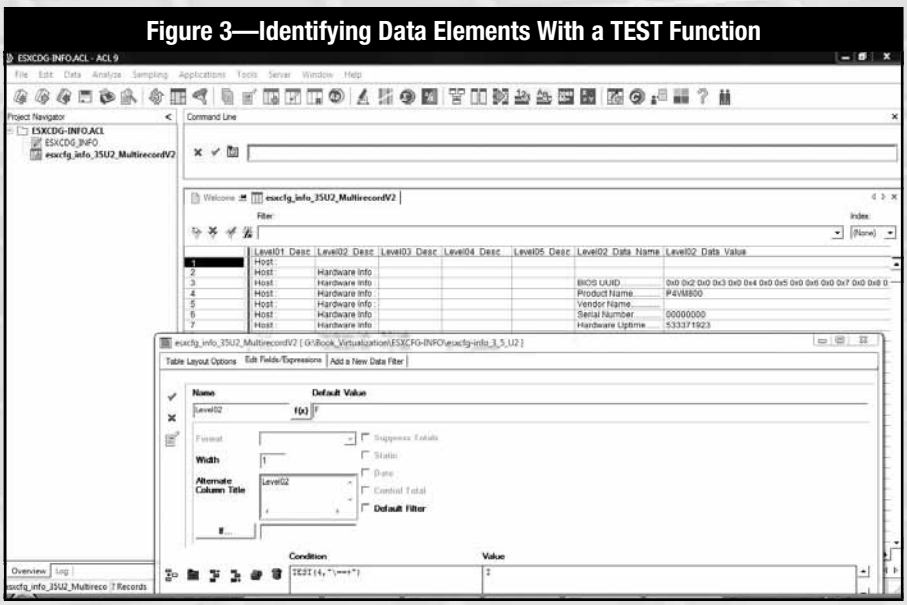


Figure 3—Identifying Data Elements With a TEST Function



may be used to assess Payment Card Industry (PCI) Data Security Standard (DSS) compliance[8] for wireless access points is Kismet.[9]

Kismet's XML output can be read by most tools that automatically interpret XML metadata and flatten the stacked data as shown in **figure 4**. Once in an analytic tool, the data can be reviewed for the presence of encryption or other settings, as shown in **figure 5**.

Adopting techniques from the next section, joining the Kismet results with the known, approved inventory of wireless access points, would help determine unauthorized devices and authorized devices that are dormant. Some may say a dormant device poses little risk; however, should that device be reactivated, the IS auditor may wish to understand the process used to patch and configure a reactivated device to the organization's current standards, if the requirements have changed during the period of inactivity.

Figure 4—Interpreting XML Data

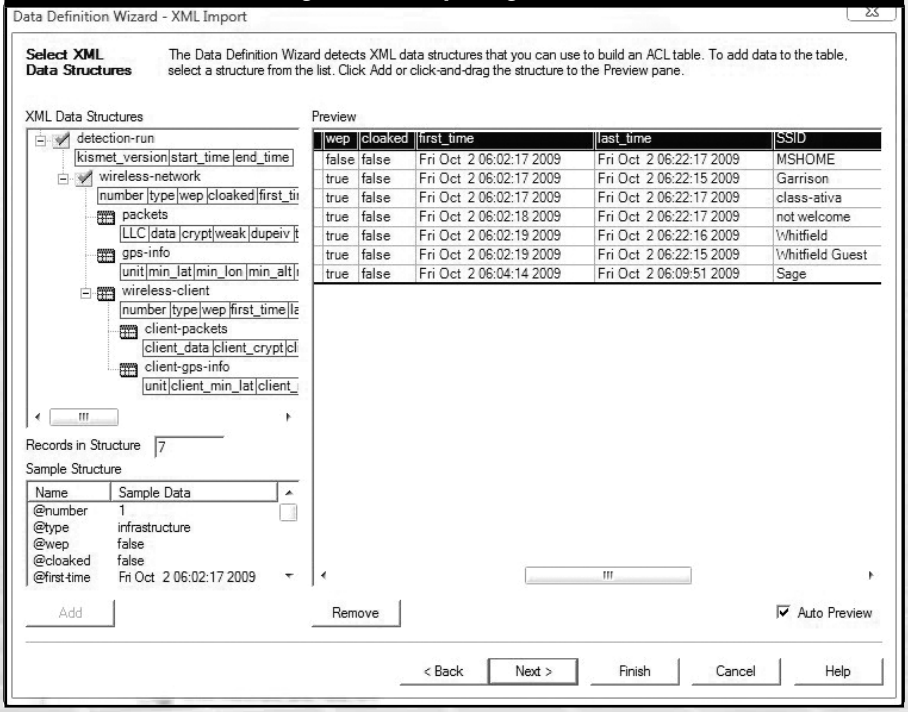
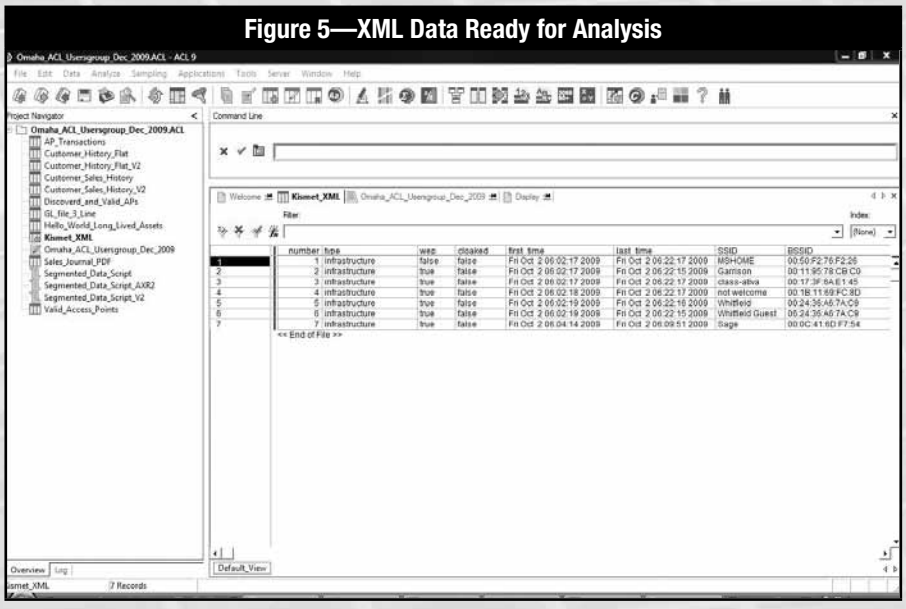


Figure 5—XML Data Ready for Analysis



Determining Context Using Multiple Files

Analyzing a single file, even when given unlimited time to do so, may not produce useful audit information. Many times additional data in secondary files are needed to fully understand and give context to the original file. Examples applicable to IS audit include:

- **Logical access**—Combining a logical access file with an HR employment file will help determine which user accounts are aligned with valid employees. Also, if the access functionality of the user's profile can be combined with the initial two files, comparison of a user's work department with the capabilities granted to him/her can be evaluated for appropriate segregation of duties.

- **Change control**—File library listings can be combined with data from the change management system and dates of file changes can be matched to dates of authorized events in the change management database to help identify any unauthorized changes.
- **Physical security**—Matching of ingress records to egress records could be performed to help identify tailgated access to sensitive IT areas. If both records are in one file, analytic tools may be used to reorganize the data by badge number and by date/time stamp using sorting tools, and then the current record can be compared to the previous record with a RECOFFSET function to determine if egress succeeded an ingress record.

REPORTING

Analytic tools have extensive capabilities to group and summarize data, bringing perspective to the analysis. The list of devices found in the scan results shown in **figure 5** could be combined with an organization's actual inventory of authorized devices. Reports could then be derived showing unauthorized devices, or authorized devices that were not on and active and were not discovered by the scan. The former contains the risk of unauthorized and possibly incorrectly configured devices, and the latter has the risk that dormant authorized devices are not enabled and are not receiving current patches. This detail will help IS management or the audit committee to understand the magnitude of the rogue device and configuration problems and the effort needed for correction.

Also, with the event tickets system noted previously, IT management, which, like IT audit, has too much to do in too little time, may be focused on solving the current high risk-rated events. This leaves little time for analysis. If the auditor can provide additional insights to the event data (volume by platform, volume by application, volume by time period), management may be able to address some recurring topics and reduce the event rate.

Summarization and grouping of data may also be used by the auditor when reporting to the audit committee. Showing the committee that the auditor wishes to place platform X on next year's schedule, instead of platform Y, due to the number of critical applications supported, the number of significant event tickets issued and IT staff hours devoted to support, based on verifiable metrics, may be better received by the committee (and the chief information officer), than more subjective measurements.

CONCLUSION

The ideas presented here can be expanded almost infinitely based upon the environment and needs of the auditor's organization, and are limited only by one's imagination. IS auditors can continue to fulfill the data analytic support role to financial or operational audits, but should also apply those skills to completing their own portion of the audit plan.

ENDNOTES

[1] ISACA IT Audit and Assurance Standards that help to explain where to use data analytics include: S5 Planning, S6 Performance of Audit Work, S11 Use of Risk Assessment in Audit Planning, S12 Audit Materiality, S14 Audit Evidence, *www.isaca.org/standards*. Also, IT Audit and Assurance Guideline G3 Use of Computer-assisted Audit Techniques (CAATs) provides guidance on how to use data analytics in IS audits.

[2] Institute of Internal Auditors, *International Standards for the Professional Practice of Internal Auditing (Standards), www.theiia.org*

[3] Ott, John; Andrew MacLeod; Kevin Mar Fan; "Computer-assisted Audit Techniques: Value of Data Mining for Corporate Auditors," *Information Systems Control Journal*, vol. 3, 2008

[4] Sayana, Anantha; "Using CAATs to Support IS Audit," *Information Systems Control Journal*, vol. 1, 2003

[5] During the course of the past couple of years, the author has had the opportunity as a software trainer and educator to discuss IS audit techniques, data analytic techniques and other techniques with a variety of IS audit professionals across the country. While it would be impossible to match names with each technique presented here, the author would like to extend a collective "thank you" to all whose path he has crossed and who have provided him with their thoughts.

[6] Baker, Neil; "Software Trend Spotting," *Internal Auditor*, August 2009. While not all-inclusive, according to the IIA survey, top data analysis tools include ACL, Microsoft Excel, Microsoft Access and IDEA.

[7] While not all-inclusive, a Google search produces commercial tools such as Hyena, GFI lanGard and ScriptLogic.

[8] Payment Card Industry (PCI), Data Security Standard (DSS) 1.2.1, section 11.1, "Test for the Presence of Wireless Access Points," July 2009, *www.pcisecuritystandards.org/security_standards/pci_dss_download.html*

[9] Kismet Wireless Assessment Software, *www.kismetwireless.net*