BUILDING NEW DATA ENGINES FOR INSURERS

By Rash Gandhi, Hanno Ketterer, Jonathan Koopmans, and Venkat Siva

IN TODAY'S INSURANCE INDUSTRY, every link in the value chain—whether it be product development, marketing, underwriting, pricing, client servicing, or claims management—is becoming a data-driven play. That isn't surprising. According to BCG studies, data-driven decision making creates an upside of from 6% to 9% of an insurer's combined ratio (a key profitability measure calculated by dividing the sum of an insurer's net claims, commissions, and expenses by its net earned premium income). Systematically using data yields several strategic benefits, too, helping insurers engage more closely with consumers and become more responsive to their needs. Indeed, insurers that don't focus on data-driven decision making will soon be overtaken by rivals that do.

The desire to use data strategically triggers the need for a next-generation data engine, by which we mean the technology stack that allows the capture of data, the generation of insights through analytical models, and the integration of those insights in real time with core processes. Unfortunately,

although data-driven decision making requires insurers to fundamentally overhaul their data engines, many organizations pay little attention to their data engines' capabilities. As a result, those organizations can't go beyond setting up pilots and experiments because their data engines are inadequate. Studies show that while 30% of insurers have started pilots, only 15% have been able to scale them. As the technology head of an industry leader in Europe ruefully told us, "We have more pilots than Lufthansa does."

Our research suggests that the data engine is critical to scaling data-driven strategies. In addition to investing in technology that matches its vision and future needs, and deploying the right operating models, every insurer must develop a data engine that will enable it to rapidly scale the pilots and experiments it has set up. (See "Rewiring Decision Making in Insurance with Data Science," BCG article, October 2018.) Doing so is crucial to success in today's marketplace, as insurance industry leaders such as China's Ping An and Europe's Aviva have

demonstrated. Ping An has built and used a massive data platform to develop a wide range of analytics, such as online credit scores and disease patterns. Aviva has used data to revolutionize the quote-and-buy experience, shifting from asking customers hundreds of questions to asking none, which has resulted in the rapid growth of its digital business.

Where Traditional Data Engines Falter

Traditional database management, which relies on using rigid technology infrastructure to capture mostly internal data in predefined formats, has become insufficient. Most insurers have set up structured data stores and data warehouses, but those are costly to maintain and can't handle the rising number of heterogeneous data formats that insurers must use. As a result, insurers often take three or more months to incorporate new and varied sources of data into their core processes. One insurer doing business in Australia took more than a year to get its data engine to function properly, owing to inadequate knowledge about its data sources and the absence of a common structure in those sources.

Many insurers lack mechanisms by which to seamlessly integrate data from their ecosystem partners into their systems. Moreover, several manual data checks are necessary to ensure consistency, and the legal necessity to comply with global data regulations slows down the process even more.

Most insurance companies have data scattered across and siloed in their product lines or channels: 75% of insurers lack a common data-storage system or an appropriate taxonomy to combine diverse types of data. Time lags associated with batch processing and dependence on IT teams delay the generation of insights from data. In fact, most insurers can integrate data-driven analytical insights into core operational processes, such as pricing and underwriting, only offline—not in real time.

For pricing, insurers still use traditional mathematical models and risk parameters,

but these have rather limited predictive powers. In addition, inflexible legacy systems—constrained by static, hard-coded business rules—prevent insurers from rapidly reconfiguring and customizing the customer journey, pricing, and underwriting.

Designing the End-to-End Data Engine

On the basis of our research into more than 20 insurance carriers, as well as solutions providers and insurtechs such as Earnix, Guidewire, HexScore, RiskGenius, and SAP, we've identified a simple way to group the main challenges to—and the best practices in—developing the next generation of insurers' data engines.
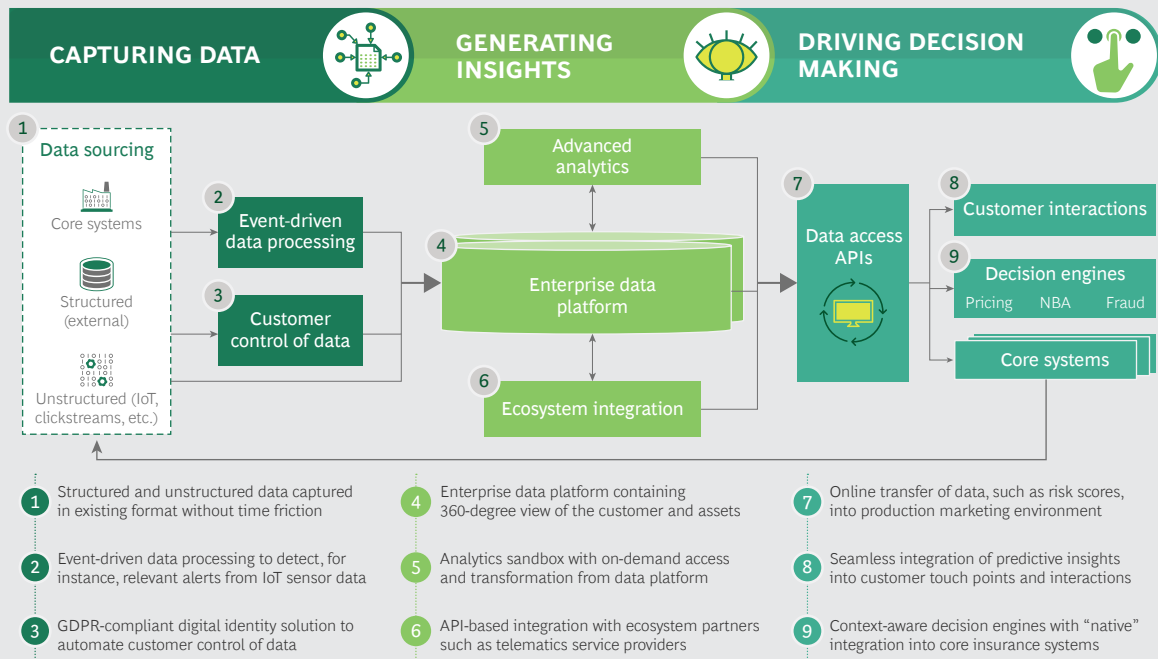
Our approach covers the entire life cycle, from the sourcing, processing, and analysis of data to the use of data-based insights in key processes to, finally, the integration of data-driven decision making into operational systems. Consequently, rather than offering a solution to deal with just one pain point, the approach provides general principles that can underly the building of an end-to-end data engine. Such data engines allow data-driven decision making and innovation at every step of the value chain.

Tomorrow's data engines must, we believe, perform three essential actions: capture diverse types of data, generate comprehensive insights, and integrate data with decision making. (See the exhibit.)

Capture diverse types of data. Data engines of the future must be able to capture heterogeneous data, in whatever the data's original format may be, in order to speed up data collection and analysis. This capability is not always easy to achieve.

Insurers are starting to access structured and unstructured data from multiple external and internal sources. For example, a direct writer in Europe accesses social and demographic data, vehicle and driver data, and weather data from open databases. It taps into commercial databases for vehicle safety data, competitive benchmarks, and company profiles. It obtains customer data

Designing Insurers' New Data Engines

CAPTURING DATA — **GENERATING INSIGHTS** — **DRIVING DECISION MAKING**

1 Data sourcing
- Core systems
- Structured (external)
- Unstructured (IoT, clickstreams, etc.)

2 Event-driven data processing

3 Customer control of data

4 Enterprise data platform

5 Advanced analytics

6 Ecosystem integration

7 Data access APIs

8 Customer interactions

9 Decision engines — Pricing · NBA · Fraud

Core systems

1 Structured and unstructured data captured in existing format without time friction

2 Event-driven data processing to detect, for instance, relevant alerts from IoT sensor data

3 GDPR-compliant digital identity solution to automate customer control of data

4 Enterprise data platform containing 360-degree view of the customer and assets

5 Analytics sandbox with on-demand access and transformation from data platform

6 API-based integration with ecosystem partners such as telematics service providers

7 Online transfer of data, such as risk scores, into production marketing environment

8 Seamless integration of predictive insights into customer touch points and interactions

9 Context-aware decision engines with "native" integration into core insurance systems

Source: BCG analysis.
Note: API = Application programming interface; GDPR = General Data Protection Regulation; NBA = Next best action.

(such as online journeys, clickstreams, and driving behavior) and customer engagement data (such as customers' records, service requests, email messages, and social media interactions with the insurer) from private databases. The insurer deploys content management platforms to access pictures and forms, and it retrieves data on policies and claims from its core insurance systems. Being able to gather and use all of that data in an integrated way gives the insurer a distinct edge in identifying customers and understanding their behavior.

Similarly, an Asian insurer collects and stores data on more than 200 customer attributes, including a number of idiosyncratic factors such as cultural norms that could affect their decision making. (See "The Trust Advantage," BCG article, November 2013.)

New data engines must also support the processing of data from live events; that is, they must be able to capture and analyze data in real time. They should permit companies to detect alerts—and, if necessary, act on data—from sensors that transmit

data over the Internet of Things. For instance, a car sensor could alert the insurer when the vehicle's airbag has been activated; and that, in turn, could trigger a telephone call to the customer to determine whether the vehicle has been involved in an accident and whether the customer needs immediate assistance. Market leaders also use machine learning to continuously fine-tune their predictive models as fresh data becomes available.

Insurers walk a fine line with regard to privacy, so the new data engines must contain legally compliant solutions that give customers complete control over their personal data. The European Union, for instance, is promoting Personal Information Management System (PIMS) rules in the context of the recently passed General Data Protection Regulation. PIMS rules allow individuals to manage their data in secure local or online storage systems and to limit sharing to the persons, entities, and occasions of their choice. Going forward, many European insurers will use only cloud-based service providers that comply with the EU's privacy laws.

Generate comprehensive insights. The new data engines must give insurers 360-degree views of their customers and assets. About 80% of insurers are now in the process of generating a single view of each customer, based on a wide array of internal, external, and inferred attributes, and consisting in broad outline of the individual's profile, financials, interactions, and risk parameters. Customer profiles rely on several types of data, including transactional data, interaction history, behavioral data, channel preferences, and social identifiers and preferences.

To facilitate the creation of new data engines, insurers are shifting from traditional data warehouses to data lakes, which can store vast amounts of structured and unstructured data for analytical purposes. The data lakes allow insurers to build rich customer profiles that capture behavioral attributes on the basis of click behavior. (See "Look Before You Leap into the Data Lake," BCG article, September 2016.) Data lakes are low-cost data storage environments that use commodity hardware and an integrated technology stack of open-source data and analytical tools.

Increasingly, insurers are using onsite or cloud-based integrated technology stacks to store and analyze data. Because cloud-based stacks are scalable, they require minimal upfront investment, and they allow integration with such third-party analytical tools as Google Cloud Machine Learning Engine. For instance, the Australian insurer that we described earlier has implemented a data engine based on a multichannel platform and marketing cloud, and decoupled from the core system. Because the engine doesn't have to conform to several rigid IT controls, the carrier can deploy new analytical insights quickly—sometimes in a matter of hours.

The new breed of data engines also allows integration with insurers' ecosystem partners through an application programming interface (API). Insurers typically have limited interactions with their customers—less than two per year, on average—so they need to obtain and use customer insights

from as many partners as they can. Such partners may include e-commerce platforms, retailers, and banks, among others. Most insurtechs offer API-based analytics services in various segments of the insurance market.

An insurer's new data engine must also provide an analytics sandbox that offers on-demand accessibility and goes beyond being a mere data collection platform. To compete effectively, insurers are learning to use such advanced analytical techniques as predictive modeling (for example, regression analysis, decision trees, and fuzzy inference), optimization (including linear programming, factorial design, and combinatorics), simulation and visualization (data visualization, Monte Carlo simulations, queuing theory, and so on), and artificial intelligence (swarm intelligence, random forests, and convolutional neural networks, among others).

Integrate data with decision making. Modern data engines enable organizations to experience the power of data-driven decision making. Companies can use them to generate granular, precise, and timely data-based insights that transform the speed and sophistication of decision making.

The best-in-class data engines are context aware, so they can optimize customer interactions across all touch points in real time. For instance, the system can alert an insurer to a customer's location, enabling the company to offer travel insurance over a mobile device the moment the customer crosses a national border. Modern insurance suites offer native integration with decision engines—in the form of out-of-the-box or custom algorithms embedded in platforms—throughout the customer life cycle, so insurers can use them for data-driven decision making.

Such data engines facilitate the creation of horizontal models of decision making that will enable insurers to optimize customer lifetime value by cutting across products and channels. They allow the transfer of data, such as risk scores, to key processes online. They permit the integration of predic-

tive insights into systems during interactions with customers. And they can personalize customers' experiences by managing a company's interactions with customers across all touch points in real time.

Deploying New Data Engines

Two factors determine whether insurers can successfully deploy new data engines: managing change, especially the process of redesigning decision making processes; and executing decisions in a disciplined manner. The latter requires companies to take several operational steps.

Focus on internal unicorns. Insurers would do well not to try to build data engines capable of handling all manner of experiments from the get-go. Such an endeavor may start well, but the process will never end. Instead, companies should focus on developing the data engine to handle one concrete initiative that will move the needle in both process and financial performance. We call such experiments internal unicorns. Adopting that focus will enable insurers to rapidly test new data sources and to quickly learn new analytical approaches.

Companies should resist investing time and money in experiments that go nowhere. Consequently, from the get-go, they should design data engines to be scalable, fully integrated, and operational for all initiatives. Once the process improvements and financial results in the pilot become visible, and the business case is evident, insurers can scale the data engine to other pilots and experiments.

Strengthen data capabilities. After building an end-to-end data engine for one initiative, companies can strengthen it by hiring and training the right kind of people, such as data engineers, to ensure that capabilities take root inside the organization, and by developing data-engineering tools to manage data pipelines extending from the source to the curated data repository.

Pay attention to data governance. Insurers can enforce consistent data standards and quality controls by creating a dedicated data organization that includes a chief data officer, data stewards or owners, and data scientists. It is also important for insurers to institute data policies and set up operating models that can help break silos and ensure the smooth flow of data. By focusing on metadata management, insurers can build a catalog of data assets and a business dictionary.

Stage the process. Finally, executives must plan the implementation in phases. A step-by-step approach, starting with one experiment, ensures that companies prioritize the scope of the process. An incremental approach permits gradual investments while delivering value along the way. Although the first data-driven solution can be up and running in six to nine months, it needs to be scaled rapidly over the next two or three years even as the insurer sets up other pilots.

Insurers can rapidly deploy next-generation data engines for data-driven decision making, and the returns, too, will flow in quickly. Consider a large European insurer that has successfully built rich profiles of customers by complementing its internal data with data from external systems and open public data platforms. Until 2013, the carrier stored its data in three silos: motor, nonmotor, and life insurance. Since then, it has consolidated its data in a data lake; and it has developed a data engine, based on an open-source software framework, for storing data and running applications off clusters of commodity hardware. Its goals were to decouple the data lake, which provides individual customer profiles and their portfolios, from the vertical system (motor, nonmotor, and life) and to store all of the information in the same place.

The insurer quickly learned the importance of focusing on the quality of data, which is a major problem worldwide. (See "How to Avoid the Big Data Trap," BCG article, June 2015.) The system stores all of the data necessary for price calculations—such as the number of kilometers driven, the speed, and the number of accelerations—in a structured format. In response to privacy concerns and regulatory restrictions,

the company stores the data in aggregated, anonymized form, rather than as detailed data on each trip that a driver takes.

The company started by setting up just two use cases; but in the medium term, it plans to integrate its systems to become more proactive, to develop event-driven applications, and to support cases in real time. Because it has a data lake, the company will be able to inject enriched data into its decision making and work on predictive models and algorithms in its data laboratory. The results are encouraging. The company now sells auto insurance policies that leverage telematics with decent margins, has better clients (as measured by renewals, repeat business, and purchase of other products) today than it had previously, and experiences no delays in settling claims.

The carrier's analytics-based algorithm is superior to its actuarial algorithm for car insurance, and its customer churn ratios have fallen to historic lows since it switched to the new data engine.

AS THE VOLUME and the variety of data available to insurers rise exponentially, companies have to figure out how to gather, store, combine, and study all manner of data, so they can develop data-based insights that will help them serve customers better. That's why every company needs to focus on designing a new data engine to serve as a critical enabler of data-driven decision making. Only insurers that make the right design and implementation choices with regard to data engines will succeed in the future.

About the Authors

Rash Gandhi is a principal in the London office of The Boston Consulting Group. You may contact him by email at gandhi.rash@bcg.com.

Hanno Ketterer is a senior partner and managing director in the firm's Amsterdam office. You may contact him by email at ketterer.hanno@bcg.com.

Jonathan Koopmans is an associate director in BCG's Amsterdam office. You may contact him by email at koopmans.jonathan@bcg.com.

Venkat Siva is a principal in the firm's London office. You may contact him by email at siva.venkat@bcg.com.

The Boston Consulting Group (BCG) is a global management consulting firm and the world's leading advisor on business strategy. We partner with clients from the private, public, and not-for-profit sectors in all regions to identify their highest-value opportunities, address their most critical challenges, and transform their enterprises. Our customized approach combines deep insight into the dynamics of companies and markets with close collaboration at all levels of the client organization. This ensures that our clients achieve sustainable competitive advantage, build more capable organizations, and secure lasting results. Founded in 1963, BCG is a private company with offices in more than 90 cities in 50 countries. For more information, please visit bcg.com.