

Holistic Approach to Big Data #4: 5 High Value Big Data Use Cases

1

At IBM, our product management, engineering, partner enablement, marketing, and other teams have all been working together to help to better understand the big data market. We've done surveys, met with analysts and studied their findings. We've met in person with end-user customers and prospects (over 300 meetings) and are confident that we found market "sweet spots" for big data.

The 5 Use Cases that we will study in this video are our sweet spots.

These Use Cases will resonate with the majority of the people that you meet with. In the coming slides we'll cover each of these in detail. We will walk through the needs, the values, and provide a number of specific user examples of the implementation of these ideas.

2

Big Data Exploration:

What's out there?, What does it mean?, and How do I use it?

Enhanced 360 degree view of a customer

Every company wants to know more about their customers, and they probably have some kind of a Customer Relationship Mgmt (or CRM) system that tracks info about the customer: Who is the customer? What are his/her preferences? What are their buying habits, saving habits? What do they buy, when, how....all with the aim of deriving more profit from that knowledge.

As we think of customer, don't limit yourself to people. By customer, we also mean corporate or business customer.

We desire to increase customer loyalty and satisfaction by understanding what meaningful actions are needed

But we are challenged getting the right information to the right people to provide customers what they need to solve problems, to cross-sell, and to up-sell.

Gaining a full understanding of customer—what makes them tick, why they buy, how they prefer to shop, why they switch, what they'll buy next, what factors lead them to recommend a company to others—is strategic for virtually every company.

IBM's own Institute for Business Value report "Real-world use of big data" cites as its #1 recommendation that organizations should focus their big data efforts first on customer analytics that enable them "to truly understand customer needs and anticipate future behaviors."

In addition to the analytics that give strategic insights into customer behavior, the importance of the 360° view extends to the 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 need to be able to navigate large amounts of information quickly to zero in on what's needed for a particular customer.

As you'll see in a moment, this 360° view is very synergistic with IBM's data governance story, especially Master Data Management (MDM).

The third High Value Big Data Use Case is: Security and Intelligence Extension

Unfortunately there are people out there trying to get to your bank accounts. ... It's in everyone's mind.

We only need to mention

- Target and credit cards
- Heartbleed
- And many other situations

The fourth High Value Big Data Use Case is: Operations Analysis

Walmart is a good example here.

Another one is Splunk, a software package that tracks how you are using your IT resources

Splunk Enterprise is a fully featured, powerful platform for *collecting, searching, monitoring and analyzing machine data*. Splunk Enterprise is easy to deploy and use. It turns machine data into rapid visibility, insight, and intelligence.

The fifth High Value Big Data Use Case is: Data Warehouse Augmentation

We will talk about this in more detail later.

3

Let's look at 360 degree view of your customer.

This diagram shows how core elements of IBM's Big Data Platform—Big Insights, Streams and Data Explorer—work together to enable

- Greater efficiencies in business processes;
- New insights from combining different data types
- And new business models made possible by new ways to view and use data.

At the bottom layer of the diagram, Data Explorer's Connector Framework provides secure access to a broad range of enterprise systems, containing both structured and unstructured data. Streams, BigInsights and Warehouse all have their own built-in capability to consume information.

At the same time, integration points among all of the components of the IBM Big Data Platform enable transfer of data as needed by the requirements of each use case and customer situation. As a result, BigInsights can serve as a "landing area" for output from Streams. Data Explorer can be used to navigate information managed by BigInsights, combined with data from enterprise systems. And BigInsights might be used to pre-process data before loading into a data warehouse.

IBM's enterprise class integration and governance capabilities apply across all of these operations, and the Data Explorer App Builder can be used to create compelling applications that incorporate data and analytics from all of the platform components. This diagram illustrates the strength of our platform story—no other vendor addresses the big data challenge with such a comprehensive platform which incorporates full stack from data access and integration, through proper management and governance, all the way up to the application layer.

While we want in these presentations to present a vendor agnostic approach to Big Data, we also wanted to show you how IBM is addressing issues of Big Data through a single platform.

4

Airbus is an InfoSphere Data Explorer customer with more than 4 TB of data indexed, including file systems, SharePoint, an SAP system and a Siebel CRM System. They have used this capability to provide unified information access enterprise-wide, and also for specific targeted business applications like their support “war room” where they handle calls from airline maintenance departments to help them get airplanes back in the air as quickly as possible. This one application accounted for more than \$36 million in savings in a single year.

To find other customers like Airbus, look for organizations whose business is information-intensive, such as those that build and support highly complex systems, or who have a compelling need to improve their operations through better use and re-use of information. It is critical for such organizations to equip their employees with access information across all of their silos, regardless of format, or where it is managed.

How does one identify the need for this type of application? A direct line of questioning works best, asking them to describe what steps their organization is taking to provide a unified view of information to employees; determine how they extract insights from their unstructured content; etc. Very few organizations are doing this successfully or efficiently.

Data Explorer is the typical starting point for this use case because it is the low-risk rapid path for an organization to get its arms around its data. This allows the organization to gain immediate business value through better information visibility, as well as prepare for the next stage of big data deployment.

5

Information about a client is viewed in an application built with the Data Explorer Application Builder.

The Data Explorer application combines information *in context* from customer relationship management (CRM), content management, supply chain, order tracking database, e-mail, and many more systems

*

to give a 360° view of the client so the user doesn't have to log in to and search multiple different systems.

*

In this one view the customer-facing support agent can see all of the contact's information -- the products used, any recent support incidents, news about the company, recent conversations, and more.

An “activity feed” in the center of the screen shows up-to-the-moment updates about the customer, the product, or any other entity that is being viewed.

*

Analytics from BigInsights, Streams, data warehouse / business intelligence (BI) products can also be shown, with the context of the analytics defined by the information displayed in the application. This frees the customer-facing professional to interact with the customer and leverage this complete view to improve customer loyalty and increase revenue.

As I mentioned a moment ago, this use case is very synergistic with Master Data Management. MDM provides a single, consistent view of data across all of the organization's various systems. This consistency ensures that the view created by Data Explorer will incorporate consistent and accurate data about an entity.

In one sense, Data Explorer provides a business user interface to trusted master data combined with related content from other structured and unstructured data sources. The availability of MDM accelerates implementation of the Data Explorer 360° application and ensures its accuracy and consistency.

6

What is Security/Intelligence Extension?

It's using big data technologies to augment and enhance traditional security solutions by analyzing new types (e.g., unstructured or streaming data) and sources of under-leveraged data to significantly improve intelligence, security, and law enforcement insight.

It helps answers questions such as:

- Why are huge amount of data outbound on Sun 2:00am
Answer: Shut down the router
- How to react to email with "Confidential" info sent outbound?
Answer: Block it

Enhanced Intelligence & Surveillance insight analyze data-in-motion and at rest to:

- Find associations
- Uncover patterns and facts
- Provide up-to-date information about systems, networks, and facilities

It can be used to analyze network traffic to support real-time cyber attack prediction & mitigation:

- Discover new threats early
- Detect known complex threats
- Take action in real-time

It can be used to analyze telco & social data to support:

- Crime prediction & protection
- Gather criminal evidence
- Prevent criminal activities
- Proactively apprehend criminals

What are the drivers for an Security/Intelligence Extension use case?

- Need to analyze from existing and new sources (data in motion and at rest) to find patterns and associations
- Need for more up-to-date intelligence information
- Ability to predict, detect, act on network/computer security threats sooner
- Ability to analyze telco and/or social data to track criminal or terrorist activity
- Need to analyze data that is continuously generated such as video, audio, smart devices

Organizations already want to enhance their existing security/intel platforms to improve local & national security, protect their borders, and prevent criminal/terrorist activity. This is another tool in the toolkit.

7

This technology provides an efficient implementation of the multiple hypothesis tracking (MHT) algorithm proposed by Donald Reid that can be used to “reconstruct events,” integrating and correlating multimedia content from diverse source, and thus provide a k-best hypotheses approach to reducing complexity.

8

TerraEchos is a world leader in advanced security solutions employing streaming data analytics for advanced cyber-physical infrastructure protection. Embedding IBM’s InfoSphere Streams inside a tactical appliance, the Kairos™ Solution processes extreme volumes of variable data at line speeds to run an array of advanced analytics from an aggregated set of continuous and discrete data sources

TerraEchos deploys a security surveillance system to detect, classify, locate, and track potential threats at highly sensitive national laboratories. The software was developed for the US Government and high-value civilian targets such as nuclear facilities, the Smart Grid, and mission critical data centers. Kairos is the world's first compute-appliance capable of addressing streaming security threats to cyber-physical infrastructure.

Kairos is a rack mounted compute-appliance customized to each client's specific location and needs. It taps directly into the backbone of the IT network — localized or cloud-based — and works like a central nervous system for security. It has been successfully used to:

- Reduce the time to capture and analyze 275MB of acoustic data from hours to one-fourteenth of a second
- Analyze real-time data from different types of sensors and 1,024 individual channels to support extended perimeter security
- Analyze structured and unstructured data such as voice, text, network packets, video, email, imagery, and audio — simultaneously.
- Detect in real time, anomalies in data streams moving at thousands of megabytes per second.
- Conduct computationally challenging analytics in milliseconds

This enables a faster and more intelligent response to any threat. As an appliance, it requires just three days to fully install, integrate, and test the core appliance in a new situation.

9

What is Operations Analysis?

It’s using big data technologies to enable a new generation of applications that analyze large volumes of multi-structured, often in-motion, *machine data* and gain insight from it, which in turn improves business results

There are a number of drivers for an Operations Analysis use case. In its raw format, businesses are often unable to leverage machine data, because:

- It is growing at exponential rates
- It comes in large volumes, variety of formats, often in-motion
- It needs to be combined with existing enterprise data
- It requires complex analysis and correlation across different types of data sets
- It requires unique visualization capabilities based on data type and industry/application

Organizations want to leverage machine data to improve business results and decision-making.

10

This slide lists some of the reasons for organizational concern about operational data. When systems go down:

- Sales and other processes stop, and revenue is lost
- Work in progress may be destroyed, and user productivity is destroyed
- Failure to meet service level agreements (SLAs) and contractual obligations can result in damages, fees, adverse publicity, and damage to reputation
- Customers are lost to competitors, some permanently
- Productivity suffers and remediation costs additional \$\$\$'s

11

Cisco is a client that is leveraging multiple big data capabilities to develop an intelligent infrastructure management solution.

Using its intelligent networking capabilities, Cisco launched a Smart+Connected Communities (S+CC) initiative to weave together people, services, community assets and information into a single pervasive solution. There are two initial use cases out of a total of 15 planned solutions:

- Intelligent Infrastructure Management Service (IIMS). An S+CC service that enables centralized monitoring and control of building systems through an integrated user interface while providing real time usage information to optimize building energy resource consumption
- Intelligent Maintenance Management Service (IMMS). An S+CC service that automates preventive and corrective maintenance of building systems and enhances lifetime of the equipment while reducing overall maintenance cost.

In these use cases, the following types of applications are being leveraged:

- Log Analytics
- Energy Bill Forecasting
- Energy consumption optimization
- Detection of anomalous energy usage
- Presence-aware energy management
- Policy management / enforcement

The BigData platform at Cisco has:

- 1) Stream computing software (IBM InfoSphere Streams) that allows user-developed applications to rapidly ingest, analyze and correlate information as it arrives from thousands of real-time sources.
- 2) An Hadoop system (IBM BigInsights) to efficiently manage and analyze big data, digest unstructured data and build environmental and location data. –
- 3) Business analytics software integrated into the platform to generate solution-relevant dashboards and reports to explore data in any combination and over any time period

At Cisco, this platform increased operating efficiency and enhanced service levels. Cisco significantly reduced costs, increased its revenues and improved its competitive position.

12

What is data warehouse augmentation and what are the drivers? Data Warehouse augmentation builds on an existing data warehouse infrastructure, leveraging big data technologies to “augment” its value.

The two main drivers are:

- (1) The Need to leverage variety of data
 - Structured, unstructured, and streaming
 - Low latency requirements (hours not weeks or months)
 - Requires query access to data
- (2) Optimize warehouse infrastructure
 - Warehouse data volumes reaching big data levels
 - Large portion of data in warehouse not accessed frequently
 - Need to optimize warehouse investment

Note: this is not meant to imply that data warehousing solutions are expensive, but instead that augmenting them with big data technologies can make the warehouse a more optimal investment since one no longer has to attempt to store and analyze *everything* within the warehouse, as this can strain it from a performance and cost perspective. Storage in the Hadoop Distributed File System (HDFS) provides a cost effective mechanism for storing less used data.

13

Working with big data is becoming an integral part of the enterprise data strategy at many firms. Indeed, a number of organizations are looking to deploy a software platform such as BigInsights so that they can manage big data from the moment it enters their enterprise. After storing the raw data in BigInsights, firms can manipulate, analyze, and summarize the data to gain new insights as well as feed downstream systems. In this manner, both the original (raw) data and modified forms are accessible for further processing.

One potential deployment approach involves using Hadoop as a source for a data warehouse. IBM’s implementation of Hadoop (InfoSphere BigInsights) can be used sift through large volumes of unstructured or semi-structured data, capturing relevant information that can augment existing corporate data in a warehouse.

The figure illustrates such a scenario, which offers firms the ability to broaden their analytic coverage without creating an undue burden for their existing systems. Once in the warehouse, traditional business intelligence and query/report writing tools can work with the extracted, aggregated, and transformed portions of raw data stored in HDFS.

14

Another potential deployment approach involves using BigInsights as a query-ready archive for a data warehouse. With this approach, as illustrated in the figure, frequently accessed data can be maintained in the warehouse while “cold” or outdated information can be offloaded to HDFS and BigInsights. This allows firms to manage the size of their existing data management platforms while servicing the well-established needs of their existing applications. Offloading rarely queried data to HDFS allows that data to remain accessible to applications that may have an occasional or unpredictable need to work with it.

—

You have now completed this video.