Analyst Insight

Aberdeen Group
A Harte-Hanks Company

June, 2011

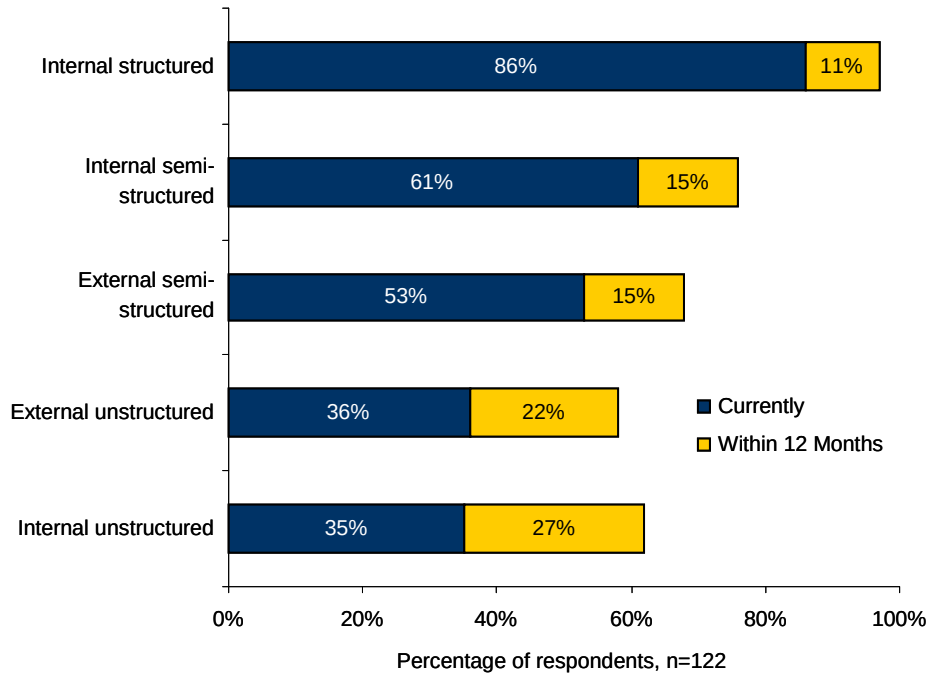
Future Integration Needs: Embracing Complex Data

Aberdeen Group conducted research into the use of data integration technologies in June 2011. This research investigated the changing nature of data integration needs, and the strategies and tactics that companies have enlisted to maximize their gains from data integration. The findings in this research are based on data collected from 122 organizations currently engaged in data integration.

Current and Future Integration Needs

The Aberdeen Group has researched the data volumes used by organizations for analytics in each of the last three years. Not only does the volume of data continue to grow, but that growth itself is accelerating. While organizations were experiencing an average growth in data volumes of 30% in 2009, the volume growth had risen to 40% by the end of 2010. Data volume growth is one thing, but organizations are also looking to integrate more complex data to help run their businesses (Figure 1).

Figure 1: Organizations Moving to Integrate Complex Data



Source: Aberdeen Group, June 2011

Analyst Insight

Aberdeen's Insights provide the analyst perspective of the research as drawn from an aggregated view of the research surveys, interviews, and data analysis

Fast Facts

Complex data volumes are growing fast:

√ 24% of organizations are experiencing their largest growth in data volume in internal semi-structured data

√ 22% of organizations are experiencing their largest growth in data volume in internal unstructured data

Aberdeen Group
A Harte-Hanks Company

Not surprisingly, most organizations (86%) current use structured data from internal sources. Structured data would include virtually anything that is sourced from a database system. With this foundation, this would include major applications used by many large and mid-sized enterprises today - Enterprise Resource Planning (ERP), Customer Relationship Management (CRM) and Supply Chain Management (SCM), for example (Table 1).

Table 1: Top 5 Sources of Internal Structured Data

Source Data System	Percentage of Survey Respondents
Enterprise resource planning	73%
Customer relationship management	70%
Supply chain management	49%
Non-ERP Accounting Applications	47%
Project Management	47%

n=122
Source: Aberdeen Group, June 2011

Internal structured data sources encapsulate the lifeblood of many organizations, capturing the millions upon millions of transactions that collectively represent the book of business. However, semi-structured data also has an important role to play. Semi-structured data is less formally organized than structured data, but may be interpreted and understood through the use of labels and tags embedded within the data. EDI is a very enduring example of a semi-structured data format used between trading partners, although a number of proprietary protocols also exist. In the last 10 years or so however, use of the Extensible Markup Language (XML) has flourished in parallel with growth in corporate usage of the internet. In fact, XML is now the most widely used protocol for the exchange of semi-structured data. Seventy-four percent (74%) of organizations surveyed integrate XML data from external sources, while 66% of all organizations integrate internally originated XML data. Organizations use XML and other semi-structured data feeds to plug into their business ecosystem. For example, a manufacturer of consumer electronics needs to exchange data with both upstream and downstream supply chain partners. If accurate, up-to-date information can be exchanged, supply of its new tablet computer (the legendary HokiKoki 4000) can be matched to demand more accurately. This can result in lower overheads since less inventory needs to be carried, sales opportunities can be maximized as stock-out situations are avoided, and price discounting necessary to liquidate excess stock can be eliminated.

When it comes to adding future data sources, more enterprises plan to introduce complex unstructured data sources in the next 12 months than any other type (Figure 1). Unstructured data includes data types such as images, audio and video. Also included would be any form of free-form text - such as email, office productivity documents, instant messaging, blogs and web content. This focus on including unstructured data reflects a growing

Fast Facts

XML is the most common semi-structured data source that organizations integrate

√ 74% of organizations are integrating XML from external sources

√ 66% of organizations are integrating XML originating from internal sources

interest in understanding customer sentiment through social media channels, as well as tapping the vast wealth of tacit corporate knowledge that is captured by office productivity tools.

Fast Facts

On average, organizations are integrating 14 external data sources, up from 11.4 on year ago

The Cost of Data Integration

Data integration is frequently an overlooked factor in business intelligence or business process improvement projects. Without question, other aspects of BI - such as dashboards and visualization - can seem more glamorous. However, integration cannot be overlooked - nor should it be under-resourced. Mobile business intelligence is a case in point. Much of the attention is paid to dashboards on smartphones and tablets that can be manipulated elegantly with simple gestures. However, Aberdeen research in December 2010 (*Mobile BI: Actionable Intelligence for the Agile Enterprise*) found that the organizations that get the most out of their mobile BI investments are far more likely to use data integration and data cleansing tools than other organizations. Without the solid foundation of trusted data provided by these tools, slick presentation counts for nothing. Aberdeen's research found that organizations are making significant investments to integrate and transform data in order to make it useful within the enterprise (Table 2).

Table 2: Average Spend on Integration in Last 12 Months

Component	Cost
Software Licenses	$278k
Services & Support	$287k
Internal Support	$555k
Fraction Spent on Integrating External Data Sources	25%

n=122
Source: Aberdeen Group, June 2011

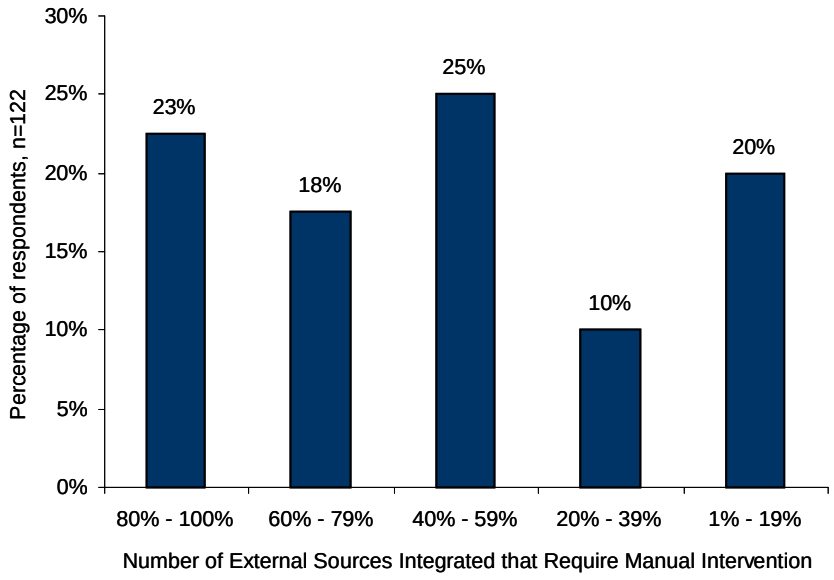
Table 2 shows the average cost of providing integration and transformation for the prior 12 months. A couple of facts are noteworthy. First, organizations are spending roughly the same amount on internal support as they do on software licenses, professional services and support services combined ($555k vs. $565k). The cost of internal support was derived by multiplying the number of Full Time Equivalents (FTEs) working on integration projects by $150,000 (the estimated annual cost of employment for each FTE). On average, survey respondents have 3.7 FTEs working on integration initiatives.

Secondly, organizations that are integrating external data sources spend 25% of their total integration budget on transformation in this area. These organizations are integrating 14 external data sources on average, and bring 29GB of data into the company each day. Prior research by the Aberdeen Group into supply chain management (*The 2011 State of Retail Supply Chain Collaboration, Visibility and Integration*) has illustrated the performance gains that can be achieved by integrating external data. For example, those

suppliers that had the ability to integrate and collaborate electronically with their key retailers were able to deliver perfect orders to customers 78% of the time. Organizations without such sophisticated integration capabilities could deliver perfect orders only 69% of the time.

One of the problems often cited by technology users when interviewed by Aberdeen is the particular difficulty of integrating external data sources. By and large, corporations have at least some control over the nature of the internal data sources they integrate. Typically, internal structured data will be sourced from commercially available off-the-shelf application packages or homegrown applications built on top of a relational database. This data will also be extracted from the source system by the in-house IT team. And, for better or worse, the quality of internally sourced data lies entirely in the organizations own hands. For example, the processes and discipline employed around data entry functions for internal applications has an overwhelming impact on the quality of the data coming out.

Figure 3: Integration of External Data is Labor Intensive



Source: Aberdeen Group, June 2011

"We are trying to create a comprehensive view of spend. But, we have contractual data in multiple systems with multiple schemas, integrating with payables in many different ways. Then non-contractual spend has multiple different processes, and there is no transactional integrity around purchase cards."

~ Logistics Manager, US Postal Service

None of those three statements need hold true for data originating from external sources. The organization receiving the data doesn't necessarily know which applications the data originated from, how it was extracted from the application, whether the extract is complete, or whether the data is clean. Further, depending on the relationship between data supplier and data consumer, the organization receiving the data might get very little help in understanding the data or in fine tuning the format of the data to best suit their needs. As a result, almost all organizations have some external data sources that require manual intervention (Figure 3).

In fact, 10% of organizations in Aberdeen's survey required some manual tasks for all of the external data sources they integrated. Only 5% of organizations were able to integrate all external data sources without any type of manual intervention. Manual processes are both time-consuming and error-prone. Re-keying data (or cut-and-paste operations to lift it from one data store and deposit it in another) are tedious, repetitive and littered with opportunities to make mistakes. With few organizations able to automate the integration of external data to any significant extent, this remains a strong opportunity to drive business process improvement.

Fast Facts

- √ Organizations with strongly integrated supply chains delivered perfect orders 78% of the time
- √ Those without strongly integrated supply chains delivered perfect orders only 69% of the time

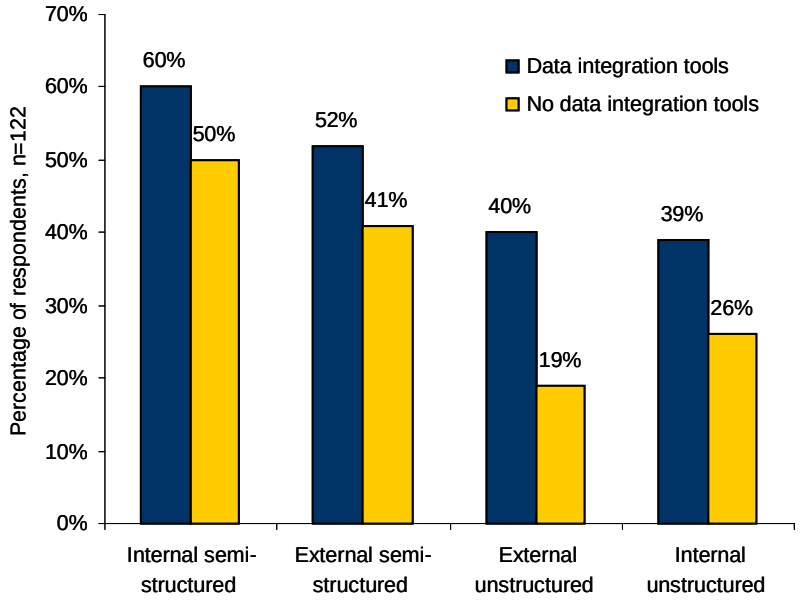
Getting the Most from Data Integration Assets

While rising volumes of data and the growing complexity of data will continue to pose challenges for organizations, there are some clear steps that can be taken to improve the productivity of integration and transformation. Aberdeen's data shows how organizations can get the most from the resources - both funds and people - that they commit to integration projects.

Not all companies that took part in Aberdeen's survey were currently using specialized integration tools. For some, the tool pressed into use is that great jack-of-all-trades, the spreadsheet. However, those organizations that do have dedicated integration tools can gain a number of advantages. For example, Aberdeen's data shows that organizations that have integration tools in place are able to use larger data sets for business intelligence and analytics - over 50% larger. More importantly however, organizations that have dedicated integration tools at their disposal are able to make a much richer set of data available to business users (Figure 4).

Despite the magic that sometimes appears to go on behind the scenes, IT departments are no different from any other. Staff in the IT organization have a finite amount of time to devote to projects and - all other things being equal - using the right tool allows people to accomplish more in less time. And so it is with integration tools. Enterprises that invest in the most appropriate tool for the task at hand are able to achieve more. This is reflected in the volumes - and types - of data that they are able to integrate, aggregate and make available to business users. For example, organizations that use integration tools are over twice as likely to be able to integrate external unstructured data into business processes (40% vs. 19%).

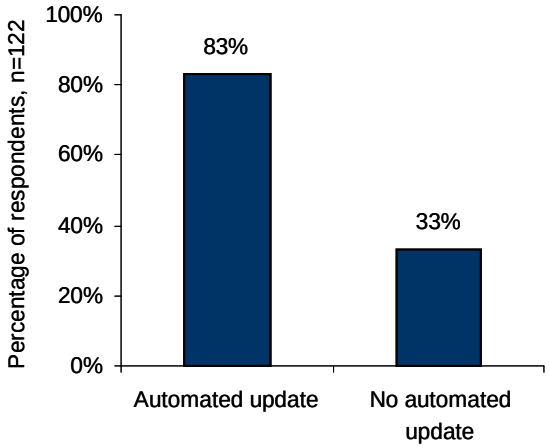
Figure 4: Integration Tools Provide Access to Richer Set of Data



Source: Aberdeen Group, June 2011

With strong integration tools in place, it then becomes possible for new data updates to be integrated and refreshed automatically. Automation of the integration and transformation process is essential to make sure that the right information gets to the right people at the right time (Figure 5).

Figure 5: Automation Delivers Timely Information More Often



Source: Aberdeen Group, June 2011

"We're trying to analyze performance trends at the product level. But, there is precious little standardization across the source systems with most using free-text fields for product identification."

~ Mary Dedrick, President, Performark

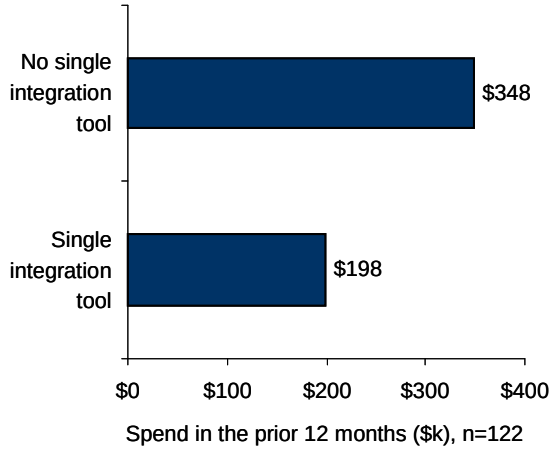
To understand why, let's revisit the consumer electronics supplier mentioned earlier. Suppose that when the product is launched on December 1, 2011, demand for the HokiKoki 4000 tablet computer is far

stronger in California than it is on the eastern seaboard. What if demand outstrips supply in that state and the shelves may be bare on December 3? It would make perfect sense to divert inventory from the east to the west to meet that higher than expected level of demand. However, if the integration process is so poor that demand data generated on December 1 is not available until December 7, which is just too late for the right decision on inventory distribution to be made.

As Figure 5 shows, 83% of the organizations that have automated data refresh are able to deliver information to managers in the time required at least 80% of the time. Without a process for automated updates, timely information can be delivered only a third of the time. Data is only valuable if it can be delivered in the timeframe that matches the needs of management decision makers. Organizations using automated data refresh are 2.5 times as likely to be able to achieve that goal. In addition, automation reduces the errors that are often introduced into data when manual processes are involved. Organizations with automated integration report a data error rate almost half that of enterprises that do not.

There are also cost benefits that accrue to companies that are able to consolidate their integration projects too - both from a product and staffing perspective. Organizations that are able to use a single tool for all integration needs spend much less on software licenses (Figure 6).

Figure 6: Cost Benefits of Single Integration Tool



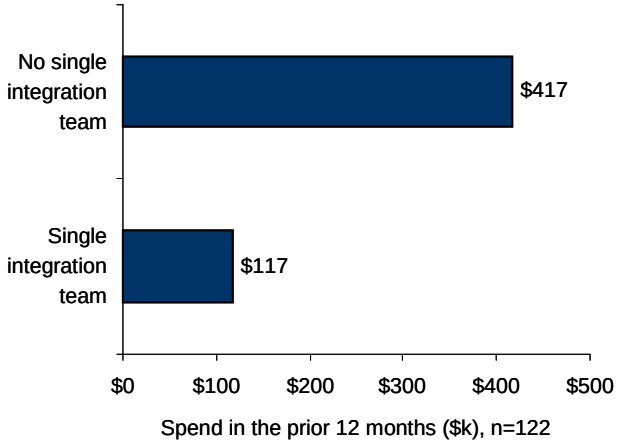
Source: Aberdeen Group, June 2011

In the last 12 months, organizations that used a single integration tool for integration spent 43% less on integration software last year than those that did not. That reduced spend is not just a factor of being part of a smaller company or having less complexity in the data. Organizations that use a single tool are similar in size to those that do not, and also integrate a similar number of external data sources. Most enterprises require relatively few licenses for integration tools - they are a tool that is used by a handful of skilled IT staff, not the general business community. As a result, if a single

tool or suite can be acquired that encompasses all integration needs, then it should be relatively easy to consolidate costs and negotiate with a single supplier. Focusing on a single tool that embraces both structured and complex data (semi-structured and unstructured) can enable competence in integration to be grown quickly so that business benefits can be delivered sooner.

In a similar vein, those firms that have a single individual or team responsible for all enterprise data integration tasks spend much less on services and support (Figure 7).

Figure 7: Concentration of Integration Skills Reduces Costs



Source: Aberdeen Group, June 2011

Whenever a single person or focused team can be brought to bear on a task, economies of scale can be gained. Skills, education and experience can be leveraged more widely across the organization and a center of excellence for integration developed. This concentration and reinforcement of skills can reduce the need for professional services and outside support, hence reducing costs in those areas. That level of consolidation is perhaps easier to achieve in smaller organizations. But some large enterprises have also strived for this degree of focus in order to reap these benefits.

Recommendations and Key Takeaways

Aberdeen's research has highlighted the benefits that can be gained from a strong and consistent focus on data integration. Aberdeen recommends the follow actions for organizations currently involved in data integration, or contemplating the adoption of such practices:

- Organizations should strive to fully automate the integration and refresh of data. Timely management decisions depend on both high quality data and timely data. That is, data needs to be delivered in the timeframe that business managers require in order to make decisions. Organizations that are able to automate the process of

updating data are far more successful at getting information into the hands of managers in the timeframe required. Increasingly, those integration needs are going to incorporate more complex data forms, which are often more difficult to integrate automatically. Data integration tasks may occur far from the customer-facing world inhabited by business managers. However, companies should ensure that the business impact of poor integration practices is understood so that the right level of resources and commitment are devoted to integration.

- Organizations with dirty data should focus on the root of the problem. For internal data sources, that root is data entry. For external sources, the root lies with partners. Organizations that required manual intervention to perform updates and refresh of data have almost twice as many errors in their data when compared to those organizations that have an automated process. Manual integration is often necessary when there is little standardization in the source data. Manual data entry, transcribing data from paper to computer, or where some degree of freedom is granted in how to enter data into a computerized form, all provide opportunities for data to get dirty. Dirty data often requires manual work to fix. This is likely to delay data updates and introduce errors into the data, which is a roadblock to automation. Robust integration and transformation capabilities can ensure that data purity is high at the start and remains that way throughout its lifecycle.
- Organizations should aim to achieve their data integration tasks using a single tool. Likewise, they should also consider consolidating their expertise into a single individual or team. Aberdeen's research shows that organizations that can take these steps are able to significantly reduce their overall integration costs with no loss of output or productivity. By embracing a single tool, businesses can ensure that the full integration and transformation process is automated, not just parts of the cycle.

As the growth in data accelerates, the integration needs of companies will change to accommodate more complex data forms, such as unstructured data. For this change to be successful, corporate practices around integration, transformation and data cleansing will need to change too.

For more information on this or other research topics, please visit www.aberdeen.com

Telephone: 617 854 5200
Fax: 617 723 7897

Aberdeen Group
A Harte-Hanks Company

Related Research	
The 2011 State of Retail Supply Chain Collaboration, Visibility and Integration; March 2011	*Turning Pain into Productivity with Master Data Management*; February 2011
Business Answers at Your Fingertips: The Real-Time Value of BI; March 2011	*Data Management for BI: Fueling the Analytical Engine with High-Octane Information*; December 2010
Agile BI: Three Steps to Analytic Heaven; March 2011	*Mobile BI: Actionable Intelligence for the Agile Enterprise*; December 2010

Author: David White, Senior Research Analyst, Business Intelligence (david.white@aberdeen.com)