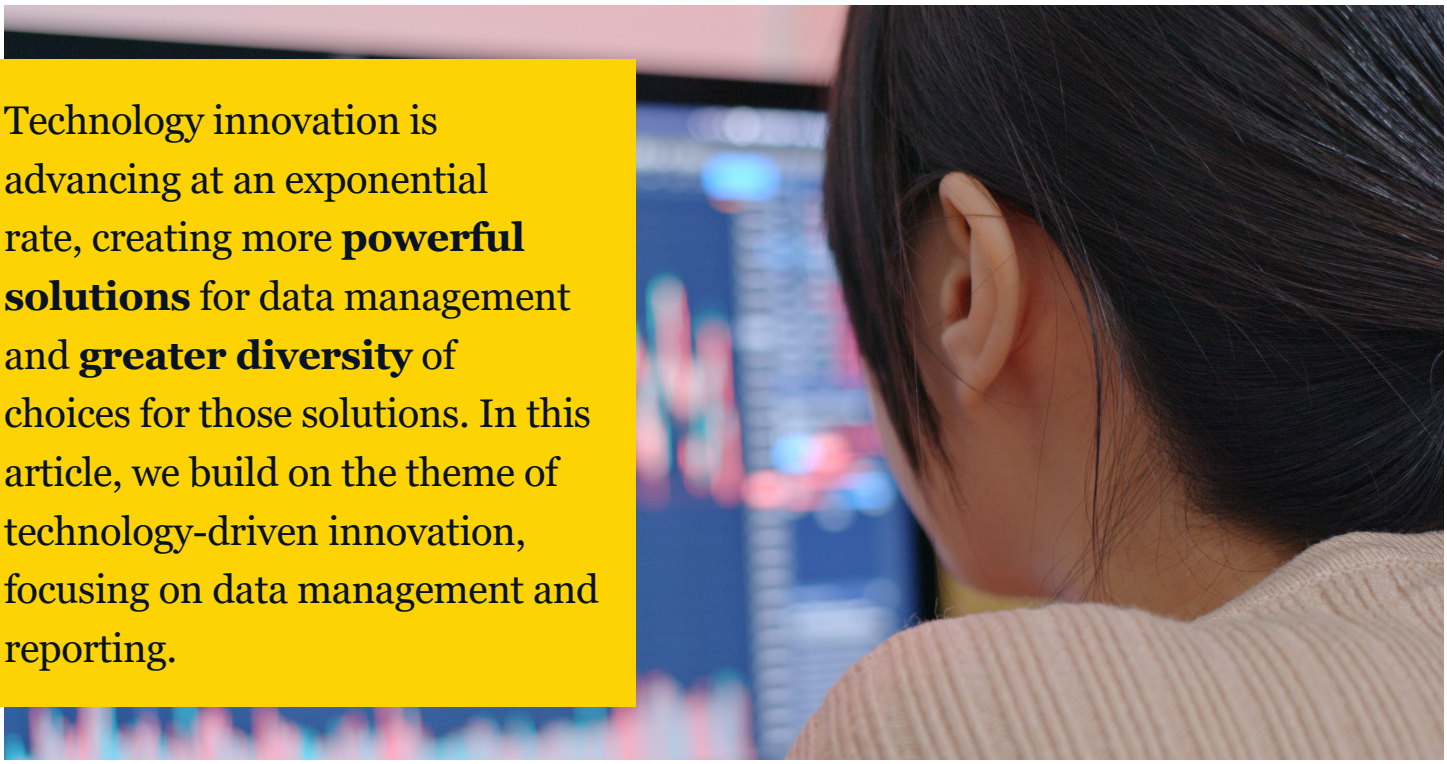
The future of actuarial modeling

How data management plays a crucial role in the end-to-end process

Building a better working world

I. Introduction

Technology innovation is advancing at an exponential rate, creating more **powerful solutions** for data management and **greater diversity** of choices for those solutions. In this article, we build on the theme of technology-driven innovation, focusing on data management and reporting.



Did you know ...

that an average autonomous vehicle is expected to produce between **300** to **5,000** terabytes of data per year?[1]

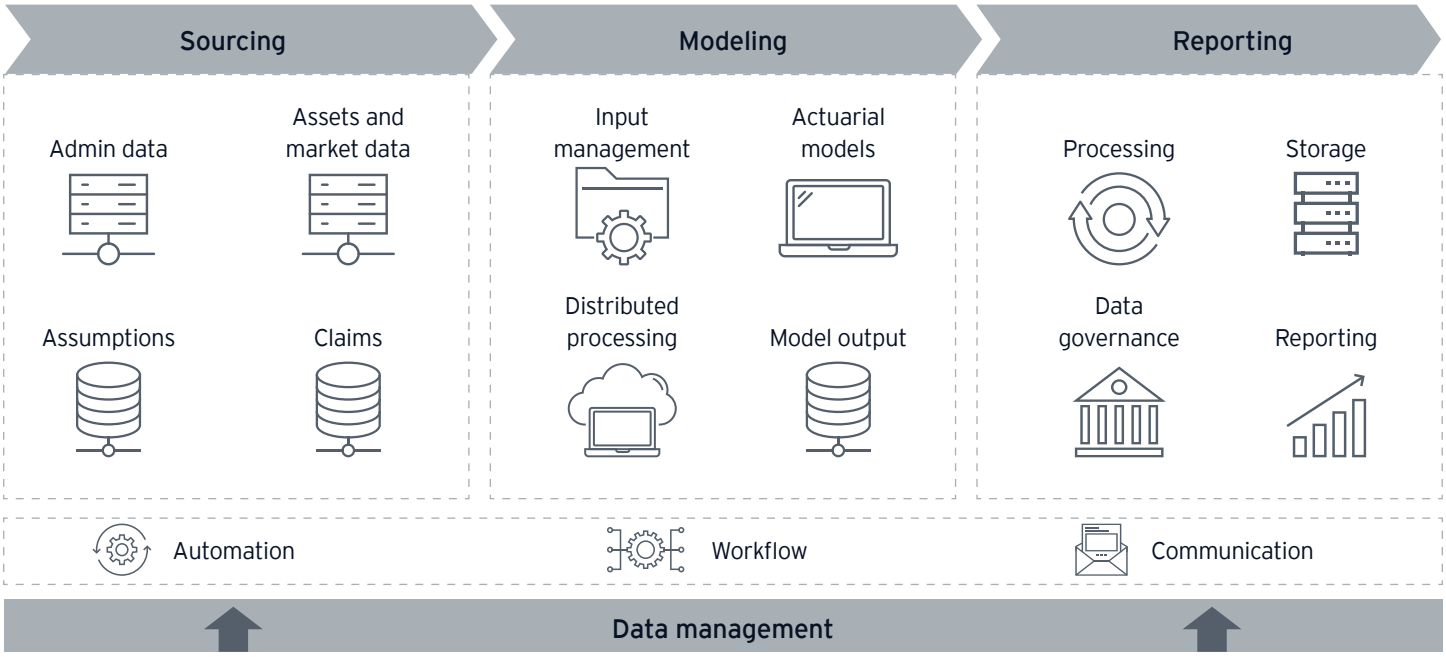
This data is streamed immediately, managed, and used to make high-stakes decisions in real time as the car is driving. In a similar way, the insurance company of the future must be able to harness large amounts of data efficiently, manage it, and subsequently make well-informed decisions. But how can an insurance company benefit from emerging data technology, such as that powering autonomous vehicles?

As the industry moves deeper into implementation of regulatory and accounting changes, the need to establish efficient and seamless technology solutions for analysis and reporting is highlighted. With the evolution to principles-based reserves (PBR) and accounting change, the mandate is clear: insurance companies must manage, process and report on large data sets for valuation, projections and risk management.

[1] Data storage is the key to autonomous vehicles' future," blog post by Mark Pastor, IoT Now Transport, February 2019, accessed June 2021 via iotnowtransport.com; Data is the New Oil in the Future of Automated Driving, blog post by Intel CEO Brian Krzanich, June 2016, accessed June 2021 via newsroom.intel.com; Autonomous cars will generate more than 300 TB of data per year, blog post by Tuxera spokesman Stan Dmitriev, November 2017, accessed June 2021 via tuxera.com; "Just one autonomous car will use 4,000 GB of data/day," blog post by Patrick Nelson, Network World, December 2016, accessed June 2021 via networkworld.com.

While reporting is the eventual outcome of an actuarial modeling process, the end-to-end process requires data sourcing and integration, model processing and reporting. In addition, companies must perform experience studies and analytics to drive business decisions. To support these, and reporting in particular, it is vital for the insurance company of the future to have a robust data management system.

Data management and governance is at the heart of the end-to-end actuarial modeling process and supports each component of the process shown below:



Amid daily conversations, we commonly hear questions of relevance posed across the following categories:

1

Database management systems – how do I manage all of the data we need to store and create for PBR and US GAAP long-duration targeted improvements (LDTI)? How do I manage this volume of data so that it is easily reportable? What new technology exists in the world of "big data"?

2

Data governance – how do I keep our data standardized, governed and quality controlled as it grows in size and complexity? What is the source of truth that should be used enterprise-wide?

The foundation of the end-to-end actuarial modeling process is the capability to gather, manage and produce data. A robust data management solution is required across the board for uses such as policy inventories, assumptions, economic inputs provided by different sources, and results sent to various customers, including financial ledgers. Statutory and accounting changes such as VM-20 and LDTI require numerous sets of model runs, potentially each with their own sets of input files, output results and transformation rules. Increased reporting and disclosure requirements necessitate more robust data management. The data model should be scalable for these changing regulatory requirements, and it should be able to support these changes with a quick turnaround. In today's world where data volume is ever increasing, this translates to managing billions of data points and is commonly measured in terabytes of storage. With ever-increasing data volume, the importance of quality controls and data lineage from trusted sources grows as well. Traditional methods and systems are rapidly being replaced by new solutions, which we will examine here.

This article will explore how actuaries can leverage emerging technology to elevate their data management and reporting capabilities and provide the flexibility to handle changing business needs. By exploring these questions in greater detail, we hope to empower actuaries with the knowledge to work cross-functionally within their organizations to build this future state. As always, we welcome feedback and suggestions for future articles.

II. Database management systems (DBMS)

Overview of DBMS types

There are several ways to store and organize the data coming from multiple source systems before transforming it for consumption and analytics. An enterprise data warehouse (EDW) is one of the more popular methods that has been in the technology space for decades, and insurance companies are in various stages of adoption. EDWs typically use *relational database management systems (RDBMS)* and store vast volumes of data in structured, regimented form. A predefined schema (think of this as very well-organized shelves in a physical warehouse) must exist before loading data. This system means that data is highly organized. This level of organization does mean that changes in data structure or additions of new data attributes (e.g., when new products are added to the business) can be a lengthy implementation effort, given the rigidity of the EDW architecture.

EDWs use extract, transform, load *(ETL)* techniques, which move data across systems, enrich it with additional information, prepare model input files and make model results reportable. Actuaries may rely on EDWs or *data marts* (smaller versions of EDWs) for financial reporting and may expect to have data refreshed once per day when working in legacy systems. However, this ETL methodology in EDW is not highly efficient or cost effective when managing an increased velocity and volume of data.

In order to address some of the challenges of an EDW, organizations started moving toward *data lakes*. Data lakes revolutionized the way data was ingested — by changing the ETL to ELT (extract, load, transform) — allowing data to be "loaded" at the velocity and volume it was generated and allowing the "transformation" to follow. Data lakes do not have any predefined schema, and therefore can intake semi-structured and unstructured data sets. The schema is applied "on read" when the user request originates. Data lakes are designed primarily on cloud technology, which reduces the cost of storing data exponentially, making it cheaper and providing virtually unlimited storage that can be scaled up/down based on demand. Data lakes address the data ingestion and storage issues observed in EDW systems, helping support data volume needs in the face of new statutory and accounting change regulation.

Table 1: Comparison of EDW and data lake

	EDW	Data lake
DBMS type	Traditionally use *relational DBMS*	Can use *non-relational or relational DBMS*
Format	Relational database management systems handle large volumes of structured data very well — however they are not well tuned for different data types or formats (such as unstructured) or frequently changing data.	Data lakes are optimized and flexible for ingesting large volumes of structured and unstructured data, but if not managed carefully can devolve.
Structure	Relational (often referred to as "SQL") databases are table based and use predefined schemas.	Non-relational databases (sometimes called "NoSQL") use alternative methods of storing data. Non-relational databases include document, key-value, graph and wide-column stores.
Access	Relational databases use SQL (a language) to access data.	SQL (the language) may still be used to retrieve data from these databases.
Schema and sequencing	Extract, transform, load (schema on write)	Extract, load, transform (schema on read)
Hosting	Can use cloud or on-premises storage.	Primarily uses cloud services with the option to separate cloud compute and cloud storage — providing flexibility in performance and cost.

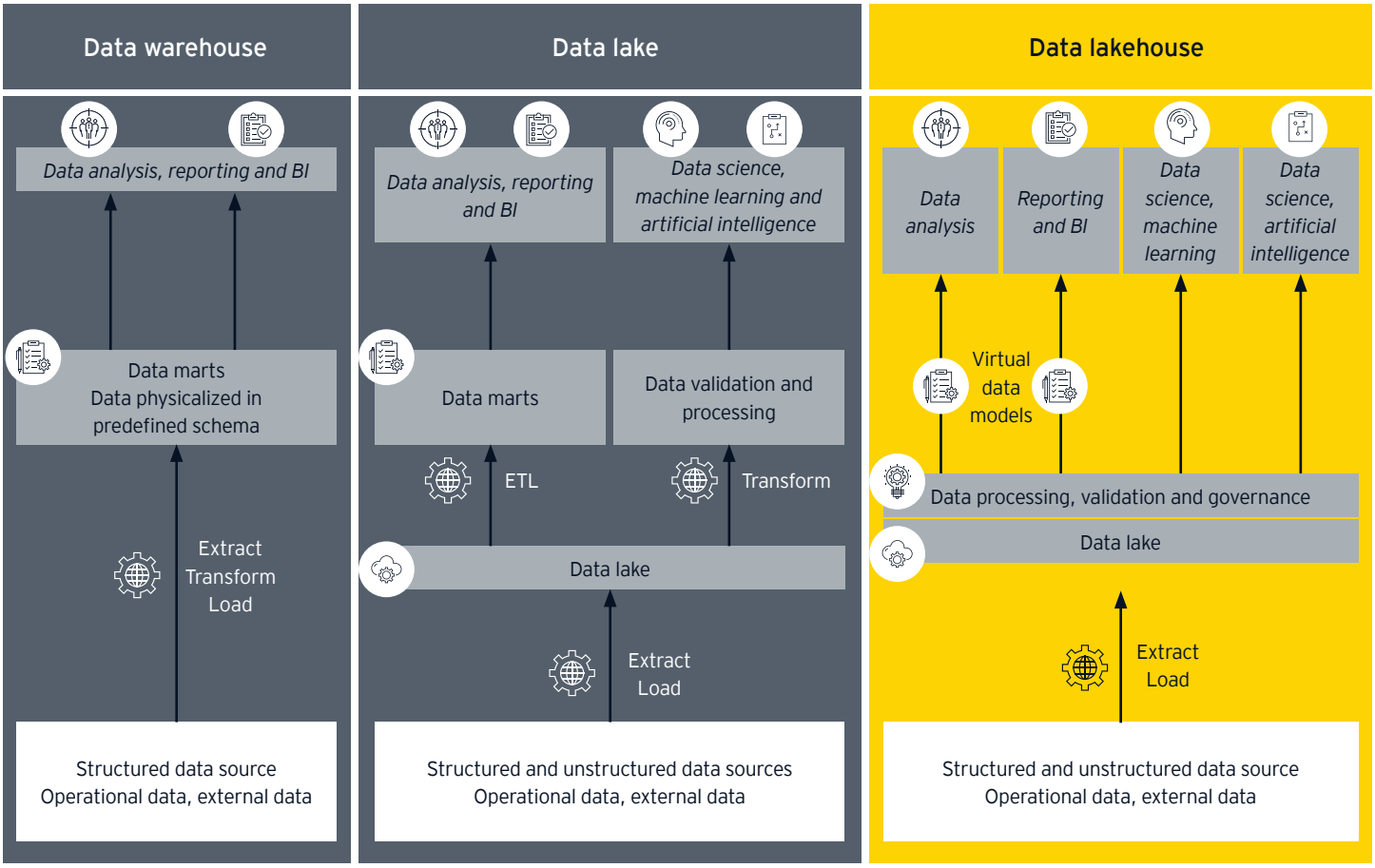
With the right technology and implementation, data lakes can dramatically improve performance for data processing and reporting. As input sources and data needs change, reporting capabilities are resilient and can stay consistent or shift as needed. The ELT method has proved to be a better approach when performing calculations, aggregations, and applying analytical functions, since transformations can be broken down and executed in parallel, thus allowing for faster execution. This is particularly useful in the face of accounting changes when increased disclosures or sensitivities are required.

Though data lakes provide cost-effective data storage and extraction capabilities, it is important to remember that data lakes allow any data format with no quantity limit, which could prevent prioritizing the data and its value. Without the ability to categorize or prioritize data, chaos can set in. If proper data governance is not in place, querying data back out of data lakes could become challenging and expensive. Additionally, a potential challenge observed for both enterprise data warehouses and data lakes is the need for complex data integration patterns to maintain data integrity during concurrent reading and writing of data.

Both EDWs and data lakes have strengths and weaknesses, and each organization will need to assess against its anticipated use cases. The next advancement in the world of data supply chain is a new data management paradigm that combines the capabilities of data lakes and data warehouses. This new generation of technology that captures the benefits of EDWs and data lakes is referred to as a "data lakehouse." Lakehouses leverage both EDW and lake-like components, implementing data structures and management features consistent with those in an EDW, but using low-cost storage solutions such as those used for data lakes. In a data lakehouse, all types of data can be stored at low cost like in a data lake; however, a layer providing a traditional "table" abstraction sits on top of this low-cost base store. Many of the valuable EDW relational database management features such as SQL-based access, query optimization, indexing and caching are available through this abstraction layer to reporting applications. The lakehouse also allows data science and advanced analytics to directly access raw data.

The following figure contrasts a data lakehouse with a data warehouse and data lake:

Figure 1: Comparison of database management solutions



A differentiator for a data lakehouse is the concept of virtual data models, which define the schema based on the consumption needs and enable it on the underlying data, which has been physicalized only once. As memory and computing costs have declined, data solutions now allow for the in-memory storage and processing capabilities that are needed to implement this architecture. This approach also allows for greater flexibility during initial and ongoing development. The system is more nimble and can handle changes, additions and deletions to the data more efficiently. Reporting capabilities are enhanced by this and rely on the processing layer, which is more dynamic and flexible if reporting requires change. This increases the ability to capture historical data needs by minimizing the amount of data model changes needed.

With many insurance companies invested in legacy data warehouses, data lakes have often been used as feeder databases into existing data warehouses. However, with the advent of data lakehouses and the supported integration and data management technologies, existing data structures can be moved into the new lakehouse architecture to provision the data in a more robust, governed way.

Data integration solutions for the lakehouse provide inbuilt capability to ensure data consistency, which was a challenge observed for legacy EDWs and data lakes. The central, robust and simplified data integration and data governance technologies in data lakehouses significantly enhance the value derived from underlying data. Because the data is physicalized only once, it is easier to trace and audit. Controls are also easier to maintain because they are primarily needed at the single point of the physicalization, not at multiple points throughout the process.

Design and implementation considerations

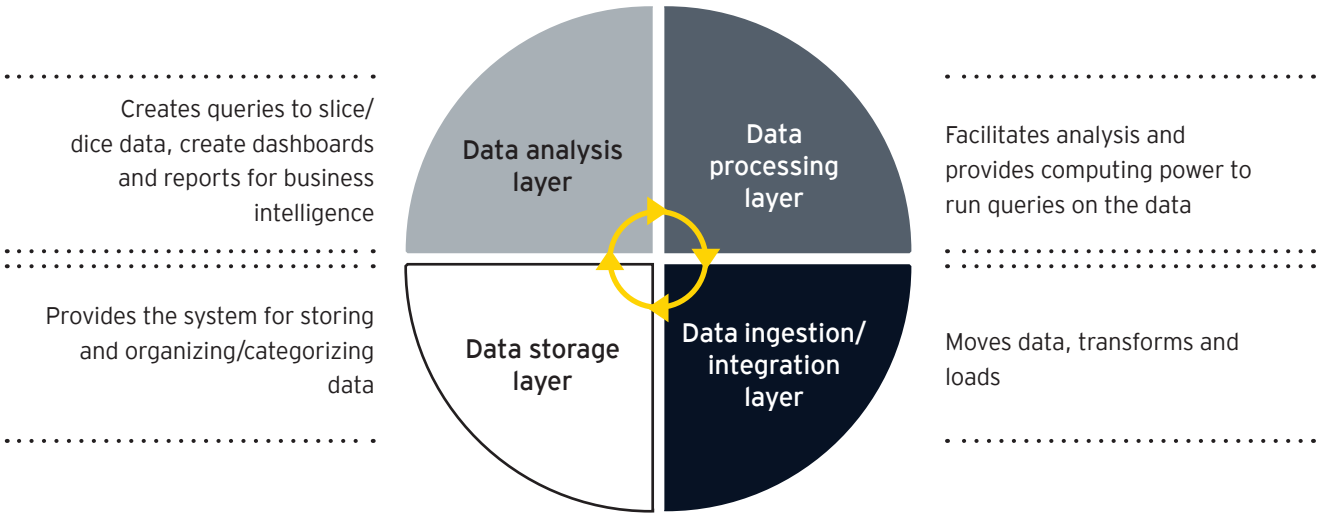
With the data lakehouse approach, the "data lake" component handles data ingestion and storage. Data lakes, which often are associated with the term "big data," can employ non-relational database management systems such as Hadoop Distributed File System (HDFS), but other solutions such as MongoDB, PostgreSQL and Cassandra are also considered non-relational databases, or sometimes hybrids. Solution providers are now expanding to introduce other non-relational database models such as document store, graph, and key-value store, which brings flexibility to the ability to store and access data in different shapes and sizes.

Traditionally for data consumption and reporting analysis, most actuaries are familiar with relational database management systems such as Oracle, MySQL, Microsoft SQL Server and MS Access. New RDBMS solutions gaining traction include SAP HANA, Snowflake and Amazon Redshift, which enable in-memory computation, schema enforcement and virtual data model implementation far more effectively than ever. These features allow for increased data storage capacity and faster data reporting and analytic capabilities.

The insurance industry at large is in the early stages of adoption for data lakes and data lakehouses. A cloud-based architecture provides the foundation for this data lakehouse, and for companies that have not yet started implementing a cloud system, establishing a data lakehouse will require additional effort and cost. Implementing database management systems requires the right architecture, sometimes referred to as a technology stack.

The data architecture required for a database management system typically consists of several layers, which are outlined below with examples of solutions:

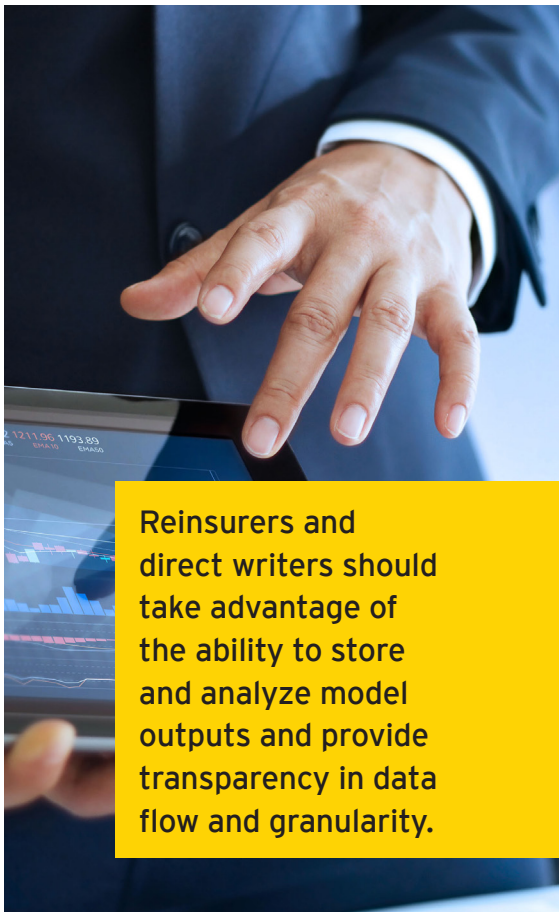
Figure 2: Illustration of data architecture layers



Implementing future state data management solutions requires a holistic approach, in partnership with both technology and business teams, focused on scalability, feasibility and cost to enhance reporting and exploration capabilities. Robust database management is complex. There are many offerings from vendors, and knowledge of data architecture is critical when implementing these solutions. Some companies elect to build their own solutions leveraging technology stacks, while others are adopting vendor ecosystems with inbuilt capabilities. When considering next steps, the organization must determine if it has the talent and desire for *custom-built* vs. *vendor-managed* solutions. Re-engineering database management architecture can be a lengthy and costly process, one that requires deep involvement with technology partners and a thorough examination of current architecture with detailed plans for future state transformation.

Costs of data storage continue to drop as solution choices increase. Cloud data storage solutions may charge on demand or per query or per execution. This option is scalable and inexpensive and offers a compelling edge over on-premises solutions. Carriers, particularly medium to large carriers with scalable capacity, should explore enhancing their database management systems and think of data as an asset. Reinsurers and direct writers should take advantage of the ability to store and analyze model outputs and provide transparency in data flow and granularity. Cost-benefit analysis should be performed at the enterprise level as many functions will stand to gain from this endeavor.

Actuaries looking for flexibility in their model inputs and outputs should be open to exploring new solutions. As we journey through accounting change and regulatory reform, it is becoming obvious to all that flexibility is paramount, especially during development and test cycles. The long-term benefits of managing data strategically as an organization outweigh the shorter-term implementation costs. As the market moves to this future state, now is the time to adapt.



III. Data governance

Description

The variety of suppliers and customers coupled with the volume of data required for valuation and projections necessitates strong data governance protocols. Risk analysis and reporting are only as good as the underlying data. Strong data governance ensures the integrity of data across all consumption points, enables and maintains data quality, and catalogs data to ensure reusability. The stronger the governance, the greater the value.

Some common governance principles of data management are helpful to define up front, which we will explore:

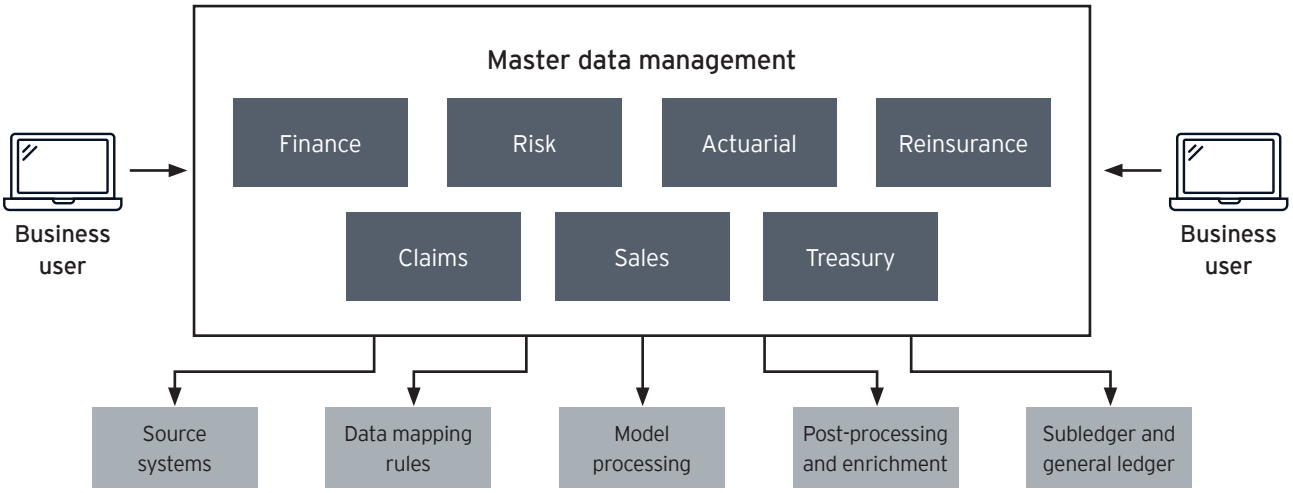
Data lineage	is implemented to know where, when, how and by whom data was created. It provides visibility to the origin of data, how it is changed and where it is sent and by whom.
Master data management (MDM)	defines and maintains consistent definitions of reporting entities, such as liability type or reserve type, and shares them across multiple systems in the enterprise. MDM standardizes data hierarchies, names and descriptions across functions including finance, actuarial, risk and reinsurance to create common definitions. Using leading practices of MDM, relationships among data sets can be defined and controlled to reduce errors and enable communication across systems in a common language.
Data catalogs	are based on the metadata an organization defines and create an organized inventory of data assets in the organization. Data dictionaries further augment the catalog by providing collections of the descriptions of data elements in a data model for business and IT professionals.

Real-world applications

Data lineage helps the enterprise understand the flow of data, connecting downstream consumption systems to source systems. Actuarial functions particularly benefit from this since many actuarial calculations depend on key assumptions and inputs, and in turn, model output is used across the company, from forecasting to earnings attribution. Actuaries must keep track of sources of record and maintain the chain of custody in order to provide reliable outputs for downstream uses.

As actuarial model inputs move through transformation processes and into the model, and results come out, *master data management (MDM)* can be used to ensure that each policy is mapped consistently with a set of rules and definitions common to actuarial, reinsurance, and finance. Business users maintain these rules and have a means of updating them as needed under governance protocols. For reporting and accounting, actuarial model results of each policy may be enriched with key data fields required to create journal entries and financial reports for regulators and investors. This is done by standardizing the data mappings and enrichments that occur from source systems to the actuarial model's raw outputs, through post-processing and into the general ledger. Audits, enterprise policy compliance and business rules validation demand strong data governance, which can address issues and respond quickly. Data lineage provides a full view of the data flow across the enterprise, ensuring policy adherence, controls and validations. In scenarios where data issues or performance impacts are observed, a data lineage framework can assist in identifying anomalies and bottlenecks.

Figure 3: Master data management illustration



Data catalogs are useful as a company seeks to define and utilize its metadata. Metadata can be used to categorize, describe and analyze model inputs and outputs, so a strong data catalog and associated data dictionaries ensure consistency and transparency across the organization.

The use of subledgers and actuarial data repositories provides great value to valuation and projection actuaries. However, it requires significant pre- and post-processing of model inputs and results to bring this data to a common transaction-based view. With proper MDM, finance, actuarial and reinsurance administration can view actuarial postings from the model to the GL or statutory exhibit lines consistently.

Implementation considerations

Key data management principles must be employed for model processing. The following items should be considered when implementing these principles:

Data lineage

Companies should first prioritize developing an industrialized end-to-end process that maintains an appropriate chain of data custody across each distinct data user. Then, data lineage can naturally be implemented with open-source tools such as CloverETL and Kylo or vendor tools such as Collibra, Talend and Octopai.

Master data management (MDM)

The organization should develop data tables, transformations and rules, housed in a centralized database, that are used consistently across the enterprise. This can be implemented with many tools, including both open source such as Talend and vendor managed such as SAP Master Data Governance.

Data catalogs

The company should develop data catalogs early in the development of their future end-to-end process and seek to maintain consistent metadata across the organization. This can be implemented with software that will crawl databases and record the data tables and metadata, allowing users to visualize and understand them.

Data dictionaries

Can be implemented by coordinating across business and IT to define the data. Something as simple as a shared Excel spreadsheet can be a starting point. Similar to data catalogs, these should be developed in the process of requirements and design.

When implemented correctly, the process will be adaptable for change and governed in a transparent manner with clear data lineage. This goes alongside a strong and robust data model, centered around improving data consumption performance and scalability for future reporting and analytics needs. Imagine having a centrally controlled and governed common set of definitions across finance, actuarial, risk, reinsurance and treasury functions with the ability to view data flow end to end. Building this is now possible and will be common in the future state.

Cost-benefit analysis

Simple data management techniques such as creating catalogs and dictionaries are low-cost, high-return investments of time and resources. Understanding data relationships is essential for actuaries in order to work with technology, administration and finance professionals in the digital world. Implementing data lineage and master data management tools will vary in cost and scope. Approaching implementation at an enterprise level creates operational efficiency, improves data quality and compliance and enables the company to transition to the end-to-end data management of the future. Without adopting some basic governance practices, a company looking to compete in the coming years will be left behind. Those ahead of the curve will adopt leading practices and generate more value as a result.

Conclusion

Actuaries have a unique opportunity to step in at this pivotal time to reimagine their future, working closely with technology and data partners. Investing time to learn about new solutions and then working with dynamic teams will yield high returns. The companies that embrace the change and understand their data management options will capture more value as they design and build for the future.

As we have discussed, systems will be connected, data will be easily managed, models will scale and automation will be front and center. But before all that gets here, actuaries must be front and center working with cross-functional teams to design and build these solutions. The future shines brightest for those who invest in expanding their knowledge and understanding of what will come.

Contacts

Insurance and Actuarial Advisory Services

Dave Czernicki
Principal
Ernst & Young LLP
dave.czernicki@ey.com

Eric Wolfe
Senior Manager
Ernst & Young LLP
eric.wolfe@ey.com

Vikas Sharan
Senior Manager
Ernst & Young LLP
vikas.sharan@ey.com

Kevin Kang
Senior Manager
Ernst & Young LLP
kevin.kang@ey.com

Philip Sandager
Manager
Ernst & Young LLP
philip.n.sandager@ey.com

Data and Analytics

Peter Corbett
Managing Director
Ernst & Young LLP
peter.corbett@ey.com

Bharath Raju
Senior Manager
Ernst & Young LLP
bharath.raju@ey.com

Prateek Agarwal
Manager
Ernst & Young LLP
prateek.agarwal3@ey.com

EY | Building a better working world

EY exists to build a better working world, helping create long-term value for clients, people and society and build trust in the capital markets.

Enabled by data and technology, diverse EY teams in over 150 countries provide trust through assurance and help clients grow, transform and operate.

Working across assurance, consulting, law, strategy, tax and transactions, EY teams ask better questions to find new answers for the complex issues facing our world today.

2106-3755345
US SCORE no. 12996-211US
ED None

ey.com