IBM Software
White paper

Information Management

Using IBM Insurance Information Warehouse & Big Data to Augment a Data Warehouse

IBM

Big Data and Insurance

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

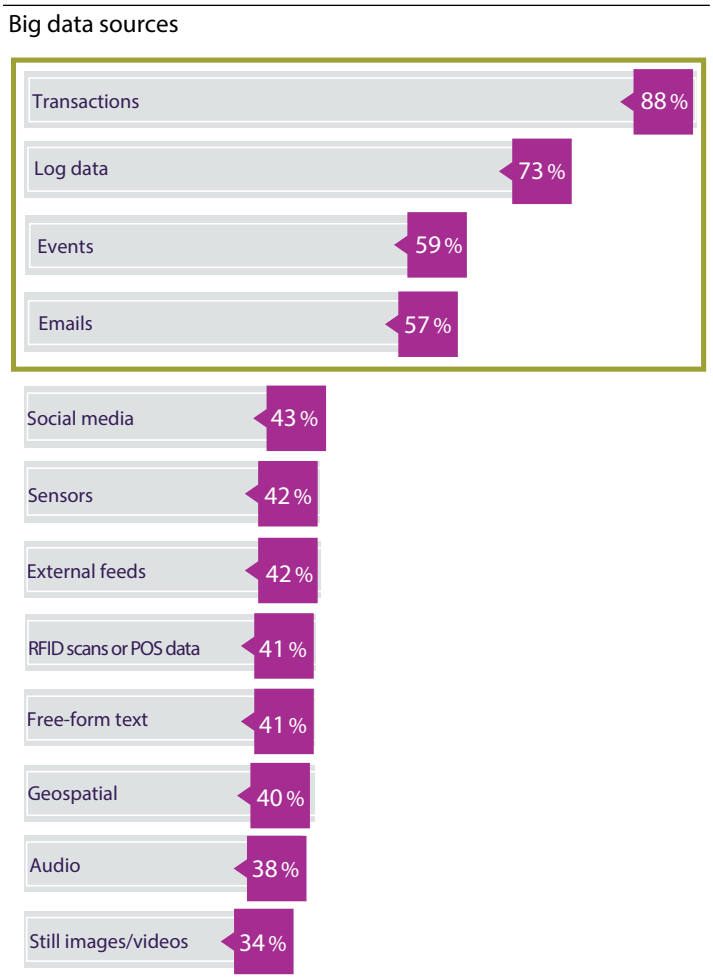
The growth in availability of new information sources such as sensor data (for example vehicle telematics) and social media is providing new opportunities for insurers to gain new insights. Text analysis unlocks the hidden meaning and context found in insurance documentation and communication (email, web, voice). This same information when aggregated and combined with external sources such as fraudster registries, shared information databases, and geo spatial data supports new levels of customer insight and transform how insurers understand and serve their customers.

IBM Insurance Information Warehouse 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.

Insurers can now use the comprehensive traditional data warehouse coverage in IBM Insurance Information Warehouse alongside new information sources for big data to develop new customer insights, support new information needs, innovate new products and manage ever increasing data volumes.

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]

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, geospatial, weather and claims transaction event data there is a relentless business demand for data analytics and exploration capabilities that work across all data.

There is 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 publicly available geospatial and claims transaction data makes it possible to extract insight about an insurers customers, third parties 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 Insurance Information Warehouse provide finical institutions with a blueprint for combining, describing and governing big data for analytics.

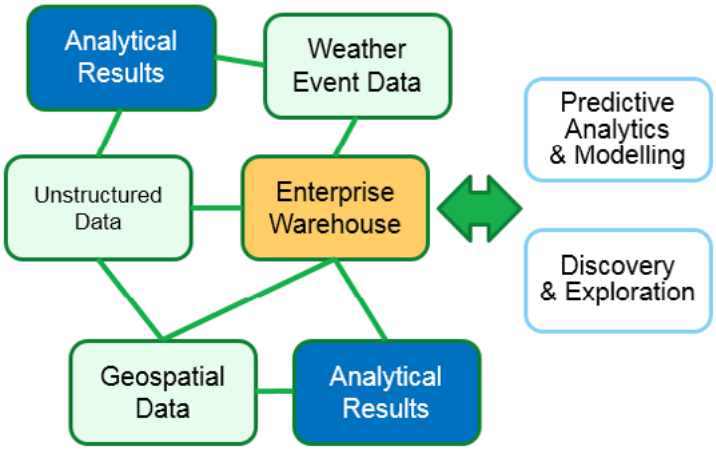


Figure 2. Combination of data

Business and Technical Priorities

The business and technical priorities for big data projects include:

Customer centered analytics

Understanding what information you hold on a customer enables the development of a 360 degree view of the customer. IBM Insurance Information Warehouse is used as a reference point for the development of an enterprise catalogue of often disparate data models and business vocabularies.

Developing strong analytical capabilities

The ability to generate business value and insight from big data requires the fusion of business expertise, data analysis skills and robust data management. IBM Insurance Information Warehouse provides a combination of integrated business, analytical and data warehouse models that help business and technical analysts plan the exploitation of data assets.

Unlocking the value in existing data

There is untapped potential locked away in both internal operations and business intelligence systems. Initial efforts are focused on using big data to gain insights from internal data sources, both new and existing. IBM Insurance Information Warehouse provides structure for the storage and analysis of transaction data such as claims and website logs.

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 (RDBMS), analytical appliances, Hadoop and NoSQL databases. IBM Insurance Information Warehouse can be used to deploy a logical data warehouse that spans various data technologies.

Using IBM Insurance Information Warehouse and Big Data to Transform Insurance

IBM Insurance Information 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.

Creating new insurance products with vehicle telematics data

The potential of big data to transform the insurance product landscape can be illustrated with a vehicle telematics and usage-based insurance use case.

Analysis of data such as customer demographics, vehicle attributes and claims history stored in the traditional data warehouse provides a basis for underwriting of vehicle insurance policies. Business users and data scientists are curious to explore how this existing analysis could be combined with information from big data sources such as vehicle telematics. This approach is illustrative of the type of prototyping that is being undertaken by early adopters.

Telematics data provides second-by-second information on vehicles and how they are being driven. Within Hadoop, the information could remain unstructured but also integrated with more structured modeled information. This can be analyzed to provide a detailed picture of how, when and where a vehicle is being driven, by aggregating the data and deciphering the driver patterns and driving rating. Ultimately this driver scoring is be used to determine the usage-based premium.

Creating the logical data warehouse

The data required for usage-based insurance requires data architectures that span 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 Insurance Information Warehouse is ideally suited to this task of data warehouse augmentation as it is implementation neutral and can be deployed on a range of platforms.

Business Example Diagrams

IBM Insurance Information Warehouse includes new big data business overview diagrams and associated entity-relationship scenario diagrams covering three uses cases: Customer Analytics and Insight, Vehicle Telematics, and Catastrophe Modelling.

The business overview diagrams provide a design-level business view that enables users to get started with Big Data. The view shows how organizations can augment traditional data with data from new Big Data sources, and allows users to visualize the main data concepts being addressed and shows how data items from the use case align with the overall model structure

The entity-relationship scenario diagrams provide a technical view and represent a subset of the model that can support a specific scenario described in the documentation. The diagrams feature a color-coded background showing which entities are suggested Apache Hadoop or similar deployments.

Vehicle Telematics Use Case

Business context

Telematics insurance is a disruptive product innovation offering early adopters an opportunity to establish/increase market share. The growth in consumer willingness to purchase more equitable motor insurance products, based on driver behavior could see 'good risk' drivers transition over to new vehicle telematics insurance products. IBM Insurance Information Warehouse helps insurers overcome the following challenges associated with developing competitive advantage from insurance telematics data:

- Integrating data from multiple telematics providers/platforms in different formats from sensor devices or mobile phones
- Large data volumes with varying levels of quality
- Relating new data sources with traditional data warehouse information
- New models for underwriting motor insurance risks based on driver behaviour are required
- Ability to store the results of pattern analysis of vehicle telematics data in your data warehouse
- Consumer demand for transparency and access to near real time driver behaviour information
- The transitioning from pilot telematics projects to production ready information solutions for telematics which can be aligned to your warehouse model but flexible enough to be deployed across multiple deployment options as needed.

Further applications of vehicle telematics include understanding relationship between claims and driver behaviour and identifying value add location based services based on insight from data.

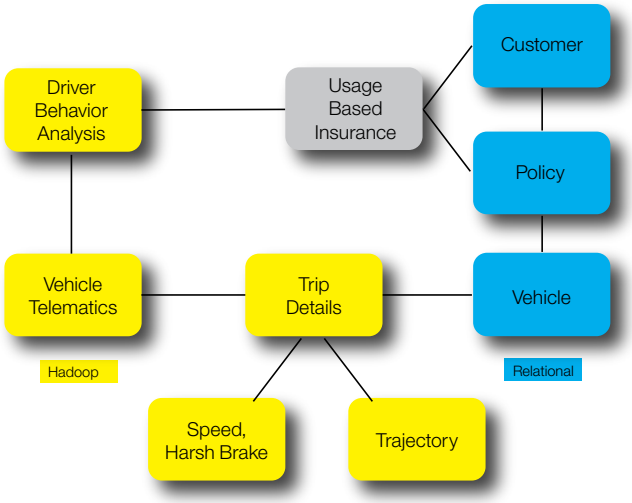


Figure 3. An insurance data model that includes both Hadoop and relational data

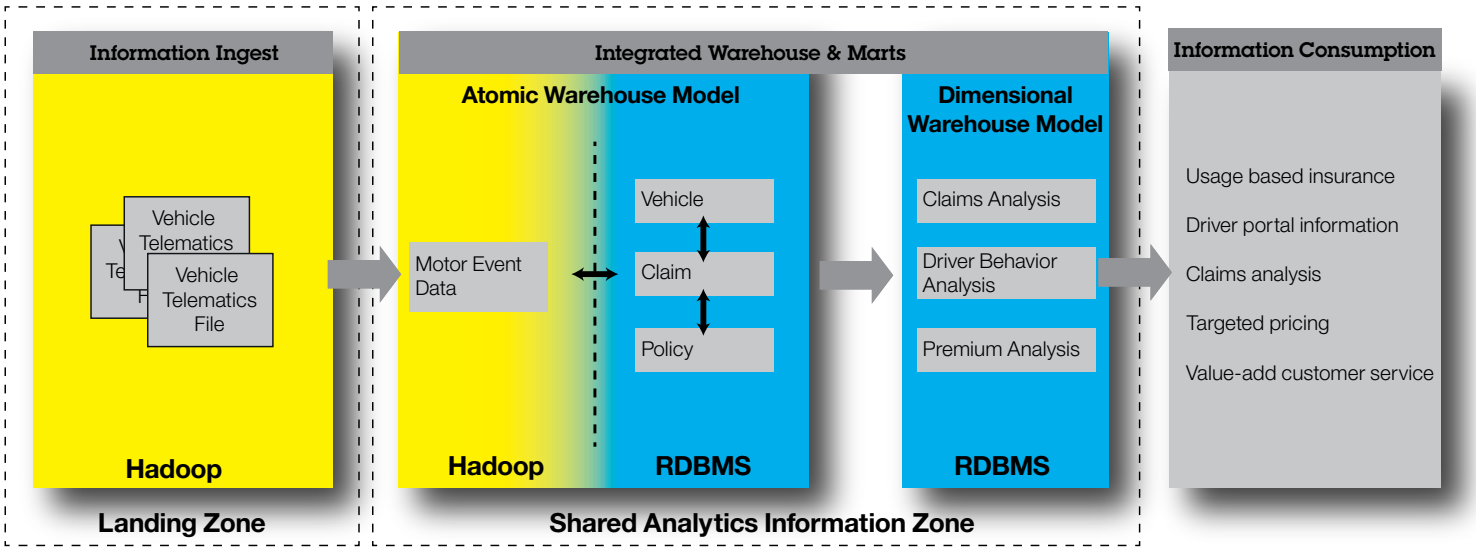


Figure 4. Vehicle telematics scenario

IBM Industry Model content

IBM Insurance Information Warehouse Business Vocabulary now includes vehicle telematics business terms with support for sensor and device data, vehicle information, pre and post crash data standard terms, driver scoring, driver activity, and road infrastructure.

There are new Analytical Requirements including claims investigation, driver behaviour and motor insight analysis. Updates to Projects Scopes and data model extensions have been made to support vehicle telematics. These accelerate requirements definition and data warehouse development. IBM Insurance Information Warehouse can support your vehicle telematics initiative throughout the data lifecycle from source to report.

From Vehicle telematics data to patterns to insight

Vehicle telematics data originates from a variety of different sources including third party providers, streamed directly from the vehicle, and sourced from mobile phone data. The quality, complexity and format may vary widely depending on the source and may even depend on the vehicle. IBM Insurance Information Warehouse promotes a flexible approach to capturing and making use of vehicle telematics big data:

Data Sources

- Storing the originating data files in multiple formats in the landing zone in Hadoop for data exploration
- Deploying IBM Insurance Information Warehouse on Hadoop and/or on your relational database management system
- Leveraging purpose built entity components in the IBM Insurance Information Warehouse Atomic Warehouse Model for vehicle data, which can capture the raw vehicle telematics data provided with flexible data structures.

- Connecting telematics data to vehicles, policies, claims and other enterprise data using in built associations in IBM Insurance Information Warehouse, within Hadoop or between Hadoop and your RDBMS.

Patterns

- Geospatial data and vehicle telematics data in IIW can be analysed with analytical tools such as IBM SPSS® or IBM InfoSphere® BigInsights to provide insight and identify patterns in driver behaviour e.g. harsh break events, journey times, night driving etc. and to score driver behaviour. Using IBM InfoSphere Streams, this analysis could also be performed in real time.
- The results of these risk assessments including driver scoring and could be captured within IBM Insurance Information Warehouse in the Atomic Warehouse Model or Dimensional Warehouse Model using flexible model content.

Insight

- Driver behaviour can be related to premium and claims to determine premium to be charged for the next period based on usage/driver behavior.

Benefits

Product Development

Use well defined IBM Insurance Information Warehouse content to scope out a usage-based insurance data project.

Data Automation

IBM Insurance Information Warehouse provides the path to data automation, once big data data exploration and ad-hoc analysis identifies need for meaningful reporting.

Near Real Time Vehicle Telematics

Hadoop-based IBM Insurance Information Warehouse content supports real time scoring and pattern analysis by using IBM InfoSphere Streams on the vehicle telematics, as the data is received. The turnaround times to driver scoring and calculation of usage based premiums can be reduced.

Claims and Incident Pattern Analysis

- Relating vehicle telematics data with IBM Insurance Information Warehouse enterprise data can support improved data profiling and understanding of the impact of driving behaviour on claims leading to more accurate predictions of claims
- The modeling of more meaningful vehicle telematics data leads to targeted pricing and products
- Vehicle telematics data, when combined with customer and policy information, provides insight into new sets of value add location based services which may be offered to driver, for example discounts on motor maintenance

Customer Information Portal

Supporting policyholder demand for readily accessible information on driver behaviour by policyholder will require the well-defined structures provided by IBM Insurance Information Warehouse.

Dimensional Warehouse Model Support

Includes new dimensional model logical design artefacts for vehicle telematics which can accelerate delivery of vehicle telematics

Data governance

Leverage IBM Insurance Information Warehouse content in IBM InfoSphere Business Glossary for data governance. Telematics data can be mapped to the insurers own enterprise business glossary or to new standards for motor vehicle data into the future.

Fraud Use Case

Business context

The continuing growth of the volume and complexity of insurance fraud requires insurers to find new ways of more effectively detecting fraud. Insurers need to look at new ways of using the data available so they can more effectively use their fraud resources and detect fraud in a more systematic way. .

IBM Insurance Information Warehouse helps insurers overcome the challenges of implementing robust fraud detection systems within their organisation by:

- Integrating data from disparate sources, for example from shared information databases for fraudsters to emails, claim folder documentation, web and call logs, and operational data
- Harnessing operational data for fraud detection and integrating with case management and investigation tools
- Developing a cost effective data architecture incorporating internal information not traditionally available for fraud detection
- Developing a real time fraud detection capability to maximise return on investment from fraud detection systems
- Leveraging social media and unstructured data to establish identity or uncover fraud networks and relationships between individuals linked to an insurance transaction
- Flexible implementation approaches to ensure the security of sensitive information

Other applications include:

- For social media, campaign analysis could use IBM Insurance Information Warehouse social media content to predict or report on campaign outcomes or to profile customers for product positioning
- For unstructured data, communication analysis could determine next best action and improve customer satisfaction

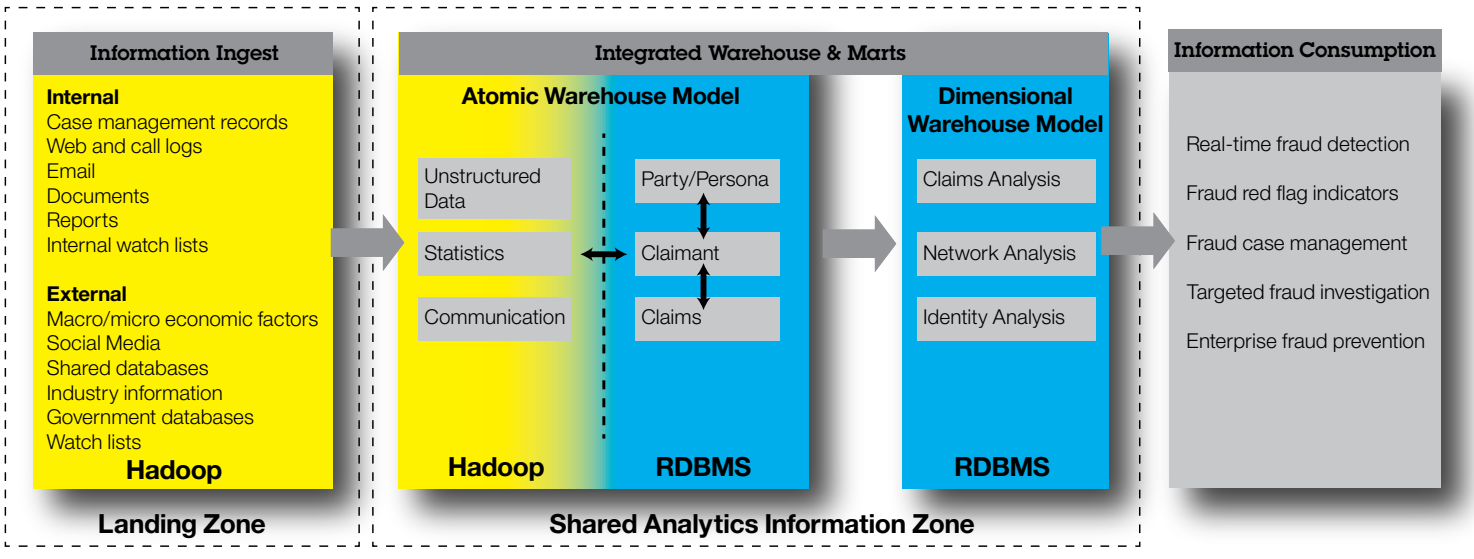


Figure 5. Fraud usage scenario

Consideration for the sensitivity of fraud information

The flexible approach provided by IBM Insurance Information Warehouse means that sensitive information can be isolated within Hadoop, RDBMS, or both. Also, fraud investigation teams can still use the associations between their data and enterprise data to collect the invaluable insight this enriched information contains.

IBM Industry Model content

IBM Insurance Information Warehouse Business Vocabulary now includes insurance fraud terms with support for:

- Unstructured data including documents, emails, call logs, web logs, shared information databases from industry and governing bodies
- Social media, communication and network activity
- Information and fraud propensity scoring

IBM Insurance Information Warehouse Analytical Requirements now include:

- Fraud claims investigation, fraud identity analysis and fraud network analysis.
- Fraud red flag indicators (based on IAIS guidelines) from different fraud perspectives:
  - Losses
  - Claimant behavior
  - Document Characteristics
  - Line of Business Characteristics covering property, motor claims/events and life
- Projects scopes and data models extensions to support fraud and social media, which accelerate requirements definition and data warehouse development

Fraud usage scenario

IBM Insurance Information Warehouse supports your insurance fraud initiative for both claims and policy life cycles so fraudulent payments or policy issuances are detected before they occur. Effective insurance fraud programmes need to focus on early detection in order to be successful.

IBM Insurance Information Warehouse helps design a data architecture that supports this early detection as it will expose fraud investigation teams to a much wider and richer data set. This includes traditional enterprise data but also new data sources such as email, web and call logs and documentation, external sources and even past fraud investigations. Fraud investigation teams using IBM Insurance Information Warehouse can move away from ad-hoc analysis and data exploration to an enterprise, systemic fraud prevention approach.

IBM Industry Models provide a flexible approach to capturing new data sources relevant for fraud and supporting fraud projects.

- Big data sources can be loaded directly onto Hadoop
  - Operational/transaction data from source systems not currently available in the RDBMS to uncover new data for new insights
  - Historical data previously considered too costly to maintain in the RDBMS but which might be useful to detecting repeat fraudsters, fraud networks or patterns
  - Social media from online social and professional communities can help make the connection between parties to a claim and also provide insight on background and behavior which might indicate possible fraud
  - Shared information databases including fraudster watchlists, criminal records
  - External data including statistics on macro and micro economic factors by geographic distribution, injury award guidelines
- Deploy IBM Insurance Information Warehouse on Hadoop, your RDBMS.
- Use fraud specific and standard entity components in the Atomic Warehouse Model
- Connect unstructured content such as claim descriptions, documents, unstructured social media personas with policies, claims and other enterprise data using in built associations in IBM Insurance Information Warehouse within Hadoop or between Hadoop and your RDBMS.
- Using analytical tools such as IBM SPSS or IBM InfoSphere BigInsights to provide data insight
  - Analysis of unstructured data and identify patterns indicating possible fraud for example, claimant behavior, multiple policies issued from same IP address
  - Text analysis including anlaysis voice to text descriptions from call logs or inconsistent and matching version of claim events
  - Scoring information for fraud propensity
  - The results of pattern and scoring analysis can be captured within the IBM Insurance Information Warehouse in the Atomic or Dimensional Warehouse Model using flexible model structures
  - Early detection can be supported by analyzing the policy and claims data in motion, which could initiate systemic fraud detection routines when a quotation is first created or a claim notified and at every step after this
  - Fraud analysis can be integrated with fraud case management tools to provide a more targeted approach to fraud investigation reducing the amount of time spent on false positive investigations and the amount lost to fraudulent claims

Benefits

Fraud Case Management

IBM Insurance Information Warehouse broadens the spectrum of information available to fraud teams from beyond even an enterprise view to one which excludes external sources providing context in which fraud might take place.

IBM Insurance Information Warehouse content

Business terms, analytical requirements and data models have been expanded to incorporate the wide variety of information sources in a fraud project scope.

Real-time Fraud Detection

IBM Insurance Information Warehouse helps in the development of a big data architecture, which supports early detection but is flexible enough to handle new and varied data sources as future need arise.

Dimensional Warehouse Model support

Includes new design artifacts for fraud, which can be deployed to accelerate delivery of fraud projects.

Data governance

Use IBM Insurance Information Warehouse content in InfoSphere Business Glossary for data governance and map new fraud data sources to your own business glossary.

IBM Insurance Information 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 Insurance Information 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. 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 Insurance Information Warehouse and Hadoop Hive and the data can then be queried by data professionals using BigSQL or analyzed using BigSheets.

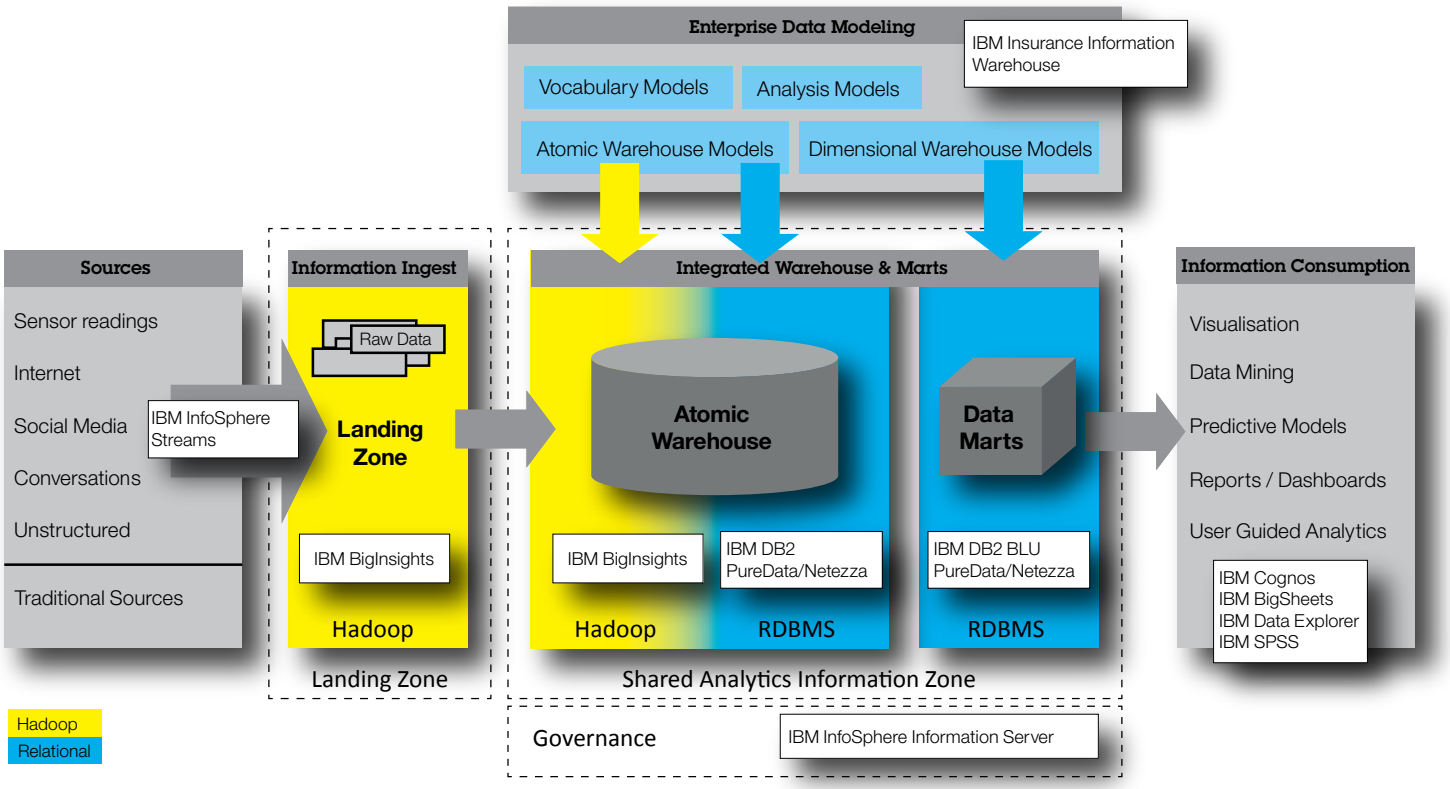


Figure 6. How data models are deployed using the big data platform

IBM DB2, IBM DB2 with BLU Acceleration, and PureData
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 such as geospatial analysis 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 insurance 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 identi-

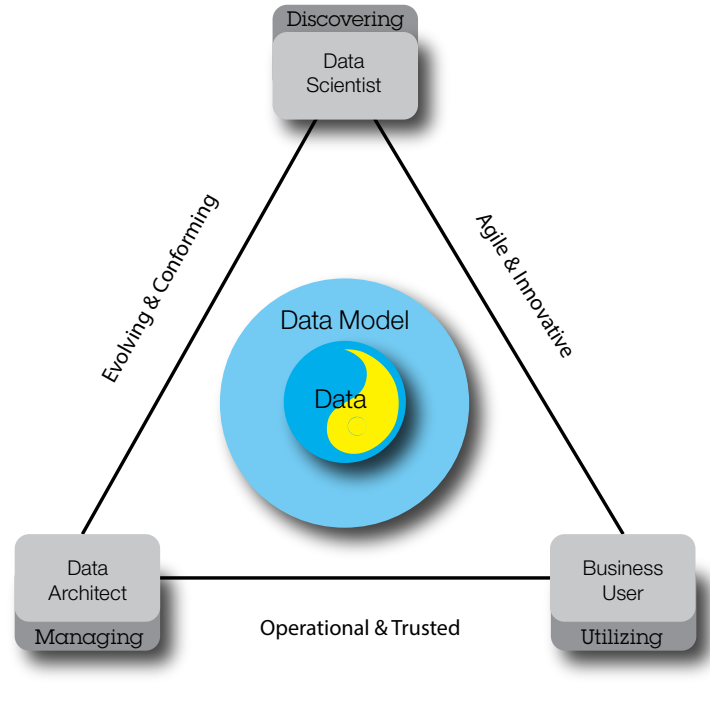


Figure 7. The different actors in a big data project have different governance agendas

fied, 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

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 Insurance Information 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 Insurance Information Warehouse means that even though the data is rapidly available, it conforms to the enterprise data model.

Conclusion

The insurance 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 Insurance Information 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.

Usage-based insurance and fraud detection are examples of how IBM Insurance Information Warehouse can help insurers take advantage of the opportunity presented by big data by providing a complete picture of the data available across the enterprise.

[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.

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
November 2014

IMW14748-USEN-01