The Road to Embracing App-Driven Analytics, AI/ML, and LLMs for Insurance

Start by Simplifying Your Data Estate

June 2023

Content

Out With The Old

In an age of generative AI and Large Language Models (LLMs), the race is on to exploit technology and data to create operational efficiencies and competitive advantage. Traditional data analytics strategies used in insurance, however, are too slow for the digital age.

This paper will present an alternative, App-driven Analytics, that empowers insurers to deploy the vast amounts of existing and new data at their disposal to streamline the way they do business and transform the customer experience.

This paper will explain why:

- *The data warehouse is too slow for today's business environment. While a relic from a past age, data warehouses aren't going anywhere, and we don't want them to - we still need the intuition of underwriters and reports on the variables that are traditionally employed in underwriting.*

- *Analytics, AI-powered insights, and automated decision making is "shifting left" - from back-end processes into business applications - and why developers can move quickly to adapt to this change by adopting the document model.*

- *Machine learning platforms need a simplified data estate to realize their full transformative potential. However, the complexities of legacy processes often diminish, and sometimes completely negate, the value of AI/ML efforts. Organizations may see pockets of ML research and development experiments, but struggle to make significant strides in fusing these ideas with actual business workflows.*

Why Traditional Analytics Practices in Insurance are Too Slow

Insurance companies are, by nature, data driven. They produce little to no tangible goods, but rather ingest, analyze, and turn data into decisions. Amongst the many decision-making processes within an insurance company, perhaps none are as critical to profitability and stability as underwriting.

Whether it's conducting analysis across thousands of data points in order to steer the company's portfolio, pricing products adequately, or designing coverages to effectively manage risk, underwriting is the heart of an insurer and it relies heavily on data analytics.

Historically, insurance has relied on a relatively small amount of data variables to make underwriting decisions. For decades, past performance could be used to accurately predict future liability.

The last few years, however, have rendered traditional methods used to quantify risk inadequate. Economic and market uncertainties, climate impacts and real-time ESG data, cyber threats, and supply chain volatility all defy traditional risk models. In addition, there is rising consumer demand for real-time digital servicing, as well as embedded and bundled insurance products that hinge on accurate risk assessment in seconds.

As a result, underwriters are being forced to scrutinize existing methods, as well as develop and monitor new models - all in increasingly shortened timeframes. The underwriting feedback loop that these models now demand shifts us away from weeks and even days, to hours, minutes, and seconds.

At the same time, the number of sources of information and data points that underwriters must interrogate and analyze is increasing. Usage-based insurance products, for example, rely on the collection and processing of real-world telemetry data. Some carriers are even leveraging telemetry usage data to shift from yearly written premiums, to monthly, to offer more competitive products to their customers.

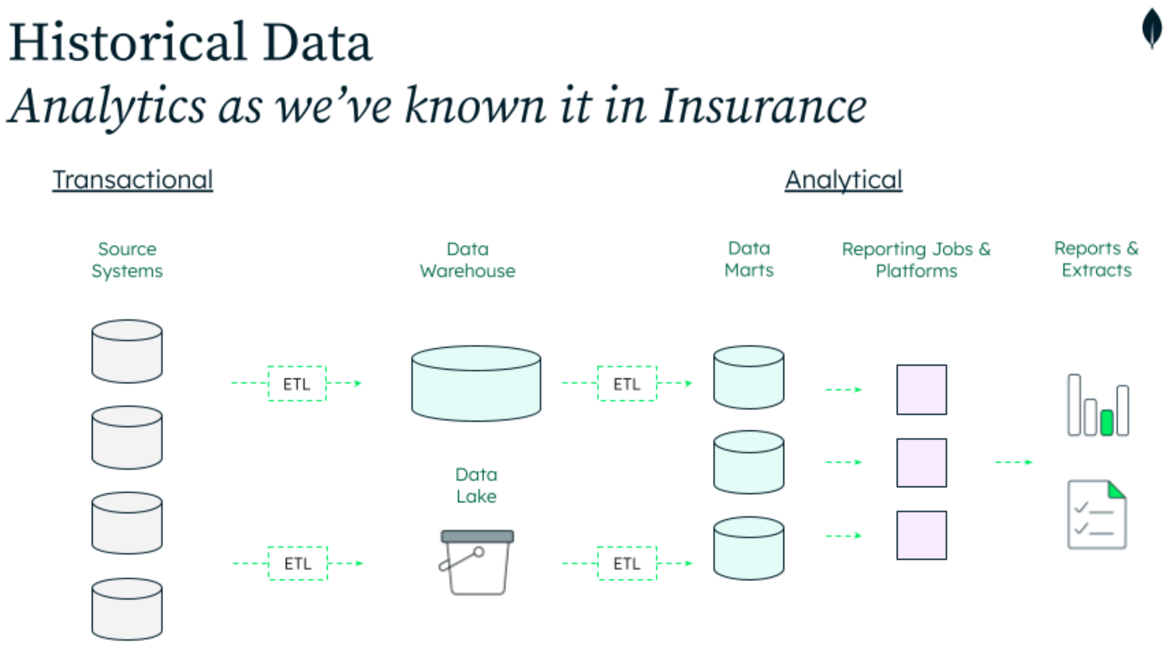
For many carriers, this increase in the amount of underwriting complexity and activity is at odds with broader organizational goals of reducing total operating expenses. While an insurer may want to be in the market with newer, data-driven solutions and products, the cost, time, and effort to bring those products to market may be high due to complexities associated with legacy systems and processes.

Big Data, Even Bigger Problems

We've had more than a decade to judge the impact of Big Data on the insurance industry. One thing clearly stands out; Big Data burned through big dollars, for some very big organizations, without necessarily delivering the big efficiency returns originally hoped for.

In many ways, the prevailing methods insurers have relied on to collect and analyze data have changed very little since the 1980s, when the era of data warehousing was born. Data is transacted against databases known as OLTP, or Online Transactional Processing. Nightly batch jobs then copy deltas from OLTP to OLAP, or Online Analytical Processing databases.

Then there's the EDW, or Enterprise Data Warehouse, a relational repository where the historical data of an organization is stored. Often, the volume and variety of data in these EDW datastores is high enough that some of the most critical and repetitive queries to support operational reporting do not perform well. The unique combination of data needed to satisfy these critical queries is often copied, once again, into data marts, unique data structures designed to answer specific questions, quickly. Power users can run ad-hoc queries against them, and business intelligence platforms can ingest the data from them, so that reports and extracts can be created.

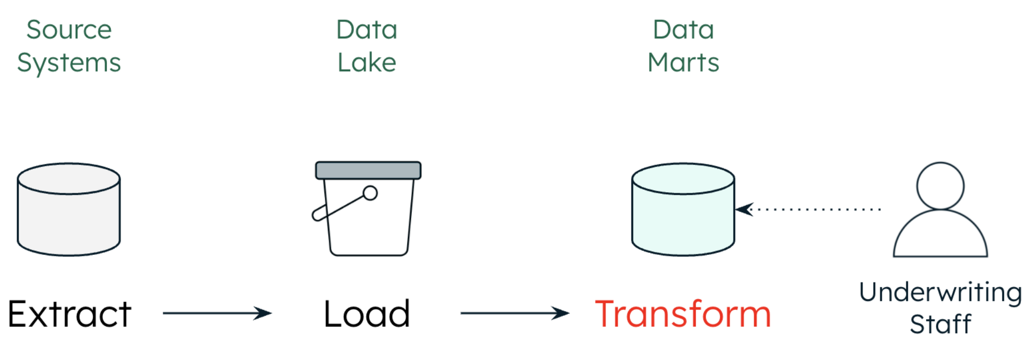


Data warehouse and data mart reports give underwriters key insights for how to modify coverages and price products effectively by analyzing past trends. Just ask an underwriter how critical claim-loss history is for understanding how risk, profit, and loss is trending within a portfolio.

And yet despite the criticality of these insights to the business, the speed at which IT support staff can move when it comes to implementing change, is limited by legacy processes and technologies. We know that a large swath of the analytics data estate runs on Relational technology, which requires structural changes to the underlying database objects before new items can be added. With that, comes the need for discipline and repeatable processes to ensure that changes, after they are developed, are vetted in QA environments before structural changes are applied to mission-critical production systems. In short, loading new data and attributes is not trivial when it comes to most relational data warehouse systems.

Data lakes, in contrast to Relational Data Warehouses, offer greater flexibility in terms of adding change. Thanks to object storage, the technology that underpins data lakes lets you move far more quickly by not requiring pre-defined structure. Unless you want a "data swamp", with vast amounts of confusing, unknown data and objects, you need similar upfront diligence and structure defined. Additionally, if you're an underwriter and you aren't a data engineer with programming skills like Python, getting data and insights *out* of data lakes can prove challenging. It's likely that you may still need the data to be copied from the data lake into a SQL-based data mart or data warehouse, so that you can make sense of it.

It seems the elephant in the room is that large amounts of data, when collected, still need to be transformed and structured in order for it to be consistently actionable and useful.



Data in the data lake may still need to be transformed for it to be accessible and usable by typical underwriting staff

How App-Driven Analytics and Machine Learning is Transforming Insurance

In today's digital economy, the speed at which you can present customers with the right information is key; instant quote comparisons, real-time premium calculations, straight through processing for underwriting.

To win in this brave new world of instant decision making, analytics processing must "shift left" from the data warehouse backend to the transactional business applications themselves.

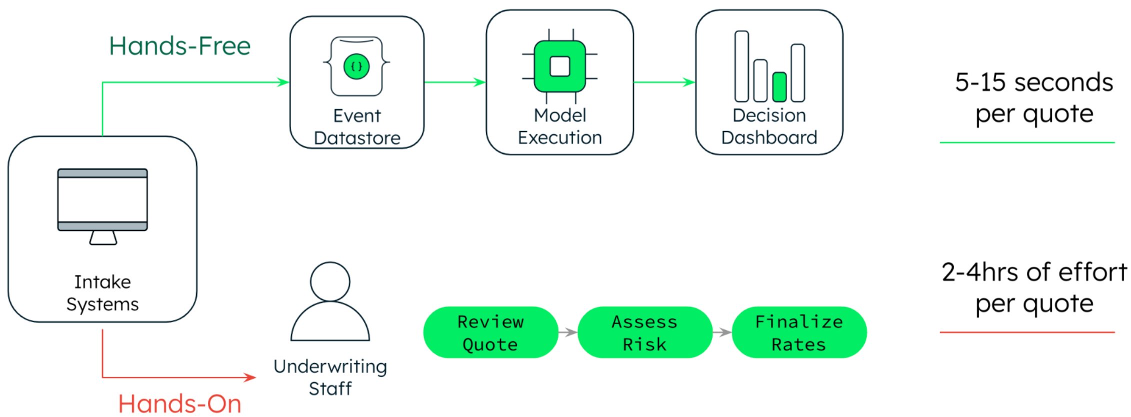
That means business applications need to work with embedded analytics and machine learning to bring insights to users in real time. This new breed of applications also needs to leverage an ever-growing volume of low-latency, live data, directly from its system of origin. This joined relationship between applications and real-time analytics is what we call application-driven analytics.

App-driven analytics stands in contrast to traditional applications that rely on much smaller data sets and often only create, rather than analyze, that data. In addition, traditional applications reserve analytics processing for dedicated analytics environments, such as the enterprise data warehouse.

In addition, the power of Machine Learning, embedded within an organization's transactional business processes and workflows, offers the possibility of doing far more with less time and people required.

Imagine for example, that your organization shifts from yearly to monthly written premiums for insurance policies. Will you hire twelve times the underwriting staff? Will you inundate your organization's most valuable decision makers with twelve times the workload?

Underwriting: Straight-Through Processing



Hands-Free vs Hands-On Undwriting example

It's a common phrase in some insurance carriers that "if an underwriter touches it, we lose money." The time spent by underwriting staff is incredibly costly, given how specialized and important the work is to the overall business. Machine learning and Artificial Intelligence can ease the burden on staff by intelligently triaging and routing work, helping your staff focus on more high value-tasks.

Reducing operating expenses at a time when investments must be made in modernizing technology and workflows is a challenge. Fortunately, the technology now available to insurance organizations can help tackle this challenge and propel the organization into an efficient and competitive digital future.

Simplifying the Data Tier Makes Application-Driven Analytics and AI/ML Work for Insurance

While insurers want to adopt app-driven analytics and AI/ML into their business applications, the challenges associated with bringing these solutions to market are significant.

Firstly, it is hard to build modern, mobile-first, real-time applications. It's also hard to train and deploy machine learning models. Bringing the two together to deliver demonstrable business results in a scalable manner is an even greater challenge.

Insurers face an uphill battle against:

- Siloed data in legacy systems
- Complex legacy data models
- Slow-to-change data analytics shops and processes

At the heart of this challenge is the need to work with vast amounts of varied data, and do so in real time. Working with data has always been the hardest part of building and evolving applications, as well as managing change in the data warehouse. Why? Because while how we use and interact with data has changed significantly over the last 40 years, the underlying data infrastructure has not.

- On the transactional side, a software delivery team may struggle to meet an SLA (service level agreement) due to the complexity and number of systems they must interact with to retrieve information, not to mention the various formats of data found within each.

- On the analytics side, a data engineering team may struggle to reduce the time and effort needed to add new objects into the data warehouse, due to the structural constraints of the underlying relational technology. And yet, it's the combination of transactional and analytics data that is now at the heart of what it takes to modern digital applications.

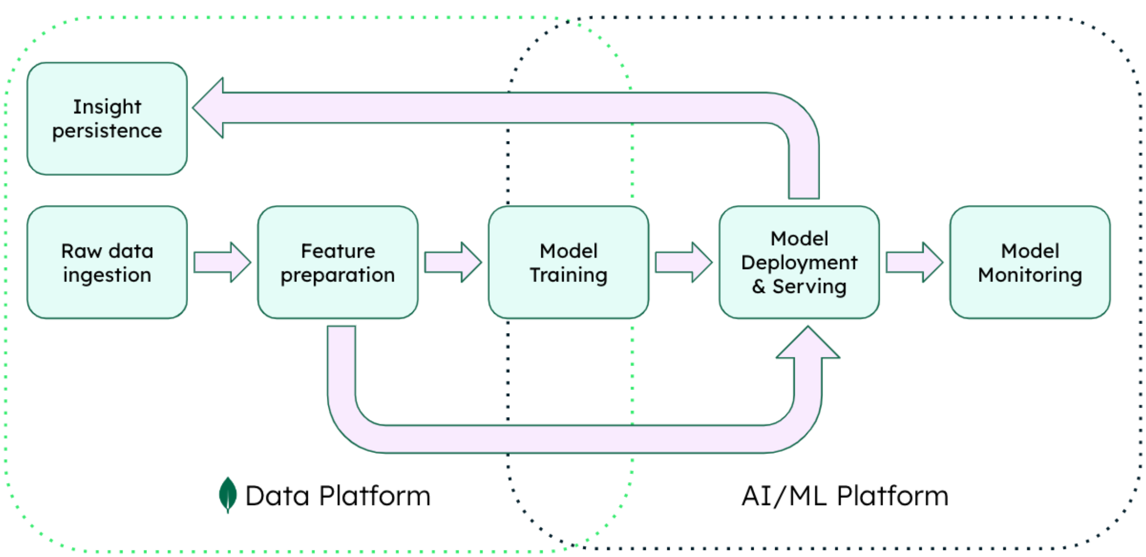
Real Time Brings Real Problems

The difficulties mentioned above are compounded when decisions need to be made in real time with "live" data, such as claim fraud detection or straight-through processing for underwriting automation.

To achieve real-time results, it's required to integrate an operational database into the architecture to stream real-time data and requests into an AI model and to persist the model output. In this hybrid system, we have both operational and analytics data requirements co-existing; this interaction adds to the overall architectural complexity of the system.

When it comes to leveraging low-latency, real-time data and events, it's important to note that raw data is typically not used "as is" when it comes to *training* AI models. First the data must be cleaned, potentially deduplicated and turned into features. Standardized techniques are required to do this, including binning, normalization, standardization, and one-hot encoding. MongoDB's aggregation pipeline provides powerful data processing capabilities that assist with this process.

MongoDB Atlas, the developer data platform, is capable of handling each of the above requirements from a single platform. Its analytical nodes and Data Lake allow for massive amounts of historical data storage, and service to the model for training purposes.



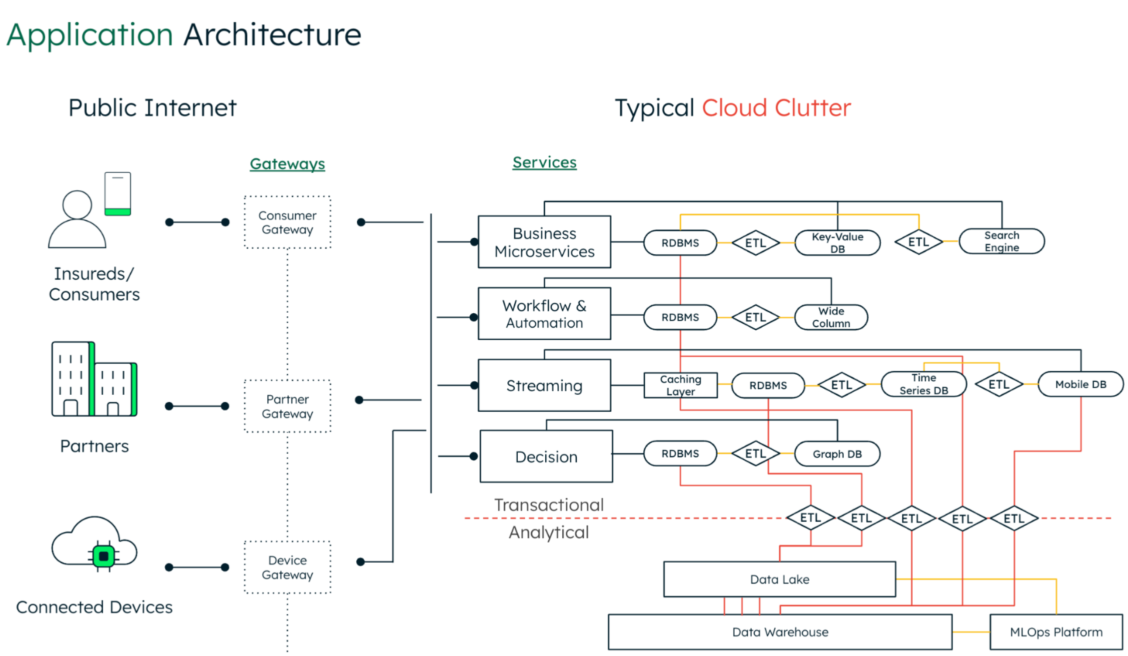
Real-time data can also be ingested and served through MongoDB Atlas via change streams, triggers, and integrations. MongoDB's powerful [aggregation framework] is capable of transforming raw data into usable features.

Lastly, integration patterns based on Spark, Kafka, and HTTP are supported out of the box, which greatly reduces overall architectural complexity.

Once decision-making models produce the new data output, it can be persisted back into the transactional database automatically actioned by additional automation tooling.

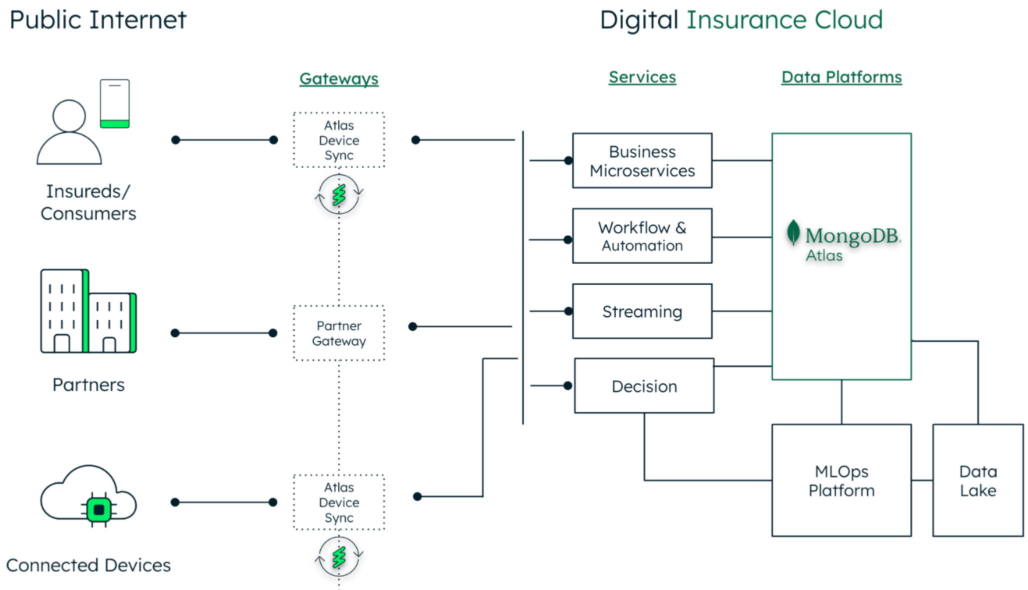
Complexity Kills Speed

MongoDB allows software developers to tame complexity at the data tier, so that they can build modern, digital applications that leverage AI-based decision making inside of new and existing business workflows.



The diagram above shows the sprawling 'spaghetti' architecture of niche databases and machine learning tools typically deployed at an insurance company.

Application Architecture



The second diagram (above) shows cloud data and solution architecture leveraging MongoDB at the data layer (which includes built-in search, mobile sync, mobile database, time series data etc). In this scenario MongoDB is paired with a dedicated AI/Ml platform. One database, no ETL, and faster results for customers and reduced time to value for insurers.

Using MongoDB, delivery teams can:

- Work quickly with highly varied and complex data sources and formats, thanks to the flexibility and power of the MongoDB Document Model. Ingest data from multiple siloed and disparate legacy data models and sources, as well as stream in and store vehicle and property telematics events from vendors, or IoT gateways directly. Whether it's Time Series, Geo-Spatial, Document, Relational, Graph, or Search, the Document Model is a superset of them all.
- Free themselves from the slow ETL change cycles of enterprise data warehouses and lakes, so they can better analyze and act on information in a truly real-time manner.

- Significantly reduce the time and complexity involved with modernizing away from legacy systems.
- Tame large amounts of information in the Data Lake, so that data scientists can spend more time training models and building features and less time trying to make sense of what might otherwise be data chaos.
- Architect and deploy comprehensive cloud solutions that are elegant and maintainable, and do not add additional complexity or burden into an already complicated landscape.

Half of the App-driven Analytics equation is the data itself, and how quickly your developers can work with frequently changing, complicated information. Developers using MongoDB Atlas can easily integrate the oldest, as well as the most modern data sources into a single version of the truth.

In addition the suite of features and tools built in to MongoDB Atlas, including search, mobile sync, charts, and many other leading security, scaling, and performance capabilities, empower developers to spend as much of their time as possible on the work that truly matters - building applications that deliver exceptional value and experiences to customers.

Paired with the ability to deploy across Google Cloud, Microsoft Azure, and AWS with the click of a button, MongoDB offers unparalleled flexibility and options for the developers of the digital era.

Learn More: Build an ML-Powered Underwriting Engine in 20 minutes with MongoDB and Databricks

About the authors

Jeff Needham

Jeff Needham is Principal of Industry Solutions, and Global Insurance Lead for MongoDB. Jeff brings industry thought leadership, as well as practical solutions and strategies to MongoDB customers. Prior to joining, he enjoyed a successful career as a data architect, working for both large insurance carriers, as well as independent software vendors.

Resources

For more information, please visit mongodb.com or contact us at sales@mongodb.com.

MongoDB for Insurance (mongodb.com/industries/insurance)
Case Studies (mongodb.com/customers)
Presentations (mongodb.com/presentations)
Free Online Training (university.mongodb.com)
Webinars and Events (mongodb.com/events)
Documentation (docs.mongodb.com)
MongoDB Atlas database as a service for MongoDB (mongodb.com/cloud)
MongoDB Enterprise Download (mongodb.com/download)
MongoDB Mobile (mongodb.com/use-cases/mobile)

Legal Notice

This document includes certain "forward-looking statements" within the meaning of Section 27A of the Securities Act of 1933, as amended, or the Securities Act, and Section 21E of the Securities Exchange Act of 1934, as amended, including statements concerning our financial guidance for the first fiscal quarter and full year fiscal 2021; the anticipated impact of the coronavirus disease (COVID-19) outbreak on our future results of operations, our future growth and the potential of MongoDB Atlas; and our ability to transform the global database industry and to capitalize on our market opportunity. These forward-looking statements include, but are not limited to, plans, objectives, expectations and intentions and other statements contained in this press release that are not historical facts and statements identified by words such as "anticipate," "believe," "continue," "could," "estimate," "expect," "intend," "may," "plan," "project," "will," "would" or the negative or plural of these words or similar expressions or variations. These forward-looking statements reflect our current views about our plans, intentions, expectations, strategies and prospects, which are based on the information currently available to us and on assumptions we have made. Although we believe that our plans, intentions, expectations, strategies and prospects as reflected in or suggested by those forward-looking statements are reasonable, we can give no assurance that the plans, intentions, expectations or strategies will be attained or achieved. Furthermore, actual results may differ materially from those described in the forward-looking statements and are subject to a variety of assumptions, uncertainties, risks and factors that are beyond our control including, without limitation: our limited operating history; our history of losses; failure of our database platform to satisfy customer demands; the effects of increased competition; our investments in new products and our ability to introduce new features, services or enhancements; our ability to effectively expand our sales and marketing organization; our ability to continue to build and maintain credibility with the developer community; our ability to add new customers or increase sales to our existing customers; our ability to maintain, protect, enforce and enhance our intellectual property; the growth and expansion of the market for database products and our ability to penetrate that market; our ability to integrate acquired businesses and technologies successfully or achieve the expected benefits of such acquisitions; our ability to maintain the security of our software and adequately address privacy concerns; our ability to manage our growth effectively and successfully recruit and retain additional highly-qualified personnel; the price volatility of our common stock; the financial impacts of the coronavirus disease (COVID-19) outbreak on our customers, our potential customers, the global financial markets and our business and future results of operations; the impact that the precautions we have taken in our business relative to the coronavirus disease (COVID-19) outbreak may have on our business and those risks detailed from time-to-time under the caption "Risk Factors" and elsewhere in our Securities and Exchange Commission ("SEC") filings and reports, including our Quarterly Report on Form 10-Q filed on December 10, 2019, as well as future filings and reports by us. Except as required by law, we undertake no duty or obligation to update any forward-looking statements contained in this release as a result of new information, future events, changes in expectations or otherwise.

Appendix

The Two Types of App-Driven Analytics

There are two classes of Application Driven Analytics, each delivering unique value to insurers:

1. In-app analytics: Developers infuse analytics on live data directly within the operational flow of the application, enhancing user experiences and driving immediate user or app actions.

2. Real-time business visibility: Analytics teams generate insights directly against up-to-date application data using tools they are already familiar with and without causing any disruption to the application.

The table below differentiates the two classes of Application Driven Analytics across technical requirements, business outcomes they drive, and the personas that care.

	In-app analytics	Real-time business visibility
Query Latency	<1 second	Seconds to minutes
Data Latency	Seconds or less	Seconds to hours
Decision Scope	Operation (i.e. select next best offer, trigger preventive maintenance)	Process (i.e. run a promo, upgrade equipment)
Built by	Developers	Data Engineers
Consumed by	Application users (humans, algorithms)	Business Analysts, Data Scientists, and LoB decision makers