IBM Software
White paper

Information Management

Using IBM Banking Data Warehouse & Big Data to Augment a Data Warehouse

IBM

Big Data and Banking

The ability to use big data is a differentiating factor for enterprises in the banking sector. The adoption of big data technologies has triggered an evolution in the way existing information management challenges are tackled.

IBM Banking Data Warehouse contains content to support the analysis of many different sources of data including transactional, clickstream, call center, social media and unstructured content.

IBM Banking Data Warehouse plays a significant role in addressing big data challenges by providing both big data specific content and by accelerating the creation of the logical data warehouse that encompasses both new and traditional data structures.

The IBM IBV study "Analytics: the real world use of big data"[1] found that organizations are being practical about engaging with big data as they work to understand how it can be of value to their business. Most are educating themselves on the key use cases, defining a big data roadmap or are conducting pilot implementation activities.

This finding is echoed by IBM Industry Model customers who are looking to extend their existing investments in data architecture tools and processes to help harness the opportunities of big data. A common theme is the use of big data to enhance and augment existing business intelligence solutions by increasing the volume and variety of data available for analysis.

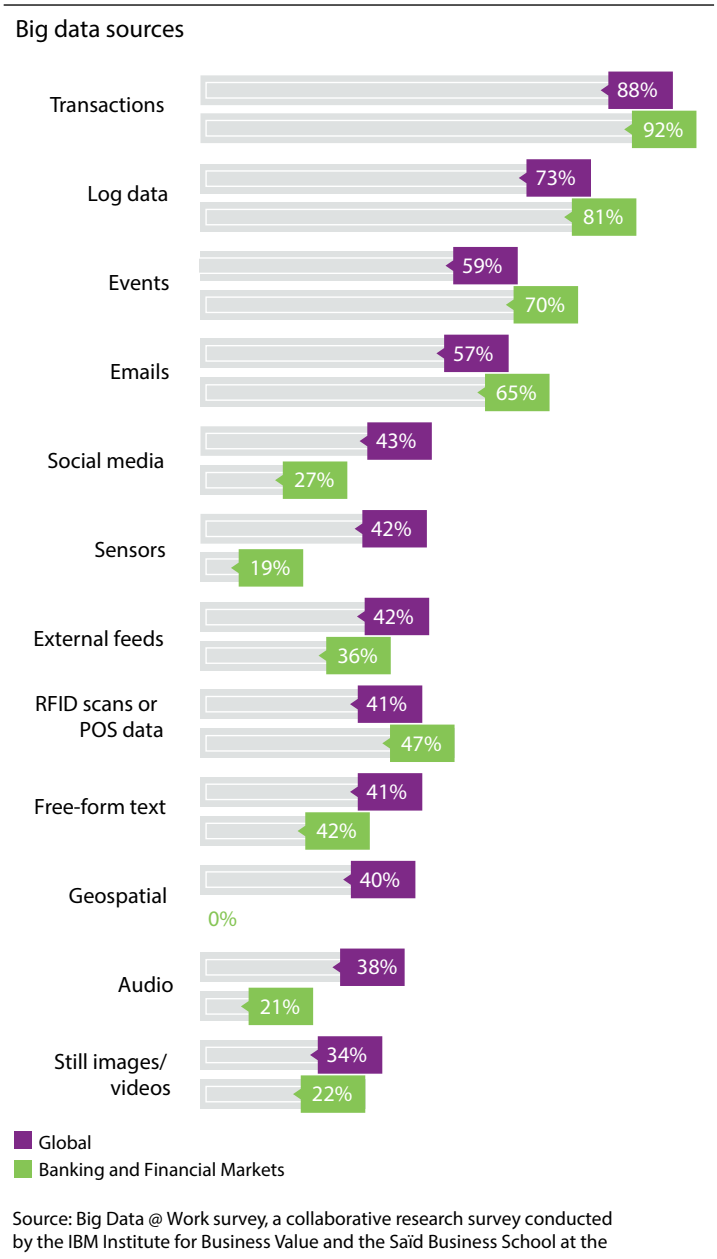

Figure 1. Organizations are mainly using internal sources for big data efforts [1]

Information Management for Big Data and Analytics

There is an emerging challenge for information management that is being encountered by enterprises that are seeking to combine big data and analytics. As enterprise data warehouses are expanded and augmented with a variety of new data such as unstructured, social and transaction event data there is a relentless business demand for data analytics and exploration capabilities that work across all data.

There is a critical need for the analytical results and insights from predictive and explorative analysis to be safely integrated with existing trusted information used by decision makers. For example, while the combination of social and finical transaction data makes it possible to extract insight about the financial institution's customers, suppliers or competitors – care is required to ensure that the data is combined accurately and in accordance with both the institution's and regulator's data and privacy guidelines.

The data models and business vocabularies provided IBM Banking Data Warehouse provide financial institutions with a blueprint for combining, describing and governing big data for analytics.

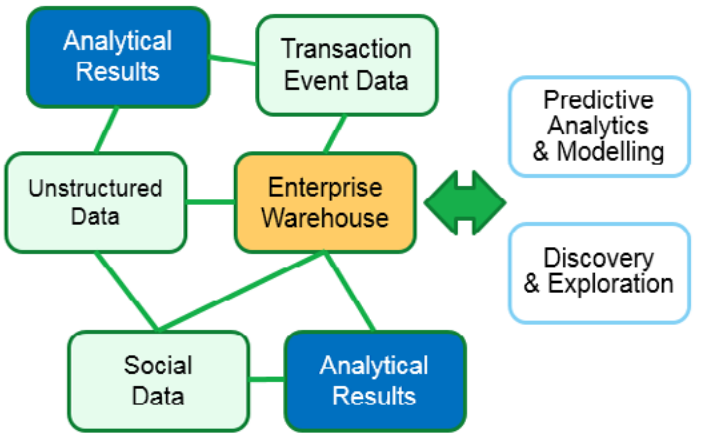

Figure 2. Combination of data

Business and Technical Priorities

The business and technical priorities for big data projects include:

- Enabling business users and analysts to understand what information is held on the individual customer. This facilitates the development of customer-centric analysis and insight such as the 360 degree of the view of the customer. IBM Banking Data Warehouse can be used to bring enterprise-wide consistency to often disparate customer-related data models and business terminologies.
- Developing the combination of business expertise, analytical skills and robust data management to generate business value and insight from big data. IBM Banking Data Warehouse provides a combination of integrated business, analytical and data warehouse models that help business and technical analysts plan the exploitation of data assets.
- Targeting the potential that is locked away in existing internal operational data and data warehouses. Initial efforts can be focused on using big data to gain insights from internal data sources including log data, detailed transactions, email, and content management systems. IBM Banking Data Warehouse provides structures for the storage and analysis of transaction data that builds up over time to form a rich query-able archive.
- Enhancing information governance as the existing challenges of managing and connecting data across the organization become more complex with the addition of new internal and external sources. Data architectures must be able to support a combination of relational database systems, analytical appliances, Hadoop and NoSQL databases. IBM Banking Data Warehouse can be used to deploy a logical data warehouses that span various data technologies.

Using IBM Banking Data Warehouse and Big Data to Transform Banking

IBM Banking Data Warehouse provides a flexible and scalable data warehouse design, enabling organizations to build a comprehensive data warehouse solution through phased development. This allows for rapid delivery of high-business-value deliverables by initially focusing on the business areas offering the greatest returns and feasibility, while building within a proven technical warehousing architecture.

Improving FATCA compliance with unstructured data

The role of IBM Banking Data Warehouse in combining data from the traditional data warehouse and the big data platform is illustrated in the use of unstructured data to support FATCA compliance activities. FATCA requires foreign financial institutions to monitor report and withhold on certain qualifying U.S customers and non-complaint organizations. Analysis of the data stored in the traditional data warehouse might not be sufficient for FATCA identification, classification and due diligence processes. The necessary data might not be stored in the warehouse or might have been captured on application forms many years previously.

To address this, the information stored in the warehouse can be augmented with various unstructured data for example, documents, withholding forms, e-mail logs, images. This provides a more complete, holistic client picture, which in turn reduces the FATCA compliance costs significantly.

Improving Customer Insight with Social and Transaction data

A similar challenge is faced by organizations who seek to augment traditional customer relationship information with data from social, financial transaction and email sources in order to provide a 360-degree view of the customer. Business and technical users need to have a common understanding on the types of data that should be used in customer analytics and how such information can be combined and explored to provide insight.

Creating the logical data warehouse

The types of structured and untrusted data a being introduced by big data projects is different to what typically be stored in the enterprise data warehouse. However, the data warehouse can be aug- mented with various unstructured data for example, documents, with- holding forms, e-mail logs, images. It provided a more complete, holistic client picture, which in turn reduces the costs and effort required for challenges such as FATCA compliance and Customer Insight.

The data available for these activities is managed with a combination of the big data platform (Hadoop) and traditional data warehouse. This combined data resource is called the logical data warehouse. As with all complex databases it is vital that a single data model is used to understand, access and govern. IBM Banking Data Warehouse is ideally suited to this task of data warehouse augmentation because it supports many implementations and can be deployed on a range of platforms.

IBM Banking Data Warehouse provides a technical view that represent a subset of the Atomic Warehouse Model component that can support Customer Insight.

Compliance Use Case

Business context

Many financial institutions today are still struggling with their structured, conventional and traditional information. With only about 20 percent of all the available information today being structured, this leaves about 80 percent of unstructured data that is often being ignored or not fully used. The challenge, and opportunity, for the banking industry is to combine the traditional structured data with various unstructured data. For example, this unstructured data can include: e-mails, documents, customer feedback/complaints, audio, images. Integrating these data sources with traditional data sources is where IBM Banking Data Warehouse can offer value.

IBM Industry Model content

IBM Banking Data Warehouse content enables the storage of non-traditional data sources such as call center notes, teller notes, client e-mails, click stream data, voice recordings. The analysis of these data sources can help identify changing client behaviors and aid in better decision making. Unstructured data sources historically stored on file systems can also now be used in client identification processes.

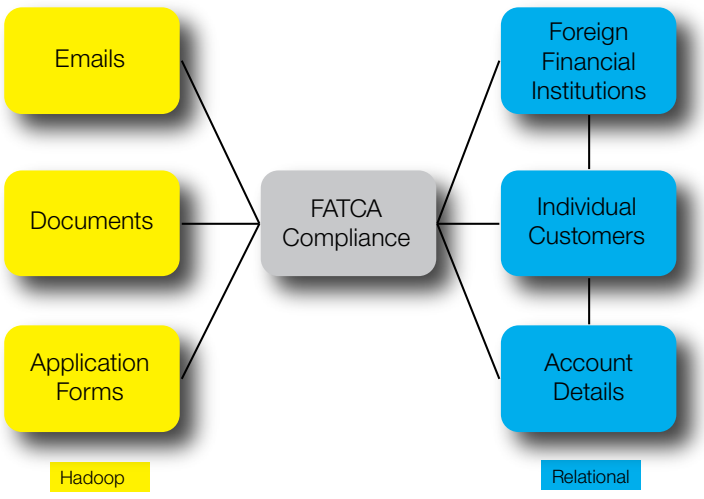

Figure 3. FATCA analysis combining both structured and unstructured data

Supportive Content for FATCA

- Identifies and defines the critical data elements under FATCA legislation
- Comprehensively groups FATCA-specific terms, which includes mappings to both the Business Terms and Atomic Warehouse Model
- Indicates the correspondence between the source terminology and the enterprise terminology as defined in the Business Terms

IBM Banking Data Warehouse covers four key aspects of FATCA compliance:

- U.S. Taxpayer compliance
- Foreign Financial Institution Compliance
- U.S. Withholding Agent Compliance
- Inter Governmental Agreements (IGAs)

Analytical Requirements for FATCA

What is the impact to the Financial Institution?

IBM Banking Data Warehouse supports the following FATCA analysis:

- FATCA Implementation Analysis
- FATCA Compliance Analysis
- FATCA Withholding Analysis

Project Scopes for FATCA

IBM Banking Data Warehouse contains a FATCA scope, which identifies the data elements that are required under FATCA.

Customer and Market Insight Use Case

Business context

Consumers are increasingly turning to social media to engage with their financial institution and to tweet or blog about products and services. This has the capability of damaging or enhancing the reputation of a financial institution and creates an imperative to be active in these social channels. Having a successful social media strategy can also help to attract new customers while retaining loyal and profitable customers.

Social media analytics also makes it possible for a financial institution to extract intelligence from the vast amount of data that is posted on social media platforms. This might include what is being said about the financial institution's customers, suppliers or competitors. The goal of this analysis is to derive business value through better decision making.

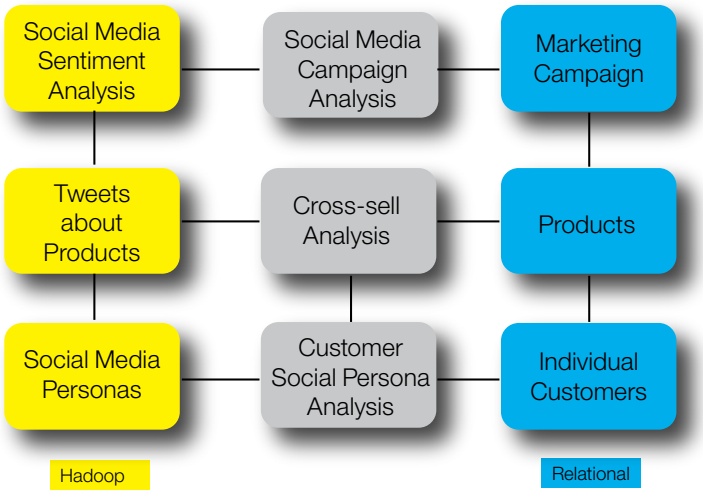

Figure 4. Customer and marketing insight

A financial institutions use of social media has possible legal and compliance considerations that need to be understood. The model implementation is independent of jurisdiction and so any specific guidance is out of the scope of this paper. For more information on guidance specific to the United States jurisdiction the Federal Financial Institutions Examination Council (FFIEC) has issued proposed guidance.

IBM Industry Model content

IBM Banking Data Warehouse contains support for social media through the requirements, analysis and design models. The data models allow a financial institution to harness social media data into a form that allows for real business insight to be derived.

IBM Banking Data Warehouse social media content can have various different business applications. For example, a financial institution might want to track social media sentiment to measure the impact of new product launches and marketing campaigns. They might also want to calculate various social media metrics such as share of voice and reach or distinguish between positive, negative and neutral postings on the company's page. A financial institution might also want to understand what trends and patterns are emerging so better business decisions can be made.

The model content also allows a financial institution to make the link between a customer and a social media persona, bringing together the traditional sources of data with the new social media data that's available. This extended view of the customer could be used to identify any concerns early that might affect a customer's creditworthiness. Social persona analysis also allows new relationships to be uncovered and identifies possible cross selling opportunities linked to life events uncovered though social media.

Gain insight and analyze consumer attitudes toward your financial institution's brand, products and services

Measure the effectiveness of campaigns by assessing consumer sentiment expressed in social media channels

Glean more information about customers at the individual level by analyzing social media combined with traditional data

Transactional Data Volumes Use Case

Business context

Recent banking regulations require financial institutions to deal with larger volumes and scaling calculations relating to transactional data. The Dodd Frank Wall Street Reform and Consumer Protection Act is an example of how new regulatory requirements are creating a need for traditional transactional data to be stored in greater volume and reported on more frequently. There is also a need to report risk measures at an enterprise level and not just account level in order to identify systemic risk.

Using big data technologies with IBM Banking Data Warehouse model content can help financial institutions meet these new enterprise-level, close-to-real-time financial reporting requirements.

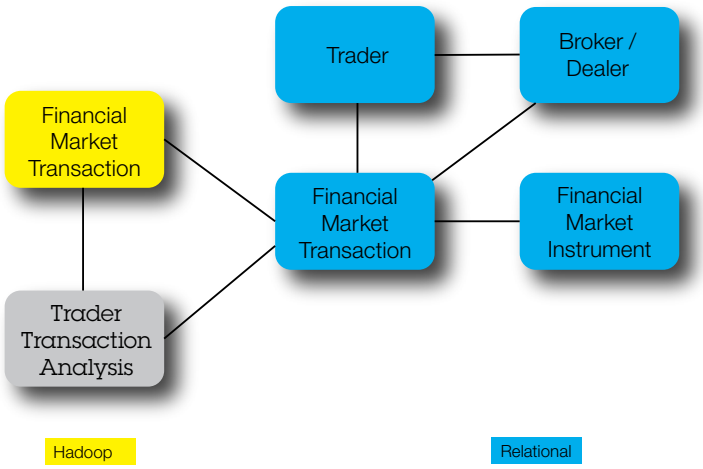

Figure 5. Trader transaction analysis

IBM Industry Model content

In addition to summarized and aggregated data, IBM Banking Data Warehouse also contains atomic transactional data. This data is traditionally stored for a limited time period because of the cost of storing these large data volumes in the data warehouse. However, with the emergence of big data technologies such as Hadoop it has become more cost effective to store larger volumes of transactional data for a longer time period.

IBM Banking Data Warehouse content contains support for banking regulation throughout the different models, including, Dodd-Frank, FATCA and Basel. Using the data models with big data can help reduce the time and cost to support these evolving financial regulations.

For example, to comply with the Dodd-Frank Large Trader Reporting rule, a significant volume of transactional data is captured as part of record keeping and reporting requirements. IBM Banking Data Warehouse provides support in this area and it might be more cost effective to use it to offload some of the transactional data to a Hadoop environment.

IBM Banking Data Warehouse and the Big Data Platform

Deploying the logical data warehouse to the big data platform

The logical data warehouse can be deployed to an integrated architecture composed of a combination of the big data platform components.

IBM Banking Data Warehouse can be deployed to technologies and data appliances such as BigInsights BigSQL, DB2® with BLU Acceleration and IBM PureData™ powered by Netezza, which supports high performance for complex analytic workloads. Big SQL supports federation to many data sources, which allows users to send distributed requests to multiple data sources within a single SQL statement.

IBM InfoSphere® Streams enables continuous analysis of massive volumes of streaming data. The logical data warehouse model can be used to impose a consistent structure on the data as it used for real-time analysis and business decisions.

IBM InfoSphere BigInsights is an enterprise-ready Apache Hadoop-based solution for managing and analyzing massive volumes of structured and unstructured data. Databases can be implemented using IBM Banking Data Warehouse and Hadoop Hive and the data can then be queried by data professionals using BigSQL or analyzed using BigSheets.

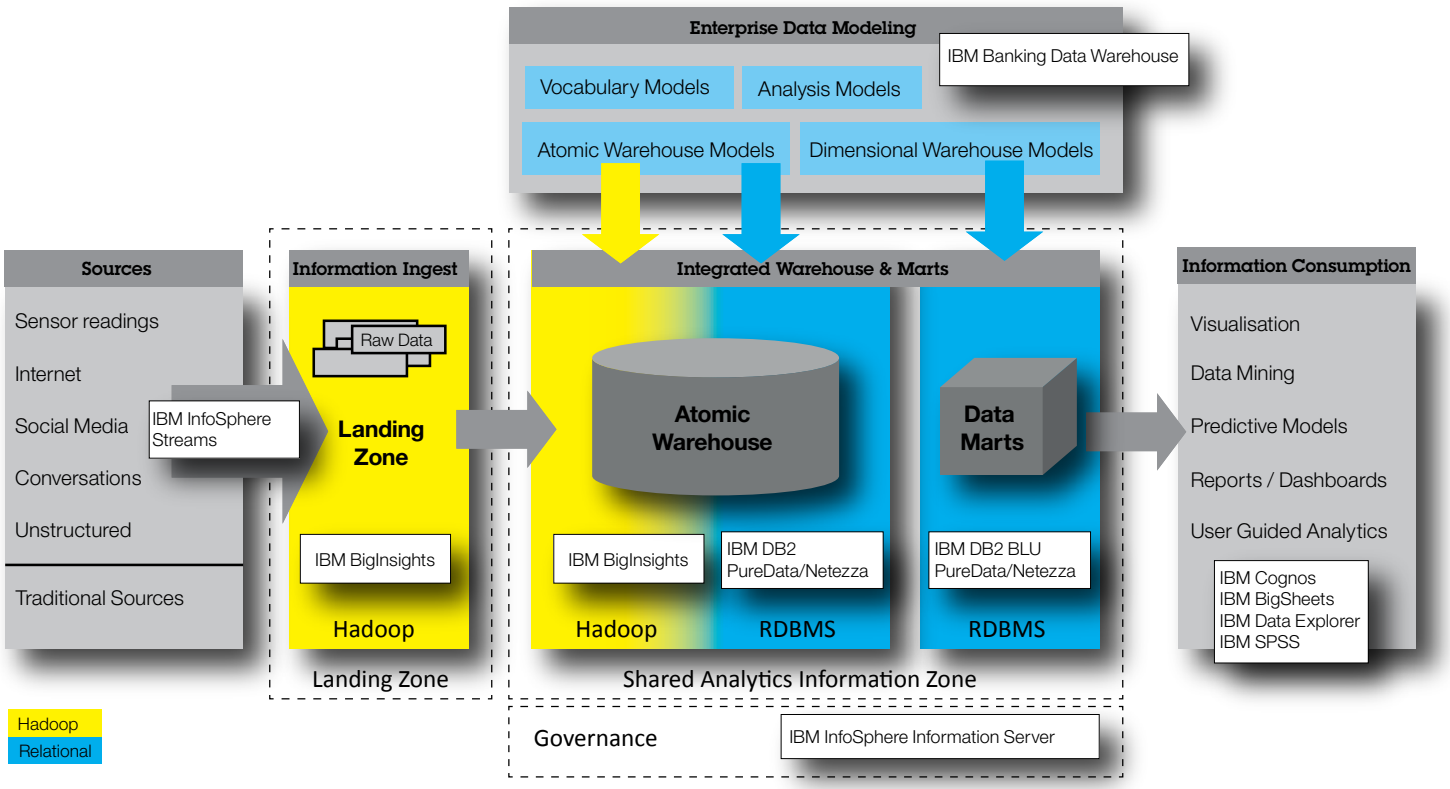

Figure 6. How data models are deployed using the big data platform

IBM DB2, IBM DB2 with BLU Acceleration, PureData (Netezza) IBM DB2 database software offers industry leading performance, scale, and reliability on your choice of platform from Linux®, UNIX® and Windows® to z/OS®. IBM DB2 with BLU Acceleration speeds analytics and reporting using dynamic in-memory columnar technologies. IBM PureData System for Analytics is a purpose-built, standards based data warehouse appliance that integrates database, server, storage and advanced analytic capabilities into a single system.

IBM InfoSphere Information Server can deploy business terms into metadata repository where they are used to understand, govern and deliver trusted information to business users.

IBM Business Analytics has created a number of solutions that address the unique needs of banking providers. Whether it's streamlining operations, improving the customer experience or identifying new opportunities and markets - IBM has the analytic capabilities you need to drive better outcomes. For example InfoSphere Data Explorer discovery and data virtualization can be used for real-time access and fusion of data from across the logical data warehouse and enterprise applications and analysts can use modeling and statistical tools such as IBM SPSS® to create predictive models using data from both Hadoop and the RDBMS.

Managing new data governance challenges on big data projects

The businesses value of analytics, based on an increased volume and variety of data, can be quickly undermined if confidence in the veracity of information is lost. The use of a single data model that spans all the data being used in the project can help avoid or mitigate these risks. The data governance challenges encountered on big data projects are illustrated in the interaction between the key roles of business user, data scientist and data architect.

Data architects are focused on the day to day challenge of ensuring that the Business Users have the correct data to make business decisions. They appreciate the need to for a common model that documents the data architecture. Data architects need to work with data scientists to evolve the methodology by which data risks are identified, avoided, accepted or mitigated. The use of a common vocabulary and related model helps the understanding of how data is being used and if it is appropriate to do so. For example, is it always appropriate to use social media data in all business decisions? Data architects:

- Work within enterprise architecture
- Focus on supporting day-to-day business operations
- Use data models to manage data
- Maintain enterprise data models to support many and varied users and applications

Business users trust that data architects, who are responsible for ensuring existing policies for information collection, use and security will also apply these policies to big data. Business users need a common language for both business and data terms to that they can clearly specify the information they want data scientists to provide. The link between the business vocabularies and data models ensures that business users and technical analysts are aware of how data fits together. Business users:

- Work within a line of business
- Focus on delivering business results
- Use business models to understand data
- Specify requirements in functional terms rather than data

Data scientists work with business users to create value though the development of analytics and business models. They expect that this is an innovative and creative process that reacts to the rapidly changing business conditions. Data scientists:

- Work on projects with varied analytical technologies
- Focus on discovering novel business value in varied data sources
- Develop solution specific data models/structures
- Develop models which are optimized for the solution function and technology

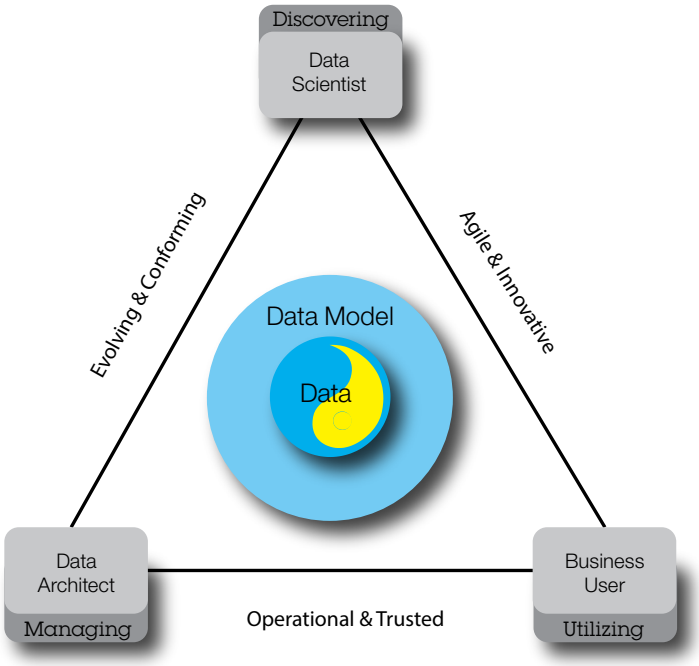

Figure 7. The different actors in a big data project have different governance agendas

Determining which data is deployed in Hadoop or RDBMS

The choice of which parts of a logical data warehouse model are deployed onto which component of a physical architecture must take account of a number of factors

Data volumes and archiving

It is economical to persist large data volumes on Hadoop that are currently archived to tape. Historical medical device transaction data that builds up to a massive volume over time can be stored in Hadoop Hive to provide a deep query-able archive. Using IBM Banking Data Warehouse the structure of a Hive table would be the same as that of most recent transaction stored in the RDBMS. While the latency of queries on Hadoop might be higher than that of the RDBMS, the data is more accessible than if it was archived to tape.

Data access

Data persisted on Hadoop facilitates massive volume data analysis and aggregation tasks using map-reduce. High performance queries of individual rows are still best deployed on RDBMS.

Data management

Simple structures in Hadoop such as HIVE are quick to implement allowing the rapid deployment of data for analysis. The use of IBM Banking Data Warehouse means that even though the data is rapidly available, it conforms to the enterprise data model.

Conclusion

The banking industry is facing an ever increasing volume and diversity of data. This presents both a technical challenge and an enormous opportunity for improvements in the operational efficiency and the development of new products.

IBM Banking Data Warehouse Model can help address the technical challenges by supporting the deployment of a logical data warehouse architecture that encompasses both new and traditional data structures and deploys across multiple technology platforms.

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
November 2014

[1] Schroeck, Michael; Rebecca Shockley, Dr. Janet Smart, Professor Dolores Romero-Morales and Professor Peter Tufano. "Analytics: The real-world use of Big Databig data. How innovative organizations are extracting value from uncertain data." IBM Institute for Business Value in collaborations with the Saïd Business School, University of Oxford, October 2012. http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-big-data-at-work.html. ©2012 IBM.

Please Recycle

IMW14749-USEN-01