Aité

PARTNER. ADVISOR. CATALYST.

Big Data in Capital Markets: At the Start of the Journey

Prepared for:

THOMSON REUTERS

TABLE OF CONTENTS

LIST OF FIGURES

EXECUTIVE SUMMARY

Big Data in Capital Markets: At the Start of the Journey, commissioned by Thomson Reuters and produced by Aite Group, explores the development of big data strategies and technologies across the buy-side and sell-side capital markets communities. It identifies use cases, challenges, and opportunities for big data technology in the sector and is based on surveys with 22 capital markets firms conducted during the months of May and June 2014, representing an even split between buy-side and sell-side participants. Participants are a subset of 423 firms contacted for the survey and represent firms that have some experience or knowledge of big data; the majority of those contacted indicated that they do not.

Key takeaways from the study include the following:

- The majority of firms active in the capital markets do not have a big data strategy in place at an enterprise level—only 5% of the 423 firms contacted felt they had enough knowledge of the subject to participate or were willing to talk about their big data programs (some feel it is too commercially sensitive an area to discuss)—most firms implementing big data are doing so in specific functional areas. The most popular use cases for big data within respondent firms are analytics for trading and quantitative research.

- Half of respondents have invested in big data already. Most firms that have implemented a big data project hail from the banking and hedge fund communities, though a handful of asset managers have invested in such projects.

- Half of respondent firms either currently employ or plan to hire a data scientist in the next 24 months, correlating directly with these firms' use of big data.

- The aspects of most importance to respondent firms in terms of big data deployments signal that actionable information and insight are equally pegged with scalability for future data volume increases.

- Respondents cite inadequate technical knowledge as the most commonly encountered challenge during a big data project, highlighting the need for appropriately skilled staff to implement such a strategy.

- As more capital-market-specific use cases for big data become prevalent in the market, firms will become more comfortable with these strategies.

- Current investments in big data are largely focused on revenue generating opportunities in the front office, but the future is likely to see much more focus on client retention, compliance function support, and enterprise risk management and governance.

Figure 1 provides an overview of the key findings of the survey, highlighting the big data dynamics within the 22 capital markets firms interviewed.

Figure 1: Highlights of Big Data Survey
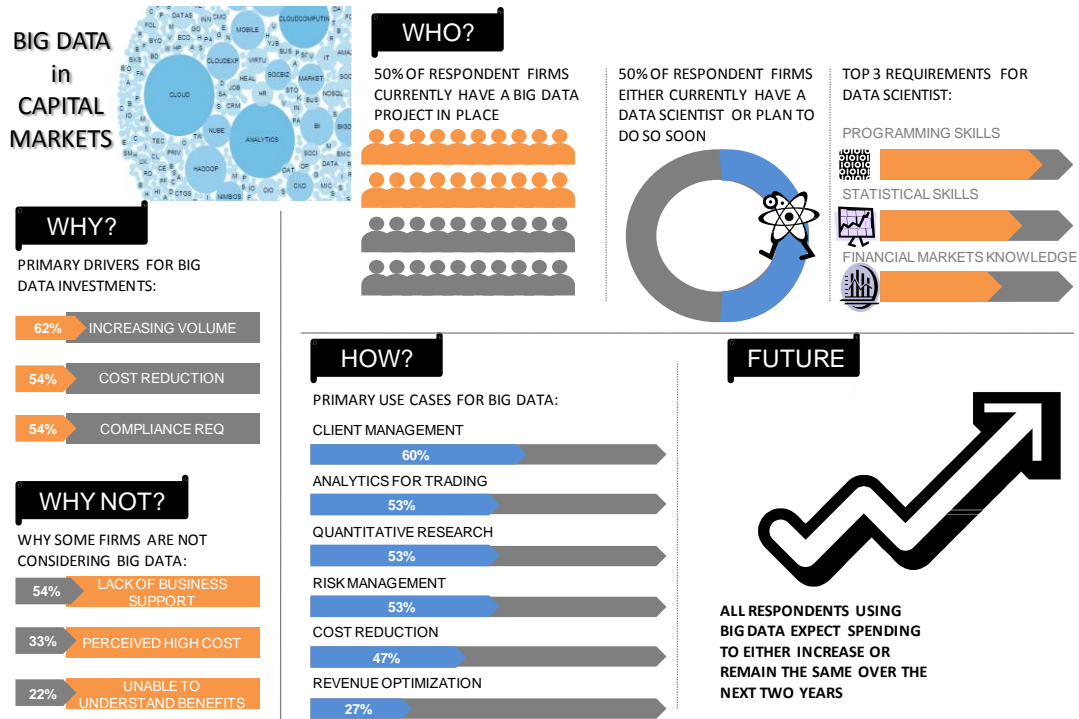
Source: Aite Group survey of 22 capital markets participants, May and June 2014

INTRODUCTION

Over the last decade, the concept of big data has been applied to a whole host of industries but the capital markets have been relatively slow to adopt these strategies. Within the financial services sector, big data has gained far more traction within retail banking due to the increasing desire of these financial institutions to profile their customers in a similar manner to early adopters of big data such as Amazon or Google. On the institutional side of the capital markets, on the other hand, there has traditionally been far more customer stickiness; hence there has been less incentive to apply big data in this manner.

Big data strategies have, however, begun to make some impact in a select few areas of the capital markets over recent years, including within sentiment analysis for trading, risk analytics, and market surveillance. This study identifies challenges and opportunities in current and future big data initiatives and analyzes the potential benefits of adopting big data strategies and technologies within the capital markets community.

METHODOLOGY

This study is based on Aite Group surveys conducted with leading market participants representing banks, asset and wealth managers, third party administrators, broker-dealers, and hedge funds that have knowledge of big data strategies and technologies. During the months of May and June 2014, Aite Group reached out to a total of 423 firms and interviewed 22 firms, representing an even split between buy-side and sell-side participants (Figure 2), to capture their views on data management strategy and the position of big data within these strategies. Given the size and structure of the research sample, the data provide a directional indication of conditions in the market for those firms that are aware of big data strategies.

The firms that did not take part in the survey reflect that there is a high percentage of the capital markets community that does not have familiarity with big data strategies or does not have such a strategy in place. The interviewed respondent firms represent 5% of the overall populace of firms included in the outreach. Those that did not participate indicated that they:

- Were not aware of big data strategies in their organizations

- Did not have sufficient experience of big data to comment

- Viewed it as a competitive differentiator and therefore did not wish to discuss future plans or current investment

6

Figure 2: Respondent Firm Type

Respondent Firm Type
(N=22)
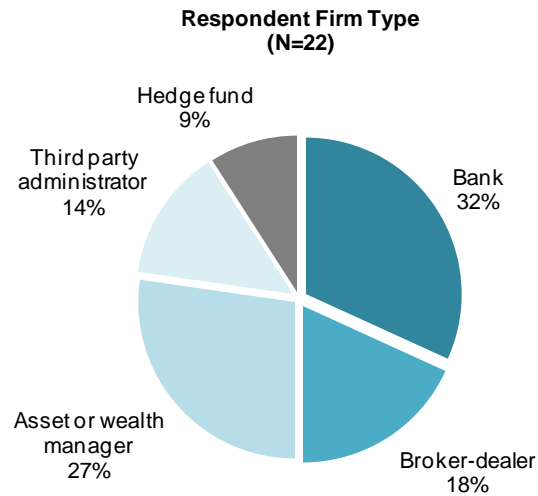
Source: Aite Group survey of 22 capital markets participants, May and June 2014

In terms of individual respondents, data management, IT, and operations roles dominate (Figure 3), but there are a couple of respondents from the trading and research and development functions. This relatively broad range of functions reflects the scope of applications for big data in the capital markets sector and the lack of centralized ownership and oversight of these implementations or evaluation efforts at most firms. Big data is therefore not the sole remit of data management at most capital markets firms. This particular dynamic represented a hurdle to be overcome in the survey outreach process because it was challenging to find a single individual within financial institutions that was aware of all of the applications of big data across the business.

Figure 3: Role of Respondents

Respondent Job Function
(N=22)

Research and development 9%

Trading 9%

Data management 37%

Operations 18%

IT 27%
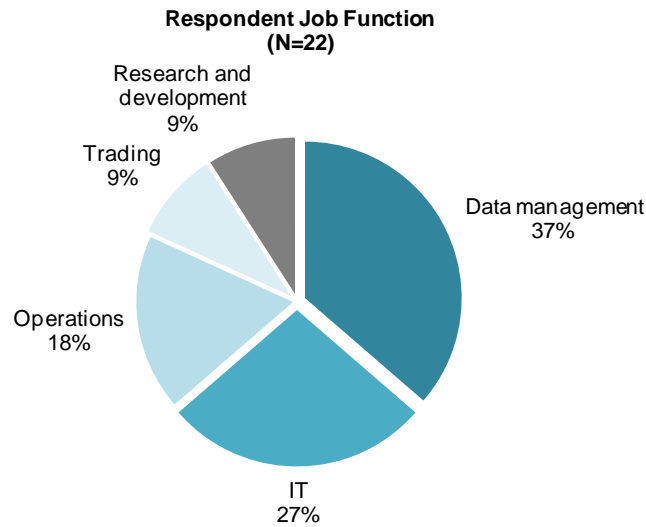
Source: Aite Group survey of 22 capital markets participants, May and June 2014

Figure 4 shows the regional presence of respondent firms, indicating the high level of market penetration in North America and the developed Asia-Pacific regions. They also have a strong presence in emerging markets, signifying the global nature of their businesses.

Figure 4: Active Regions for Respondent Firms

Geographic Presence of Respondent Firms
(N=22)

North America	82%
The developed Asia-Pacific	64%
Developed Europe	59%
The emerging Asia-Pacific	50%
Latin America	45%
Emerging Europe	45%
The Middle East/Africa	41%
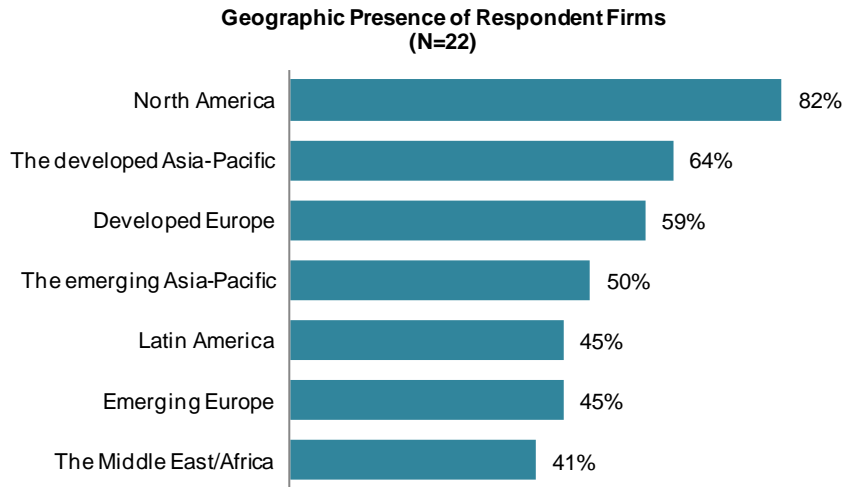
Source: Aite Group survey of 22 capital markets participants, May and June 2014

Respondent firms are very active in most major asset classes, including a very high use of derivatives and FX, reflecting the higher level of market penetration for big data in coping with

complex data related to these instruments (Figure 5). There is some correlation between the range of instruments supported and the use of big data by these financial institutions; the higher the number of complex instrument types, the higher the use of big data across various use cases and data types.

Figure 5: Asset Classes Supported by Respondent Firms

Asset Class Coverage of Respondent Firms
(N=22)

Asset Class	Percentage
Derivatives	82%
FX	82%
Equities	77%
Swaps	77%
ETFs	73%
Exchange-traded equity options	68%
Exchange-traded futures	68%
Fixed income	68%
Funds	59%
Other (Please specify)	9%
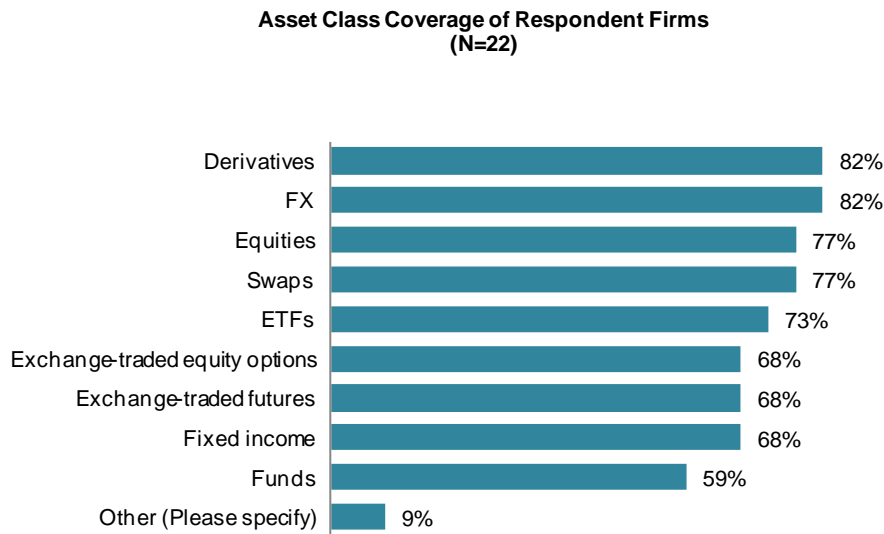
Source: Aite Group survey of 22 capital markets participants, May and June 2014

Only half of respondent firms have already got big data initiatives in place and a minority is actively considering investment in such an initiative in the near future (Figure 6). This indicates that despite the hype that surrounds big data, penetration in the capital markets is still rather patchy. It should again be noted that this group represents a particularly educated subset of the market at large.

Figure 6: Current Investment in Big Data

Q. Does your organization currently have a big data initiative or strategy in place?
(N=22)
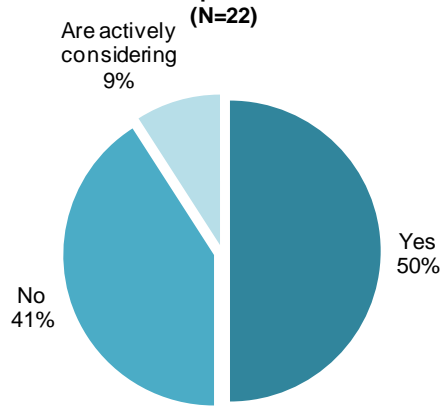
Source: Aite Group survey of 22 capital markets participants, May and June 2014

DEFINING BIG DATA

Big data has been a much misused and misunderstood term within the financial services industry for some time, applied to everything from traditional relational databases (RDBMS) to web-based sentiment analysis tools. It is frequently linked to the underlying technology that is being marketed as "big data" (hence the aforementioned references to databases and tools), rather than introduced as a standalone concept, but when it is discussed in a theoretical manner, it is often poorly defined. The judgment of what exactly constitutes as "big" is also somewhat of a contentious matter; are we talking in terabytes, petabytes, or exabytes? Does the strategy necessarily have to be focused on a project related to a high volume of data, or can other factors determine its application?

The concept of "big" in a retail or scientific environment is different from what is considered to be big in a capital markets context. The profiling of consumers within the retail markets, for instance, involves analysis of unstructured data collected from a whole host of sources such as a broad range of social media, whereas capital markets tend to deal largely in structured data sets from a more limited set of defined sources—market data vendors, market infrastructures, and counterparties, for example. Some unstructured data sets have, however, become important to capital markets institutions in areas such as sentiment analysis and market surveillance for profiling certain trends and activities in the market and within firms, but these have not traditionally been the data sets of primary importance to the business.

In order to conceptualize the scale of a big data problem, it is worth looking at exactly how big a petabyte actually is in relative terms. One petabyte is 1,000,000,000,000,000 bytes of digital information and it is enough to store the DNA of the entire population of the United States (Figure 7). An exabyte is 1000 petabytes.

Figure 7: The Relative Size of a Petabyte
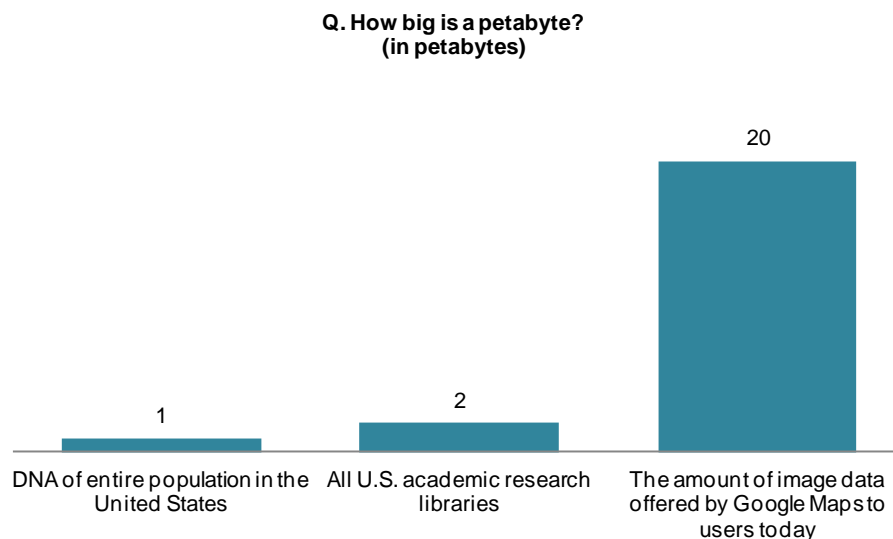
Source: Aite Group

A common definition across sectors for big data is a strategy or technology deployment that deals with data problems that are too large, too fast, or too complex for conventional database or processing technology. The concept of big data is therefore often linked to the three Vs, which were identified by Gartner in 2001:

- Velocity—the speed of data delivery and processing

- Volume—the amount of data that must be managed or processed

- Variety—the range of different data sets that must be dealt with, including both structured and unstructured formats

Another V has been added to the list by some academics over the last few years; veracity, which relates to measurement or labeling of the integrity or quality of the data being processed or stored; thus a reflection of the rising importance of data transparency. The manipulation, management, and interpretation of these data sets—complex, high volume, or varied—therefore constitutes a big data challenge, according to this logic.

There are significant grey areas remaining however, due to the fact that not all analytics or data processing efforts can be categorized as big data, but are increasingly being included in that bracket by financial institutions and software vendors. In order to understand the concept of big data and its related terminology, it is best to consider the strategic elements separately from the technology elements. Given that the industry tends to link these together, there will continue to be some degree of confusion about whether use of a big data technology necessarily entails the solving of a big data problem. By separating out the strategic elements of big data from the technology associated with the concept it is possible to reduce some degree of misinterpretation, though the grey areas do persist.

STRATEGY AND TECHNIQUES

It is questionable whether any of the three Vs alone can be considered to represent a big data problem in the capital markets. Strategies that focus only on coping with a high volume of data have existed for some time and have not previously been labeled as big data strategies. The gradual increase in trading volume across financial markets over recent years, for example, has not posed a significant threat to current front, middle or back office technology (aside from areas where manual intervention or workarounds are rife). Tick data has always been a large data set and firms have been dealing with that in a structured manner for years. They may not be perfect, but trading and securities processing technologies have generally been able to scale to meet increased electronic flows of data resulting from market structure change and increased electronic activity.

Related to this, in the front office, high frequency trading (HFT) technology has adequately coped with much higher velocity of data—moving the industry from dealing in seconds to milliseconds to microseconds—but this is not traditionally considered solely the realm big data (there are plenty of other buzzwords for that space). Analytical manipulation of complex and varied data sets related to instruments such as over-the-counter derivatives has also been in

existence for much longer than big data has existed as a concept. Complexity or variety of data alone is therefore not sufficient to warrant being tagged a big data problem.

So what exactly constitutes a big data challenge in the capital markets context? There appears to be the most take-up of the term in areas related to projects that involve multiple variables, such as high volumes of complex data that must be cross-referenced in a specific timeframe. These tasks do not necessarily need to be performed in real time, though they can be in the context of sentiment analysis for trading, but there is a focus on reducing the latency of data aggregation tasks for ad hoc regulatory reporting or risk analysis, for example. As large firms focus on trying to eliminate silos, especially in response to regulatory mandates, this aggregation has presented significant challenges. Siloed data sets must now be combined from heterogeneous asset class, product, and risk information. Big data strategies in these markets tend to be synonymous with analytical tasks or those related to reporting or governance functions. The consumption of text-based unstructured data has also been a significant driver for some projects.

Currently, there are no viable public examples of enterprise-wide big data projects in place within financial institutions active in the capital markets; hence it is very early days overall for the development of these strategies and techniques within the sector. There are plenty of projects being conducted at the business unit level but no public case studies at a group-wide level.

THE EVOLVING DATA FUNCTION

The big data story in the capital markets is directly tied to the rising importance of data management as a function within financial institutions. Regulatory, client, and internal drivers have forced most firms to re-evaluate the core reference data sets on which they are basing their trading, risk management, and operational decisions. The establishment of C-level executive positions dedicated to championing data management and data governance—a chief data officer or something similar—is proof of this enhanced focus.

As well as staffing changes, there is also a gradual shift going on behind the scenes in the shape and structure of data management within many firms. There is much more impetus being placed on transparency within the financial markets overall and on proving the reliability of data sets on which strategic and day-to-day decisions are being made. Providing an audit trail for data underlying risk analytics or pricing of trades, for example, is considered to be of utmost importance by both institutional investors and regulators. Transparency requirements have therefore meant data must be tagged with metadata to prove lineage and provenance of individual items. Some firms are working on centralization programs in order to store and scrutinize this data in a comprehensive manner, as well as to bring down data costs in the long term.

Figure 8 shows respondents' views on the level of importance that is being placed on data management and analytics internally within their firms, reflecting that most place at least a medium priority on the function. More than half of respondents believe their firms have placed a high priority on data management and only a minority feels it is a low priority. The firms that place a low priority on the data both hail from the hedge fund sector, thus reflecting the propensity within the alternative funds segment for some firms to outsource middle and back

office functions to a third party; hence responsibility for data management may sit outside of these firms' internal teams.

Figure 8: Level of Importance Placed on Data Management by Respondent Firms

Internal Priority Level for Data Management and Analytics
(N=22)

Low priority
9%

Medium priority
36%

High priority
55%
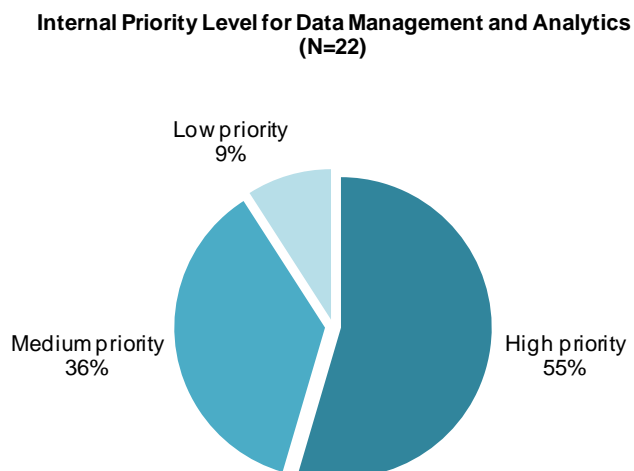
Source: Aite Group survey of 22 capital markets participants, May and June 2014

Priority levels for data management and analytics have risen as adoption of electronic trading has spread across different regions and asset classes, diversity of data sources and sheer volume of data have increased substantially over the last decade. In addition to traditional market data, growing interest around non-traditional, unstructured data has also added more complexity in terms of firms' ability to deal with data. On top of all these challenges, regulatory compliance requirements around data integrity, storage, and accessibility have added more pressure on market participants to develop a viable long-term data management strategy to meet both competitive and regulatory needs.

Along with an increased priority on data-related functions, these firms have engaged in more actively evaluating their internal strategies related to data management and analytics with a view to improving them. This evaluation process tends to be based on a relative assessment against peers in the market—firms of a similar size and focus. Figure 9 indicates that half of respondents feel their data management strategy is on par with their peers in the market, though the other half display a broad range of perceptions about their strategy—everything from significantly behind to significantly ahead. There is a correlation between respondents that consider their firms to be significantly ahead of their peers and those that have invested in big data initiatives. This reflects respondents' beliefs that the majority of their peers have not invested in big data or are at an earlier stage in the investment process.

Figure 9: Perception of Internal Data Management Strategy

Q. How would you rate your organization's data management strategy?
(N=22)

Significantly behind peers
4%

Don't know
4%

Ahead of peers
14%

Significantly ahead of peers
14%

On par with peers
50%

Behind peers
14%
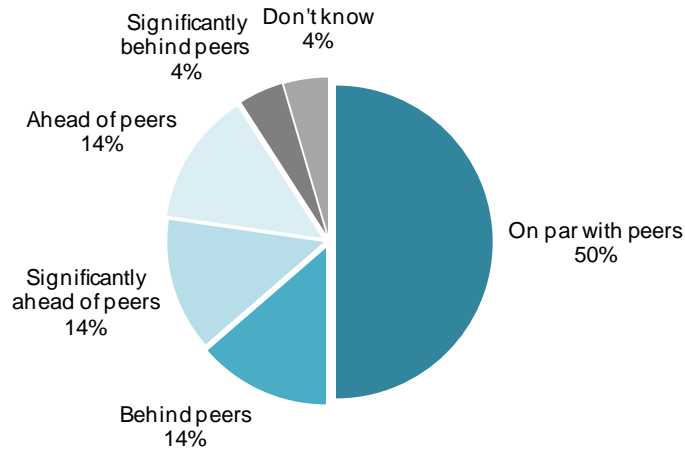
Source: Aite Group survey of 22 capital markets participants, May and June 2014

More respondents are confident in their firms' data analytics capabilities than their data management strategies (Figure 10), likely due to higher focus and spending on risk analytics support in the post-Lehman environment and a focus on gaining an edge in performance by adding layers of analytics to key front-office functions. Perception of data analytics sophistication, unsurprisingly, correlates closely with usage of big data strategies among respondent firms. This does not, however, mean that firms are limiting analytics to big data implementations—structured data analytics are also included.

Figure 10: Perception of Internal Data Analytics Strategy

Q. How would you rate your organization's data analytics strategy?
(N=22)
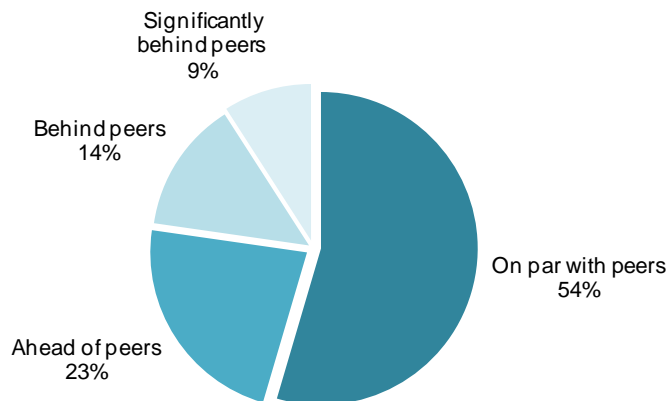
Significantly behind peers
9%

Behind peers
14%

On par with peers
54%

Ahead of peers
23%

After five years of maintenance and stagnant IT spending, some firms are finally focusing on innovation and revenue growth to capture the next generation of market opportunities. To this end, an increasing number of firms are focused on data-centric improvement programs, looking to discover unique data-driven initiatives that can help them find alpha.

THE RISE OF THE DATA SCIENTIST

One of the key factors in the establishment and development of a big data strategy is the hiring and retention of sufficiently qualified staff members to lead the project and one such role is a data scientist. Data science is a discipline that encompasses the range of techniques and processes used to generate business value from data. Use of the word 'science' in this circumstance relates to the scientific processes applied to data analytics such as the development of a hypothesis, the subsequent experimentation process, and the formulation of a conclusion.

The role of a data scientist has appeared within capital markets as C-level executives have begun to seek revenue-generating opportunities by mining their firms' data assets and to search for opportunities to realize operational efficiencies. Data management was once considered to be primarily the realm of the back office and IT, but is now increasingly considered to be a vital source of value for the business. Nevertheless, there is some ambiguity surrounding the role of a data scientist within an organization.

Traditionally data analysts and managers tend to look at a single category of data such as securities reference data, whereas data scientists are focused on analyzing information from numerous disparate sources with the objective of unlocking insights that will either create value or provide a solution to a business problem. The job of a data scientist is more broadly defined now and goes beyond just analytics to include consultancy services, research, enterprise-wide taxonomy, automating processes, ensuring the firm keeps pace with technology development and managing analytics vendors.

Half of respondent firms either currently employ or plan to hire a data scientist in the next 24 months, correlating directly with these firms' use of big data (Figure 11)—all of the firms with these plans currently have a big data strategy in place. The firms with no plans to employ a data scientist are either at an early stage in their evaluation of big data or have no plans to use such a strategy in the near future. Only one respondent admitted to not recognizing the data scientist function, though it should be noted that the total community of respondents have particularly high awareness of big data overall comparative to the general capital markets populace.

Figure 11: Respondent Firm Views on the Data Scientist Role

Q. Do you employ a data scientist?
(N=22)

No, but plan to
employ in the next
24 months
9%

What is a data
scientist?
5%

No, no plans to
employ
45%

Yes, currently
employ
41%
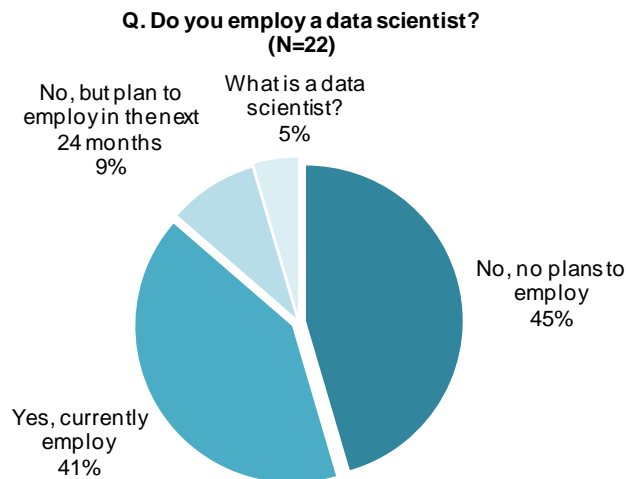
Source: Aite Group survey of 22 capital markets participants, May and June 2014

Even though there is confusion regarding the role of a data scientist, there is some consensus across various other industries about the core competencies a big data professional should have. A statistical and programming ability is essential to first determine which software packages are suitable to run and develop new analytics, but also to understand the correct techniques to best establish relationships between large volumes of heterogeneous data sets and help define common taxonomies among them. In addition to a quantitative skill set, employers are now emphasizing the need for 'softer' skills such as visualization and communication proficiency, which are important in the process of transforming analytical data into a human-consumable form.

The ability to simplify complex analytical data outputs into usable insights for the business requires an understanding of the business context and the end functions into which this data must be fed. Accordingly, data scientists that have knowledge of key challenges facing the financial institution in question and even the industry are better able to deliver insights that generate value or solve a specific problem—this is why respondents highlight financial knowledge as the second most important requirement for a data analyst (Figure 12).

It is no longer sufficient to have a data scientist with a strong mathematical ability, communication skills and business aptitude are also very important. This is where the lack of data scientists at an industry level within capital markets proves problematic because though applicants may have statistical and programming skills, they are less likely to have relevant industry experience. If the data scientist does not understand the industry dynamics and drivers, then there is much higher chance of data misinterpretation as they are unable to properly sanity check the models and assumptions they are working with.

Figure 12: Respondent Perceptions on Role Requirements for a Data Scientist

Essential Requirements for a Data Scientist
(n=9)

Requirement	Value
Statistical skills	8
Financial knowledge	7
Moderate programming skills	6
Advanced programming skills	4
Machine learning	4
Knowledge of data architectures	1
Ability to work under pressure	1
Data administration	1
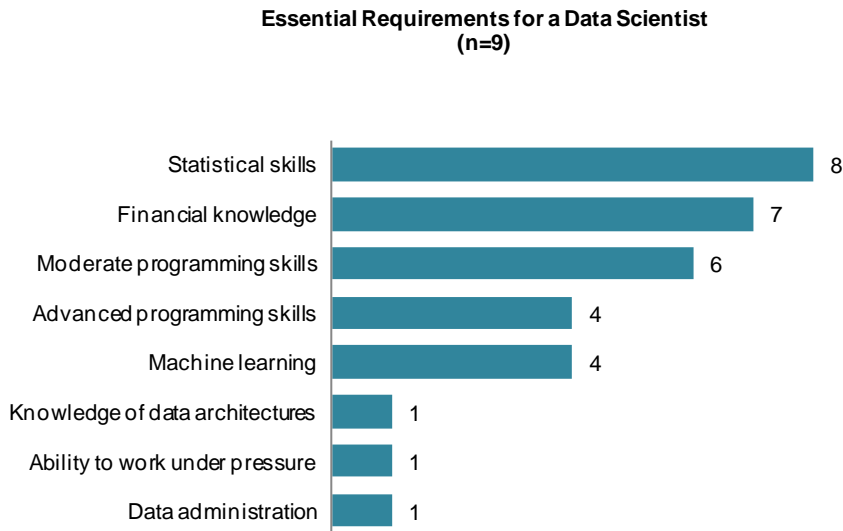
Source: Aite Group survey of 22 capital markets participants, May and June 2014

From 220 jobs pertaining to a search for the role 'data scientist' on the job website Indeed on June 5, 2014, 12 were roles at companies active in the capital markets industry. This is indicative of the lack of penetration of this role in financial markets comparative to other industries and the slow rate of adoption of the function overall. The most common skill requirements stated on each job specification included:

- Computer science fundamentals (algorithms, data structures etc)

- Programming ability(knowledge of SQL, C++)

- Experience within financial services (especially capital markets)

- Strong communication and interpersonal skills

Figure 13 shows the three main requirements for a successful data scientist within the capital markets, highlighting why it is so challenging to find the right individuals. The requirement for relevant domain knowledge makes it harder to move from a field like genetics to finance without significant education and retraining.

Figure 13: The Key Requirements for a Successful Capital Markets Data Scientist
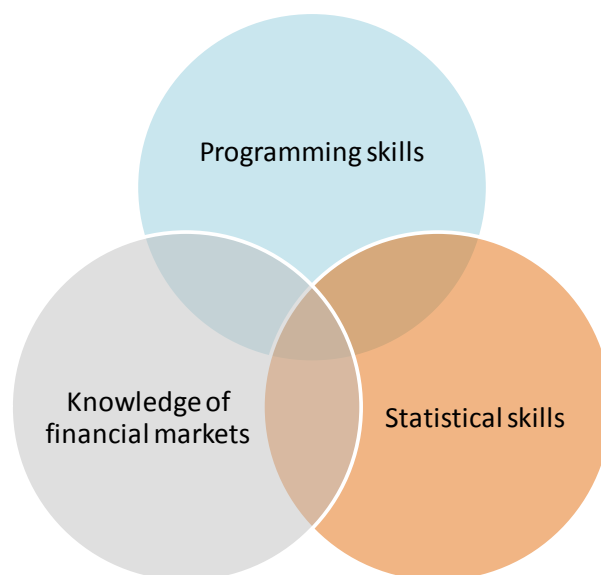
Source: Aite Group survey of 22 capital markets participants, May and June 2014

DATA MANAGEMENT STRATEGIES

The data management function within financial institutions is in a state of continuing evolution due to factors such as regulatory and market infrastructure change, and an increase in requirements related to transparency overall. This has meant that rather than residing purely under the auspices of IT, data management now has more of a business focus and a wider range of stakeholders from within a capital markets firm have input into the function overall. Communication skills and the ability to coordinate multiple teams is therefore a much more important skill set for those engaged in data management overall, regardless of whether it is related to big data or small data.

Figure 14 indicates that there is no one way to structure data management because of the range of individuals under which the function sits in respondent firms. Half of respondents have a senior level individual (C-level in some firms) that is in charge of data management, highlighting the independence and importance of the function overall. Aite Group estimates that there currently around 90 chief data officers in the financial services sector, though there are many more firms with individuals that are chief data officers by another name. Moreover, these roles may not necessarily be new and could just be firms rebranding executives that used to sit under IT or the chief information officer (CIO) role.

Figure 14: Functional Responsibility for Data Management

Responsibility for Data Management Function
(N=22)

CEO 4%
CTO 5%
COO 9%
CIO 14%
Head of data management or chief data officer 50%
Other 18%
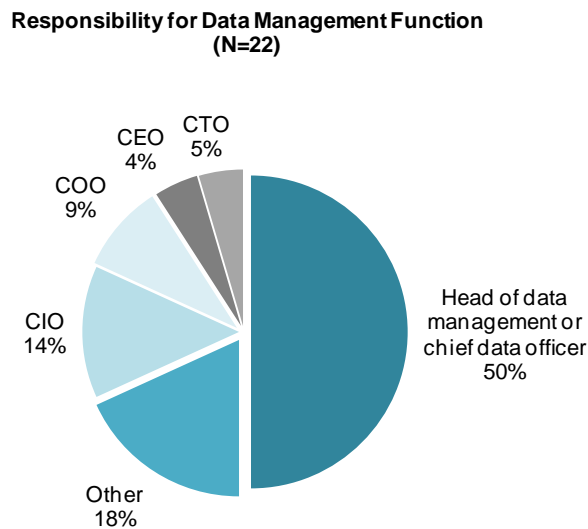
Source: Aite Group survey of 22 capital markets participants, May and June 2014

The chief data officer role is currently in a state of early evolution and the intent is for these individuals to focus on developing firms' overall data strategies, rather than purely focusing on technical challenges. This necessarily involves looking at data from a business value perspective rather than purely as a control function; hence big data should comfortably sit in this purview in future.

Most firms are dealing with a fragmented technology environment and this means that data management teams are faced with a significant challenge to aggregate internal data and respond to internal business, client, and regulatory demands. As part of an industry-wide push to reduce operational risk, these firms are also attempting to improve the manner in which they measure the maturity of their key internal data sets. These assessments include a focus on areas including:

- Data-quality metrics such as accuracy, completeness, timeliness, integrity, consistency, and appropriateness of individual data items

- How satisfied the business end users and clients are with regard to data transparency and availability overall

- The amount of manual reconciliation or data cleansing that must be done to get data into the right state for consumption, including the cost and number of employees dedicated to these tasks

- The ability of the business to respond to data requirements (internal and external) in a timely manner and not be hampered by scalability issues

- The elimination of legacy applications and systems via the consolidation of data environments

TECHNOLOGY AND TECHNICAL ARCHITECTURE

The label "big data" has been misapplied to a whole host of technologies including traditional RDBMS, but there are a number of characteristics of big data technologies that set them apart from more traditional data management technologies. These characteristics tend to relate to speed of processing and the ability to handle large volumes of data such as column-oriented or schema-less databases. To this end, technologies such as Hadoop, MapReduce, and Not Only SQL (NoSQL) are most often cited as examples of big data in action.

- **Hadoop**. Apache's Hadoop is an open-source Java-based software technology borne out of a project to create an open-source search engine to crawl all the web pages on the internet. It has become synonymous with the term big data. Two of its most well-known main components are MapReduce and the Hadoop Distributed File System (HDFS). These are tools that enable data processing in parallel and storage of blocks of subdivided data across servers. They are two of the many featured in the Hadoop ecosystem, which is continually evolving:

  - **Hadoop YARN.** This is a sub-project that was introduced in Hadoop 2.0 that separates the resource management and processing components. It was born of a need to enable a broader array of interaction patterns for data stored in HDFS beyond MapReduce.

  - **Hadoop Spark.** This sub-project is also part of the next generation of Hadoop but is not tied to the two-stage MapReduce paradigm. The focus is on improving performance up to 100 times faster than Hadoop MapReduce, for certain applications. Spark provides primitives for in-memory cluster computing that allows user programs to load data into a cluster's memory and query it repeatedly.

- **NoSQL and schema-less databases**. Structured query language (SQL) is a programming language used to manage data held in a RDBMS and is already established in the industry as an effective tool for data management. On the other hand, NoSQL databases classify a broad range of databases that depart from RDBMS and can be schema-less or feature a soft schema. These databases lend themselves to semi-structured and unstructured data, are able to scale horizontally (whereas RDMS must scale vertically), and/or can store whole documents (without the data therein first being individually parsed). Most, however, give up the atomicity, consistency, isolation, and durability (ACID) taken for granted in transactional RDBMS in exchange for scalability and distributed processing capabilities. It should be noted that these are gradually evolving however, and some of the NoSQL offerings allow SQL-like querying and ACID in addition to the more NoSQL-like access techniques.

  - **Column-oriented databases.** RDBMS are row-oriented and are ideal to support transaction processing, but column-oriented databases (also referred to as time-series) instead focus on supporting data aggregation, compression, and rapid query times. The downside to these columnar databases is that they will generally only allow batch updates and therefore have a much slower update time than traditional models. Kdb+ is the most prevalent commercial example

21

and it is widely used in cash equities trading for high-speed access to market and tick data. Cassandra is an open-source example that is more often used in big data projects. Key-value databases are closely related, with Redis being prominent in financial services.

- **Document-store databases.** RDBMS store data granularly and this data has already been broken down and transformed from a larger data set. Document-store databases are able to load documents in whole for subsequent processing and querying of the data once inside the database. MongoDB (for humongous data) is the most prevalent open-source example, with wide application for consumption of XML data (FpML for OTC derivatives being an obvious use-case).

Although RDBMS and kdb+ have been traditionally implemented within financial services firms, newly hired technologists more familiar and enthusiastic with big data technologies are bringing some of the newer databases to bear. Most of them are open-source and considered to be more economical by technologists. MongoDB and Cassandra are being deployed for many new projects; the former is being used most-often for read-optimized applications (e.g., Twiiter sentiment analysis) and the latter for write-optimized applications. Redis is more popular for use with general middleware. The fact that some of these technologies lack ACID support doesn't seem to be an impediment to their usage as technologists have no problem writing their own transaction layers around them.

Technologies that have been around for some time can also reside happily alongside big data deployments such as parallel processing. The sharing of a computing workload amongst multiple processors, all executing at the same time is a useful tool for big data as it has the ability to cope with high volume and complex data sets by splitting up the work across multiple nodes. Given that most firms are unlikely to have deployed technologies such as Hadoop across the gamut of their internal business functions, compatibility of big data with current technology environments is of significant importance.

The technology environments that data management teams are currently dealing with are dominated by standard RDBMS, though there is a fairly high level of data-related tooling, including those focused on integration, business intelligence, and data quality (Figure 15). There are also some big data technologies in place, with 45% using Hadoop, 32% using NoSQL, and 27% using other big data analytical tools and complex event processing (CEP) technology.

Figure 15: Internal Data Management Technology Environments

Q. What kinds of data management technology does your organization have in place?
(N=22)

Technology	Percentage
Standard relational databases	77%
Business intelligence tools	59%
Standard data integration tools	59%
Data quality monitoring tools	55%
CEP technology	50%
Time series database	45%
Open-source databases	45%
Hadoop	45%
NoSQL databases	32%
Big data analytical appliance	27%
HDFS	9%
Proprietary database software	5%
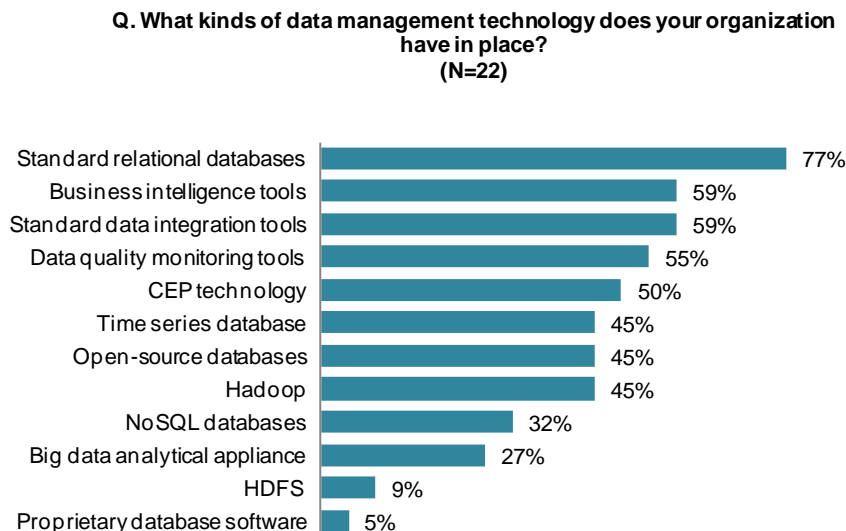
Source: Aite Group survey of 22 capital markets participants, May and June 2014

INTERPLAY WITH CLOUD

In contexts outside of the capital markets industry, big data technology is often closely interlinked with cloud technology. For example, Cloudera was founded in October 2008 to deliver the first enterprise-class implementation of Apache Hadoop. An obvious benefit of using cloud computing as a solution to support a big data strategy is its cost effectiveness. Clouds essentially eliminate the upfront cost of IT investment and the ongoing maintenance costs of on-premise hardware and software; a key dynamic when you consider the resource-intensive nature of most big data implementations. By using cloud, businesses are able to scale up or down on a 'pay as you go' basis, rather than being reliant on internal IT resources.

The vast majority of the capital markets is, however, rather cautious about the use of public cloud technology in commercially sensitive areas. Security remains a concern for most firms and as big data is used to deliver insights for revenue-generating functions, senior management may decide against handing over sensitive information to cloud providers. Private clouds tend to be the norm for top-tier sell-side firms, but these involve higher costs than public cloud deployments and therefore may not bring down the costs of big data support in any significant manner.

BIG DATA IN CONTEXT

The parameters around big data strategies across industries are relatively fluid but when considered in the context of a particular industry segment, it is easier to determine whether data management efforts can actually be categorized as big data projects. No doubt, big data has made the biggest visible impact in the business-to-consumer retail space, where internet giants such as Amazon and eBay have used these strategies and technologies to profile consumer behavior and better target end-consumers. To this end, Amazon deals with millions of back-end operations every day and queries from more than half a million third-party sellers, eBay operates two data warehouses at 7.5 petabytes and 40 petabytes, as well as a 40-petabyte Hadoop cluster for search, consumer recommendations, and merchandising.

When discussing the concept of high volume of market data in a capital markets context, however, it tends to be in the realm of terabytes rather than petabytes or exabytes, which would be more commonly seen in applications in the retail environment. The addition of unstructured data to the mix, however, can increase volumes but not by an exponential magnitude because of the relatively limited number of data sources that the market uses for tasks such as sentiment analysis.

Within the capital markets sector overall it is relatively early days for the adoption of big data strategies, which is reflected by the fact that only half of respondent firms have such a strategy in place and the overall respondent group is a particularly educated subset of the capital markets community with regard to big data. A handful of top-tier banks and hedge funds have been the frontrunners in implementing big data strategies and technologies and these deployments are in very specific areas.

These strategies are also being applied to a whole range of functions, everything from front office trading to back office reference data management. The most popular use cases for big data within respondent firms are analytics for trading and quantitative research, both of which sit together in the front office and are tied to revenue generating opportunities (Figure 16). After five years of maintenance and stagnant IT spending, firms are finally realizing that investments must be made in innovation and revenue growth to capture the next generation of market opportunities. An increasing number of firms are focused on data-centric initiatives, looking to discover unique data-driven initiatives that can help them find alpha. As well as looking at gaining insight from newer data sources such as Twitter or other news sites, firms have primarily been mining their own internal data sets, particularly in the area of regulatory compliance and market surveillance.

Figure 16: Current Applications of Big Data Within Respondent Firms

Q. In which areas is your firm using or planning to use big data technology?
(n=13)

Application	Value
Analytics for trading	8
Quantitative research	7
Revenue optimization	6
Client management	6
Risk management or modeling	6
Cost reduction	5
Market surveillance or fraud	5
Regulatory reporting support	5
Marketing	5
Strategy development and..	4
Positional or transactional..	4
Reference or market data	3
Log data	1
Product design and testing	1

Source: Aite Group survey of 22 capital markets participants, May and June 2014

Because these strategies are being applied in so many different areas, it is often challenging to bundle them together under the banner of big data internally. On this note, a clear indicator about the confusion surrounding big data in a capital markets context and the immaturity of its application overall in this sector is the challenge in finding a centralized team that is responsible for big data applications or strategies. There tends to be rather disparate application of big data across firms, with projects signed off at the business unit level rather than under the remit of a dedicated big data team.

The popularity of big data within the realm of big data is unsurprising, given the growing importance of unstructured data to this community. Aite Group estimates that annual spending on unstructured data for electronic trading will reach US$228 million by the end of 2014, up from US$202 million in 2013 (Figure 17).

Figure 17: Estimated Global Spending on Unstructured Data for Trading

Estimated Global IT Spending on Unstructured Data for Electronic Trading, 2009 to e2014 (US$ million)
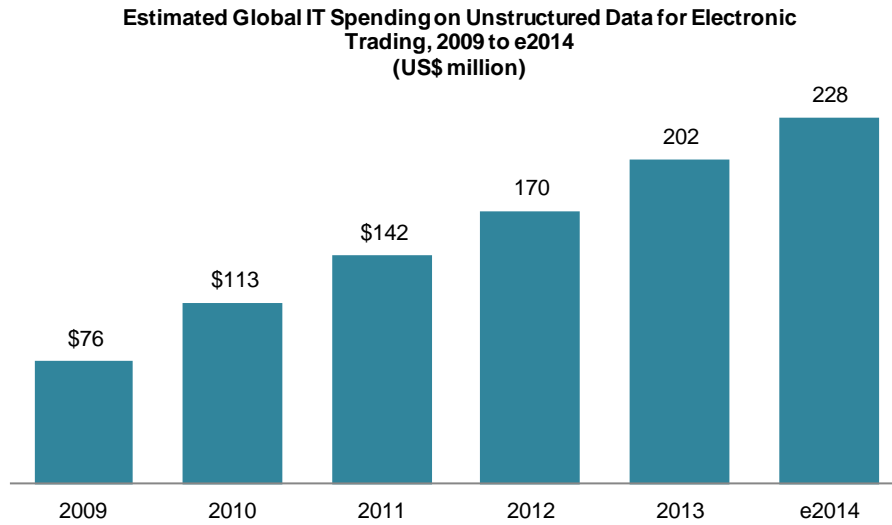
Source: Aite Group survey of 22 capital markets participants, May and June 2014

The firms using big data strategies are generally not new to the area; the majority of respondents have been working on these projects for more than 12 months and over half have been doing this for over three years (Figure 18). After all, the term big data has been around for quite some time, even if only a handful of capital markets firms have opted to develop such a strategy.

Figure 18: Length of Time Using Big Data

Q. When did your organization begin to consider working with big data? (n=13)
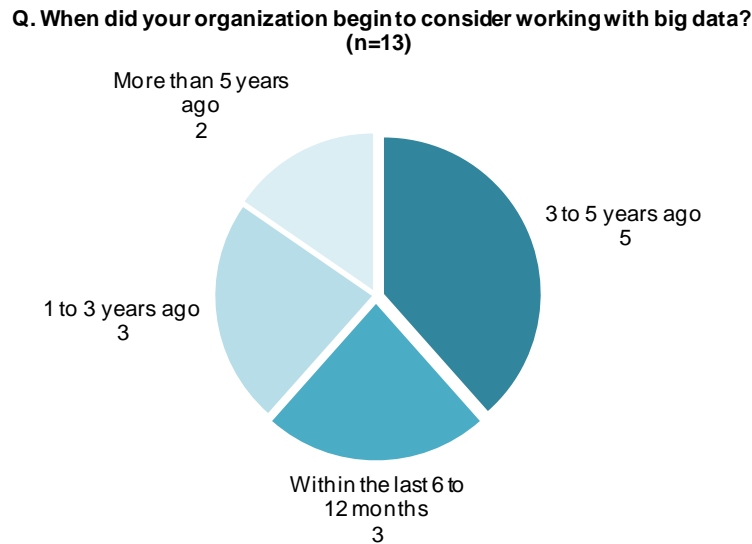
Source: Aite Group survey of 22 capital markets participants, May and June 2014

There is also a wide range of data sets that respondents believe could be suited to a big data approach; the most popular of which is market data (Figure 19). This type of data it being prioritized because it can be enhanced with sentiment analytics for trading purposes, but also can be better managed using a big data approach overall via rationalization across a business line and cross-referencing with other key data items for risk or regulatory reporting purposes (the next two data sets of importance). The size of the market data set is therefore expanding due to the addition of other linked data sets, including text-based sentiment data and metadata tags.

Figure 19: Data Considered to be Suited to a Big Data Approach

Q. Which data sets do you believe are best suited to a big data approach
at your organization?
(N=22)

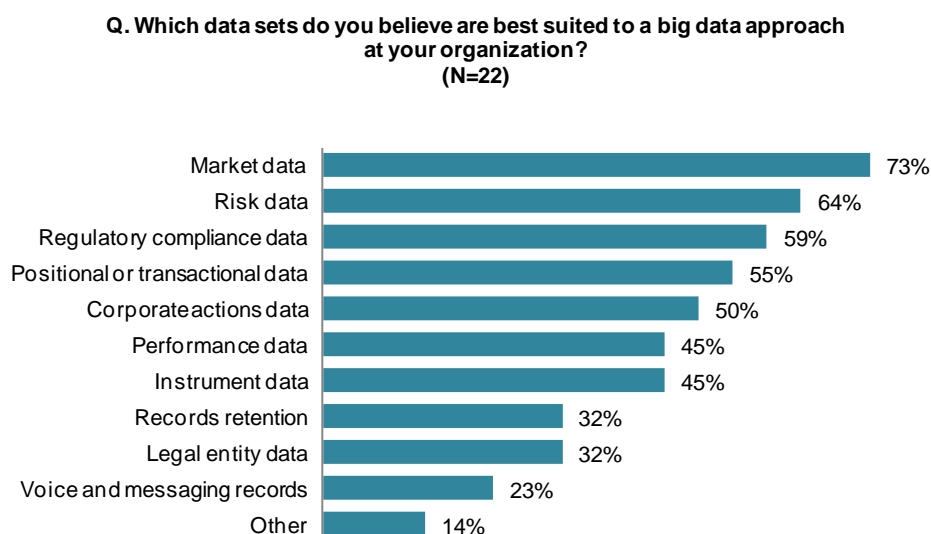| Data set	Percentage
Market data	73%
Risk data	64%
Regulatory compliance data	59%
Positional or transactional data	55%
Corporate actions data	50%
Performance data	45%
Instrument data	45%
Records retention	32%
Legal entity data	32%
Voice and messaging records	23%
Other	14%

Source: Aite Group survey of 22 capital markets participants, May and June 2014

CAPITAL MARKETS USE CASES

Big data projects may be frequently implemented in a sporadic manner across capital markets firms, but they can be categorized into three main types of implementation—those focused on revenue generation, those aimed at meeting compliance or risk requirements, and those focused on cost reduction and operational efficiency.

REVENUE GENERATION: SENTIMENT ANALYSIS

A good example of a revenue generating focus for big data is in the area of sentiment analysis. To this end, a big data strategy can be used to gather and process information surrounding specific markets to create a clear understanding of market sentiment in order to drive front office trading strategies, as well as to determine the valuation of individual securities. Using this information, traders are able to determine whether various market participants and commentators, including those on social networks such as Twitter or blogs, are bullish or bearish and formulate investment strategies accordingly. On a microeconomic level, sentiment and news surrounding an individual company can be incorporated into valuation methodologies to produce a fundamental price for a security. By comparing this to the market value, investors can

more effectively gauge whether a security is undervalued or overvalued, thereby highlighting potential opportunities for arbitrage.

Behavioral analysis is an increasingly important input for modeling and predicting the movement of financial assets. Market sentiment is most commonly measured using a range of statistical tools focusing on stock market highs and lows to produce a relative comparison between the number of advancing versus declining stocks. More recently however, firms have focused on increasing their capabilities to perform news analytics, thus entailing the use of big data strategies.

HFT as a strategy is less profitable than it was a few years ago, as traders have pushed speed of execution to its limits, leaving hedge funds and institutional asset managers to look at smart trading to gain a competitive advantage. Smart trading makes effective use of data management strategies to accumulate and integrate data of various types. Information flowing from social media networks, news articles, weather reports and other unstructured data formats can potentially be a source of revenue, thereby justifying the use of sentiment analysis at capital markets firms. The end goal is to use big data to view the market state holistically and extract usable trading signals, an aspect which HFT lacks.

Hedge funds are now using systematic trading strategies and have developed 'scanner algorithms' which trawl the universe of Twitter and news for trigger words to generate trading signals. This could potentially be dangerous for the market as the 'hash crash' of April 2013 proved ("hash" referring to a Twitter hashtag), where programs examining Twitter picked up on a tweet from a hacked account stating explosions at the White House had injured the President, which automatically prompted a selloff in the stock market, knocking 145 points of the Dow Jones Industrial Average. A hoax tweet in the future could activate a similar reaction. Regulators now have to focus on reckless behavior in addition to exploitation of the market. To tackle this particular issue, the Commodity Futures Trading Commission (CFTC) has contemplated introducing requirements for algorithms to be certified, but has since backed away from this due to issues related to intellectual property issues.

In 2013, State Street Global Advisors announced the launch of a big data initiative, dubbed State Street Global Exchange, which is focused on providing a centralized analytical data platform for its quantitative equity teams for portfolio modeling, investment analytics, and data management. To this end, the platform is aimed at integrating and enriching data from a range of sources, including both structured and unstructured data sets, in order to enable researchers and portfolio managers to make better strategic decisions.

REGULATORY COMPLIANCE

There are many areas where big data can play a role in helping financial institutions better meet their compliance requirements—from dealing with the unstructured data contained within emails via text mining for market surveillance purposes, to enabling the cross-referencing of key sets of internal data related to derivatives instruments in order to facilitate trade reconstruction and reporting. In an environment of heightened scrutiny of data quality for regulatory reporting (even if some regulators have not got this quite right yet) and in a market where ad-hoc reporting is on the increase, speed and accuracy have lent themselves toward an argument for big data.

In the realm of regulatory reporting, linking data sets across a firm can be particularly challenging, especially if you are a top-tier sell-side firm with more than 100 data warehouses storing your key reference data sets in a siloed manner. Trade reconstruction reporting under Dodd-Frank has recently proved a good use case for big data because of the requirement to respond to regulatory investigations within a 72-hour period and to deal with data including voice-recorded and text-based information, as well as tagging this data with items such as legal entity identifiers.

Fraud detection and suspicious activity monitoring have also been key areas of deployment for instances of big data within the universe of retail banking and the capital markets are following suit. Big data analytics enable businesses to overlay existing information such as client transactions with data gathered from unstructured and semi-structured data sources to deliver greater insights into fraud. Moreover, it is essential for these businesses to define a standard for 'normal' so they can easily flag anomalous behavior but the whole dataset is necessary to be able to do this. For example, some anomalous activity could be identified by studying a few weeks of transactions but with the majority of cases the insight or pattern will come from months of data so volume and being able to run analytics on large datasets is also important.

COST RATIONALIZATION

Big data technology can also be used to bring down the cost of data storage over time and, although this is a far less glamorous use of such a strategy, it has the potential to enable long-term system and data feed rationalization. Capital markets firms often struggle with data aggregation for ad-hoc reporting to feed both internal and external reporting functions. The process of extraction, transformation, and aggregation of data can therefore detract from core competencies by taking valuable manpower and resources from business functions. The appeal of big data in this context is that firms are able to tier storage—older or less well-used data sets can be stored in cheaper data containers—and they can establish a long-term data retention facility that will not be challenged by increasing volumes of input over time. This strategy allows for the deletion of records from operational systems to free up computation resources and storage, as well as restricting operational access to these records.

Historical data needs to be maintained for regulatory reporting purposes—hence investments are being made in records and retention databases to act as permanent archives of relevant business records—but it is not the sole application for this data. As firms' risk management and performance measurement techniques mature over time, they will be increasingly reliant on scrutinizing historical data and combining it with real-time tick data to compare exposure information and apply portfolio weights, model liquidity, and performance attribution over a determined period.

It is very early days for these use cases within the capital markets segment therefore it is hard to estimate actual figures for cost savings. Respondents indicate that these savings are expected to be realized over a longer time horizon as a result of resource reallocation and increased efficiencies.

PITFALLS AND CHALLENGES

One of the biggest reasons for the failure of big data projects in the past has been the lack of compatibility between capital markets business requirements and the capabilities of big data technology. For example, Hadoop implementations have sometimes been unsuccessful because the technology is reliant on offline batch processing via MapReduce, which is incompatible with real-time analytics. Furthermore, Hadoop closely couples resource management with data processing; hence it has not previously been possible to prioritize tasks when running multiple applications simultaneously. Hadoop 2.0 is, however, aimed at tackling these resource allocation challenges.

Firms also need to recognize that their data storage strategy needs to change via the addition of a focus on tiered storage, placing the data sets of most importance on faster devices and allowing other sets to be less readily accessible but more cheaply stored. The prioritization of these data items is a significant challenge when considered against the background of a functionally siloed capital markets firm, with different priorities and budget sensitivities in each silo. Another component of this is a more proactive approach to data retention, where firms are able to retire and delete data after it has reached the end of its retention timeframe.

The strategic intent of the big data project overall also needs to be carefully considered. If an aim (or aims) relevant to the business is not identified, then big data strategies will not deliver return on investment. To this end, analysis could result in unforeseen biases in outputs via data gaps or a lack of relevant contextual information.

Respondents cite inadequate technical knowledge as the most commonly encountered challenge during a big data project (Figure 20), highlighting the need for appropriately skilled staff to implement such a strategy. The biggest issue in this regard is finding employees or consultants that understand the business *and* the technology—a pool of individuals that is particularly hard to locate. Some firms have opted to hire a team of individuals with the combined skills of a data scientist due to the lack of a single available individual with all of the required capabilities.

Data privacy is also a key concern for big data projects—a challenge that is shared with cloud technology implementations. There must be a sufficient level of protection for commercially sensitive data and information that should be kept private by organizations due to legal and regulatory requirements, which can vary from jurisdiction to jurisdiction. Some countries are more stringent than others about the treatment and storage of data such as Korea and Switzerland; hence these factors must be carefully considered before any implementation can take place.

30

Figure 20: Common Challenges Faced During Big Data Rollouts

Challenges Faced During Big Data Projects
(n=13)
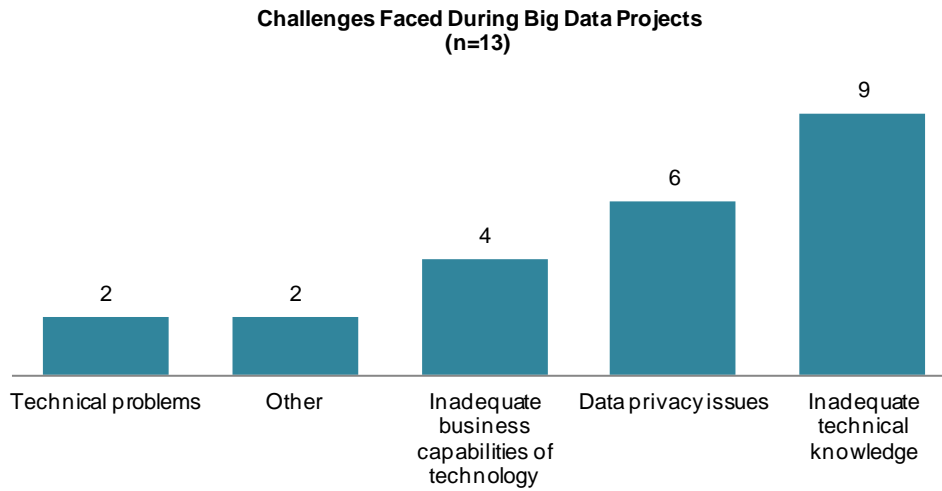
Source: Aite Group survey of 22 capital markets participants, May and June 2014

MAKING THE CASE FOR INVESTMENT

The building of a business case for big data investment can be challenging because of the immaturity of the strategy in the capital markets overall. Disparate deployments across financial institutions mean that there is not an easily compiled list of criteria against which a business case can be built; they tend to vary from deployment to deployment. Moreover, understanding what your peers are doing with big data is not an easy task when firms are reluctant to share details that may entail ceding competitive edge.

A first step toward making a case for investment is understanding the drivers that other firms considered when assessing big data (Figure 22). The most frequently cited driver is coping with increasing data volume, which is unsurprising, given that we are talking about "big" data. There are no reliable statistics to prove the growth of capital-markets-specific unstructured data over time but looking at the growth of Twitter data, which is frequently used for sentiment analysis, is a relatively good indicator of this growth (Figure 21).

Figure 21: Growth of Twitter Data and Users Over Time

Average Number of Tweets Per Year (in Billions) and Total Number of Active Twitter Users Worldwide (in Millions)
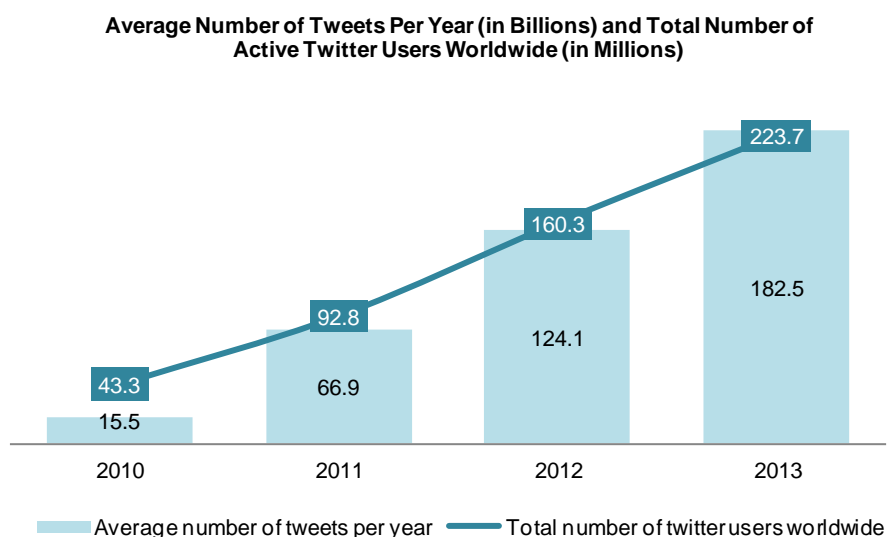
Source: Twitter

Other important considerations include cost reduction drivers—though these can be tough to sell internally because the length of time to deliver return on investment can vary from project to project. Cost savings can be realized in areas such as:

- Direct cost savings via the tiering of data storage

- The consolidation of data sources and the elimination of legacy applications over time

- Improved efficiency in market or client responsiveness—faster and more accurate analysis means less client attrition over time

- Avoidance of fines and reputational damage due to a lack of regulatory compliance in areas such as trade reconstruction and reporting

- Faster responsiveness in identifying fraudulent behavior and market abuse

Risk reduction metrics and revenue generating opportunities can also be hard to measure in a tangible manner, depending on the area of deployment. As with many similar infrastructure projects, it is easier to measure the costs of big data—storage, computation power, transmission capacity, and direct energy costs, as well as requisite big data experts (in-house and consultants)—than the benefits.

Figure 22: Drivers for Usage of Big Data

Q. What were the drivers behind your organization's decision to look at big data technology?
(n=13)

Driver	Value
Increasing data volume	8
Cost reduction	7
Compliance and surveillance	7
Software scalability	6
Risk management support	6
External reporting requirements	6
Migration from legacy systems	5
Strategic data quality	5
Client data challenges	5
Unstructured data support	5
Open-source software.	4
Coping with complex products	4
E-discovery	3
Internal reporting requirements	3
Other	2

Source: Aite Group survey of 22 capital markets participants, May and June 2014

The aspects of most importance to respondent firms in terms of big data deployments signal that actionable information and insight are equally pegged with scalability for future data volume increases (Figure 23). Frontrunners in big data deployments are therefore firmly fixed on future profitability and growth via the identification of potential opportunities and threats in the markets.

In terms of insight, this can be related to the identification of trading opportunities in the market, the running of market exposure analytics for portfolio managers and traders examining data on a real-time basis, running stressed scenarios via interactive scenario modeling, and responding to these results. It could also be related to running faster pricing calculations, calculating consolidated global positions, and producing on-demand value-at-risk calculations as part of a risk modeling improvement program. Infrastructure-related factors are also fairly important, with operational risk, data storage efficiency, and monitoring tools cited as important by more than half of respondents.

Figure 23: Aspects of Big Data of Most Importance

Q. Which aspects of a big data management solution are most important to your organization?
(n=13)

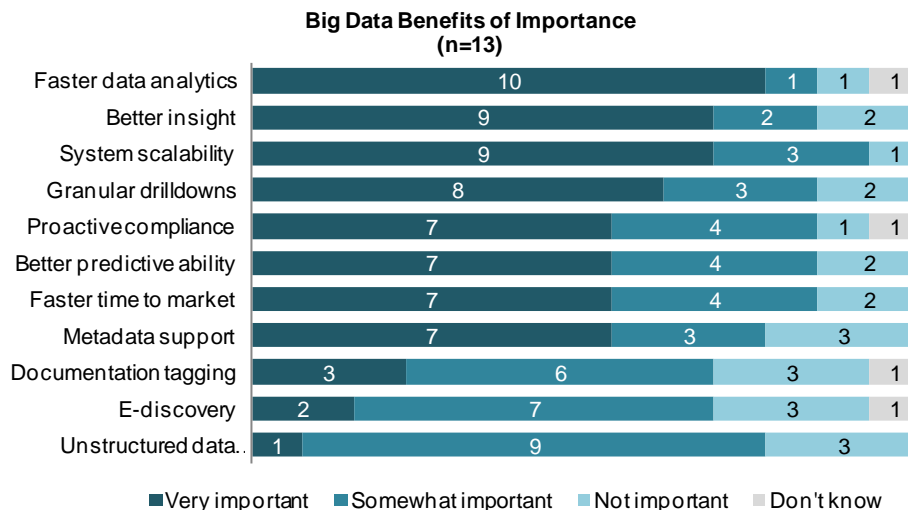| Aspect	Value
Analytical insight	10
Scalability for future	10
Data storage efficiency	8
Operational risk control	8
Better monitoring tools	8
Facilitating data integration	6
Visual analytical support	6
Support for key data sets	6
Tiered storage	4
Support for bitemporality	3
Downstream support	3
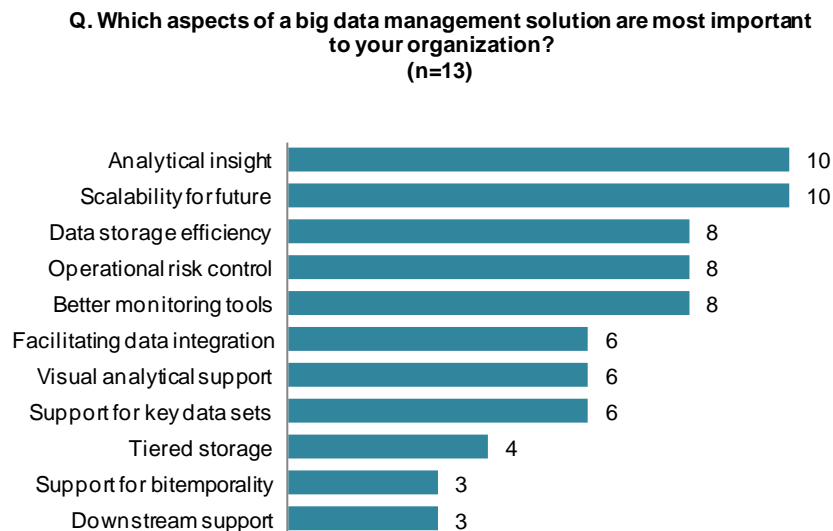
Source: Aite Group survey of 22 capital markets participants, May and June 2014

These dynamics are again present in respondents' views on the benefits of big data, with speed, insight, scalability, and the ability to drill down into data at a granular level ranking highest on the list in terms of firms identifying these items as "very important" (Figure 24). The importance of "better insight" relates to the ability of big data to support complex analytics on both internal and external data in a whole range of contexts across the business.

Slightly lower down the list is regulatory compliance, but this is an area of increasing importance for firms and is likely to be one of the key business cases for big data investments in future. The prohibitively high cost of maintaining multiple systems for reporting and compliance over time may cause firms to turn to big data strategies in a bid to join these efforts together. Currently, the primary use cases in this regard tend to be around KYC and fraud detection; a dynamic shared with the retail banking community. Though unstructured data support is only identified by one firm as "very important" and sits at the bottom of the list, it is "somewhat important" to the majority of firms.

Figure 24: Benefits of Big Data

Big Data Benefits of Importance
(n=13)

	Very important	Somewhat important	Not important	Don't know
Faster data analytics	10	1	1	1
Better insight	9	2	2	
System scalability	9	3	1	
Granular drilldowns	8	3	2	
Proactive compliance	7	4	1	1
Better predictive ability	7	4	2	
Faster time to market	7	4	2	
Metadata support	7	3	3	
Documentation tagging	3	6	3	1
E-discovery	2	7	3	1
Unstructured data..	1	9	3	

■Very important ■Somewhat important ■Not important ■Don't know

Source: Aite Group survey of 22 capital markets participants, May and June 2014

All this must be considered against the fact that not everyone is yet convinced of the return on investment when it comes to big data strategies. Those respondents that are not currently considering big data primarily cite a lack of business buy-in to the concept and the high cost of such a deployment as the detracting factors (Figure 25). At a time when budgets are fairly restrictive overall for investment in new technology, projects without a thoroughly convincing business case will continue to flounder.

Figure 25: Reasons Why Firms are Not Considering Big Data

Q. Why is your organization not considering a big data technology
approach?
(n=9)

Lack of business support/executive commitment	5
Perceived to be too expensive	3
No need to use it	2
Don't understand the benefits	2
Budget allocated elsewhere	1
Don't know enough about it	1

Source: Aite Group survey of 22 capital markets participants, May and June 2014

In order for a firm to successfully build a big data strategy, a number of important steps must be taken:

- A preliminary assessment of whether a big data strategy is actually required—could this function or task be solved with existing resources? If a big data strategy is appropriate, what kinds of benefits are expected? Those with a clear target to aim for are able to deliver value to the business in a faster manner.

- Identification of the key data assets and the number of current systems involved—if data needs to be extracted and transformed, what kinds of challenges could lie ahead? What kind of IT architecture is currently in place?

- Identification of the right business functions to lead the project and the definition of roles and responsibilities in the program of work. Will the team be extracting data directly from end-user systems or can business lines deliver files of data (dependent on the use case in question)?

- The building of a team with the right expertise in big data strategy and technology—how many data scientists does it take to transform a data architecture successfully?

- Establishing the hardware horsepower required in order to complete the project, including the amount of memory required to run the technology—will the data expand as it is extracted? How much computation power will be required for analytics?

THE FUTURE OF BIG DATA

Big data in capital markets is a long way off being considered mature; it has yet to reach a tipping point in usage within top-tier sell-side firms, let alone across the industry as a whole. Though there are viable reasons for investment, these tend to be focused on specific use cases rather than an over-arching big data strategy. Those that are currently aware of the benefits of big data and have successfully realized some of these internally (in spite of the hurdles), are certainly keen to extend their usage of these strategies and technologies. These implementations will, however, likely remain piecemeal for the near future.

Figure 26 shows that 68% of respondents expect to start actively evaluating big data technology in other areas, or for the first time in the case of those that have yet to invest. A reticent 14% feel they will not evaluate big data technology in the next two years and 18% don't know if their firm has any definite plans to conduct such an exercise. Given this respondent group is an educated subset of the market at large, most firms in the capital markets sector are likely to be in the "not likely" and "don't know" categories for some time to come, though this will change as more use cases for big data become prevalent in the market.

Figure 26: Planned Future Investment in Big Data

Q. Over the next 24 months, what are the chances your organization will
start actively evaluating big data technology?
(n=22)

Not likely 14%
Very likely 41%
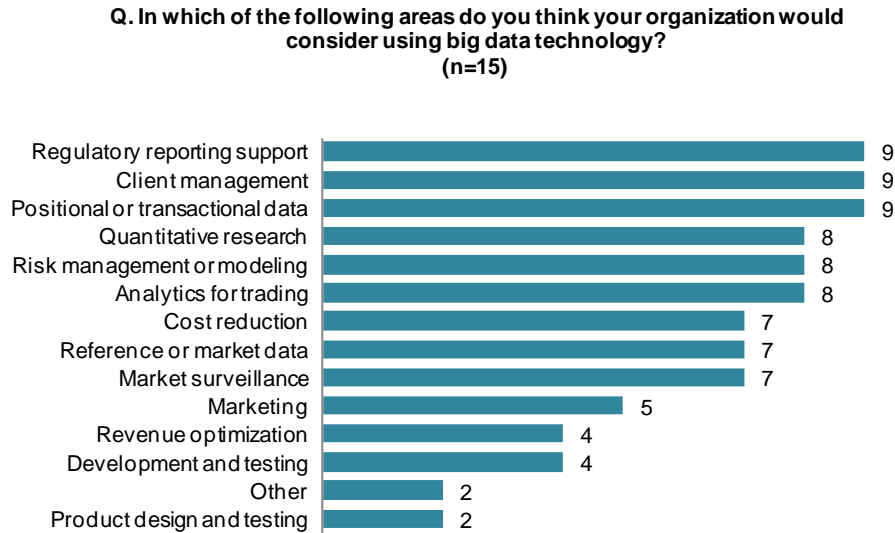Don't know 18%
Somewhat likely 27%
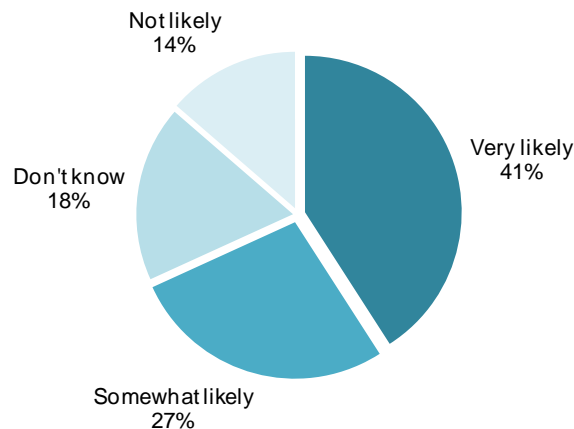
Source: Aite Group survey of 22 capital markets participants, May and June 2014

The areas of focus for future investment are different from the areas of current investment, with more of a slant toward regulatory compliance and client data than trade support and sentiment analysis (Figure 26). Current investments in big data may largely be focused on revenue generating opportunities in the front office, but the future is likely to see much more focus on client retention and compliance function support. As institutional clients begin to demand more transparency and greater responsiveness from their service providers, firms will be forced to evaluate newer strategies such as big data, which have been successfully used in this manner within the world of retail banking. In terms of compliance, firms are increasingly being required

to store a high volume of historical data; hence data storage costs over time will play a bigger role in business case building for big data projects.

Figure 27: Areas in Which Firms are Planning to Extend Big Data

Q. In which of the following areas do you think your organization would consider using big data technology?
(n=15)

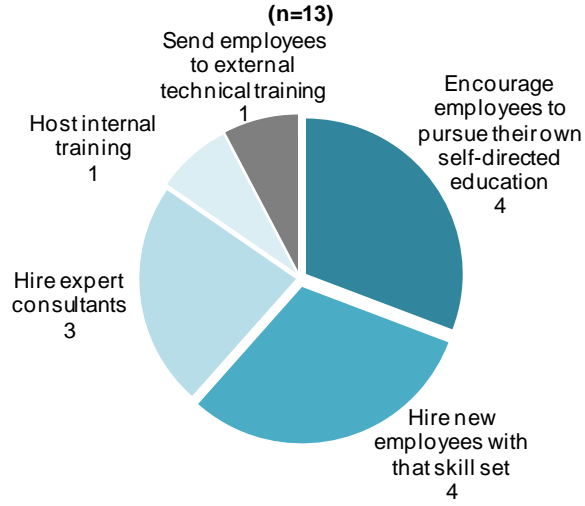| Area	Value
Regulatory reporting support	9
Client management	9
Positional or transactional data	9
Quantitative research	8
Risk management or modeling	8
Analytics for trading	8
Cost reduction	7
Reference or market data	7
Market surveillance	7
Marketing	5
Revenue optimization	4
Development and testing	4
Other	2
Product design and testing	2

Source: Aite Group survey of 22 capital markets participants, May and June 2014

A key part of the future growth of big data as a strategy within the capital markets overall will be the continued education of internal staff about its benefits and uses. Respondent firms currently using big data strategies tend to rely on self-directed education and hiring expert knowledge into the firm in order to grow their internal knowledge base (Figure 28). A danger with the approach of hiring in expertise is that firms are opened up to key person risk if those skills and knowledge are not passed on to a wider demographic of in-house staff. Getting this balance right will be key to future successful deployments and maintenance of big data strategies and technologies over time. As with other data management programs, big data strategies do not have an end point—they must be constantly refined and evolved as market and competitive dynamics change over time.

Figure 28: Education Programs for Big Data

Q. How does your organization grow its big data knowledge among its employee base?
(n=13)

Send employees to external technical training
1

Host internal training
1

Hire expert consultants
3

Hire expert consultants
3

Encourage employees to pursue their own self-directed education
4

Hire new employees with that skill set
4

Source: Aite Group survey of 22 capital markets participants, May and June 2014

CONCLUSION

- **It is still early days for big data in the capital markets.** Some firms have acted as early adopters but even they often do not have comprehensive big data strategies in place. A handful of top-tier banks and hedge funds have been the frontrunners in implementing big data strategies and technologies and these deployments are in very specific areas such as trade analytics, market surveillance, and risk modeling.

- **Education and viable use cases will drive adoption**. As more capital-market-specific use cases for big data become prevalent in the market, firms will become more comfortable with these strategies. A tipping point in adoption will require firms to get a better handle on what comprises a big data strategy and how such an approach can deliver return on investment over time.

- **Big data usage will reflect the continued rising importance of data management as a function.** The data management function within financial institutions is in a state of continuing evolution due to factors such as regulatory and market infrastructure change, and an increase in requirements related to transparency overall. Firms are now looking at data from a business value perspective rather than purely as a control function, which plays directly into the strengths of big data.

- **Firms are looking for insights, speed of response, and future scalability from big data.** The aspects of most importance to respondent firms in terms of big data deployments signal that actionable information and insight are equally pegged with scalability for future data volume increases. Delivering tangible returns on investment in these key areas will enable further investment in big data across other areas of the business.

- **Many current deployments may be in the front office but the future will see more middle and back office implementations of big data.** Current investments in big data are largely focused on revenue generating opportunities in the front office, but the future is likely to see much more focus on client retention and compliance function support. High volume historical data storage requirements will compel firms to examine the relative cost of big data to current database technologies.

ABOUT THOMSON REUTERS

Thomson Reuters is the world's leading source of intelligent information for businesses and professionals. We combine industry expertise with innovative technology to deliver critical information to leading decision makers in the financial and risk, legal, tax and accounting, intellectual property and science and media markets, powered by the world's most trusted news organization. Thomson Reuters shares are listed on the Toronto and New York Stock Exchanges.

CONTACT

For more information on Thomson Reuters Big Data solutions, please contact:

Stephen Malinak

starmine.quantconsulting@thomsonreuters.com

For all media inquiries, please contact:

Grace Keith

+1.(646).223.5532

Grace.Keith@thomsonreuters.com

For more information, go to:

http://thomsonreuters.com

41

ABOUT AITE GROUP

Aite Group is an independent research and advisory firm focused on business, technology, and regulatory issues and their impact on the financial services industry. With expertise in banking, payments, securities & investments, and insurance, Aite Group's analysts deliver comprehensive, actionable advice to key market participants in financial services. Headquartered in Boston with a presence in Chicago, New York, San Francisco, London, and Milan, Aite Group works with its clients as a partner, advisor, and catalyst, challenging their basic assumptions and ensuring they remain at the forefront of industry trends.

AUTHOR INFORMATION

Virginie O'Shea
+44.(0)207.092.8129
voshea@aitegroup.com

Rishi Shah
+44.(0)207.092.8138
rshah@aitegroup.com

CONTACT

For more information on research and consulting services, please contact:

Aite Group Sales
+1.617.338.6050
sales@aitegroup.com

For all press and conference inquiries, please contact:

Aite Group PR
+44.(0)207.092.8137
pr@aitegroup.com

For all other inquiries, please contact:

info@aitegroup.com