insideBIGDATA

InsideBIGDATA Guide to

Big Data for Finance

by Daniel D. Gutierrez
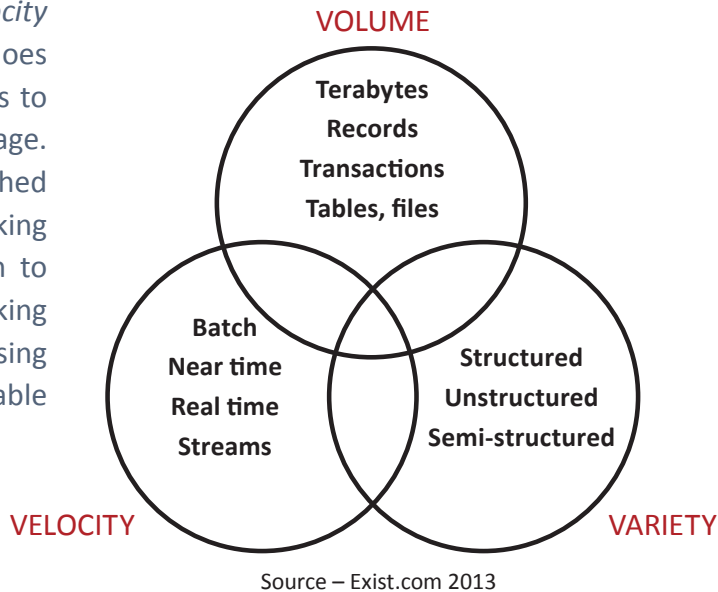
BROUGHT TO YOU BY

DELL | (intel)

Introduction to Big Data for Finance

According to the 2014 IDG Enterprise Big Data research report, companies are intensifying their efforts to derive value through big data initiatives with nearly half (49%) of respondents already implementing big data projects or in the process of doing so in the future. Further, organizations are seeing exponential growth in the amount of data managed with an expected increase of 76% within the next 12-18 months. With growth there are opportunities as well as challenges. Among those facing the big data challenge are finance executives, as this extraordinary growth presents a unique opportunity to leverage data assets like never before.

As the 3 V's of big data: *volume, velocity* and *variety* continue to grow, so too does the opportunity for finance sector firms to capitalize on this data for strategic advantage. Finance professionals are accomplished in collecting, analyzing and benchmarking data, so they are in a unique position to provide a new and critical service—making big data more manageable while condensing vast amounts of information into actionable business insights.

Contents

VOLUME

Terabytes
Records
Transactions
Tables, files

Batch
Near time
Real time
Streams

Structured
Unstructured
Semi-structured

VELOCITY   VARIETY

Source – Exist.com 2013

The proverbial 3 V's of big data: volume, velocity and variety and what constitutes each classification for the finance sector.

Although many perceive a financial industry mindset that's wary of change — it is the cautious adoption of big data technologies such as cloud solutions and Hadoop using the open source distribution model that will be transformative in many fundamental ways. Further, financial services firms that use business analytics efficiently and effectually will thrive in the current uncertain and fluctuating markets, while those that do not are likely to founder.

Financial services firms are discovering many innovative and strategic directions by consolidating data traditionally managed in departmental silos in order to more holistically analyze risk exposure, comply with regulatory mandates, and perform enterprise-level analytics. Traditional technologies such as relational database management systems make it challenging, many times impossible, to process growing volumes of data and make them accessible, actionable and elastic to changing requirements in terms of advanced analytics. Big data solutions that support evolving business and regulatory requirements by maintaining a network of growing data stores will become invaluable in their ability to be used for multiple purposes and to answer any query months or years from now.

In a recent research report, SunGard has identified ten trends shaping big data initiatives across all segments of the financial services industry:

1. Larger market data sets containing historical data over longer time periods and increased granularity are required to feed predictive models, forecasts and trading impacts throughout the day.

> Big data solutions that support evolving business and regulatory requirements by maintaining a network of growing data stores will become invaluable in their ability to be used for multiple purposes and to answer any query months or years from now.

2. New regulatory and compliance requirements are placing greater emphasis on governance and risk reporting, driving the need for deeper and more transparent analyses across global organizations.

3. Financial institutions are ramping up their enterprise risk management frameworks, which rely on master data management strategies to help improve enterprise transparency, auditability and executive oversight of risk.

4. Financial services companies are looking to leverage large amounts of consumer data across multiple service delivery channels (branch, web, mobile) to support new predictive analysis models in discovering consumer behavior patterns and increase conversion rates.

5. In post-emergent markets like Brazil, China and India, economic and business growth opportunities are outpacing Europe and America as significant investments are made in local and cloud-based data infrastructures.

6. Advances in big data storage and processing frameworks will help financial services firms unlock the value of data in their operations departments in order to help reduce the cost of doing business and discover new arbitrage opportunities.

7. Population of centralized data warehouse systems will require traditional ETL processes to be re-engineered with big data frameworks to handle growing volumes of information.

Big data is an emergent trend driving investments in enterprise analytics, and correspondingly, analytic excellence is central to much needed innovation in today's financial services marketplace.

8. Predictive credit risk models that tap into large amounts of data consisting of historical payment behavior are being adopted in consumer and commercial collections practices to help prioritize collections activities by determining the propensity for delinquency or payment.

9. Mobile applications and internet-connected devices such as tablets and smartphone are creating greater pressure on the ability of technology infrastructures and networks to consume, index and integrate structured and unstructured data from a variety of sources.

10. Big data initiatives are driving increased demand for algorithms to process data, as well as emphasizing challenges around data security and access control, and minimizing impact on existing systems.

Big data is an emergent trend driving investments in enterprise analytics, and correspondingly, analytic excellence is central to much needed innovation in today's financial services marketplace. Business analytics applied to capital management, regulatory compliance, corporate performance, trade execution, security, fraud management, and other instrumental disciplines is the principal innovation platform to improving strategic decision making. Analytics and the ability to efficiently and effectively exploit the big data technology stack, advanced statistical modeling, and predictive analytics in support of real-time decision making across business channels and operations will distinguish those companies that flourish in uncertain markets from those that misstep.

The goal for this Guide is to provide direction for enterprise thought leaders on ways of leveraging big data technologies in support of analytics proficiencies designed to work more independently and effectively in today's climate of working to increase the value of corporate data assets.

Retail Banking

Thanks in large part to the availability of data and the movement from in-person to online banking, today's banking institutions look very different from those of just a decade ago. As the scale of data is overwhelming traditional systems, banks must adapt to new technologies to unlock the power of their data.

Regulatory demands for banks have become much more stringent and the increased level of historical storage requirements have meant that banks must keep data for longer periods of time.

Banks have long been considered innovators when using data analytics to tackle numerous business challenges such as risk management, fraud detection and price discovery. Yet today, the volume of data is much bigger and more diverse than ever before. At the same time, regulatory demands for banks have become much more stringent and the increased level of historical storage requirements have meant that banks must keep data for longer periods of time. This complexity has necessitated new approaches and technologies leading to big data infrastructure.

There are a number of motivating factors for engaging big data technology in support of retail banking:

- The availability and scale of data is extraordinary and requires a new technology mindset.

- A significant transition from in-person to online banking resulting from the ease and affordability of executing financial transactions.

- A distinct need to collect and analyze this information in order to accurately assess risk and market trends.

- The availability of new data sources such as data from social media, blogs and other news feeds offer significant new opportunities. As with all online markets, banking is competitive and banks are interested in using any opportunity, identified through data, to cross sell and upsell customers.

Coupled with the above motivations, there are a number of potential application areas for big data in retail banking:

- Since banks weigh the risk of opening new accounts versus the opportunity to hold deposits, big data can screen new account applications for risk of default. In addition, big data can identify high-risk borrowers for auto loans.

- Maximizing customer touch points leading to better customer service.

- Churn detection and avoidance.

- Cross-sell/up-sell using recommender systems.

- Monetize anonymous banking data in secondary markets.

Solving the challenging problem of customer churn is one good example of how big data can make a real difference for retail banking.

Banking on Big Data

76 Percentage of banks that say the business driver for embracing big data is to enhance customer engagement, retention and loyalty

71 Percentage of banks that say that to increase their revenue, they need to better understand customers and big data will help them do so

55 Percentage of banks that say that having a real-time view of data provides a significant competitive advantage and believe that batch mode data is ineffective

BASED ON A SURVEY OF 183 BANKERS WORLDWIDE; SOURCE: NGDATA IN CONJUNCTION WITH FINEXTRA AND CLEAR2PAY

Results of a global banking industry survey showing how banks see big data helping with business engagement.

"In 2012, over 50% of customers either changed their bank or were planning to change."
– Global Consumer Banking Survey by Ernst & Young

Research shows that service quality is one of the most prominent reasons why customers switch banks. With credit card transactions, loyalty programs, and many other customer touch points, banks possess far more data about customers than any other industry, but in spite of all the data, customer-centric companies like banks are often unable to deliver effective personalized service. The main reason is the low level of customer intelligence.

The world is increasingly interconnected, instrumented and intelligent and in this new world the *volume, velocity,* and *variety* of data being collected is unprecedented. As the amount of data created about a consumer is growing, the percentage of data that banks can process is going down fast. Based on their engagement with

popular consumer destinations like Amazon or Yelp, customers have expectations about similar experiences from the banking applications. Without deep know-how about their customers, banks may not be able to meet these expectations. The result? Lost revenue opportunities, low coupon redemption rates, lower share of customer's wallet and lost competitive agility. In a nutshell, not being able to gain insights from the goldmine of data means banks are allowing their competitors to identify critical business trends and act on those before they can, ultimately losing business.

In summary, in order to advance the level of customer intelligence banks must:

- Leverage big data to get a 360 degree view of each customer.

- Drive revenues with one-to-one targeting and personalized offers in real-time.

- Reduce business risk by leveraging predictive analytics for detecting fraud.

- Achieve greater customer loyalty with personalized retention offers.

- Employ the power of big data without worrying about complexities and steep learning curves.

There are many quality software tools allowing banking institutions to reap the benefits of big data. For example, the Kitenga Analytics Suite from Dell is an industry leading big data search and analytics platform designed to integrate information of all types into easily deployed visualizations. With this content mining and analytics solution, banks are

> Kitenga also helps to reduce the pain of solving big data problems by providing out-of-the-box capabilities for building search and analysis capabilities that work over big data assets.

able to transform complex and time consuming manipulation of web-scale data resources into a fast and intuitive process. Banks can harvest sentiments from Twitter feeds, blogs, news reports, CRM systems, and other sources, and combine them with demographic and regional data to better understand market traction and opportunities. Kitenga can be characterized with the following high-level feature set:

- **Analytics –** Sophisticated natural language processing (NLP), machine learning, predictive modeling, sentiment analysis, social network analysis, and visualization.
- **Search –** Interactive and intuitive. Search interface allows business analyst to explore and exploit all data resources.
- **Visualization –** Interactive web-based authoring empowers business users to perform analysis, visualize results and make decisions.

Kitenga also helps to reduce the pain of solving big data problems by providing out-of-the-box capabilities for building search and analysis capabilities that work over big data assets. These capabilities enable data analyst personnel to:

1. Author big data solutions by dragging-and-dropping analysis components that operate over Hadoop for solving everyday problems.

2. Execute analytical pipelines that operate over Hadoop to scale up by scaling out

3. Monitor the jobs as they run

Kitenga is Hadoop-enabled for big data scalability and allows for integration of disparate data sources and cost efficient storage of growing data volumes. Kitenga can directly analyze Hadoop results using information visualization tools that bind directly to Hadoop Distributed File System (HDFS) files, as well as index the created data and metadata into a searchable form with embedded visualization capabilities. Kitenga acts as a wrapper of libraries on top of Hadoop, with a drag-and-drop interface to query, search, and analyze data directly against HDFS, without the need to create and script MapReduce jobs.

Kitenga enables financial institutions to integrate structured transaction data with a variety of unstructured private and public information sources to quickly discern patterns associated with fraud and immediately identify new transactions that meet the model and require additional review before being approved.

> With the Dell Boomi master data management (MDM) solution, you can take the complexity out of integration. The software solution offers on-premise and cloud integration without appliances, software or coding for easier transfers between CRM and back-office applications.

In the banking industry, it's easier to provide customers with a more personalized experience when you have the latest data. But since so much of that data lives in so many places throughout a business, it's hard to make the best decisions quickly. With the Dell Boomi master data management (MDM) solution, you can take the complexity out of integration. The software solution offers on-premise and cloud integration without appliances, software or coding for easier transfers between CRM and back-office applications.

Credit Scoring

Historically, loan and credit scoring methodology employed by credit bureaus and used by banks and other financial institutions has been based on a five component composite score including (i) past loan and credit applications, (ii) on time payments, (iii) types of loan and credit used, (iv) length of loan and credit history and (v) credit capacity used. Until the big data revolution, this approach has seen little innovation in making scoring a commodity.

Today, new technology platforms have opened the doors for change in credit scoring and big data scoring services are beginning to be available. Loan and credit decisions are determined in seconds using automated processes based on machine learning algorithms. The breadth of data that can be used for credit scoring has expanded considerably. For each scoring decision, big data applications collect data from a broad range of external data sources ranging from social networks, e-commerce data, economic databases, micro geographical statistics and other sources. In some cases, big data scoring technology can use upward of 10,000 data points in real-time to asses a customer's creditworthiness.

As an extension to traditional scoring services, new technology companies using big data scoring are providing scoring-as-a-service options for online loan and credit decisions. This type of solution is provided to banks, debt collectors, e-commerce sites, leasing and other financial companies. These systems can integrate into the customer's existing systems and/or website.

For a valuable use case example of how big data has transformed the credit scoring arena, see the side bar "Credit Scoring at Novum Bank." To assess credit applications, Novum Bank in the Netherlands recently started using the Dell STATISTICA, the analytical software solution. In the interview, Chief Credit Risk Officer, Joop Bruinzeel talks about micro-credit, the importance of credit scoring and the use of analytical software.

CASE STUDY

Credit Scoring at Novum Bank
by Marcel Wiedenbrugge

Imagine you are active in the provisioning of (micro) credit and a customer wants to borrow temporarily a few hundred euros from you. How do you determine whether it makes financial sense to do business with this customer? For Joop Bruinzeel, Chief Credit Risk Officer (CCRO) at Novum Bank, this question is just another day's issue. As a provider of micro-credit, Novum Bank daily provides relatively small amounts (from €100 to €600) to customers where traditional banks have no interest due to a high risk profile. Properly set up and tuned credit risk management is essential.

For assessing credit applications, Novum Bank recently started using STATISTICA, the analytical software solution from StatSoft (now a part of Dell). In this interview I speak with Joop Bruinzeel about micro-credit, the importance of credit scoring and the use of analytical software.

Why did you choose STATISTICA?

Joop: *"Before I started working for Novum Bank, I immersed myself in modelling. I got back in touch with another company that was specializing in short-term loans and had (successfully) made use of STATISTICA. As the shareholders of Novum Bank required mature risk management, STATISTICA was perceived as a logical choice to go with."*

What do you like about STATISTICA?

Joop: *"The beauty of STATISTICA is that you can build decisioning models which you can test on older portfolios (also called backlog or backtesting). The workbench also offers the possibility to add your own insights to the models. That allows us to refine the models, so that we can achieve better results."*

Read entire case study[1]

[1] http://www.statsoft.com/Company/Press-Room/STATISTICA-News/EntryId/579/Credit-Scoring-at-Novum-Bank-Data-Mining-defines-success-in-high-risk-lending

Back Trading & Testing

Another area of opportunity using big data technology is building back-testing software solutions. Back-testing refers to the process of testing a trading strategy, investment strategy, or predictive model using existing historic data. Back-testing is considered a special type of cross-validation applied to time series data. The goal of back-testing is to estimate the performance of a strategy as if it had been employed during a prior period. This requires simulating past conditions with sufficient detail, making one limitation of back-testing the need for detailed historical data.

Since major markets produce massive amounts of messages per day, US financial markets for example produce around 50 billion data points per day, it is extremely computationally intensive task to process them. While back-testing is a computationally intensive task it is also easy to parallelize. Multiple trading days can be back-tested simultaneously making it an ideal candidate for big data techniques such as MapReduce. In order to complete back testing within a reasonable time in an environment like the mentioned example of US markets, big data architectures like Hadoop are an invaluable tool.

The Dell™ | Cloudera™ Solution powered by Intel is a good option for addressing the needs of back-testing software. Cloudera Apache Hadoop Distribution (CDH) Enterprise consists of tools for institutions that gather insights from vast data volumes and varied data types and find that managing large volumes of unstructured data exceeds the capacity and capabilities of traditional data intelligence systems. Aside from MapReduce, financial industry firms can perform interactive analysis on any data stored in Hadoop HDFS and HBASE with Cloudera Impala.

Intel has a long history with Hadoop — release 2.0 of its own Hadoop distribution in 2012. By aligning the Cloudera and Intel roadmaps in 2014, Intel created the platform of choice for big data analytics in order to help accelerate industry adoption of the Hadoop data platform and enable companies to mine their data for insights that inform the business.
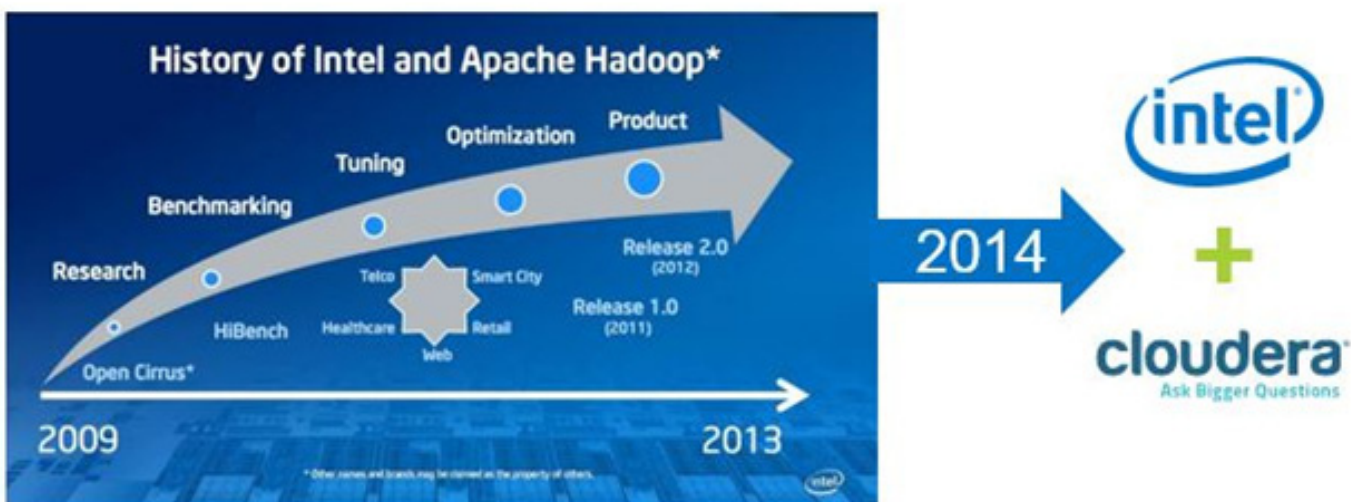
Intel has played an important role in the Hadoop industry, culminating with its alignment with Cloudera in 2014

Adopting Big Data for Finance

The big data revolution has dramatically changed the financial services industry. The following driving factors have motivated the need for financial sector companies to collect, store and analyze massive volumes of data:

- **Changes in the delivery of financial products and services.** Customers no longer need to physically visit their local bank to make deposits, make investment decisions or complete their banking transactions. Buyers and sellers of stocks execute their trades online instead of relying on brokers. Individuals file their taxes using online tools versus meeting with a tax accountant to prepare and file. As the industry has increasingly moved online, it's become faster, easier, and more affordable for consumers to handle their own banking and financial transactions.

- The result of these trends is that financial services and products have become commoditized. Instead of establishing relationships with a local service provider, customers often choose the most convenient and inexpensive online offering available. Personal connections as a requirement for customer loyalty has changed dramatically. However, every action made by a customer can be captured and analyzed by organizations seeking to understand the behaviors and preferences of their customers as they would have traditionally done through in-person interactions — now however, the process has become digitized.

- **Increased volume of activity.** The ease and affordability of executing financial transactions via online mechanisms has led to ever-increasing activity and expansion into new markets. Individuals can make more trades, more often, across more types of accounts, because they can do so with online tools in the comfort of their own homes, or on the go from a mobile device. Increased access and ease of use translates into increased volume of activity, which in turn translates into rapidly growing data volumes.

- It is important for banks, investment firms, and other financial services organizations to be able to collect and analyze this information in order to accurately assess risk and determine market trends. This became apparent during the market downturn of 2007-2008, when banks and brokerage houses scrambled to understand the implications of massive capital leverage and their ability to model and refine liquidity management. A single bank might capture internal transactions exceeding two billion per month, in addition to collecting public data of over a billion monthly transactions. These tremendous transaction volumes have made it nearly impossible to create models that take into account multi-year data sets using detailed data. Financial firms manage anywhere from tens to thousands of petabytes of data, yet most systems used today build models using only samples as small as 100 gigabytes. Relying on data samples requires aggregations and assumptions, resulting in inaccuracies in projections, limited visibility into actual risk exposure, instances of undetected fraud, and poorer performance in the market.

- **New sources of data.** The increased pace of online activity has led to the availability of new sources of data that are complex to consume, such as unstructured social media data. This information, if combined with individual financial transactions and history, can help to paint a holistic picture of individuals, families, organizations, and markets. Big data technologies are required to process these structured and unstructured data sources since using traditional relational database and data warehousing technologies for this purpose is no longer possible.

- **Increased regulatory compliance.** As result of more rigorous regulatory compliance laws, the financial services industry has had to store an increasing amount of historical data. New technology tools and strategies are needed to address these demands.
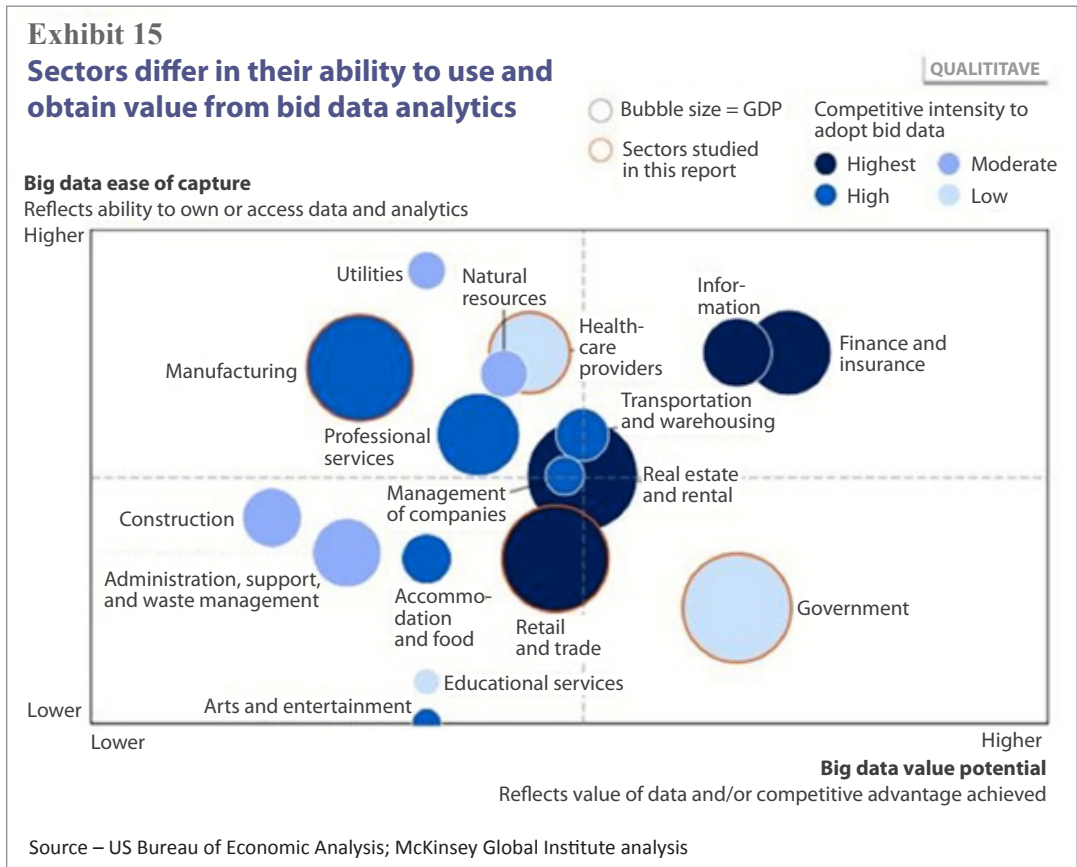
Not all industries are created equal when it comes to being able to put data to use. Some, like the construction industry, are constrained by the amount of data they can capture and even more hamstrung by their ability to get value from it. Others, like the finance industry, both generate a lot of data and can put it to use. In the graphic below from the U.S. Bureau of Economic Analysis, it is clear that the finance industry ranks highest in terms of its ability to use and obtain value from big data.

Determining the appropriate level of engagement for a big data deployment project is an important consideration toward insuring the success of the project. For example, department-level big data projects generally are more successful than large-scale initiatives which routinely fail. An incremental approach is better.

Exhibit 15
Sectors differ in their ability to use and obtain value from bid data analytics

QUALITITAVE

Bubble size = GDP
Sectors studied in this report

Competitive intensity to adopt bid data
Highest   Moderate
High   Low

Big data ease of capture
Reflects ability to own or access data and analytics
Higher

Utilities   Natural resources   Information   Finance and insurance
Health-care providers
Manufacturing   Transportation and warehousing
Professional services   Real estate and rental
Construction   Management of companies
Administration, support, and waste management   Accommodation and food   Retail and trade   Government
Educational services
Arts and entertainment
Lower

Lower   Higher
Big data value potential
Reflects value of data and/or competitive advantage achieved

Source – US Bureau of Economic Analysis; McKinsey Global Institute analysis

Finance industry is highest ranking in terms of its ability to use and obtain value from big data.

Here is a short list of guidelines toward the adoption of big data for finance:

- Develop methods and services for the valuation of data — and extend their role in compliance and internal control to the ethical and effective stewardship of data assets.

- Unite disparate data from a variety of systems designed to meet the diversity across regions regarding language, regulations, currency, time zone, etc.

- Use big data to offer more specialized decision-making support — often in real time — and decide when data can most usefully be shared with internal and external stakeholders and monetized as new products or services.

- To offset flat or declining revenue streams, financial services firms need to develop new big data centric products while also targeting existing products to new audiences.

- Use big data and its associated tools not only to identify risks in real time and improve forensic accounting abilities but also to evaluate the risks and rewards of long-term investment in new products and new markets.

- Gaining agility starts with an assessment of existing processes and systems. Financial industry firms must identify what existing practices will not support the progress they need. To get significant advantage in today's competitive landscape, they should pursue technologies that support new, innovative practices.

Spark is an open-source data analytics cluster computing framework built on top of HDFS. Spark serves as evidence of the continuing evolution within the Hadoop community — away from being a batch processing framework tied to the two-stage MapReduce paradigm to a more advanced in-memory, real-time platform.

Hadoop represents a good path for financial sector firms to adopt big data. With Hadoop, firms have access to a powerful platform providing both highly scalable and low cost data storage tightly integrated with scalable processing. Financial firms are now able to tackle increasingly complex problems by unlocking the power of their data. The capability to understand and act upon their data opens the door to a richer and more robust financial ecosystem.

A recent addition to the Hadoop ecosystem is Spark. Spark is an open-source data analytics cluster computing framework built on top of HDFS. Spark serves as evidence of the continuing evolution within the Hadoop community — away from being a batch processing framework tied to the two-stage MapReduce paradigm to a more advanced in-memory, real-time platform.

Now, these firms can better serve their customers, understand their risk exposure and reduce incidents of fraud. Here is a short list of benefits afforded by Hadoop to the financial services industry:

- Scalability to capture and analyze data previously untapped
- An economical way to store and process data
- A central repository for various data structures from existing and new data sources
- Operational efficiency by moving jobs to technology designed to process multiple data types
- The ability to ask different questions to improve decision making

By implementing a Hadoop solution, an organization processing approximately 20 billion events per day is able to significantly decrease operational costs: decreased storage costs from ~$17/GB to ~21 cents, experience a one year payback on the entire initiative.

The Intel Powered Dell™ | Cloudera™ Apache Hadoop Solution is a superb choice for lowering the barrier to adoption for financial institutions intending to use Apache™ Hadoop® in production. Dell Xeon based PowerEdge servers, Force10 networking, the linux operating system and the Cloudera Manager tools make up the foundation on which the Hadoop software stack runs.

You can start planning out your Hadoop infrastructure using one of the global network of Dell Solution Centers, technical labs that enable you to architect, create a proof-of-concept

Using StackIQ's Cluster Manager software the customer was able to rapidly provision and re-provision the servers to run the two different applications with various configurations. The use of Cluster Manager enabled the customer to complete more and higher quality tests than originally expected.

(POC), validate, and build data center solutions. In a recent POC for a large provider of financial data, Dell partnered with StackIQ to configure a 60-datanode cluster to compare two different big data technologies, Cassandra and HBase. Using StackIQ's Cluster Manager software the customer was able to rapidly provision and re-provision the servers to run the two different applications with various configurations. The use of Cluster Manager enabled the customer to complete more and higher quality tests than originally expected.

Security Considerations

An important application area where big data is taking a firm foothold in many financial industry firms is information security. In conjunction with the traditional 3 Vs of big data, financial industry firms must consider a fourth V: *vulnerability*. To manage big data effectively, you must keep it secure and compliant with regulatory requirements at all times (vulnerability). Protecting a vast and growing volume of critical information — and being able to search and analyze it to detect potential threats — is more essential than ever. As the software platforms (e.g. Hadoop) supporting this quantity of data move to mainstream use, managing their security and availability becomes a big data challenge in and of itself, requiring continuous diagnostics and monitoring.

> The increasingly global nature of the financial services industry makes it necessary to comprehensively address international data security and privacy regulations.

Banking and financial institutions need to secure the storage, transit and use of corporate and personal data across business applications, including online banking and electronic communications of sensitive information and documents. The typical IT environment consists of a mix of new and legacy systems and applications across highly distributed networks of branch offices, call centers and web portals. Many of the traditional point security solutions that are deployed add complexity and management costs, and leave gaps between systems and applications that are highly vulnerable to attack. The increasingly global nature of the financial services industry makes it necessary to comprehensively address international data security and privacy regulations.

Financial institutions are top targets of cybercrime. While all types of businesses are vulnerable to attacks by criminals, it's the security breaches at financial firms that elicit the most media attention, public scrutiny and legislator consternation. When threats occur, it's more than financial loss at stake.

Customers question their trust in their bank's ability to provide security and protect their privacy. Hard-earned customer loyalty diminishes. How can you detect fraud and stop attackers before they threaten your financial institution and its customers? Big data technologies can help by enabling financial firms to not only capture in near real time every event that occurs across the entire organization but also provide context to understand these events so information can be shared to better issue alerts of potential and actual threats.

Many finance industry firms are using big data to detect and/or prevent fraud. Big data supports what's known as continuous or behavioral authentication, a process that can help prevent fraud. Further, detecting security breaches using huge volumes of security data along with unstructured social media data, combined with new big data tools such as Hadoop, enables financial industry firms to be more proactive about security. Big data can enhance data security for the finance industry through:

- Understanding activity patterns among customers and the broader industry.
- Sharing of data – critical especially about emerging attack vectors and threats.
- Increasing reliance on data to predict attacks, based on trends that are targeting the industry.

One particularly good solution for finance industry data security requirements is Dell's SharePlex Connector for Hadoop. Proactive security requires data analytics for a business intelligence advantage and essential decision-making insight. The Hadoop framework gives you that, but integrating data can be time consuming, providing only snapshots that quickly become out of date. SharePlex Connector for Hadoop loads and continuously replicates changes from an Oracle database to a Hadoop cluster — in near real time to Hive and HDFS, and in real time to Hbase. This gives you all the benefits of maintaining a real-time or near real-time copy of source tables, so your organization can efficiently and cost-effectively perform big data analytics in support of enterprise security.

Regulatory Compliance Considerations

The regulatory environment under which banks operate adds substantial operational concern for the business. Stringent regulatory compliance laws have been put in place to improve operational

> Financial industry companies have recognized that the key to optimizing their business operations in today's regulatory environment is a matter of maintaining an efficient and large-scale data management infrastructure.

transparency. With government regulation driving many business processes, banks have apprehension leading up to a national election because they don't know what direction the regulatory wind might blow. Currently, financial services organizations are held much more accountable for their actions, and are required to be able to access years of historical data in response to regulators' requests for information at any given time, specifically:

- The **Dodd-Frank Act** requires firms to maintain records for at least five years.
- **Basel** guidelines mandate retention of risk and transaction data for three to five years.
- **Sarbanes-Oxley** requires firms to maintain audit work papers and required information for at least seven years.
- **FINRA/Tradeworx Project –** SEC requires the creation of a real-time transaction monitoring system to detect potentially disruptive market activity stemming from high-frequency trading. A consolidated audit trail or CAT (for day after monitoring) by FINRA is to include orders, quotes, updates and cancellations, plus a real-time system by Tradeworx are two initiatives for market surveillance.

Further, these records must be available on demand, or in some cases must be normalized and sent to regulators proactively. Financial industry companies have recognized that the key to optimizing their business operations in today's

regulatory environment is a matter of maintaining an efficient and large-scale data management infrastructure. This level of compliance can be addressed via big data technologies.

The wide reaching effects of the economic meltdown in 2008 were largely due to the lack of visibility into consumer actions and groups with related risk profiles, combined with the increased flow of consumer funds into financial firms. Financial industry firms, as well as the underlying economy, were more susceptible than statistical models had predicted. As a result, the finance world set out to find new classes of technologies that would allow them to manage and take advantage of ever-growing data sets. In the process, they evaluated the technologies deployed by large web concerns like Facebook and Google. In short, they found Hadoop: an open source software architecture that enables distributed parallel processing of huge amounts of data across inexpensive, commodity servers. With Hadoop, no data set is too big or complex.

> Many of today's top financial services firms have already deployed Hadoop to create centralized data-hubs combining large volumes of diverse and detailed data, crafting competitive advantage in several key business application areas.

Hadoop is a stable and dependable platform that financial organizations can trust for their big data requirements. Using big data solutions driven by Hadoop allows financial services firms to optimize capital leverage while maintaining the reserves required by regulators. Many of today's top financial services firms have already deployed Hadoop to create centralized data-hubs combining large volumes of diverse and detailed data, crafting competitive advantage in several key business application areas.

The Hadoop software stack is a good choice to unlock the power of the financial firm's data assets including but not limited to compliance requirements:

- **Modeling market risk –** using detailed data to give banks better insight into the material behavior of complex financial instruments and creating more predictable results from these investment vehicles. Using Hadoop, banks can build and evaluate models in an expedient manner, leading to a high degree of competitive advantage.

- **Modeling consumer risk –** giving greater insight into capital availability and liquidity. Using Hadoop-based solution, financial services organizations are able to optimize capital leverage while maintaining the reserves required by regulators.

- **Fraud detection and AML capabilities –** allowing banks to detect critical breaches faster, saving money that directly affects the bottom line. As financial firms roll out new products and services to customers, the committers of fraud and money laundering activities constantly are adapting to fraud prevention techniques.

- **Recommendation/personalization –** providing sophisticated tools for cross-sell and up-sell purposes so banks can offer new financial products and services to their existing high-value customers. With the ability to track user interactions across applications and marketing channels in real time, banks can dynamically categorize customers and experiment with selectively targeted offers, including investment strategies and savings advice to help customers improve their financial discipline. The result is an improved and more personalized client experience.

- **Mortgage portfolio valuation –** identifying which mortgages are candidates for default. The Hadoop platform engages the use of unstructured data sources that offer greater visibility into mortgage models, such as property valuation sites, local consumer offers such as crime report data and public sentiment data.

One solution to the rigors of regulatory compliance is to employ the Dell SecureWorks Security & Risk Consulting team designed to provide the expertise and analysis to help you enhance your IT security posture, reduce your information security risk, facilitate compliance and improve your operational efficiency.

One solution to the rigors of regulatory compliance is to employ the Dell SecureWorks Security & Risk Consulting team designed to provide the expertise and analysis to help you enhance your IT security posture, reduce your information security risk, facilitate compliance and improve your operational efficiency. The Dell team has helped thousands of customers design their strategic security programs, assess and test their defenses, resolve critical information security breaches, and meet their compliance mandates.

Dell SecureWorks' Compliance and Certification services offer assessments specifically tailored to address regulatory issues for financial services firms. Security consultants work with you to gauge the current risks to your mission critical IT assets, and to assess the current state of your security posture as compared to best practices and regulatory guidance.