The Modern Data Warehouse: Agile, Automated, Adaptive

Produced by David Loshin and Abie Reifer from DecisionWorx, LLC
in collaboration with The Bloor Group
December 2015

DecisionWorx   **BG** The Bloor Group

Sponsored by:

TIMEXTENDER

IBM

Cloudera®

WhereScape®

1   Table of Contents

2   Executive Summary

This research study seeks to understand perceptions of the modern data warehouse and its incorporation of adaptive and automated technologies to make design, development and operations more agile. Respondents were asked about technology choices for data warehousing, including what we term "agile technologies" such as data warehouse automation, cloud-based data warehousing, hosted data warehousing/data warehousing as-a-service, as well as newer technologies like columnar databases, in-memory computing and Hadoop. The results communicate insights about awareness, as well as adoption rates, of those newer, more agile data warehouse platforms and technologies, particularly ones that are promoted within the vendor communities to reduce development effort, speed time-to-value or provide capabilities beyond what is typically offered as part of a conventional, RDBMS-based data warehouse deployment.

The key findings reflect that organizations are open to adopting a range of auxiliary technologies to augment their data warehouse environment, creating a hybrid architecture that encompasses more than one technology. Conventional relational database management systems (RDBMSs) remain the most frequently chosen approach as the primary platform, but more mature agile technologies are gaining some momentum as the corporate experience increases.

The responses corroborated some of the agile technology vendors' claims that their solutions enabled simplified productionalization, lowered staff requirements, lowered overall costs of operations, and increased overall perception of value. Respondents who selected one of the agile technologies as their primary solution are more likely to feel that their solution provided the right tools and guidance for configuration and administration, simplified the development of ETL/data integration, better supported metadata management, and they believe the solution is able to efficiently ingest data and support publishing data to the cloud.

As data warehouse professionals become more experienced or work at organizations with more mature data warehousing programs, awareness of the agile technologies increases, particularly in the area of Hadoop adoption and data warehouse automation. Among those with fewer than four years of experience, though, there is greater familiarity with cloud-based and hosted data warehousing.

The technology that ranks highest among those actively exploring acquisition of new technologies is Hadoop, followed closely by cloud-based data warehousing, in-memory data warehousing and data warehouse automation. Adoption rates will potentially increase as awareness of the technologies increases, even in organizations that have already chosen a primary technology. By producing more thought leadership content for prospective adopters, vendors of the agile technologies have an opportunity to influence individuals to consider and possibly adopt their products.

3 Introduction – The Emergence of Agile, Automated and Adaptive Data Warehousing

As a result of data warehousing, business intelligence and analytics having matured and moved into the mainstream. Much of the architecture for data warehouses essentially conforms to accepted conventions that involve components for data ingestion, data preparation, data integration, loading into dimensional models and provisioning to downstream information consumers. In the best of all possible worlds, for business organizations this means that the implementation of a data warehouse should largely be simplified. Yet data warehouse professionals admit that challenges remain in productionalizing their reporting and analytics systems.

Fortunately, there is an emerging and growing community of vendors providing alternative platforms and tools that are intended to simplify the creation, deployment and ongoing maintenance of the data warehousing infrastructure. Some of these alternatives are meant to simplify the architecture; others are intended to speed the development and deployment, and some provide improved performance to accelerate time-to-value.

Examples of what we call "agile data warehouse technologies" include data warehouse automation, Hadoop, cloud-based or hosted data warehousing, or columnar and in-memory databases, among others. No matter what, all are designed to help streamline the environment, speed the analysis cycle or otherwise enhance agility in supporting end-user reporting and analytics needs. By simplifying the data warehouse production process, these types of tools can free the data analysts and other data professionals from the details of designing and instantiating a data warehouse environment, enabling them to focus on the more critical business aspects that include engineering/designing of reports, helping business analysts configure their queries and generally concentrating on information provisioning.

This research report is intended to help the reader better understand the concepts of data warehouse automation and show how improved data warehouse performance that increases organizational agility can add value to the organization. Importantly, we have surveyed data warehouse professionals (architects and users alike) to assess the value proposition in terms of staffing, expenses, management overhead, time-to-knowledge and any other factors that might influence the consideration, acquisition and adoption of these tools. This report details some quantitative assessments of the results of the survey and provides our qualitative interpretation of how those results can guide the reader into making the right decision regarding data warehouse architecture and engineering.

4   Objectives of the Study

The goals of this study were threefold:

First, we wanted to understand user and developer perceptions regarding the ease of development and deployment of their primary data warehousing platform/solution.

Second, where possible, we wanted to contrast the user/developer experiences between more conventional technologies with emerging agile data warehousing solutions such as data warehouse automation, cloud-based data warehousing, hosted data warehousing/data warehousing as-a-service, as well as newer technologies like columnar databases, in-memory computing and Hadoop to support the data warehousing effort.

Third, we wanted to assess awareness, as well as adoption rates, of those newer, more agile data warehouse platforms and technologies, particularly ones that are promoted within the vendor communities to reduce development effort, speed time-to-value or provide capabilities beyond what is typically offered as part of a conventional, RDBMS-based data warehouse deployment.

Demographic characteristics of the participants, such as an individual's level of experience or the years of corporate data warehouse experience, were used for analysis and differentiation.

5   Agile Data Warehouse Technologies

The word *agile* conveys a number of meanings. Its simple definition as an adjective is "quick and well coordinated," but the term has taken on additional meaning in the context of system design and development. According to the Agile Alliance, the agile software development methodology emphasizes "close collaboration between the programmer team and business experts; face-to-face communication (which is more efficient than written documentation); frequent delivery of new deployable business value; tight, self-organizing teams."[1]

For the purposes of this study, we have chosen a selection of technologies we consider to enhance corporate agility in terms of speeding design and development, improving cycle times in producing reports and analyses and strengthening the collaboration between developers and business analysts to use their data warehouse to produce actionable insights. Those technologies are:

- **Columnar databases**, which are engineered to align data in ways that speed loading and query response time.
- **In-memory databases**, which take advantage of systems that are configured with large amounts of memory for loading significant portions of data warehouses (e.g., "hot" tables that are frequent targets of queries and analytic models), resulting in faster delivery of application results.
- **Hadoop**, the open source ecosystem of tools for data distribution and parallel execution using platforms built on commodity components, which provides an environment well suited for developing predictive and prescriptive analytics applications that can consume and take advantage of massive data volumes.
- **Data lake**, which, according to TechTarget, is "a large object-based storage repository that holds data in its native format until it is needed,"[2] providing a place for collecting data sets in their original format, making those data sets available to different consumers and allowing data users to consume that data in ways specific to their need.
- **Data warehouse automation**, including tools that facilitate the end-to-end production of a data warehouse, encompassing the entire development lifecycle: source system analysis, design, development, generation of data integration scripts, building, deployment, generation of documentation and testing as well as support for ongoing operations, impact analysis and change management.
- **Cloud-based data warehousing**, referring to implementations where the corporate data warehouse is deployed on a cloud platform, which eliminates a large part of the hardware acquisition, operations and maintenance effort and cost.
- **Hosted data warehouse**, also referred to as "data warehousing-as-a-service," where a service provider not only hosts the hardware and software platform (thereby reducing and eliminating hardware and software acquisition costs as well

---

[1] "What is Agile Software Development," downloaded November 6 from

[2] See http://searchaws.techtarget.com/definition/data-lake

as operations and maintenance costs), but additionally provides consulting and guidance in the design, development and deployment of the data warehouse.

- **Data warehouse appliance**, which is a specialty hardware configuration engineered for high-performance reporting and analytics. An appliance is typically configured with multiple processing nodes, multiple storage nodes and high-speed interconnectivity that can be configured to specific data warehousing and business intelligence needs. This option is often seen as being low cost in comparison with conventional enterprise data warehouse platforms.

6   Analysis Approach – Questions

The survey was generally composed of five main sets of questions:

1. **Demographic Questions**
   a. Principal industry of the organization
   b. Job function
   c. Role in influencing technology decisions
   d. The number of years the respondent has worked with data warehousing
   e. The number of years that the organization has had data warehouses

2. **Evaluation Parameters**
   a. Perception of the degree of challenge for bringing the data warehouse into production
   b. Number of full-time equivalent (FTE) staff supporting the data warehouse program
   c. Allocation of staff time to different aspects of the data warehouse production cycle

3. **Technologies Employed**
   a. List all the data warehouse approaches and technologies used
   b. Identify the primary technology employed
   c. Identify where the data warehouse environment is deployed

4. **Opinion Questions** that gauge agreement with statements about the ease of:
   a. Data modeling
   b. Data integration
   c. Development
   d. Maintenance and support
   e. Scalability and performance
   f. Agility
   g. Costs
   h. Security and compliance

5. **Technology Awareness and Adoption**
   a. Indicate level of familiarity with new technologies
   b. Provide insight as to what level of consideration exists for the adoption of new technologies

The demographic questions and the evaluation parameters were used as differentiating filters to assess levels of satisfaction regarding the various phases of the end-to-end data warehouse development and implementation processes.

One of the key differentiating questions was whether the organization currently had a data warehouse in place or not. Of those who had completed the survey, more than 27% indicated that their organization did not yet have a data warehouse, and the focus of their responses was limited to the questions regarding awareness and consideration of data warehousing technologies for adoption.

7   Demographics of the Respondents

Overall, there were 417 individuals who responded to at least one question. Of this set of respondents, there were 233 individuals who completed the survey. The completed responses served as the pool that was used for the analysis.

7.1   Industries Represented

The industries most represented by respondents were Telecommunications, Technology, Internet/Electronics, Government, Finance/Financial Services, Insurance, Education, Healthcare and Pharmaceuticals. Other represented industries included Airlines/Aerospace (including Defense), Automotive, Business Support and Logistics, Construction, Machinery, Homes, Entertainment/Leisure, Manufacturing, Nonprofit, Retail/Consumer Durables, Transportation/Delivery and Utilities, Energy and Extraction.

7.2   Job Function

The respondents overwhelmingly exhibited technical expertise, with 39% reporting that their job function was Information Technology; 15% identified in consulting, approximately 12% identified as managers, 6% in engineering and approximately 5% identified as analysts.

7.3   Influencing Technology Decisions

Our respondents were well placed in terms of technology decision making, with 27% having authority to make technology decisions and 56% reporting that they advise management on technology decisions.

7.4   Personal Data Warehouse Experience

In general, the respondents were highly experienced in data warehousing. Over 70% have more than 4 years of experience, and more than 41% of the respondents have more than 10 years of experience working with data warehouse technologies. There were some less-experienced professionals in the crowd as well, with 16% admitting that they were just starting their data warehousing experience.

7.5   Corporate Data Warehouse Experience

Correspondingly, the organizations that were represented in the survey also exhibit some degree of longevity in working with data warehouses: 63% have had a data warehouse for more than 4 years, and 36% of the respondents reported that their organization has had a data warehouse more than 10 years. As previously noted, there is a crowd of corporate novices, with 28% indicating that their organization did not yet have a data warehouse.

7.6   Geographic Distribution

Not all respondents provided information about their country of origin, and these numbers do not sum to the total number of complete responses. However, for those who did report their country:

94 of the completed responses originated from North America.

60 of the completed responses originated from Europe.

14 of the completed responses originated from Asia/Pacific.

4 of the responses originated from South America.

1 response originated from Africa.

8 Summary of Findings

Analyzing the survey responses exposed interesting results, and a more detailed discussion of some these results follow in this report. However, here we highlight some of the key findings:

- **RDBMS is most frequent** – The conventional approach to data warehousing using relational database management systems remains the most frequently chosen approach.
- **Agile technologies are gaining momentum** – More mature agile technologies (such as cloud-based data warehousing, columnar databases and data warehouse appliances) are gaining some momentum as the corporate experience increases.
- **Agile technologies are popular with less experienced organizations** – Organizations with shorter duration of corporate data warehousing experience (2-6 years) seem to be considering the agile technologies as their primary platform at higher rates.
- **Agile technologies correlate with lowered staff allocation** – The respondents whose organizations have selected one of the agile technologies as their primary solution report generally lower allocation of full time staff members to support the data warehouse program than those who are using the conventional approach. Although other factors are likely to contribute to this difference, there appears to be a correlation between the users of the agile technologies and reduced staffing.
- **Agile technologies simplify productionalization** – The trends suggest that adopters of the agile technologies are having an easier time in productionalizing their reporting and analytics environment. For respondents whose organizations have had data warehouses for longer periods of time, the results suggest that those adopting the agile technologies reported less of a challenge than those adopting the conventional approach.
- **Agile technologies are seen to have a reasonable cost** – While in general more respondents seem to feel that the cost of their solutions are reasonable, the overall results indicate that in organizations whose primary technology is one of the agile technologies, there is a much more favorable perception of cost reasonability.
- **Agile technologies are seen to be worth the value** – Individuals at organizations that have chosen the agile technologies as their primary platform are more likely to perceive that their platform choice is worth its value than those using conventional technologies as their primary platform.
- **Agile technologies support configuration and administration** – Those who have selected one of the agile technologies as their primary solution are more likely to feel that their solution provided the right tools and guidance for configuration and administration.
- **Agile technologies are seen to simplify ETL/data integration** – Those with agile technologies are more likely to agree that the solution simplifies the development of ETL/data integration and better supports metadata management, and they believe

the solution is able to efficiently ingest data and support publishing data to the cloud.

- **Hybrid environments are growing in adoption** – Organizations are open to adopting a range of auxiliary technologies to augment their data warehouse environment, creating a hybrid architecture that encompasses more than one technology.
- **Experienced professionals are familiar with agile technologies** – As data warehouse professionals become more experienced or work at organizations with more mature data warehousing programs, awareness of the agile technologies increases, resulting in greater potential for integrating the different technologies together within the same environment.
- **More experienced people are more familiar with Hadoop and data warehouse automation** – Those at organizations with more than 4 years data warehousing experience are more likely to be very familiar with the agile technologies than those at organizations with fewer than 4 years of experience. The least recognized technologies are data lakes and Hadoop, although higher percentages of respondents indicated that they were "very familiar" with Hadoop. The most recognized technologies are Hadoop and data warehouse automation.
- **Less experienced people are more familiar with cloud-based and hosted data warehousing** – Individuals with 4 or more years of experience are more likely to be familiar with data warehouse automation and Hadoop but are less likely to be familiar with data lakes. However, among those with less than 4 years of experience, there is greater familiarity of cloud-based and host data warehousing.
- **Adoption tracks with awareness** – Adoption rates will potentially increase as awareness of the technologies increases, even in organizations that have already chosen a primary technology.
- **Awareness of data lakes is lowest** – While there seems to be a high percentage of reported familiarity with Hadoop, there is a lack of familiarity with data lakes, which are typically described as built using Hadoop and HDFS (Hadoop Distributed File System).
- **Vendors have opportunities to raise awareness** – By producing more thought leadership content for prospective adopters, vendors of the agile technologies have an opportunity to influence individuals to consider and possibly adopt their products.
- **Hadoop, cloud-based, in-memory and data warehouse automation lead active exploration** – The technology that ranks highest among those exploring new technologies is Hadoop, followed closely by cloud-based data warehousing, in-memory data warehousing and data warehouse automation.

9   Choices for Data Warehouse Technologies

Of the 168 respondents whose organizations have data warehouses, 107 (or 64%) noted that their primary data warehouse platform technology was a conventional RDBMS. 53 (or 31.5%) reported using one of the aforementioned agile technologies, and the remainder indicated using other alternatives, including specialty platforms as well as NoSQL data management systems, among others.

Although the majority of the respondents said that their organization relied on conventional RDBMS platforms, the fact that almost one-third had chosen one of the agile technologies indicates a growing acceptance of technology choices that diverge from the traditional reference data warehouse architecture.

9.1   Does Corporate Experience Influence Choice of Solution?

To what extent does data warehousing experience factor into the choices made about data warehouse platform technology? The survey collected information about the respondent's experience as well as the number of years that the respondent's organization has had a data warehouse.

For very experienced organizations (those with 7 or more years of having a data warehouse), there are three facets of experience that could influence technology decisions. The first is that more experienced organizations may have already seen a return on their investment and are potentially willing, from the financial perspective, to consider new technologies. Second, the team may be well informed about how their existing technology works, providing some appreciation of what their existing platforms can or cannot do and how the available features of newer approaches may benefit the organization. Third, past experiences in design, development and implementation provide an appreciation of the level of effort necessary to navigate the challenges of adopting a new data warehousing technology.

Of this subpool of respondents representing very experienced organizations, 29.66% had selected one of the agile technologies as their primary data warehouse platform, mostly using columnar databases or a data warehouse appliance. However (as can be seen in Table 1), among this cohort of very experienced organizations, the adoption of the agile technologies as part of the data warehouse environment is quite impressive, with rates ranging from 10.17% for hosted data warehouses to 32.30% for Hadoop and between 24% and almost 30% for columnar, in-memory, data warehouse appliance and data warehouse automation.

Technology	Percentage
Columnar database	29.66 %
Data warehouse appliance	25.42 %
In-memory database	28.81 %
Data warehouse automation tool	24.58 %
Cloud-based data warehouse	21.19 %
Data lake	22.88 %

Technology	Percentage
Hosted data warehouse solution	10.17 %
Hadoop/HDFS	32.20 %
NoSQL database	16.10 %
Graph database	9.32 %

Table 1: Adoption rates for agile technology solutions as part of the data warehouse environment for the very experienced organizations.

For respondents from experienced organizations that have used data warehouse solutions for 4-6 years, approximately 57% were using conventional RDBMS solutions while 43% used newer technologies for their primary technology. Also interesting is that in this group we found that there was a higher shift in the use of cloud-based data warehouse technologies (17.9%), which was proportionally more than twice that of any other single technology to be selected. In this subgroup of experienced organizations (as shown in Table 2), we still saw adoption of the other technologies with the highest being columnar and cloud-based data warehouse (at 25%) and data warehouse appliance (21.43%), but no others measuring higher than 18%.

Technology	Percentage
Columnar database	25.00%
Data warehouse appliance	21.43%
In-memory database	14.29%
Data warehouse automation tool	0.00%
Cloud-based data warehouse	25.00%
Data lake	17.86%
Hosted data warehouse solution	3.57%
Hadoop/HDFS	17.86%
NoSQL database	14.29%
Graph database	7.14%

Table 2: Adoption rates for agile technology solutions as part of the data warehouse environment for the experienced organizations.

Among those with less corporate experience (0-3 years having had a data warehouse), the conventional RDBMS still seemed to be most prevalent, with 59% choosing it as their primary platform technology. Proportionally (as shown in Table 3), inclusion of agile technologies was actually lower than with the more experienced organizations, with cloud-based data warehousing and Hadoop topping the charts at 22.73%, but generally around 9-14% for the other technologies.

Technology	Percentage
Columnar database	9.09%
Data warehouse appliance	13.64%
In-memory database	13.64%
Data warehouse automation tool	9.09%
Cloud-based data warehouse	22.73%
Data lake	0.00%
Hosted data warehouse solution	4.55%
Hadoop/HDFS	22.73%

Technology	Percentage
NoSQL database	4.55%
Graph database	18.18%

Table 3: Adoption rates for agile technology solutions as part of the data warehouse environment for the less experienced organizations.

These results suggest that even in those environments in which the conventional RDBMS platform remains the most frequently chosen approach as the primary technology, the more mature agile technologies (cloud-based data warehousing, columnar databases and data warehouse appliances) are gaining some momentum as the corporate experience increases. Organizations with shorter experience (2-6 years) seem to be considering the agile technologies as their primary platform at higher rates. Even more enlightening are the rates of inclusion of different data warehousing alternatives. As the level of experience increases, the different agile technologies increase in adoption (even if not as the primary platform).

10 Correlation Between Primary Technology and Staffing Levels

Agile data warehouse technology vendors often suggest that a motivating factor for adopting their technologies is that they simplify the processes of designing, developing and deploying the data warehouse environment. The claim is that reducing complexity in data warehouse design and implementation reduces their staffing requirements, and one of the survey's questions sought to explore this claim.

The survey asked respondents "How many full time equivalent (FTE) staff members support the data warehouse program?" with the choice of range options including 0-3, 4-7, 8-10 and >10 full time staff members.

The results in Table 4 show the overall distribution of answers as well as the responses broken out by those whose primary data warehouse is a conventional RDBMS and those who have chosen one of the agile technologies as their primary data warehouse platform.

Full time staff	All	Conventional	Agile
0-3	28.57%	28.04%	28.30%
4-7	22.02%	19.63%	30.19%
8-10	10.71%	12.15%	5.66%
More than 10	38.69%	40.19%	35.85%

Table 4: Number of full time staff members working on the data warehouse program.

These percentages demonstrate that those who reported using the agile technologies also reported generally lower allocation of full time staff members to support the data warehouse program than those who are using the conventional approach. Overall, 47.67% of those using the conventional RDBMS are supported with 7 or fewer FTEs, while 58.49% of those with the agile technologies are supported with 7 or fewer FTEs. Although other factors are likely to contribute to this difference, there appears to be a correlation between the users of the agile technologies and reduced staffing.

11 Assessing Challenges in Bringing the Data Warehouse into Production

One of the common experiences for data warehouse implementation is encountering unanticipated challenges during the design, development and, most frequently, the deployment stages. One might say that the development of agile data warehousing technologies was a reaction to these challenges, with the intent of providing alternatives for development teams to finesse these difficulties, either by speeding the design-develop-test cycle or by simplifying the development process through automation.

The survey explored respondents' perception of how challenging it was to bring their organization's data warehouse into production. Furthermore, the survey solicited information about specific aspects of the end-to-end process of data warehouse design, development and deployment, including data modeling, code development and data integration as well as ongoing operations and maintenance.

11.1 Perceptions of Overall Level of Challenge

The survey's investigation into the challenges of deployment began with a general question: "How challenging was it for your organization to bring its data warehouse into production?" Respondents were presented with five choices, ranging from "Not challenging at all" to "Extremely challenging."

Overall, the results are not surprising and are consistent with historical perceptions of the difficulty in moving a data warehouse into production. As can be seen in Table 5, 50.6% reported that it was challenging, and a combined 35.71% thought it was very challenging or extremely challenging.

Level of challenge	Percentage
Extremely challenging	5.95%
Very challenging	29.76%
Challenging	50.60%
Not very challenging	12.50%
Not challenging	1.19%

Table 5: How challenging was it for your organization to bring its data warehouse into production?

When only considering those respondents whose organizations used the conventional RDBMS as their primary technology, we see a shift in the level of challenge, as is shown in Table 6. In this subgroup, a lower percentage reported that it was not very challenging or not challenging at all (cumulative 11.21% vs. cumulative 13.69% overall), and similarly, the percentage reporting that it was extremely challenging increased from 5.95% to 7.48%. While the majority still reported the median level of challenge, there was a noticeable shift to the more challenging side.

Level of challenge – Conventional RDBMS	Percentage
Extremely challenging	7.48%
Very challenging	30.84%
Challenging	50.47%
Not very challenging	11.21%

Level of challenge – Conventional RDBMS	Percentage
Not challenging	0.00%

Table 6: Level of challenge for conventional RDBMS.

Conversely, when we only consider those whose organizations have adopted one of the agile technologies as their primary data warehouse solution, we see the opposite shift. As is shown in Table 7, a higher percentage of those respondents reported that it was not very challenging or not challenging at all (a cumulative 18.86% vs. 11.21% for RDBMS). At the same time, a lower percentage reported that it was very challenging or extremely challenging (cumulative 28.31% vs. 38.32% for conventional RDBMS).

Level of challenge – Agile technologies	Percentage
Extremely challenging	1.89%
Very challenging	26.42%
Challenging	52.83%
Not very challenging	15.09%
Not challenging	3.77%

Table 7: Level of challenge for agile technologies.

We interpret these findings to suggest that although challenges remain in moving a data warehouse into production using any technology, the trend suggests that adopters of the agile technologies are having an easier time in productionalizing their reporting and analytics environment.

11.2 How Does Corporate Experience Correlate to Data Warehouse Challenges?

Typically, one might expect that longer experience in data warehousing would tend to simplify the productionalization process. The percentages in Table 8 would agree with this theory, as respondents from organizations with data warehouses for 4 or more years indicated lowered perceptions of challenge overall.

Challenging? Overall	0-3 years	4 or more years
Extremely challenging	4.55%	6.16%
Very challenging	40.91%	28.08%
Challenging	45.45%	51.37%
Not very challenging	9.09%	13.01%
Not challenging	0%	1.37%

Table 8: Overall – How challenging based on organization's years of experience with data warehouse technologies?

When these responses were segregated based on the primary data warehouse technology (conventional vs. agile, shown in Table 9 and Table 10), overall the numbers reflected similar trends – the greater the corporate experience, the lower the perception of challenge. However, comparing conventional vs. agile for the more experienced organizations, it appears that those adopting the agile technologies reported less of a challenge (17.7% reporting not very challenging or not challenging at all) than those adopting the conventional approach (12.8 reporting not very challenging and 0% reporting not challenging at all).

Challenging? Conventional RDBMS	0-3 years	4 or more years
Extremely challenging	7.69%	7.45%
Very challenging	46.15%	28.72%
Challenging	46.15%	51.06%
Not very challenging	0%	12.77%
Not challenging	0%	0%

Table 9: Use of conventional technologies – How challenging based on organization's years of experience with data warehouse technologies?

Challenging? Agile technologies	0-3 years	4 or more years
Extremely challenging	0%	2.22%
Very challenging	25%	26.67%
Challenging	50%	53.33%
Not very challenging	25%	13.33%
Not challenging	0%	4.44%

Table 10: Use of agile technology – How challenging based on organization's years of experience with data warehouse technologies?

One potential extrapolation is that greater corporate experience frames the level of difficulty experienced in adopting the newer technologies, perhaps in the wake of less-than-satisfactory conventional implementations.

12 Perception of Cost/Value of the Primary Data Warehouse Technology

The survey posed a number of opinion questions regarding perceptions of costs to determine of whether respondents perceived costs of the technology to be within reason and whether the products were perceived to be worth the value.

12.1 Reasonableness of Costs

The respondents were asked how strongly they agreed whether the software and hardware costs of their primary data warehouse solution are reasonable. The summarized results from those who opted to answer this question are shown in Table 11.

Reasonableness of cost	All	Conventional RDBMS	Agile
Strongly agree	12.50%	8.41%	18.87%
Agree	55.36%	53.27%	64.15%
Disagree	19.05%	23.36%	11.32%
Strongly disagree	5.36%	5.61%	1.89%

Table 11: Are the software and hardware costs reasonable?

Across all respondents, without regard to the particular technology used, there was a favorable perception of the costs, as 67.86% either agreed or strongly agreed with the statement. When the cohort was divided into two subcategories (those whose primary data warehouse technology was the conventional RDBMS vs. those whose primary technology was one of the agile technologies), there were some clear adjustments in the perception of cost reasonability. Those using conventional technology indicated a less favorable view, with the percentage of favorable responses lowered to 61.68%. Alternatively, those with agile technologies overwhelmingly responded favorably, with a cumulative 83.02% agreeing or strongly agreeing with the statement.

Within each category of the agile technologies there was general satisfaction with the reasonableness of costs. Those whose primary technology was either columnar or in-memory generally felt their costs were reasonable (13 out of 16 agreed, 1 strongly agreed, while only 1 disagreed). While the number whose primary technology was Hadoop was small (6 in total), 4 strongly agreed and the other 2 agreed. Three out of the 4 whose primary technology was data warehouse automation agreed, and 12 out of the 14 respondents whose organization's primary technology was a cloud-based data warehouse either agreed or strongly agreed.

So although in general more respondents seem to feel that the cost of their solutions are reasonable, the overall results of this question indicate that in organizations whose primary technology is one of the agile technologies, there is a much more favorable perception of cost reasonability.

12.2 Perception of Value

Our next question focused on whether the respondents felt that they get their money's worth out of the approach that they had selected. The survey solicited the degree to which

the respondent agreed with the statement "The software and hardware costs of our data warehouse are well worth the value."

Perception of value	All	Conventional RDBMS	Agile
Strongly agree	14.88%	13.08%	16.98%
Agree	54.76%	52.34%	64.15%
Disagree	23.21%	27.10%	15.09%
Strongly disagree	1.79%	1.87%	0.00%

Table 12: Are the software and hardware costs well worth the value?

The results of this question were similar to the question about cost reasonability. When looking at all responses, irrespective of the specific technology, there was a generally favorable perception (69.64% agreed or strongly agreed) of value. However, when the cohort was divided into those choosing a conventional RDBMS as their primary platform vs. those whose primary platform was built using the agile technologies, the latter demonstrated an apparently greater appreciation of value, with 81.13% agreeing or strongly agreeing, only 15.09% disagreeing and no one strongly disagreeing.

One can interpret these results to suggest that individuals at organizations that have chosen the agile technologies as their primary platform are more likely to perceive that their platform choice is worth its value than those using conventional technologies as their primary platform.

13 Maintenance and Support – Configuration and Administration

The resource requirements for budget and staffing are not limited to the design, development and deployment phases of data warehouse productionalization. The survey solicited data about respondent perceptions associated with maintenance and support. The first question asked the participants to relate how strongly they agreed or disagreed with the statement "Our primary data warehouse solution provides the appropriate tools and documentation that minimizes its configuration, administration, and maintenance." The results are shown in Table 13.

Configuration and management	All	Conventional RDBMS	Agile
Strongly agree	12.50%	9.35%	20.75%
Agree	50.60%	48.60%	56.60%
Disagree	30.95%	36.45%	18.87%
Strongly disagree	4.76%	5.61%	1.89%

Table 13: Appropriate tools and documentation for configuration and management.

Across the entire pool of participants whose organization had a data warehouse, 63.1% had a favorable response, indicating that they agreed that the primary solution provided appropriate tools and documentation to minimize the configuration and administration efforts. Reviewing the subgroup of those whose primary solution is the conventional RDBMS, the survey responses still showed a generally favorable response (cumulative 58% agreeing or strongly agreeing), but there is a bit of a shift to the less favorable responses when compared to entire pool of respondents, with a combined 42% disagreeing or strongly disagreeing.

However, the subgroup whose primary solution is one of the agile technologies has a dramatically more overall favorable response, with 77.35% agreeing or strongly agreeing. Much of that shift is attributable to a significant increase in the number of those strongly agreeing, with 20.75% for the agile technologies compared to only 9.35% for conventional.

These results can be interpreted to suggest that those who have selected one of the agile technologies as their primary solution are more likely to feel that their solution provided the right tools and guidance for configuration and administration.

14 Data Integration and Technology Choice

Data integration has always been considered one of the more time-consuming and complex aspects of a production data warehouse, and the survey included questions soliciting participants' opinions regarding how they perceived the ease of a number of aspects of data integration. The respondents were asked to indicate the degree to which they agreed or disagreed with a number of statements.

The first statement was "The current tools we use as part of our primary data warehouse solution provide an efficient way to profile and explore data sources." There were no material differences between the positive (agree or strongly agree) of those whose primary data warehouse technology was the conventional RDBMS vs. those who had adopted one of the agile technologies.

Differences begin to appear when comparing the results of the other four questions. For each of those four opinion questions, we share the overall results, then divided the cohort into those who had reported using a conventional RDBMS as their primary data warehouse technology and a group of those who had reported one of the agile technologies as their primary data warehouse technology.

The second statement was "Our primary data warehouse solution simplifies the development of ETL/data integration processes for loading data." The results are shown in Table 14.

Simplified data integration	All	Conventional RDBMS	Agile
Strongly agree	18.45%	20.56%	16.98%
Agree	52.98%	50.47%	60.38%
Disagree	19.64%	23.36%	9.43%
Strongly disagree	5.36%	4.67%	5.66%

Table 14: Primary data warehouse solution simplifies the development of ETL and data integration.

For this question, those with the conventional RDBMS responded with a favorable response (agree or strongly agree) 71.03% of the time, while those opting for agile technologies responded favorably 77.36% of the time, although interestingly those with agile technologies overall did not overwhelmingly strongly agree. One possible interpretation for this is that in general, the agile technologies may simplify the typical aspects of data integration but may still pose challenges when it comes to more complex data sets or sources.

The third statement posed to the participants was "Our primary data warehouse solution provides strong support for the capture of, and collaborative use and management of metadata."

Metadata management	All	Conventional RDBMS	Agile
Strongly agree	13.69%	8.41%	26.42%
Agree	40.48%	42.06%	35.85%
Disagree	33.33%	37.38%	30.19%
Strongly disagree	8.33%	9.35%	1.89%

Table 15: The primary data warehouse solution supports metadata management.

As is shown in Table 15, there was a significant shift for strong agreement for those whose primary solution was one of the agile technologies (26.42% vs. 8.41%) and comparably much fewer strongly disagreeing (1.89% vs. 9.35%). In other words, those whose primary technology is one of the agile technologies consider their approach to better support metadata management.

Increasingly, organizations are deploying some of their applications using cloud-based or hosted systems, and there is an increased demand for data sourced from these and other cloud-based environments. The survey explored participants' perceptions of challenges in integrating data from cloud-based systems with the fourth statement, "Our primary data warehouse solution enables our organization to efficiently ingest data from a wide range of data sources including ones in the cloud." The results are shown in Table 16.

Data ingestion	All	Conventional RDBMS	Agile
Strongly agree	18.45%	14.02%	28.30%
Agree	42.26%	39.25%	49.06%
Disagree	29.17%	33.64%	20.75%
Strongly disagree	7.14%	10.28%	0.00%

Table 16: Primary data warehouse enables ingestion of data from many sources, including the cloud.

There is an overwhelming shift toward the favorable responses for the agile technologies. 28.30% of those at organizations adopting agile technologies strongly agreed with this statement; nobody strongly disagreed. The overall difference in favorable responses for the agile technologies was 77.36%, as opposed to 53.27% for those choosing the conventional approach. In particular, those with Hadoop solutions accounted for part of this favorable shift for the agile technologies – most of the Hadoop respondents strongly agreed with this statement.

The final statement examined the perceptions of the solution to support the "supply side" of data warehousing, specifically in publishing data to the cloud, with the statement "Our tools support the ability to publish data to cloud-based systems." The results are shown in Table 17.

Publish to cloud	All	Conventional RDBMS	Agile
Strongly agree	16.67%	13.08%	24.53%
Agree	35.12%	29.91%	43.40%
Disagree	28.57%	37.38%	15.09%
Strongly disagree	8.33%	10.28%	1.89%

Table 17: The primary solution supports publication to the cloud.

As with the previous statement, there is an apparent advantage associated with the agile technologies, with favorable responses 67.93%, compared with 42.99% for the conventional RDBMS solutions. Again, a significant portion of the favorable responses for the agile technologies is accounted for by Hadoop and those with cloud-based data warehouses.

While the responses for the first statement (about profiling the data) did not show significant differences, it appears that for the subsequent aspects of the data integration tasks (ETL development, metadata management and integration with the cloud), those participants from organizations that have one of the agile technologies as their primary data warehouse platform are more satisfied with how the solutions supported and facilitated data integration.

15 Diversity of the Technology Environment and the Emerging Hybrid Data Warehouse Environment

The results presented earlier in this study have already demonstrated that a growing number of organizations have adopted one of the agile technologies as their primary platform. However, another point of interest we uncovered is that there is also an apparent trend in which one or more of the agile technologies augment the data warehouse environment. Of the 168 respondents who reported that their organization had a data warehouse, 104 indicated that one or more of the agile technologies were used in their environment.

And this is true even for those who selected the conventional RDBMS as their primary platform. 66 respondents indicated that they are using a conventional RDBMS as their primary platform in conjunction with at least one of the agile technologies, and there are organizations that are using a combination of more than one of the agile technologies.

The number of users of each of these technologies in conjunction with another primary technology is listed in Table 18.

Agile technology	Number of respondents
Columnar database	34
Data warehouse appliance	30
In-memory database	40
Data warehouse automation	31
Cloud based data warehouse	25
Data lake	32
Hosted data warehouse solution	14
Hadoop HDFS	43

Table 18: Respondents indicating that their data warehousing environment blends more than one technology.

Of the 107 respondents who have indicated that their primary data warehouse is based on a conventional RDBMS, 66 have indicated that they are using at least one of the other agile technologies along with it. Also, we see situations where users are integrating a multitude of these technologies to work in conjunction with their primary platform technology. In support of that, we see that while 41 of the respondents indicated they are using only one of the agile technologies together with their primary platform, 28 respondents indicated they are using 2 of the agile technologies, and 35 are using 3 or more of these agile technologies along with their primary warehouse technology.

One might surmise from these results that organizations are open to adopting a range of auxiliary technologies to augment their data warehouse environment, creating a hybrid architecture that encompasses more than one technology.

Moreover, in combination with the results regarding technology awareness, it suggests that as data warehouse professionals become more experienced or work at organizations with more mature data warehousing programs, awareness of the agile technologies increases,

resulting in greater potential for integrating the different technologies together within the same environment.

16 Awareness of Agile Technologies

16.1 Does Level of Corporate Experience Impact Awareness and Familiarity of Agile Technologies?

Often, and unsurprisingly, an impediment to adopting a specific technology is simply that the designers are not familiar enough with what the technology can do to even consider it. As technology awareness is a prerequisite to adoption, it is important to assess the levels of recognition of the different technologies. Therefore, one part of the survey was focused on understanding respondent awareness and familiarity with the different agile data warehousing technologies.

The survey asked the respondents to characterize their organization's level of familiarity of newer data warehousing technologies and approaches, including data warehouse automation, cloud-based data warehouse, hosted data warehouse platform-as-a-service, data lakes, in-memory data warehouse and Hadoop. This question was posed to all respondents regardless as to whether they have a data warehouse platform in place. As a result, all 233 participants who completed the questionnaire responded to this question. The overall results are shown in Table 19.

	Unfamiliar	Slightly familiar	Very familiar
Data warehouse automation	28.33%	44.21%	27.47%
Cloud-based data warehousing	25.75%	48.50%	25.75%
Hosted data warehouse platform-as-a-service	30.04%	45.92%	24.03%
Data lakes	48.93%	32.62%	18.45%
In-memory data warehouse	36.48%	42.49%	21.03%
Hadoop	30.90%	40.77%	28.33%

Table 19: Level of familiarity with agile technologies.

Hadoop appears to have the highest level of "name recognition," with the highest percentage of "Very familiar" answers. Data warehouse automation and cloud-based data warehousing are not far behind. Conversely, the data lake appears to be the least recognized technology, with 48.93% reporting that their organization was unfamiliar with that technology.

When the pool of respondents was segmented according to years of corporate experience, we begin to see some interesting differences. First, the results from respondents at organizations that have had data warehouses for 4 or more years are shown in Table 20.

	Unfamiliar	Slightly familiar	Very familiar
Data warehouse automation	23.97%	43.84%	32.19%
Cloud-based data warehousing	20.55%	49.32%	30.14%
Hosted data warehouse platform-as-a-service	23.97%	45.89%	30.14%
Data lakes	39.04%	36.99%	23.97%
In-memory data warehouse	26.03%	45.89%	28.08%
Hadoop	25.34%	41.78%	32.88%

Table 20: Familiarity with agile technologies at organizations that have had data warehouse for 4 or more years.

No matter what has been the selected primary technology for the data warehousing platform, for more experienced organizations, across the board there is a shift of approximately 5% of the pool indicating that they are very familiar with the listed technologies. What most stands out, though, are the percentages listed for those unfamiliar with two specific technologies: data lakes and in-memory data warehouses, both of which show reductions of approximately 10 percentage points each.

These numbers can be contrasted with the smaller subgroup of those from organizations that have had a data warehouse for fewer than 4 years, shown in Table 21.

	Unfamiliar	Slightly familiar	Very familiar
Data warehouse automation	18.18%	50.00%	31.82%
Cloud-based data warehousing	36.36%	40.91%	22.73%
Hosted data warehouse platform-as-a-service	31.82%	54.55%	13.64%
Data lakes	59.09%	36.36%	4.55%
In-memory data warehouse	36.36%	54.55%	9.09%
Hadoop	40.91%	31.82%	27.27%

Table 21: Familiarity with agile technologies at organizations that have had data warehouses for fewer than 4 years.

The technology that warranted the highest level of familiarity was data warehouse automation, followed by Hadoop, and then cloud-based data warehousing. Interestingly, the data lake was the technology that garnered the most unfamiliar responses, followed by Hadoop.

The final segment that was examined included those respondents at organizations that did not yet have a data warehouse, shown in Table 22.

	Unfamiliar	Slightly familiar	Very familiar
Data warehouse automation	41.54%	43.08%	15.38%
Cloud-based data warehousing	33.85%	49.23%	16.92%
Hosted data warehouse platform-as-a-service	43.08%	43.08%	13.85%
Data lakes	67.69%	21.54%	10.77%
In-memory data warehouse	60.00%	30.77%	9.23%
Hadoop	40.00%	41.54%	18.46%

Table 22: Familiarity with agile technologies at organizations that do not have a data warehouse.

As might be expected, there is a much lower percentage of respondents indicating that they were very familiar with any of the agile technologies than those associated with an organization that does have a data warehouse. And similar to those at the organizations with fewer than 4 years of having a data warehouse, the technologies that respondents were more frequently unfamiliar with were data lakes and in-memory data warehouses.

16.2 Personal Experience and Technology Awareness

The same responses can be evaluated in terms of respondents' personal experience with data warehousing. The cohort was divided into the more experienced respondents (those with 4 or more years of experience) and the less experienced respondents (those with fewer than 4 years of experience). The results for the more experienced respondents are shown in Table 23.

	Unfamiliar	Slightly familiar	Very familiar
Data warehouse automation	21.34%	43.90%	34.76%
Cloud-based data warehousing	22.56%	51.22%	26.22%
Hosted data warehouse platform-as-a-service	27.44%	48.17%	24.39%
Data lakes	39.63%	36.59%	23.78%
In-memory data warehouse	24.39%	50.00%	25.61%
Hadoop	23.78%	42.07%	34.15%

Table 23: Technology familiarity – those with 4 or more years of personal data warehousing experience.

As in the earlier analysis using the organization's years of experience, both Hadoop and data warehouse automation closely shared first place for highest level of familiarity. Cloud-based data warehousing takes the next spot in this analysis, followed by in-memory data warehouses. Yet again, the data lake appears to be the technology where respondents have the most unfamiliarity.

The corresponding results for the less experienced respondents are shown in Table 24.

	Unfamiliar	Slightly familiar	Very familiar
Data warehouse automation	44.93%	44.93%	10.14%
Cloud-based data warehousing	33.33%	42.03%	24.64%
Hosted data warehouse platform-as-a-service	36.23%	40.58%	23.19%
Data lakes	71.01%	23.19%	5.80%
In-memory data warehouse	65.22%	24.64%	10.14%
Hadoop	47.83%	37.68%	14.49%

Table 24: Technology familiarity – those with fewer than 4 years of personal data warehousing experience.

Again, not surprisingly, much fewer of the individuals with less experience are very familiar with the agile technologies. The two highest percentages of familiarity were two similar technologies: cloud-based data warehousing and hosted data warehousing/platform-as-a-service. The technologies with the highest percentages of unfamiliarity were data lakes and in-memory data warehousing, followed by data warehouse automation.

16.3 Opportunities for Expanding Awareness and Familiarity

There are some conclusions that can be drawn from comparing these results.

First, it is not surprising that there is greater awareness of the agile technologies at the more experienced organizations. Even among those at more experienced organizations, though, the percentage of respondents who claim to be very familiar with any of these newer technologies never exceeds one-third of the population. Yet familiarity with a technology is typically a prerequisite to its evaluation and adoption within the organization. One might infer from this that adoption rates will potentially increase as awareness of the technologies increases, even in organizations that have already chosen a primary technology.

Second, while there seems to be a high percentage of reported familiarity with Hadoop, the lack of familiarity with data lakes, which are typically described as built using Hadoop and HDFS,, indicates something of a disconnect. It would be worth exploring whether the

conventional messaging around the concept of the data lake clarifies its relationship with the Hadoop ecosystem.

Third, for those who do not yet have a data warehouse, the lack of awareness of the agile technologies may hinder consideration of the agile technologies as part of the process of selecting a primary technology. This suggests that by producing more thought leadership content for prospective adopters, vendors of the agile technologies have an opportunity to influence individuals in considering and possibly adopting their products.

17 Evaluating New Technologies

The survey asked participants about where they were in the evaluation and acquisition process for the agile data warehousing alternatives and were given a choice of "No plans/very early stage," "Information gathering," "Proof-of-concept or prototyping," "Final evaluations" and "Acquired." We can characterize two categories (proof-of-concept and final evaluation) as actively engaged in acquisition, while the information gatherers are demonstrating interest in learning more.

17.1 Leading Technologies in the Acquisition Process

The results, shown in Table 25, are promising indicators of the potential for adoption.

	No plans/ very early stage	Information gathering	Proof-of-concept or prototyping	Final evaluations	Acquired
Data warehouse automation	24.89%	42.49%	12.02%	3.00%	17.60%
Cloud-based data warehousing	25.32%	42.06%	12.88%	5.58%	14.16%
Hosted data warehouse platform-as-a-service	28.33%	42.92%	12.88%	4.29%	11.59%
Data lakes	31.76%	39.06%	12.45%	4.29%	12.45%
In-memory data warehouse	27.90%	40.77%	13.30%	4.72%	13.30%
Hadoop	25.75%	35.62%	16.31%	4.72%	17.60%

Table 25: Where respondents are in the evaluation process.

These results indicate that in general, for those technologies we have identified as agile technologies, most respondents are in the information gathering phase. The technology that appears to rank highest among those exploring new technologies is Hadoop, where 17.60% have already acquired the technology, and over 21% are actively involved in the acquisition process. While data warehouse automation ranks with Hadoop in terms of those who have acquired the technology (also 17.60%), the corresponding percentages of the actively engaged are slightly lower. Cloud-based data warehousing (17.46%) and in-memory data warehousing (18.02%) also have a healthy percentage of respondents actively engaged in acquisition.

17.2 Does Corporate Data Warehouse Experience Influence Evaluations?

Are there any predispositions for evaluating agile technologies depending on organizational maturity? The pool of respondents' organizations can be divided into three subgroups: those that have had a data warehouse for more than 4 years, those that have had a data warehouse for fewer than 4 years and those that do not yet have a data warehouse.

The results for the 146 respondents from organizations with data warehouses for more than 4 years are shown in Table 26.

	No plans/ very early stage	Information gathering	Proof-of-concept or prototyping	Final evaluations	Acquired
Data warehouse automation	19.18%	41.78%	11.64%	2.74%	24.66%
Cloud-based data warehousing	19.86%	41.10%	13.70%	6.16%	19.18%
Hosted data warehouse platform-as-a-service	24.66%	41.10%	13.01%	4.79%	16.44%
Data lakes	27.40%	35.62%	13.01%	5.48%	18.49%
In-memory data warehouse	23.29%	36.30%	15.07%	4.79%	20.55%
Hadoop	19.86%	33.56%	17.81%	5.48%	23.29%

Table 26: Evaluation of the agile technologies – organizations with 4 or more years of corporate data warehouse experience.

As with the overall population, at experienced organizations the highest percentages of respondents are in the information gathering stage. However, the results definitely show an increase in the percentages of those that have already acquired the newer technologies. Data warehouse automation leads the technologies in acquisitions, followed by Hadoop. In addition, the two technologies that exhibit the highest degree of active engagement are Hadoop (23.29% aggregate), followed by in-memory data warehousing (19.86% aggregate) and cloud-based data warehousing (also 19.86% aggregate).

The results for the 22 participants with fewer than 4 years of corporate data warehouse experience are shown in Table 27.

	No plans/ very early stage	Information gathering	Proof-of-concept or prototyping	Final evaluations	Acquired
Data warehouse automation	18.18%	31.82%	18.18%	9.09%	22.73%
Cloud-based data warehousing	22.73%	45.45%	9.09%	9.09%	13.64%
Hosted data warehouse platform-as-a-service	18.18%	45.45%	22.73%	9.09%	4.55%
Data lakes	27.27%	40.91%	22.73%	4.55%	4.55%

In-memory data warehouse	18.18%	45.45%	22.73%	13.64%	0.00%
Hadoop	31.82%	22.73%	22.73%	4.55%	18.18%

Table 27: Evaluation of the agile technologies – organizations with fewer than 4 years of corporate data warehouse experience.

For this group it also appears that most respondents are in the very early stages or in the information gathering phase for nearly all technologies. As with the more experienced organizations, the leaders in terms of acquired technology are data warehouse automation and Hadoop. And although the size of this cohort is small, there does seem to be a generally high percentage of active engagement, particularly in the prototyping stage.

Finally, the results for the group of 65 respondents whose organizations currently have no data warehouse solution are shown in Table 28.

	No plans/ very early stage	Information gathering	Proof-of-concept or prototyping	Final evaluations	Acquired
Data warehouse automation	40.00%	47.69%	10.77%	1.54%	0.00%
Cloud-based data warehousing	38.46%	43.08%	12.31%	3.08%	3.08%
Hosted data warehouse platform-as-a-service	40.00%	46.15%	9.23%	1.54%	3.08%
Data lakes	43.08%	46.15%	7.69%	1.54%	1.54%
In-memory data warehouse	41.54%	49.23%	6.15%	1.54%	1.54%
Hadoop	36.92%	44.62%	10.77%	3.08%	4.62%

Table 28: Exploring agile technologies – organizations without a data warehouse.

In this subgroup, there are (understandably) smaller percentages of those in the final evaluation or acquired stage. Rather, there is a shift toward the earlier aspects of the evaluation process, with most respondents in the information gathering phase. Again, more of those evaluating cloud-based data warehousing and Hadoop, followed by data warehouse automation, appear to be further along in the process than some of the other technologies.

17.3 Organizational Maturity and the Hybrid Environment

Clearly, the increased publicity and awareness campaigns associated with Hadoop may account for its apparent popularity in addition to the conventional RDBMS approach to reporting and analytics. At the same time, it seems clear that across the board, in addition to Hadoop, three of the technologies are recognized as adding value to the organization,

namely data warehouse automation, cloud-based data warehousing and in-memory data warehousing.

That being said, there is definitely greater adoption and active engagement among those at more mature organizations. There may be a number of motivating factors for organizations with established data warehousing programs to be open to expanding the technology footprint of their environment with agile technologies. Those who have had a degree of success may be looking to improve operational performance, provide storage augmentation or improve accessibility to the business analysts. Those organizations that have struggled with the challenges of moving their warehouse into production may be exploring how other technologies may simplify the process.

18 Reflections and Considerations

18.1 The Enterprise Data Warehousing Environment Is Adapting to New Technologies

There is no doubt that the relational database remains the core of most data warehousing platform environments, as even a number of the agile technologies – columnar databases, in-memory databases, data warehouse appliances and cloud-based or hosted data warehouses – rely in some ways (or often many ways) on the relational database model. However, the results of our survey clearly indicated a strong interest in evaluating the newer agile technologies, integrating them within the reporting/BI/analytics environment and in many cases choosing them as the primary technology.

18.2 Adopting Agile Technologies

Some organizations that are beginning their data warehousing journey are going straight to the agile technologies. Practically speaking, though, any organization that has already committed a large investment in developing its existing platform is not likely to rip it out and replace it with a completely different technology. Except in a few scenarios, the level of effort involved in a complete renovation via migration to a new technology introduces risks that organizations must consider carefully. Rather, it is more likely that an organization intent on adopting new technologies will do so alongside the existing platform.

18.3 The Benefits of Agile Technologies

While there are some areas that our research did not indicate any aggregate preference between the conventional RDBMS approach and the agile technologies, it did indicate that there is general agreement with many of the claims made by vendors of the different agile technologies. Our results largely indicated that agile technologies correlate with lowered staff allocation, are perceived to simplify moving a data warehouse into production, are seen to have a reasonable cost and are perceived to be worth the value. These technologies also are seen to deliver value from an operational perspective in terms of having the right tools and guidance for configuration, management, ease of development for data integration, metadata management and interoperability with cloud-based systems.

18.4 Maturity Levels, Technology Choice and Evaluations

Those at organizations with more than 4 years data warehousing experience are more likely to be very familiar with the agile technologies than those at organizations with fewer than 4 years of experience. Individuals with 4 or more years of experience are more likely to be familiar with data warehouse automation and Hadoop, but are less likely to be familiar with data lakes. However, among those with less than 4 years of experience, there is greater familiarity of cloud-based and hosted data warehousing. Organizations with shorter duration of corporate data warehousing experience (2-6 years) seem to be considering the agile technologies as their primary platform at higher rates.

Among the respondents, Hadoop, cloud-based, in-memory, and data warehouse automation lead active exploration (that is, those who are either developing proofs-of-concept, prototyping or in a final evaluation stage). The technology that appears to rank highest among those exploring new technologies is Hadoop, followed closely by cloud-based data warehousing, in-memory data warehousing and data warehouse automation.

18.5 Awareness of Agile Technologies

As might be expected, in general there is greater awareness of the agile technologies that are more mature, such as Hadoop, data warehouse automation, and cloud-based and hosted data warehousing. In particular, more experienced people are more familiar with Hadoop and data warehouse automation. Interestingly, fewer people seemed to be aware of the concept of the data lake, and we are inferring from this that in some situations, awareness of Hadoop at a conceptual level does not necessarily translate into familiarity with its practical application.

Intellectually, it is reasonable to believe that technology adoption rates will potentially increase as awareness of the technologies increases, even in organizations that have already chosen a primary technology. Therefore, vendors have opportunities to raise awareness. By producing more thought leadership content for prospective adopters, vendors of the agile technologies have an opportunity to influence individuals in considering and possibly adopting their products.

18.6 Architecting for a Hybrid Data Warehousing Environment

As data warehouse professionals become more experienced or work at organizations with more mature data warehousing programs, awareness of the agile technologies increases, resulting in greater potential for integrating the different technologies together within the same environment. Organizations seem to be open to adopting a range of auxiliary technologies to augment their data warehouse environment, creating a hybrid environment that encompasses more than one technology.

This suggests that anyone involved in the planning and design for a data warehousing environment should presume that it will encompass both conventional and newer technologies. Therefore, the results suggest that the data warehousing environment be designed from the start with the intent of integrating different types of data warehousing technologies to address different types of users' needs over time.

18.7 Preparing for Evaluation of Alternate Agile Technologies

The agile technologies are designed to address low performance in one or more of a number of areas, including delays in design and development, inadequate response times or the need to accommodate greater data volumes and more diverse data variety. If you are considering exploring how these technologies can improve how reporting and analysis can create corporate value, there are some steps to take to properly prepare the organization.

- **Team**: Identify the key individuals who will manage and execute the evaluation. Choose individuals with a number of years of experience, as our research indicates

increased awareness correlates with greater experience. Ensure that the team members are good communicators and can share their feedback with corporate stakeholders.

- **Assessment**: Assess the areas of low performance and prioritize in terms of impediments to value creation.
- **Evaluation and synthesis**: Determine which of the technologies best addresses the most critical issues.
- **Education**: Accumulate intelligence about the candidate technologies to understand how they work, how they interoperate with your existing application landscape and what the requirements are for instantiation.
- **Evaluation criteria**: Specify the metrics that are to be used to compare how each alternative addresses the areas of low performance and how the vendors and their products within each category meet your corporate needs.
- **Laboratory**: Establish an evaluation platform that reflects how the technology would be integrated within a production environment to ensure that the performance characteristics of any pilots/proofs-of-concept can be accurately measured.
- **Practical proof-of-concept**: Design a challenge for testing the technologies using real data associated with real business challenges. Use your organization's objectives as the yardstick by which vendor products are compared.

At this point the context for evaluation of the technology is established, and experienced individuals are positioned to manage and execute the evaluation program. This allows the organization to be prepared to engage vendors within each specific technology category and have them demonstrate their products.

19 About the Authors

David Loshin, Managing Director at DecisionWorx (www.decisionworx.com), is a recognized thought leader and expert consultant in the areas of analytics, big data, high performance computing, data governance, data quality, master data management and business intelligence. Along with consulting on numerous data management projects over the past 15 years, Loshin is also a prolific author of numerous books and papers regarding all aspects of data management and business analytics, including the recently published *Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph*. He is a contributor to the searchBusinessAnalytics and searchDataManagement channels at TechTarget and is a frequent invited speaker at conferences, web seminars and sponsored web sites and channels.

Abie Reifer, Managing Director at DecisionWorx (www.decisionworx.com), is a technology and strategy leader with extensive experience in system implementations. He currently serves in a leadership position at a data collection and survey research organization. Previously Reifer served as CTO of an international telecommunications company. He has also served as an advisory strategy consultant to a leading US telecommunications carrier. Reifer began his career at Bell Communications Research and received his Master's degree in Engineering from Columbia University.

20 About The Bloor Group

The Bloor Group provides detailed analysis of today's enterprise software industry. Co-founded by Dr. Robin Bloor and Eric Kavanagh, the company leverages Web-based multimedia to deliver vendor-neutral education that is designed to reveal the essential characteristics of information technologies. The Bloor Group galvanizes the industry's independent analysts to provide valuable insights via innovative Webcasts, articles, research programs and white papers.

Visit www.bloorgroup.com and www.insideanalysis.com for more information.

21 About the Sponsors

TimeXtender Data Warehouse Automation (DWA) is revolutionizing the way a data warehouse is developed and maintained. By utilizing technology that automates all manual Data Warehouse encoding processes; from design to development and operation and maintenance to change management, TimeXtender ensures an improved and inexpensive solution that is fully documented.

TimeXtender is the global market leader in DWA. This means that the medium and large enterprises at home and abroad easily, quickly and safely can start to set up and structure their precious data in a Data Warehouse. Learn more at www.timextender.com

IBM

IBM has the most complete set of capabilities to enable today's Logical Data Warehouse. IBM is a strategic advisor that can help clients:

- Innovate by effectively addressing a broad and ever-evolving set of data management needs. Through our strategic acquisition strategy and organic development borne of IBM Research, we have amassed true best-in-class data warehousing solutions that address virtually every information need.

- Boost the success of their data warehouse initiatives. IBM combines proven innovations, a sharp focus on integration and world renowned industry experts to help ensure the success of data warehouse initiatives. Plus, IBM is an industry leader in providing a comprehensive solution portfolio that targets key aspects of data warehousing – including data integration, governance and security.

- Modernize their data warehouse to fuel real-time business. IBM's commitment to innovation focuses on designing the right mix of data platforms and integration capabilities for our clients' changing business requirements.

To learn more contact your IBM representative or IBM Business Partner, or visit: ibm.com

cloudera

Cloudera delivers the modern data management and analytics platform built on Apache Hadoop and the latest open source technologies. The world's leading organizations trust Cloudera to help solve their most challenging business problems with Cloudera Enterprise, the fastest, easiest and most secure data platform available for the modern world. Our customers efficiently capture, store, process and analyze vast amounts of data, empowering them to use advanced analytics to drive business decisions quickly, flexibly and at lower cost than has been possible before. To ensure our customers are successful, we offer comprehensive support, training and professional services.

Learn more at http://cloudera.com

WhereScape®

The pioneer in data warehouse automation software, WhereScape empowers organizations worldwide to deliver business value from their decision support infrastructure - saving time, money and reducing risk.

Working smarter and faster as new organizational demands emerge and data sources expand in scale and complexity, WhereScape RED & WhereScape 3D enable organizations to prototype, build, deploy, manage and renovate decision support infrastructures – including enterprise data warehouses, business-facing data marts, and big data solutions.

Learn more at www.wherescape.com