IBM Software
White paper

Information Management

Using IBM Unified Data Model for Healthcare & Big Data to Augment a Data Warehouse

IBM

Big Data and Healthcare

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

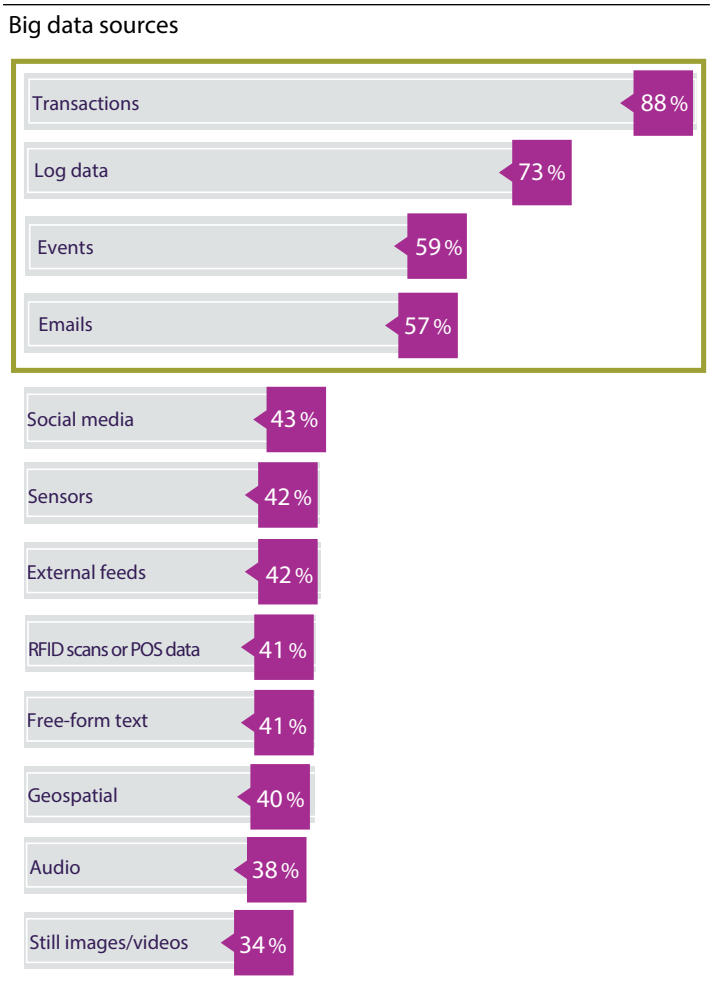
The growth in availability of information sources, such as medical device data and text analysis of written medical records, allows for the creation of longitudinal patient records which improve efficiencies and support patient-centric care. This same information is aggregated and combined with external sources such as disease registries and can support clinical studies by providing data on large numbers of patients.

IBM Unified Data Model for Healthcare plays a significant role in addressing these 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.

Developing a Big Data Roadmap

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.



Respondents with active big data efforts were asked which data sources they currently collect and analyze. Each data point was collected independently. Total respondents for each data point range from 557 to 867.

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

The business and technical priorities for big data projects include:

Patient centered analytics
Enabling business users and analysts to understand what information is held on a patient. This enables the development of patient centric analysis and insights such as the creation of the longitudinal patient record. IBM Unified Data Model for Healthcare can be used to bring enterprise-wide consistency to often disparate patient-related data models and business terminologies.

Developing the skills to use big data
Having the combination of business expertise, analytical skills and robust data management to generate business value and insight from big data. IBM Unified Data Model for Healthcare provides a combination of integrated business, analytical and data warehouse models that help business and technical analysts plan the usage of data assets.

Unlocking the value in existing data
Targeting the potential that is locked away in existing internal operational data stores. 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 Unified Data Model for Healthcare has structures for the storage and analysis of detailed records such as those generated by medical devices.

Enhancing information governance
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 Unified Data Model for Healthcare can be used to deploy a logical data warehouse that spans various data technologies.

Using IBM Unified Data Model for Healthcare and Big Data to Transform Healthcare

IBM Unified Data Model for Healthcare 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 patient care with the longitudinal patient record

The longitudinal patient record (LPR) allows healthcare providers make informed decisions about a patient's treatment plan by providing a complete picture of the patient's health status. Healthcare providers no longer have to spend significant amounts of time trying to piece together a picture of the patient's health status; the information is readily available in the LPR.

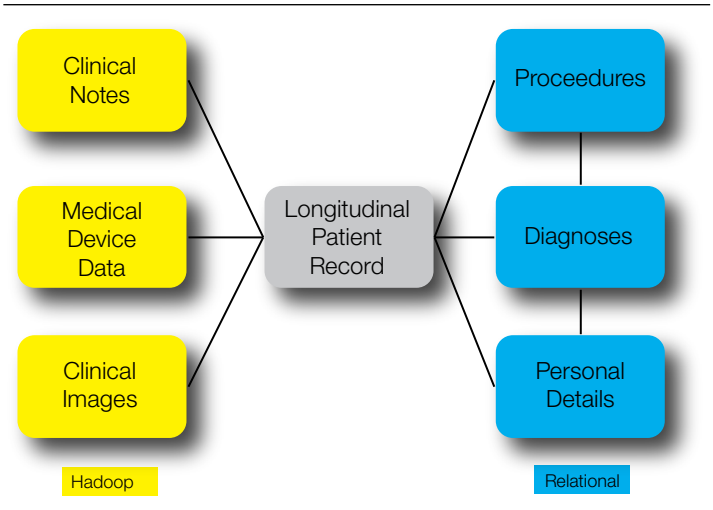


Figure 2. An LPR data model that includes both Hadoop and relational data

Clinicians are working with data scientists to develop one complete patient medical record to bring together large volumes and varieties of structured and unstructured data from different healthcare providers, for example, hospitals, family doctor practice. The most important benefit for the patient is continuity of care, for example healthcare providers in the community can access the patient's data captured during a hospital visit.

Creating the logical data warehouse

The data that makes up the LPR is managed using a combination of the big data platform consisting of both Hadoop and traditional data warehouse structures. 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 Unified Data Model for Healthcare is ideally suited to this task of data warehouse augmentation as it supports many implementations and can be deployed on a range of platforms.

Longitudinal Patient Records and the Healthcare Providers Mission

Healthcare providers are being asked to improve patient care and outcomes while reducing costs. The use of longitudinal patient records can help achieve both of these objectives.

The healthcare industry is under competitive and legislative pressure to reduce the cost of care, efficiently manage resources and improve patient care. At the same time, healthcare organizations are experiencing an explosion in the volume and variety of data relating to patients and their treatment. A considerable proportion of this data is unstructured and not integrated with the structured data in the traditional data warehouse. Existing analytics and patient records are often limited to this structured data.

Improving Patient Care and Outcomes

The LPR allows the healthcare provider access a patient's complete healthcare record, that is, all of the healthcare information captured in the primary, secondary, tertiary, emergency, home and community care settings.

Having access to a patient's complete healthcare record means that a healthcare provider has a complete picture of the patient's health status and can interpret the patient's current complaints in perspective of their known conditions and lifestyle. Often, a healthcare practitioner is dependent on the patient giving an accurate medical history, which can be time consuming and more difficult if a patient is unwell. The LPR removes this dependency by providing the patient's up-to-date health record. The healthcare practitioner is empowered to quickly create the most appropriate treatment plan for the patient to promote a better outcome.

Reducing Costs

Having access to a patient's LPR reduces the amount of unnecessary tests, procedures and more importantly patient discomfort. Previous test and procedure results can be accessed in the longitudinal record, which allows the healthcare provider request the most appropriate investigations to support the patient's treatment plan. This way, the patient receives the best care without the inconvenience and discomfort of unnecessary tests and the associated delays and costs.

Longitudinal Patient Record Use Case

Patient Encounter without LPR

John presents to a hospital Emergency Department (ED) complaining of a shortness of breath and chest pain. The ED medical providers discuss John's symptoms with him. When did they commence? Has he had any changes in diet or physical activities? Does he have any chronic conditions or allergies?

John is feeling distressed and is confused but gives the medical staff as much information as he can think of. The team conduct a physical examination and check his vital signs. Registration details are recorded including Name, Address, Age, Gender and his insurance details. If John had visited an ED before the staff might have access to these healthcare records but John has only recently moved to the area. Based on the interview and examination, the ED medical providers reach a diagnosis, order further tests, prescribe treatment including medication and decide to admit John for an inpatient stay.

This is a typical scenario of a patient encounter and is often satisfactory but it has involved the medical team making a diagnosis based on limited information. Some of this was provided by John who is clearly unwell and confused about the details.

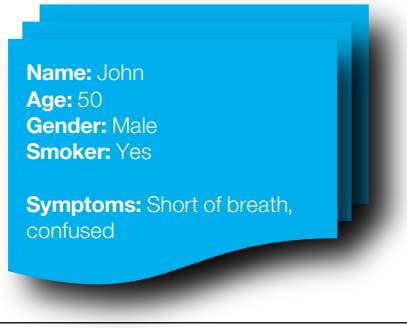


Figure 3. Patient encounter without LPR

Patient Encounter with an LPR

If the same scenario arose in a facility where access to John's LPR was available, the medical providers would have a comprehensive picture of John's health. They would be in a much better position to quickly make an accurate diagnosis and create the most appropriate treatment plan.

The medical personnel discuss John's symptoms with him, as before, and conduct a physical examination but when they collect the registration details they are alerted to information held in John's LPR. For example:

- History of the John's known medical conditions
- Details of any allergies he might have
- Previous medical procedures including surgeries
- Lab test results
- Imaging results
- Doctors notes and other observations
- Long-term medications that he is taking or is supposed to be taking
- His family medical history
- His social history including diet, exercise and smoking status
- His occupation and living arrangements

Now the doctor notices that John was referred by his family doctor to a healthcare facility for a full cardiac assessment seven years ago. This took place in another state where John lived before relocating here only six months ago. The referral notes from John's doctor indicate that he has Atrial Fibrillation (AF) and at the time he believed that John's increasing level of work related stress was a contributing factor as was his diet and irregular eating pattern. John was prescribed a low dose of a beta-blocker to slow his heart rate and encouraged to join a healthy lifestyle program as well as attending smoking cessation counseling. John's medication prescription and dispensing records indicate that he continued on this medication until six months ago and then appears to have stopped taking it.

Figure 4. Patient encounter with LPR

This new vital information completes many gaps in the story John told in his confused state and it helps the doctor understand the likely cause of his problem and decide the most appropriate treatment plan.

The ED medical staff are confident John's AF has not been managed correctly for the last six months because he has not been taking his medication. As a result, his AF is causing his chest pain and he is quickly commenced on IV medications to control the AF, which in turn reduces his chest pain.

All of the types of data mentioned in John's LPR example are supported in IBM Unified Data Model for Healthcare as are many more.

Using Big Data to Create a Longitudinal Patient Record

The purpose of the LPR is to provide a complete patient medical record by combining all of the structured and unstructured data from different healthcare providers, for example, hospitals, family doctor practice, into one record. One of the most important benefits for the patient is the all important continuity of care. Healthcare providers in the community can access the patient's data captured in the ED directly instead of having to depend on the patient to retell their ED story. The healthcare provider can establish what happened during the patient's visit to the ED themselves.

As mentioned earlier, the LPR is the amalgamation of structured and unstructured data about the patient. Areas like the following are including in IBM Unified Data Model for Healthcare:

- Demographic, personal and administrative information for example, next of kin, insurance details
- Visit history
- Allergies
- Medical history
- Family medical history
- Social history including living arrangements and supports
- Physical exam reports
- Functional assessments
- Vital signs, personal care device data and output from hospital point of care devices for example, blood sugar machines, blood gas machines.
- Chronic conditions and comorbidities.
- Previous diagnoses, treatments and outcomes
- Previous invasive/non-invasive procedures and outcomes
- Current medications and details of previous episodic prescriptions
- Laboratory results and reports
- Diet, smoking status, participation in physical activity and participation in wellness programs
- Provider clinical notes for example, nursing or physician clinical notes.
- Referral letters
- Discharge instructions and letters
- Medical images and reports
- Pathology reports
- Genomic data
- Ancillary reports for example, Physiotherapy reports
- Home health reports

Combining all of this information creates a more comprehensive view of the patient as a person including their lifestyle and any trends in their health rather than just seeing a collection of symptoms and a diagnosis to be treated as a one off event.

Much of this information lies outside of what was traditionally considered a person's electronic medical record for multiple reasons including the informal or unstructured nature of some of the information as well as the lack of a suitable and cost effective IT infrastructure to store and retrieve it. A modern healthcare data warehouse based on IBM Unified Data Model for Healthcare, which can be deployed across multiple technology platforms suited to these different data formats, can overcome the limitations that existed previously.

Making Unstructured Medical Data Accessible

Much vital healthcare-related data is still being captured in an unstructured forma on paper or electronically.

Clinical Notes

Documentation such as the patient's medical history, discharge reports, provider clinical notes and exam reports can be found in multiple formats such as scanned handwritten notes, free text electronic documents or email. It is crucial that healthcare providers have access to this unstructured data because it is frequently this information that helps a healthcare provider make a diagnosis and create or modify a treatment plan. Often this type of data can include concerns or suspicions that were captured by a healthcare provider as part of their clinical assessment and might be relevant to the patient's developing condition. Apart from providing access to the documents themselves, natural language processing solutions such as IBM LanguageWare® provide the capability to identify and extract key information from the unstructured data and suggest specific clinical observations to the users. IBM Unified Data Model for Healthcare includes structures to support the storage of the raw unstructured documents and to land the extracted clinical data content.

Clinical Images

Copies of images such as x-rays, MRIs and CTs and their reports from other facilities or departments do not need to be requested when a patient presents. These electronic images taken elsewhere can be stored and accessed in the LPR. Scanned hardcopy images converted to electronic images are also accessible.

Making full use of Detailed Medical Device Data

Medical devices generate a massive volume of data that is not currently being used. A patient might be using various medical devices in the home such as glucometers, nebulizers, heart monitors and home dialysis. These devices generate large volumes of data that might, when reviewed, have a bearing on the patient's current condition.

Often this information is only used for alerts or short-term tests and observations but having the ability to retain it in volume can provide further insights to a person's general condition and any developing health issues. For example, a community-based healthcare provider could trend a home-based patient's blood sugar levels and note that the levels are slowly increasing over the last three days but are still normal. Early intervention could prevent the blood sugar levels from becoming abnormal. This type of interaction with the LPR promotes a better outcome for the patient and can help to keep a patient's healthcare costs at a minimum by avoiding a hospital admission.

While in hospital care, a patient is regularly connected to point of care devices like a cardiac monitor, vital signs machine, infusion pump, ventilator and a central line monitor. Once again these devices can generate enormous volumes of data that is relevant to the patient's care and to the assessment of their condition. Monitors might be programmed to alert healthcare professionals when a particular reading exceeds a preset threshold but the full data set can be stored and analyzed to identify trends and patterns in the patient's condition, for example, trend the patient's temperature against the antibiotic administration.

This device output data is so great in volume that it was routinely discarded apart from specific alerts and some snapshot measurements but with new big data technologies the raw device output can be processed by data stream techniques in real or near real time while the data itself can be retained on low-cost mass storage platforms.

Governance, Collaboration and Data Exchange

In an ideal world our patient attends one healthcare facility for all tests and treatments making the collection and storage of medical information relatively simple. In the real world, however, a person has a Primary Care Provider (PCP) to consult for routine health matters. The PCP might refer them to a number of other healthcare providers in various healthcare facilities for specialist care. The patient might also find himself or herself in an emergency department in a location far from their home being treated for a sudden exacerbation of a known condition or for an accidental injury. In order to build an LPR, healthcare providers must have the ability to pool the knowledge they have and so create a complete picture of the patient as a person. Only then can they be sure that they are using the latest and most complete information to assess, diagnose and treat the patient appropriately. This improves the care and outcome for the patient while saving unnecessary expense to the provider and the eventual payer, whether it is a health plan, the state or the patient themselves.

A person's medical data must be protected and so security of storage and access must be tightly controlled. Patient confidentiality rules impose strict limits on how such information can be shared. It is, however, in the patient's own interest that medical practitioners have access to their full medical history to ensure the best care possible. Access could be provided and limited by the role a practitioner plays in the care of an individual. Hospital groups and healthcare alliances might be in a position to build collaborative data stores for LPRs to ensure that their medical practitioners are given the information they need to treat the patients quickly and appropriately.

Data security is supported in IBM Unified Data Model for Healthcare using a role-based access approach.

Looking Beyond the Individual Patient

LPRs have a value beyond the treatment of the individual patient. In the field of medical research, the more information available about patients with particular conditions or on particular treatment plans the better the quality of their conclusions. By being able to sample extensive data about large numbers of patients, a clinical study is better able to filter out any factors that would lead to false conclusions. The LPRs in this case can be de-identified and aggregated as required to protect the identities of individuals without compromising the needs of the researchers.

A Cancer Registry is a central repository of cancer data which is usually collected by Cancer Registrars. The data represents a summary of a patient's demographic information, diagnosis, tumour and treatment. Departments of Health typically mandate the collection of cancer cases so that the collected data can be used to observe cancer trends and provide a research base for studies into the possible causes of cancer and treatment of cancer.

IBM Unified Data Model for Healthcare includes business subject areas for Clinical Trials and Studies as well as supporting the concept of patient cohorts that can be built by grouping patients according to many factors including medical conditions, medications or their inclusion on a patient registry. It also includes a worked example that demonstrates how to scope a section of a typical US State Cancer Registry.

IBM Unified Data Model for Healthcare 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 Unified Data Model for Healthcare 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. BigSQL 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 Unified Data Model for Healthcare and Hadoop Hive and the data can then be queried by data professionals using BigSQL or analyzed using BigSheets.

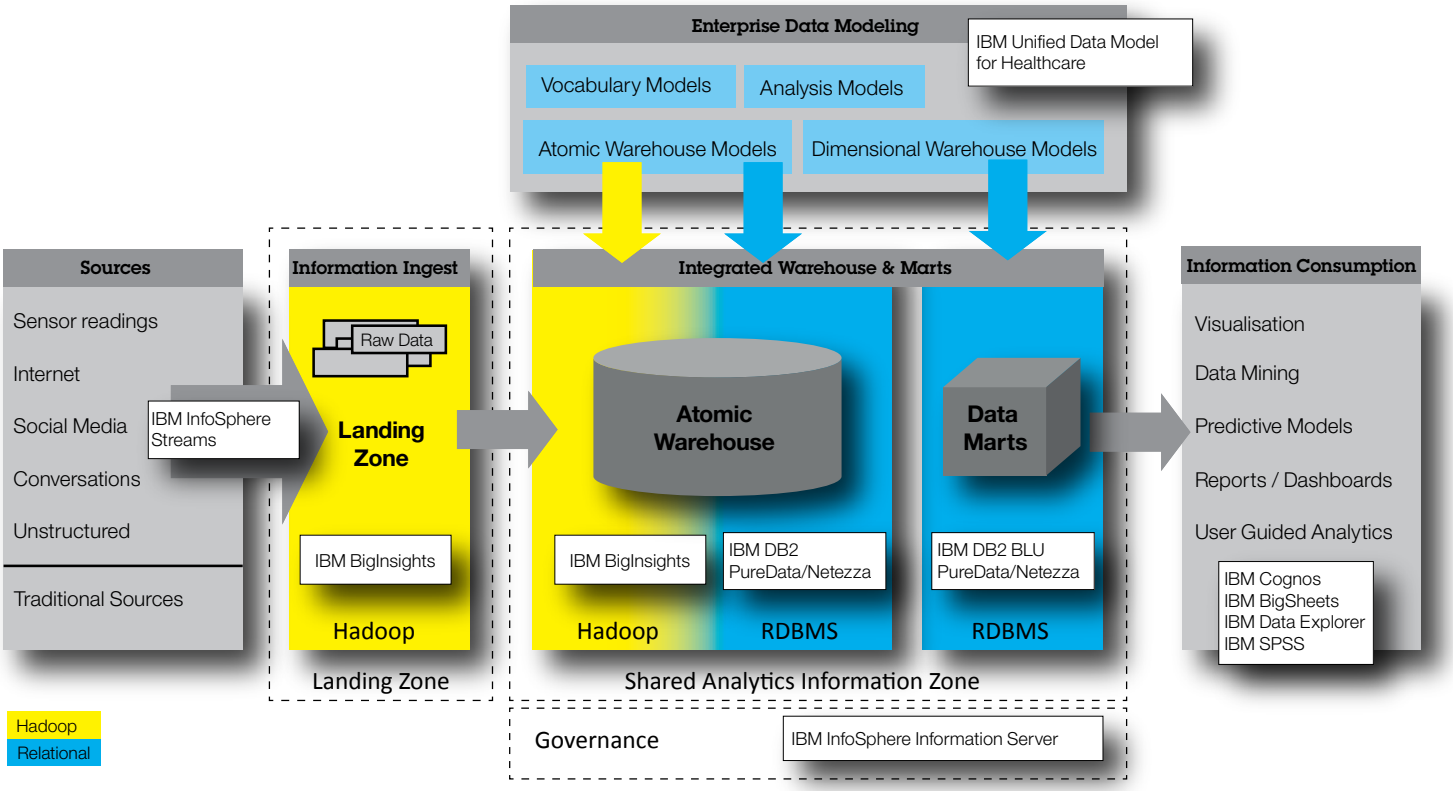


Figure 5. How data models are deployed using the big data platform

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 (Netezza) 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
Business terms are deployed into metadata repository of 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 healthcare 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

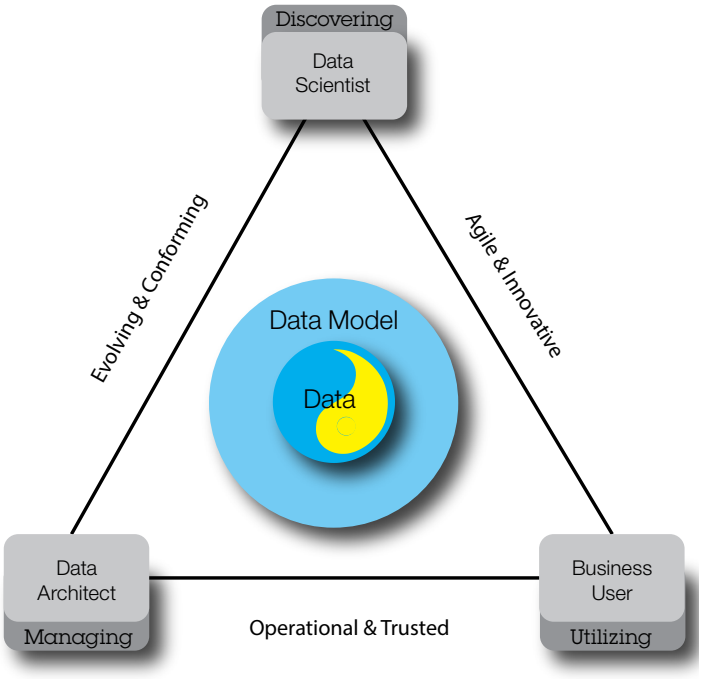


Figure 6. The different actors in a big data project have different governance agendas

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

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 Unified Data Model for Healthcare 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 Unified Data Model for Healthcare means that even though the data is rapidly available, it conforms to the enterprise data model.

Conclusion

Healthcare providers are facing an ever increasing volume and diversity of data. This presents both a technical challenge and an enormous opportunity for improvements in the delivery of patient care.

The Longitudinal Patient Record (LPR) is an example of how IBM Unified Data Model for Healthcare can help healthcare providers take advantage of the opportunity presented by big data by providing a complete picture of the patient's health status and so ensuring that the best care is delivered in the most efficient manner.

IBM Unified Data Model for Healthcare can help address the technical challenges by supporting the deployment of a 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

Please Recycle

IMW14747-USEN-01

[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.