snowflake®

HOW HEALTHCARE AND LIFE SCIENCES COMPANIES CAN LEVERAGE THIRD-PARTY DATA IN THEIR ANALYTICS

Enrich your data and unlock new insights faster with live, governed access to external data.

CHAMPION GUIDES

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Third-party data, which is data that comes from sources external to an organization, enables healthcare and life science companies to improve the quality of results from clinical trials and increase the ROI of marketing efforts through better audience segmentation and targeting. External data empowers teams to make better data-driven decisions, especially when it's integrated with first-party data. However, traditional methods for sourcing third-party data are inefficient and unsecure. Traditional data marketplaces may not scale, and legacy technologies for transferring data (FTP, APIs) can require extensive engineering work before the data can be used. This situation can result in delays, stale data, and poor data analysis.

In this ebook, you will learn how to:

- **Access live third-party data without any ETL,** making the data immediately available for analysis or to merge with your own data
- **Easily discover third-party data sets,** such as anonymized prescription data, medical sales data, COVID-19 data, and demographic data, that best fit your business needs
- **Use enrichment services to improve the quality of first-party data** by securely sharing slices of your data with providers

WHY FRICTIONLESS ACCESS TO EXTERNAL DATA CAN IMPROVE CLINICAL TRIALS AND BUILDING A PATIENT 360

Companies in the healthcare and life sciences industries are becoming increasingly data-driven, which has a number of implications. Pharmaceutical companies are optimizing drug development and commercialization through borderless data access, while payers and providers are improving patient outcomes through a more secure and seamless exchange of patient information. Meanwhile, distribution partners are sharing real-time insights, which helps to minimize supply chain disruptions.

While healthcare and life science companies have their own rich insights on patients, prescriptions, and outcomes, they're increasingly leveraging third-party data to obtain additional insights for the purposes of improving clinical trials and marketing ROI. The problem is accessing that data in a timely, cost-effective way that is compliant with industry data governance and privacy regulations such as HIPAA. Healthcare companies are often building and maintaining dozens, if not hundreds, of data pipelines and dealing with a wide range of data formats, which can result in their data scientists spending much of their time on relatively low-value data munging tasks. Complex ingestion processes can also result in dips in data quality and a high error rate.

Generated by individuals, businesses, and IoT sensors, third-party data originates from a variety of sources and exists in a wide range of categories, including but not limited to:

- **Consumers (demographics, psychographics, consumer sentiment)**
- **Health (genomic annotation, prescriptions, smoking prevalence, life expectancy)**
- **Point of sale (online, in-store)**
- **Mobility (location, in-store foot traffic, COVID-19 trends)**
- **Online behavior (searches, social media, app usage, web traffic)**
- **Media/advertising (consumption, measurement, viewership)**
- **Events (satellite and weather, event detection)**

By combining these third-party data sets with their own data, healthcare and life science companies can improve the accuracy of clinical trials and build a 360-degree view of customers to personalize content and offers and improve ad targeting.

TWO WAYS TO UNLOCK THE HIDDEN VALUE IN YOUR DATA

While the benefits of using third-party data from a data provider are clear, companies in the healthcare and life sciences industries can extract even more value from external data when they integrate it with first-party data and analytics.

Below are two examples of how third-party data, when combined with first-party data sets, can optimize clinical trials and improve the performance of marketing campaigns.

IMPROVE THE EFFICIENCY OF CLINICAL TRIALS

By leveraging genomic annotation data, real-world data from platforms such as IQVIA and Compile, and other third-party data sets, pharmaceutical companies can make their clinical trials more efficient and accelerate the usual timeframes for bringing a new drug to market.

Third-party data can also be used to identify confounding variables in a clinical trial, such as high rates of COVID-19 transmission that result in more participants getting sick and developing a fever (but not as a result of the drug they've taken). Drug makers can acquire COVID-19 epidemiological data and harmonize it with data they've collected from the trial, such as participants' average body temperature day by day, to understand where high rates of COVID may be skewing the results. From there, they can conduct COVID-19 testing in affected states or regions to rule out the possibility that COVID-19 is affecting their observations.

ENRICH PATIENT DATA TO BUILD A TRUE PATIENT 360

By leveraging third-party demographic data, life-event data, purchase data, and browsing data, healthcare and life science companies can enrich the customer information that's already in their databases and obtain a true patient/customer 360. From there, they can refine their segmentation and targeting strategies to increase the likelihood of patients, members, and customers receiving content and offers that are truly relevant to them, which makes their marketing spend more efficient.

For example, payers can advance value-based health management initiatives for at-risk member populations by ensuring that the right audience receives information about available care plans and services. This can assist in the management of chronic diseases, such as diabetes and heart disease, by ensuring that patients receive relevant information about carrier-sponsored wellness programs and benefit offerings at every stage of their health journey.

THE CHALLENGES OF SOURCING THIRD-PARTY DATA

While the reasons to source third-party data are sound, the legacy practices for doing so are inefficient. Two major obstacles stand in the way for many healthcare and life science companies:

- **Traditional methods of accessing external data require extensive time and monetary resources.**
- **Legacy file sharing methodologies are error-prone and not secure.**

Getting data from traditional data marketplaces can be overwhelming. How do you choose from the multitude of data marketplaces? How do you get started? How do you determine which vendors are reliable, what data will be most useful, and what value each data set really holds?

These questions point to the cost, time, and effort required to find and select the best-suited third-party data. It comes down to a problem of scalability. There's no efficient process for contacting each vendor, evaluating its data, and acquiring the data.

Organizations spend enormous amounts of money staffing teams to perform this time-consuming task. Some organizations let their data scientists muddle through the data acquisition process, but this takes expensive resources away from the task they were hired to do: build data models. Other organizations use data aggregators and brokers to facilitate data marketplace transactions. This solution has advantages but doesn't address the challenges presented by reliance on traditional file sharing methodologies.

Vendors continue to employ FTP, APIs, and other file downloading techniques to transfer data, which often requires them to copy files and perform intensive engineering work to extract, transform, and load (ETL) data. Even APIs are a burden, because developers need to maintain and troubleshoot multiple APIs from different data vendors. In addition, every API has different security and authentication methods, which puts the burden on an organization's security teams and poses potential risks. Additionally, APIs are inefficient for receiving large volumes of data.

But the biggest challenge with these data sharing methods is that they result in stale data copies that are nearly impossible to securely govern. Manual transfer efforts are prone to human error, and organizations face potential security and compliance issues. Worst of all, they are left with questionable data that results in poor analytics.

Taken together, these challenges raise this question: How can you use external data at scale without wasting time, money, and resources and without compromising on security and compliance?

THE DATA CLOUD SOLVES TRADITIONAL DATA SHARING ISSUES

If organizations could minimize time spent on building and managing ad hoc methods of data sharing, they could allocate more resources to extracting valuable insights from that data. But the numerous barriers presented by traditional data marketplaces and legacy data sharing practices need to be eliminated so organizations are empowered to discover and evaluate data sources with ease and combine external data with internal data for rapid analysis.

Snowflake's Data Cloud represents the modern answer for data. With its multi-cluster shared data architecture, Snowflake's platform centralizes data in a single, secure location in the Data Cloud: the network that connects Snowflake customers, partners, data providers, and data service providers across public cloud providers and regions. The result: Traditional data sharing barriers are removed and data silos are eliminated. Organizations immediately benefit from access to secure and governed data, which can be shared within and between organizations.

That's because Snowflake is built using Snowgrid: a unique technology that spans globally, connecting regions and clouds together, and enables secure and governed data sharing.

With Snowgrid, you can take advantage of Snowflake Secure Data Sharing features to share and access live, ready-to-query data across clouds and across regions, without any ETL or APIs. Anyone granted access to a data set simply references the data in a controlled and secure manner, without gaining physical custody of the data. This means that data access is revocable, enabling you to better comply with regulations such as the General Data Protection Regulation (GDPR). And because any changes made to the data are done to a single version, data remains up to date for all data consumers who have access, with minimal latency or concurrency issues.

With first-, second-, and third-party data unified in the Data Cloud, shared data can be combined instantly with existing data for faster analysis. Data is available in ready-to-query format without replication, transformation, or processing. Data analysis delays can become a problem of the past.

And, with its cloud-agnostic architecture, Snowflake enables organizations to have seamless and immediate access to shared data, regardless of cloud infrastructure, cloud provider (AWS, Azure, or Google Cloud Platform), or supported region.

SNOWFLAKE DATA MARKETPLACE: POWERED BY SNOWGRID

Snowflake's Snowgrid is the technology foundation for a new and modern data marketplace.

Data consumers can access live, ready-to-query third-party data and data services in Snowflake Data Marketplace. Rather than waste time hunting for vendors and downloading stale data, consumers can use Snowflake to easily discover, evaluate, and access live external data in a secure and compliant manner that is frictionless.

With the burden of data transformation removed, integrating external data with your existing data becomes fast and seamless. Data analysis can begin right away.

And you can improve your own data quality with enrichment services without having to go through the traditional steps of data copying and transformation, which are time-consuming and can pose security and compliance challenges. Instead, through Snowflake Data Marketplace you can share slices of your data with the data provider for enrichment and augmentation, and the enriched data is then securely shared back directly into your Snowflake account.

Whether you use external data to augment data sets for improved business analytics or to help train AI/machine learning (ML) models for data science, the benefits of Snowflake Data Marketplace are clear.

- **Ease of discovery:** Benefit from a single place to access a wide variety of data sets that can be queried, joined with internal data, used in data modeling, or added to BI tools—all with speed and ease.
- **Live, up-to-date data:** Never worry about stale data again. Without any manual intervention or scheduling required, all updates made by the third-party data provider are immediately reflected in your data sets.
- **Reduced costs:** Eliminate unnecessary data analytics expenses around data loading, transformation, and API integration and management. Because there's generally no data movement (only data access), you also don't pay storage costs for third-party data.
- **Personalization:** Request personalized, secure data feeds from data providers that are customized to your specific data needs.
- **Enriched internal data:** Use enrichment services to improve the data quality of your own data by securing sharing slices of your data with providers.
- **Global access:** Receive fast access to third-party data on multiple major cloud providers.

BECOME AN EMPOWERED DATA CONSUMER TODAY

With Snowflake Data Marketplace, the ability to securely access and quickly combine data from third-party sources can help improve the rigor of clinical trials and improve campaign performance when new drugs or medical devices are taken to market.

Discover for yourself the difference Snowflake Data Marketplace makes. Go to snowflake.com/data-marketplace to sign up for a free trial. There, you can find genomic annotation data, anonymized prescription and medical claims data, COVID-19 data, health expenditure data, and more, and you can start querying these data sets immediately.

There's only one question remaining: Are you ready to extract more value from your first-party data?

ABOUT SNOWFLAKE

Snowflake delivers the Data Cloud—a global network where thousands of organizations mobilize data with near-unlimited scale, concurrency, and performance. Inside the Data Cloud, organizations unite their siloed data, easily discover and securely share governed data, and execute diverse analytic workloads. Wherever data or users live, Snowflake delivers a single and seamless experience across multiple public clouds. Snowflake's platform is the engine that powers and provides access to the Data Cloud, creating a solution for data warehousing, data lakes, data engineering, data science, data application development, and data sharing. Join Snowflake customers, partners, and data providers already taking their businesses to new frontiers in the Data Cloud. **Snowflake.com**.