The top five ways to get started with big data

Big data: A high-stakes opportunity

Remember what life was like before big data? The term has become so prevalent in the business lexicon that sometimes it's hard to remember that big data is a relatively recent phenomenon. Some may have viewed it as a fad, but data generated by people, processes and machines is only continuing to grow. Big data is here to stay.

Make no mistake, data is an asset—but not when you're drowning in it. In the information age, one of your greatest resources can also be your biggest downfall if your organization doesn't know how to leverage it properly. So what can you do with your data?

Consider these actual scenarios:

- The healthcare industry spends roughly USD250 billion on healthcare fraud per year. By 2016, this could grow to more than USD400 billion a year.[1] The US healthcare sector could create more than USD300 billion in value every year using big data creatively and effectively to drive better efficiency and quality.[2]
- One rogue trader at a leading global financial services firm created USD2 billion worth of losses, almost bankrupting the company. Financial institutions now have a lot more data at their fingertips to help them prevent both external fraud (involving customers, account holders or policyholders) and internal, employee-related incidents.
- In Europe, governments could save more than EUR100 billion (USD149 billion) in operational efficiency improvements alone by using big data,[3] not including using big data to reduce fraud and errors and boost the collection of tax revenues.

- Retailers miss out on USD93 billion in sales each year because they don't have the right products in stock to meet customer demand. A retailer using big data could increase its operating margin by more than 60 percent.[4]
- Six billion global subscribers in the telecommunications industry—which is growing at double-digit rates each year[5]—are demanding unique, personalized and often location-based offerings that match their individual lifestyles.

With such high-stakes costs and opportunities, the market is primed for big data solutions. In a recent study conducted by the IBM Institute for Business Value in collaboration with the Saïd Business School at the University of Oxford, respondents were asked to describe the level of big data activities in their organizations today. The results suggest four main stages of big data adoption and progression along a continuum: Educate, Explore, Engage and Execute (see Figure 1).[6]
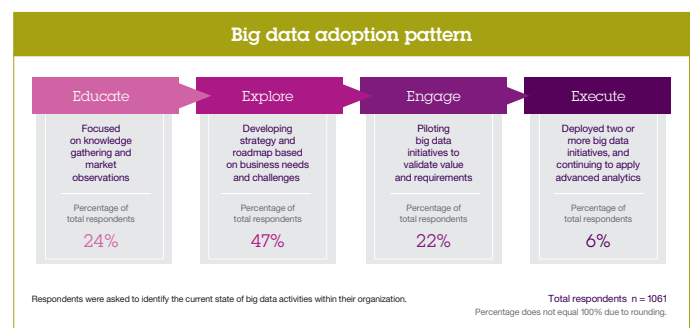

Figure 1. The four phases of big data adoption

While only 6 percent of organizations are already executing big data initiatives, about one-quarter are piloting initiatives, half are developing a strategy and will be looking to purchase soon, and a quarter more are in an information-gathering phase.

If you are not working on a big data strategy, your competitors probably are. The difficulty is figuring out how and where to get started.

Big data use cases

Because much of the big data activity in the market in the past has focused on learning about big data technologies, vendors haven't made a concerted effort to help organizations understand which problems big data can address. IBM has been the exception.

Through conducting surveys, studying analyst findings, talking with more than 300 customers and prospects and implementing hundreds of big data solutions, IBM has identified the top five high-value use cases that can be your first step into big data:

1. **Big data exploration:** Find, visualize and understand big data to improve decision making
2. **Enhanced 360-degree view of the customer:** Extend existing customer views by incorporating additional internal and external information sources
3. **Security/intelligence extension:** Reduce risk, detect fraud and monitor cybersecurity in real time
4. **Operations analysis:** Analyze a variety of machine data for improved business results and operational efficiency
5. **Data warehouse augmentation:** Integrate big data and traditional data warehouse capabilities to gain new business insights while optimizing existing warehouse infrastructure

Use case 1: Big data exploration

The first step in leveraging big data is to find out what you have and to establish the ability to access it and use it to support decision making and day-to-day operations—in other words, big data exploration.

Most discussions of big data start with three Vs—volume, velocity and variety. These identify the dimensions of the challenge that every large organization deals with daily as they struggle to extract value from their information resources, make better decisions, improve operations and reduce risk. Any important decision, customer interaction or analysis inevitably requires information from multiple data sources. IBM® InfoSphere® Data Explorer, part of the IBM big data platform, provides the capability to easily navigate information within enterprise systems as well as data from outside the organization.

The growth of so-called "raw" data from sensors, machine logs, clickstreams, websites and so on presents yet another challenge. How do organizations add context to this data to fuel better analytics and decision making? Here again, the ability of InfoSphere Data Explorer and other capabilities in the IBM big data platform to fuse information from these semi-structured sources together with enterprise data can add valuable context to help organizations gain enhanced value from this data.

IBM big data exploration capabilities also help to contain risk. Organizations that lack the ability to navigate and explore large areas of their information landscape put themselves at risk of leaking confidential information such as personally identifiable information (PII), losing important trade secrets and strategic information to competitors, and being unable to retrieve and verify information when required for litigation and other corporate governance matters.

Is big data exploration is the right use case for you?

Ask yourself:

- How do you separate the "noise" from useful content?
- How do you perform data exploration on large and complex data?
- How do you find insights in new or unstructured data types (such as social media and email)?
- Are your users exploiting information to make factional business decisions or is the inability to find information inhibiting good business practices?
- How do you enable employees to navigate and explore enterprise and external content? Can you present this in a single user interface?
- How do you identify areas of data risk before they become a problem?
- What is the starting point for your big data initiatives?

Use case 2: Enhanced 360-degree view of the customer

Gaining a full understanding of customers—how they prefer to shop, why they switch, what they'll buy next and what leads them to recommend a company to others—is strategic for virtually every business. However, this requires companies to leverage internal and external sources of information to assess customer sentiment, and understand what meaningful actions will help them develop relationships with customers.

A recent IBM Institute for Business Value report on real-world use of big data[7] recommends that organizations focus their big data efforts first on customer analytics that enable them "to truly understand customer needs and anticipate future behaviors." In this case, the term "customer" is used in a broad sense—it could mean patients in healthcare, a person of suspicion in government or suppliers in manufacturing.

In addition to these analytics that give strategic insights into customer behavior, the importance of the 360-degree view extends to front-line employees. Forward-thinking organizations recognize the need to equip their customer-facing professionals with the right information to engage customers, develop trusted relationships and achieve positive outcomes such as solving customer problems and up-selling and cross-selling products. To do this, they must navigate large amounts of information quickly to zero in on what's needed for a particular customer.

IBM InfoSphere Data Explorer works in combination with IBM InfoSphere Master Data Management (MDM) to combine information in context from all the applications and repositories containing customer information (CRM, ECM, supply chain, order tracking database, email and so on) to give a complete view of the customer—without requiring the user to log into and search multiple disparate systems.

In this one view, the customer-facing professional can see all of the customer's information—what products she has purchased, recent support incidents, news about her company, recent conversations and more. An activity feed in the center of the screen shows up-to-the-moment updates about the customer, product or other entity that is being viewed. Analytics from InfoSphere BigInsights™, InfoSphere Streams, IBM Cognos® business intelligence and IBM SPSS® products can also be shown, with the context of the analytics defined by the application. This frees the employee to interact with the customer in a more personalized fashion. By doing so, they can provide the right answer quickly while also recommending up-sell opportunities. This visibility helps drive customer loyalty, satisfaction and ultimately revenue.

As shown in Figure 2, leveraging master data management can ensure the accuracy and reliability of data across all of an organization's various systems. This consistency ensures that the view created by InfoSphere Data Explorer will incorporate consistent and accurate data about an entity. In one sense,

Figure 2. Information about a customer as viewed in an application built with the InfoSphere Data Explorer Application Builder, leveraging InfoSphere Master Data Management for a trusted view of customer data.

InfoSphere Data Explorer provides a business user interface to trusted master data combined with related content from other structured and unstructured data sources.

Is the 360-degree view of the customer use case right for you?

Ask yourself:

• How do you identify and deliver all data about a customer, product or competitor to those who need it?
• How do you combine your structured and unstructured data to run analytics and discover insights?
• How are you driving consistency across information assets when representing your customers, clients and partners?
• How do you deliver a complete view of the customer to enable your line-of-business users to ensure better business outcomes?
• How do you apply insights and take actions?

Use case 3: Security/intelligence extension

To combat new and emerging sophisticated security threats, organizations must adopt approaches that help spot anomalies and subtle indicators of attack. Many organizations today are using big data technologies to augment and enhance traditional security solutions to significantly improve intelligence, security and law enforcement insight.

With an extended security/intelligence approach, organizations can:

• Sift through massive amounts of data—both inside and outside the organization—to uncover hidden relationships, detect patterns and prevent security threats
• Uncover fraud by correlating real-time and historical account activity to uncover abnormal user behavior and suspicious transactions
• Examine new sources and varieties of data for evidence of criminal activity, such as the Internet, mobile devices, transactions, email and social media

There are three main applications for the extended security/intelligence use case:

1. **Enhanced intelligence and surveillance insight:** Organizations can analyze data in motion and data at rest to find associations or uncover patterns. This type of real or near real-time insight can be invaluable and even life-saving.

2. **Real-time cyberattack prediction and mitigation:** The growing number of high-tech crimes—including cyber-based terrorism, espionage, computer intrusions and major cyber-fraud—pose a real threat. By analyzing network traffic, organizations can discover new threats early enough to react in real time.

3. **Crime prediction and prevention:** The ability to analyze telecommunications data (for example, call detail records) and social media data enables law enforcement to pick up on criminal threats among the noise and gather criminal evidence. Instead of having to wait for a crime to be committed, they can prevent them from happening in the first place and proactively apprehend criminals.

Depending upon the scenario, organizations are likely to need one of the following security/intelligence platforms: Criminal Information Tracking System, Surveillance Monitoring System or a Security Information and Event Management (SIEM).

Today, these platforms access data from a variety of structured data sources (transactional, databases, network, firewall and others). The platform data is stored and managed in its own database or warehouse. However, these systems cannot handle new and emerging big data trends that require analysis of real-time streaming data or unstructured data types (see Figure 3).

Big data technologies such as stream computing (InfoSphere Streams) and enterprise-class Apache Hadoop analytics (InfoSphere BigInsights) enhance these traditional security and intelligence analysis platforms by natively accessing data from unstructured and/or streaming big data sources such as telecommunications records, smart devices, Twitter streams, Facebook posts, email, point-of-sale monitoring, location-based sensors, video, audio, and thermal and other machine-generated data.

Identify and protect against threats by building insights from broad data sets

Traditional security operations and technology

Logs
Events   Alerts

Configuration information

System audit trails

Identity context

Network flows and anomalies

External threat intelligence feeds

Video/audio surveillance feeds

Web page text

Business process data

Email and social activity

Customer transactions

Big data analytics

New considerations
Collection, storage and processing
• Collection and integration
• Size and speed
• Enrichment and correlation

Analytics and workflow
• Visualization
• Unstructured data analysis
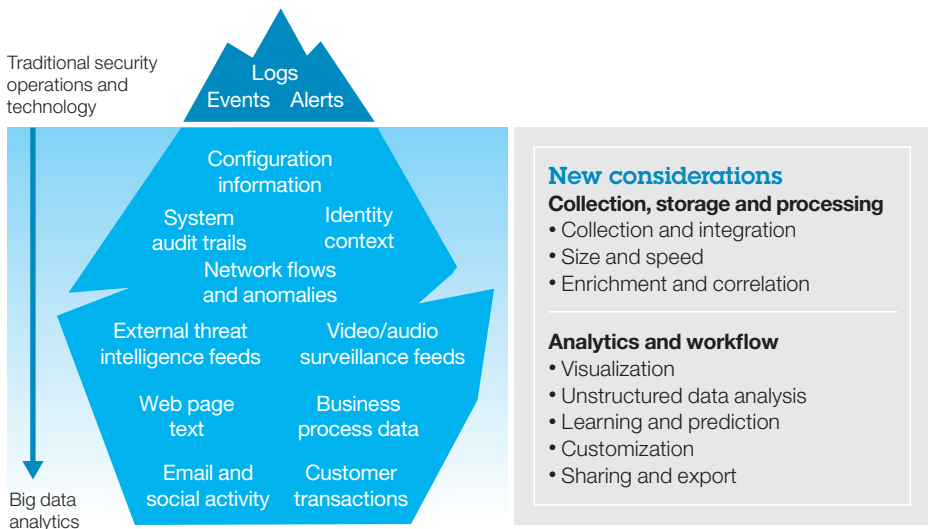• Learning and prediction
• Customization
• Sharing and export

Figure 3. Building deeper security insights from broader data sets.

Real-time data can be processed and analyzed using InfoSphere Streams and the resulting output can be stored in a data warehouse or InfoSphere BigInsights. Clients using the IBM i2® Analyst's Notebook® can directly view and analyze real-time data using the InfoSphere Streams integration.

Is the security/intelligence extension use case right for your enterprise?

Ask yourself:

- Do you need to enrich your security or intelligence system with real-time data from unused or underleveraged data sources (video, audio, smart devices, network, call data records or social media)?
- Do you need sub-second detection, identification and resolution of physical or online threats?
- Do you need to follow activities of criminals, terrorists or persons on a watch list?
- Do you need to correlate large volumes of technical or human intelligence data and sources to look for associations or patterns (big data forensics)?
- Do you need to enhance your SIEM solution with unstructured data (email, social) to improve cyberthreat detection and remediation?

Use case 4: Operations analysis

The abundance and growth of machine data—which is generated by computers and network devices as well as sensors, meters and GPS devices—is another major driver of big data solutions. This data comes in large volumes and a variety of formats, including in-motion or streaming data. It requires complex analysis and correlation across different types of data sets. It also requires unique visualization capabilities based on data type and industry or application.

Organizations that disregard this vast, rich source of information are making business decisions based only on a small subset of the data available to them. By combining machine data with existing enterprise data through operations analysis, organizations can:

- Gain real-time visibility into operations, customer experience, transactions and behavior
- Proactively plan to increase operational efficiency
- Identify and investigate anomalies
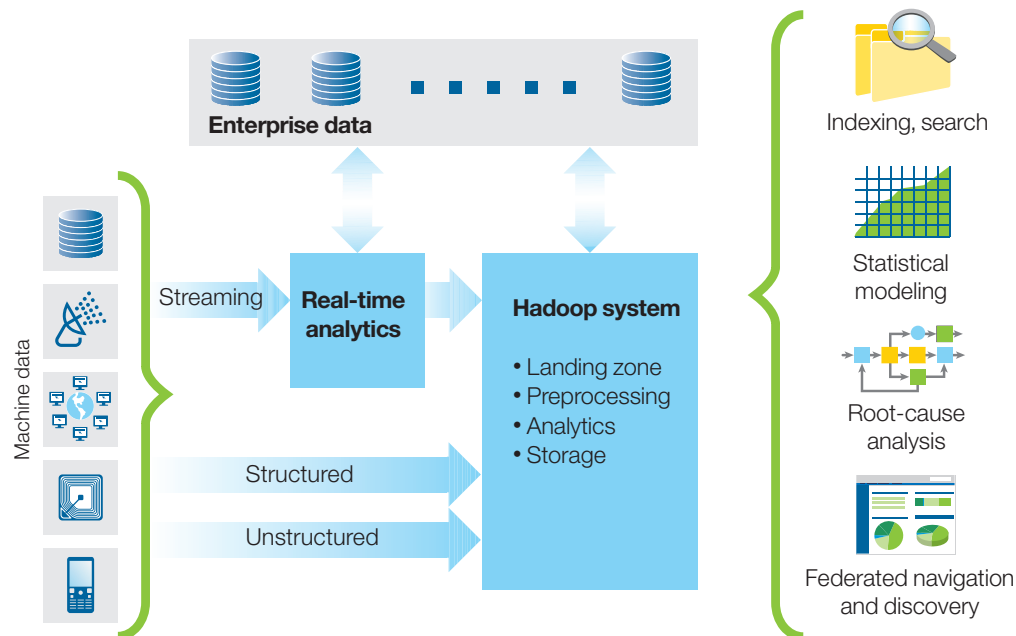- Monitor end-to-end infrastructure to proactively avoid service degradation or outages

Figure 4. Operations analysis combines machine and enterprise data for rich insights.

As shown in Figure 4, you may have large volumes of machine data, in various formats that don't work well with each other, coming into your Hadoop Distributed File System (HDFS). You may also have streaming data. InfoSphere BigInsights, which comes with a machine data accelerator built for ingesting and processing large volumes of machine data to provide in-depth business insights. The machine data can then be correlated with other enterprise data such as customer or product information.

Combining machine and business data allows you to put it into the hands of the operational decision maker, which in turn increases operational intelligence and efficiency. These decision makers can visualize data across many systems to get the most informed view and react quickly to changes and events.

Is the operations analysis use case right for you?

Ask yourself:

- **Do you deal with large volumes of machine data, such as raw data generated by logs, sensors, smart meters, message queues, utility systems, facility systems, click stream data, configuration files, database audit logs and tables?**
- **Are you able to perform the complex analysis that is required to correlate information and key performance indicators across different data sets and events in real time?**
- **Are you able to search and access all of your machine data?**
- **Do you have the ability to visualize streaming data and react to it in real time?**
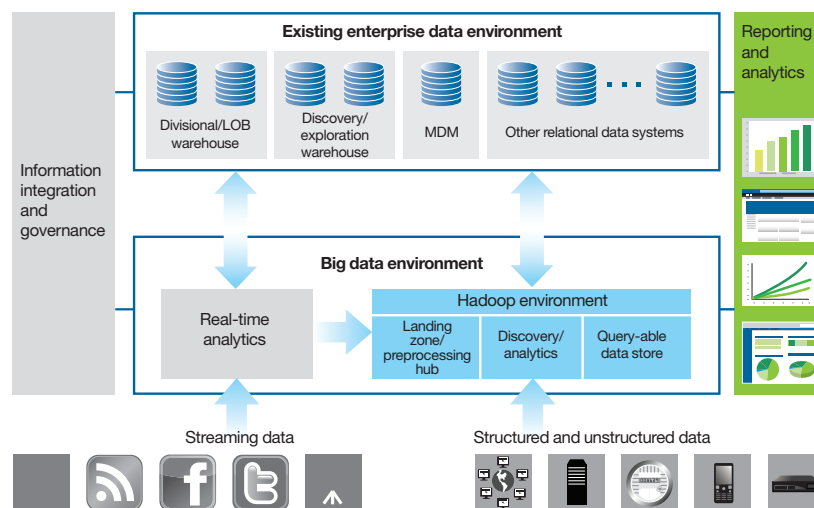- **Are you able to perform root cause analysis using that data?**

Figure 5. Data warehouse augmentation helps maximize the value of data.

Use case 5: Data warehouse augmentation

The final use case, data warehouse augmentation, builds on an existing data warehouse infrastructure, leveraging big data technologies to augment its value. It is not a replacement for your data warehouse environment—rather, it is designed to maximize the value of it.

Data warehouse augmentation stems from two basic needs. The first is the need to leverage a variety of data to gain new business insights. Organizations want to be able to analyze multi-structured data, but the warehouse isn't built for this. Relying on data warehousing alone means companies are forced to neglect valuable data. Additionally, organizations are demanding lower latency; they need information in hours or minutes, not weeks or months. Lastly, organizations require query access to data.

The second basic need is optimization of the warehouse infrastructure. Warehouse data volumes today are reaching big-data levels, putting stress on the data warehouse. The warehouse itself may not be expensive, but when you try to store and analyze everything in that environment, performance will suffer and costs will rise.

There are three types of data warehouse augmentation (see Figure 5):

1. **Pre-processing hub:** Used when an enterprise-grade Hadoop capability (InfoSphere BigInsights) is needed as a staging area or "landing zone" for data before determining what data should be moved to the data warehouse. InfoSphere Data Explorer can be used for early exploration, to determine what data you want to move to run deeper analytics or cheaper storage. This isn't a required step, but it can be used in areas where organizations want to leave some of their data at rest.

Stream computing (InfoSphere Streams) can also be used as a real-time component by processing and analyzing streaming data, without having to store it first, and determining what data should be saved—either in HDFS or the data warehouse. In some cases, data won't need to be saved; being able to process and act on information as it is happening can also reduce storage in the warehouse. With this landing zone approach, data can be cleansed and transformed before loading into the data warehouse.

2. **Discovery/analytics:** This approach uses stream computing analytics on data in motion, giving organizations the ability to perform analytics that might have previously been done in the data warehouse, therefore optimizing the warehouse and enabling new types of analysis. Different data types can be combined with warehouse data, enabling deep analytics to provide insights not previously possible. In addition, stream computing can act as an analysis filter to find the high-value nuggets of data which then can be stored in InfoSphere BigInsights or the data warehouse.

3. **Query-able data store:** In this approach, infrequently accessed or aged data can be offloaded from warehouse and application databases using information integration software and tools. This helps organizations store cold, low-touch data

in low-cost storage yet keep it accessible within InfoSphere BigInsights using query or BI tools. InfoSphere Data Explorer can be used to view and navigate all the data stored in InfoSphere BigInsights.

Is the data warehouse augmentation use case the right big data starting point for your organization?

Ask yourself:

- Are you drowning in very large data sets (terabytes to petabytes)?
- Do you use your warehouse environment as a repository for all data?
- Do you have a lot of cold, or low-touch, data?
- Do you have to throw data away because you're unable to store or process it?
- Do you want to analyze data in motion to determine, in real time, what data should be stored in the warehouse?
- Do you want to perform data exploration on complex and large amounts of data?
- Do you want to analyze nonoperational data?
- Are you interested in using your data for traditional and new types of analytics?

The IBM big data platform

The five big data use cases described in this paper provide high-value starting points for companies looking to begin their big data journey. The IBM big data platform can play an integral role in that transformation.

Big data use cases require an integrated set of technologies that are specifically designed to address the unique challenges of working with high-volume, high-variety and high-velocity data. These are not single-issue problems with single-product solutions. The IBM platform helps companies reduce the time and cost of big data projects, as well as achieve a rapid return on investment (ROI) by leveraging pre-integrated components. In addition, out-of-the-box and standards-based services offer a head start on deployment. You can start small to address an initial use case and progress to others as you proceed on your big data journey.

For more information

To learn more about big data use cases and the IBM big data platform, contact your IBM representative or IBM Business Partner, or visit: **ibm.com**/software/data/bigdata/use-cases.html

Additionally, IBM Global Financing can help you acquire the software capabilities that your business needs in the most cost-effective and strategic way possible. We'll partner with credit-qualified clients to customize a financing solution to suit your business and development goals, enable effective cash management, and improve your total cost of ownership. Fund your critical IT investment and propel your business forward with IBM Global Financing. For more information, visit: **ibm.com**/financing

[1] Financial Crimes Report to the Public: Fiscal Years 2010-2011. www.fbi.gov/stats-services/publications/financial-crimes-report-2010-2011

[2] McKinsey Global Institute. "*Big data: The next frontier for innovation, competition, and productivity*." May 2011. www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

[3] McKinsey Global Institute. "*Big data: The next frontier for innovation, competition, and productivity*." May 2011. www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

[4] McKinsey Global Institute. "*Big data: The next frontier for innovation, competition, and productivity*." May 2011. www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

[5] International Telecommunication Union. "*Measuring the Information Society*." September 2012.

[6] IBM Institute for Business Value in collaboration with the Saïd Business School at the University of Oxford. "*Analytics: The real-world use of big data.*" November 2012. http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-big-data-at-work.html

[7] IBM Institute for Business Value in collaboration with the Saïd Business School at the University of Oxford. "*Analytics: The real-world use of big data.*" November 2012. http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-big-data-at-work.html

Please Recycle

IMW14710-USEN-00