Streamlining for Growth

NOW Insurance's Data Journey with Onehouse

ONEHOUSE × NOW INSURANCE

Contents

Introduction

In the evolving insurtech sector, NOW Insurance is making significant strides with its data-driven and AI-powered approach to insurance. Founded in 2019 by industry expert Philip Cabaud, the company aims to transform the insurance experience for modern healthcare professionals. NOW Insurance emphasizes customizable, affordable, and flexible coverage options, using advanced predictive analytics and machine learning to streamline insurance processes. This focus on infusing AI and ML into its operations and products not only improves customer experience, but also gives NOW Insurance a distinct competitive advantage in the insurance market.

However, growth and innovation bring their own challenges, particularly in data management. As NOW Insurance expanded rapidly, the growing complexity of its data operations began to highlight the limitations of its existing data management systems, underlining the need for a more advanced approach to handle increasing data demands and support continued advances.

The partnership with Onehouse emerged as a pivotal solution, enabling NOW Insurance to enhance its data management strategy significantly and to build a strong foundation for data analytics capabilities. Through this partnership, NOW Insurance overcame obstacles in its data ecosystem — improving core infrastructure, and upgrading data and analytics capabilities, all while achieving new levels of operational efficiency and cost savings.

NOW INSURANCE

About NOW Insurance:

NOW Insurance, a data-driven, AI-powered Insurtech company, is designed to disrupt a centuries-old business model and deliver a range of customizable, affordable, and flexible coverage options that meet the needs and requirements of modern physicians and healthcare providers. By leveraging advanced, next-generation technologies such as predictive analytics and machine learning, NOW Insurance has transformed previously intricate and complex insurance processes into seamless, frictionless customer experiences. NOW Insurance was founded in 2019 by industry veteran Philip Cabaud to provide a best-in-class service for physicians and healthcare professionals in ways they have yet to experience with traditional, legacy insurers.

Streamlining for Growth: Tackling Data Engineering Challenges at NOW Insurance

In the initial phase of operation, NOW Insurance relied on a Ruby on Rails application hosted on AWS for its technical infrastructure, complemented by basic Google Analytics for data reporting. This foundational setup, while functional, limited the company's ability to fully use its data in pursuit of its customer-first objectives.

The main challenge arose from frequent and necessary schema changes in the rating engine, their primary business application. Early on, it was clear that these changes, critical for meeting new business needs, could cause instability in their data pipelines as data volume and complexity grew. The complex extract, transform, load (ETL) processes and the need to reconcile historical data with current records required complete table backfills for each schema adjustment over several years, a process that was both time-consuming and resource-intensive.

As NOW Insurance grew, the company took proactive steps to improve their data management systems. They focused on enhancing ETL pipelines and integrating large amounts of third-party data more effectively. NOW Insurance saw this as an opportunity to upgrade and future-proof their systems with the latest technology, best-in-class tools, and the best-performing platform for their current and future needs.

Additionally, NOW Insurance had a lean data engineering team. This created a pressing need for a data management solution that would enable the data team to achieve more in less time, without sacrificing accuracy or efficiency. The company sought a robust system capable of handling continuous, scalable data replication into their data lake—managing around 10TB of compressed third-party data, a data volume expected to double or triple each year—without the need for an overwhelming amount of manual input.

NOW Insurance's Integration of Onehouse for Data Control and Scalable Growth

NOW Insurance, under the guidance of their data specialist Jonathan Sims, undertook a strategic initiative to re-design and re-build their data management system and architecture from the ground up, using the latest technology and best platform available for their present and future needs. They sought a solution that combined the ease of use of a data warehouse with the expansive scalability of a data lake, emphasizing interoperability across diverse formats, engines, and cloud providers so as to avoid potential vendor lock-in. This journey led them to Onehouse, which stood out for its ability to execute near real-time data replication with minimal impact on NOW Insurance's operational systems, ensuring fresh data, without exorbitant infrastructure costs or extensive manual intervention.

Onehouse's integration of PostgreSQL's change data capture (CDC) feature helps NOW Insurance facilitate data replication into the data lake without burdening the online transaction processing (OLTP) databases. Most pressing to NOW Insurance's operations, the company is now able to handle schema changes confidently. With Onehouse, NOW Insurance can isolate unexpected data caused by upstream schema modifications or malformed records into quarantine tables. This ensures continuous data pipeline operations and reduces the need for manual intervention.

Onehouse's integration with technologies such as Debezium, Kafka, Spark, and Hudi not only simplifies the complexities of NOW Insurance's data infrastructure but also empowers the company to enhance its central data repository. The platform's advanced storage optimization techniques, capable of boosting storage performance by up to 100 times, address the challenges of large-scale data management. NOW Insurance uses Postgres CDC and S3 data loads with Onehouse to ensure the comprehensive preservation of their data history — encompassing all updates, deletions, and schema modifications. By using Onehouse's automated compaction process, NOW Insurance streamlines the merging of incremental updates and deletions, maintaining the integrity and organization of its data. These capabilities are particularly vital in light of the company's obligation to comply with data retention laws over extended periods.

Furthermore, with Onehouse, NOW Insurance is able to tap into a wide range of data transformation and management features — including pre-built, no-code transformations and the flexibility to incorporate custom code transformations — and leverage Apache Hudi's capabilities for table optimization, encompassing clustering, compaction, cleaning, and incremental processing. With this

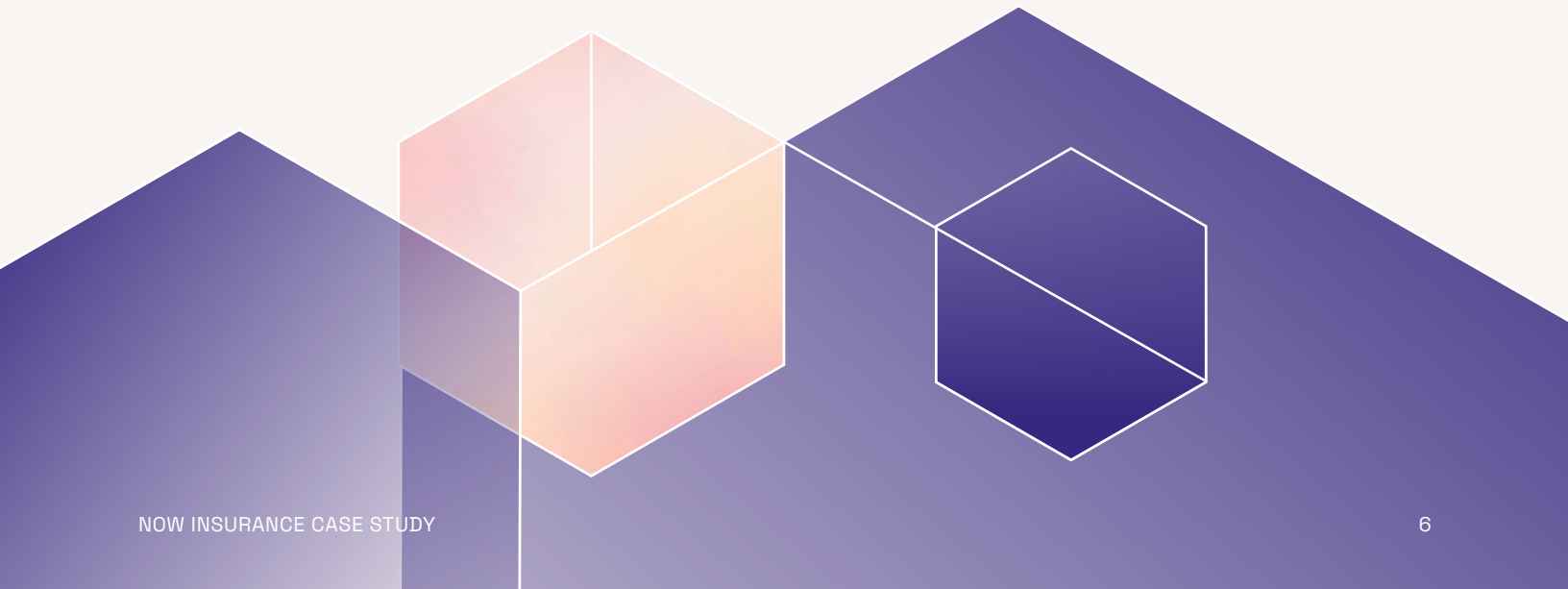
suite of features, NOW Insurance can handle high-volume data replication tasks, all while significantly improving the performance and cost-efficiency of data operations.

Some of the most useful features for NOW Insurance are:

- **Coarse-grained segmentation** to enable more efficient query filtering, greatly accelerating data retrieval and improving overall efficiency.
- **Auto file-sizing feature** to eliminate the challenges of managing small files, optimizing storage utilization.
- **Auto-cleaning functionality** to proactively manage storage by discarding outdated versions and failed commits, thereby contributing to cost savings.
- **Advanced clustering techniques**, which leverage multi-dimensional sorting algorithms to manage high-cardinality data, allow NOW Insurance to handle vast data volumes without compromising on performance or incurring excessive costs.

Finally, NOW Insurance enhances data security and governance by running Onehouse within their own virtual private cloud (VPC). This approach offers granular control over sensitive data and reduces reliance on external vendors. By effectively minimizing unnecessary data duplication and transfer, NOW Insurance ensures a more secure and efficient data management environment.

By implementing Onehouse's Universal Data Lakehouse architecture, NOW Insurance has not only navigated its initial challenges, but has also positioned itself to efficiently integrate and manage large-scale, complex data sources. This strategic implementation has set a new standard in their operational efficiency, security, and data-driven decision-making capabilities.



Harnessing Data, Cutting Costs, Driving Innovation: The Tangible Benefits of NOW Insurance's Collaboration with Onehouse

NOW Insurance's most pressing data challenge had to do with the ongoing changes to its data schemas and with the need to reconcile historical data — sometimes consisting of years-old datapoints - with current records. Onehouse's support for time travel and the ability to manage table partitions enables the teams at NOW Insurance to delve into complex data histories, including data that can arrive up to three years after a given event.

By automating the conversion of raw data into queryable tables and optimizing the management of machine learning model datasets, NOW Insurance enhances operational efficiency, even as the company navigates the complexities of handling around 10TB of compressed data spread over approximately 500 tables — all while enhancing the depth of data analysis which is essential for developing machine learning models for insurance underwriting. As a result, not only can NOW Insurance refine its machine learning models for underwriting, using detailed histories to improve both predictions and risk assessments, it has also freed up time to focus on strategic projects, such as testing and developing new machine learning applications for product innovation.

Finally, by implementing Onehouse as their managed data lakehouse, NOW Insurance achieves notable cost savings and operational efficiencies beyond what traditional data warehousing solutions like Snowflake or Redshift can offer, with the promise of even greater savings as their data volumes expand. This strategic move has significantly streamlined their data engineering processes, particularly in the maintenance of Apache Airflow on Astronomer. **Data engineering now saves around 30 hours per month on the development of new pipelines and 10-15 hours on upkeep, effectively cutting Airflow operational costs by 50%.** For an SMB with a lean data engineering team, these improvements translate into meaningful reductions in both time-to-value and costs.

Looking ahead, NOW Insurance aims to integrate data from Onehouse back into PostgreSQL databases, enhancing their insurance underwriting application. The company also plans to increase both the volume and variety of its data by three to four times over the next one to two years. This strategy is expected to significantly boost the development of machine learning and AI models, leveraging NOW Insurance's rich and diverse dataset to maintain a competitive edge in the insurtech sector.

Results

- **Enhanced Data Freshness:** For analytics on the production database, enabling effective legal compliance and strategic analytics through full data history retention, all while optimizing storage for up to 100x performance gains.

- **Streamlined Model Training and Development:** Datasets for machine learning models are now converted, stored, and organized efficiently, accelerating processing and reducing costs for about 10TB (compressed) of active data across roughly 500 tables.

- **Support for Complex Data Analysis:** Ability to handle time travel and manage table partitions for intricate histories, improving machine learning models for insurance underwriting.

- **A Universal Data Lakehouse to future proof data management and operations:** Plans for expanding the volume and diversity of data by 3-4x in the next 1-2 years to further enhance machine learning and AI model development.

- **Significant Cost Savings:**
  - Compared to traditional warehousing solutions like Snowflake or Redshift, **expecting savings to increase as data volumes grow**.
  - Monthly savings of approximately **30 hours per data engineer when adding new pipelines**, and about **10-15 hours in maintaining existing ones**.
  - **Cut Airflow operating expenses by 50%**, representing substantial savings in both time and financial resources for a small to medium-sized business.

> *"It's not just about managing data; it's about empowering our operations with efficiency, reducing costs, and maintaining the integrity and performance of our data infrastructure. The advanced features of Onehouse, including its sophisticated auto-optimization techniques, have allowed us to manage large volumes of data more effectively than ever, ensuring high performance and cost efficiency across the board. This strategic integration has not only simplified our complex data landscape, but has also set a new standard for operational excellence in the insurtech space."*
>
> **– Jonathan Sims, VP, Data & Analytics NOW Insurance**