snowflake®

BEST PRACTICES FOR LEVERAGING THIRD-PARTY DATA IN YOUR 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 your organization, enables organizations to improve data analytics and decrease time-to-insight. External data empowers teams to make better data-driven decisions, especially when it's integrated with internal data. However, traditional methods for sourcing third-party data are inefficient and unsecure. Traditional data marketplaces do not scale, and legacy technologies for transferring data (FTP, APIs) require extensive engineering work before the data can be used. This situation results in delays, stale data, and poor data analysis.

In this ebook, you will learn how to:

- Access live third-party data from its original location, which is available immediately for analysis or to merge with your own data
- Easily discover third-party data sets that best fit your business needs
- Use enrichment services to improve the quality of internal data by securely sharing slices of your data with providers

EXTERNAL DATA SOURCES ARE NOW A DATA ANALYTICS IMPERATIVE

In today's fast-paced business environment, organizations are only as strong as the data they use. Over the last decade, companies have invested time, money, and resources to capture, store, and analyze their own data.

Although internal data may demonstrate how an organization can improve operations and customer relationships, few companies capture the breadth and scope of data needed to understand market shifts, uncover emerging and existing competitive intelligence, and monitor consumer habits.

What's missing is external data. Third-party data helps organizations uncover data-driven insights beyond the data they generate on their own. This expanded data analysis empowers companies to better understand broader industry trends, identify market shifts before competitors can react, analyze customer behavior to deliver new products or services, and forecast sales more accurately.

At a time when every strategic and financial decision hinges on data analysis, the question to ask is: **Why aren't we using more data?**

It's more important than ever to augment your internal data with third-party data sets. Forrester reports that 56% of global data and analytics decision makers have expanded their ability to source external data (or are in the process of doing so), while another 21% expect to do so in the next 12 months.[1] And Gartner predicts that, by 2022, more than a third of large organizations (35%) will be either sellers or buyers of data via formal online data marketplaces, which is an increase from 25% in 2020.[2]

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

- Online behavior (searches, social media, app usage, web traffic, geo-locations)
- Consumers (transactions, consumer sentiment)
- Individuals (employment and labor, credit)
- Businesses (advertising, pricing, ratings and reviews, store locations)
- Events (trades, satellite and weather, event detection)
- Aggregated data (IoT, web-crawled, B2B, open data)

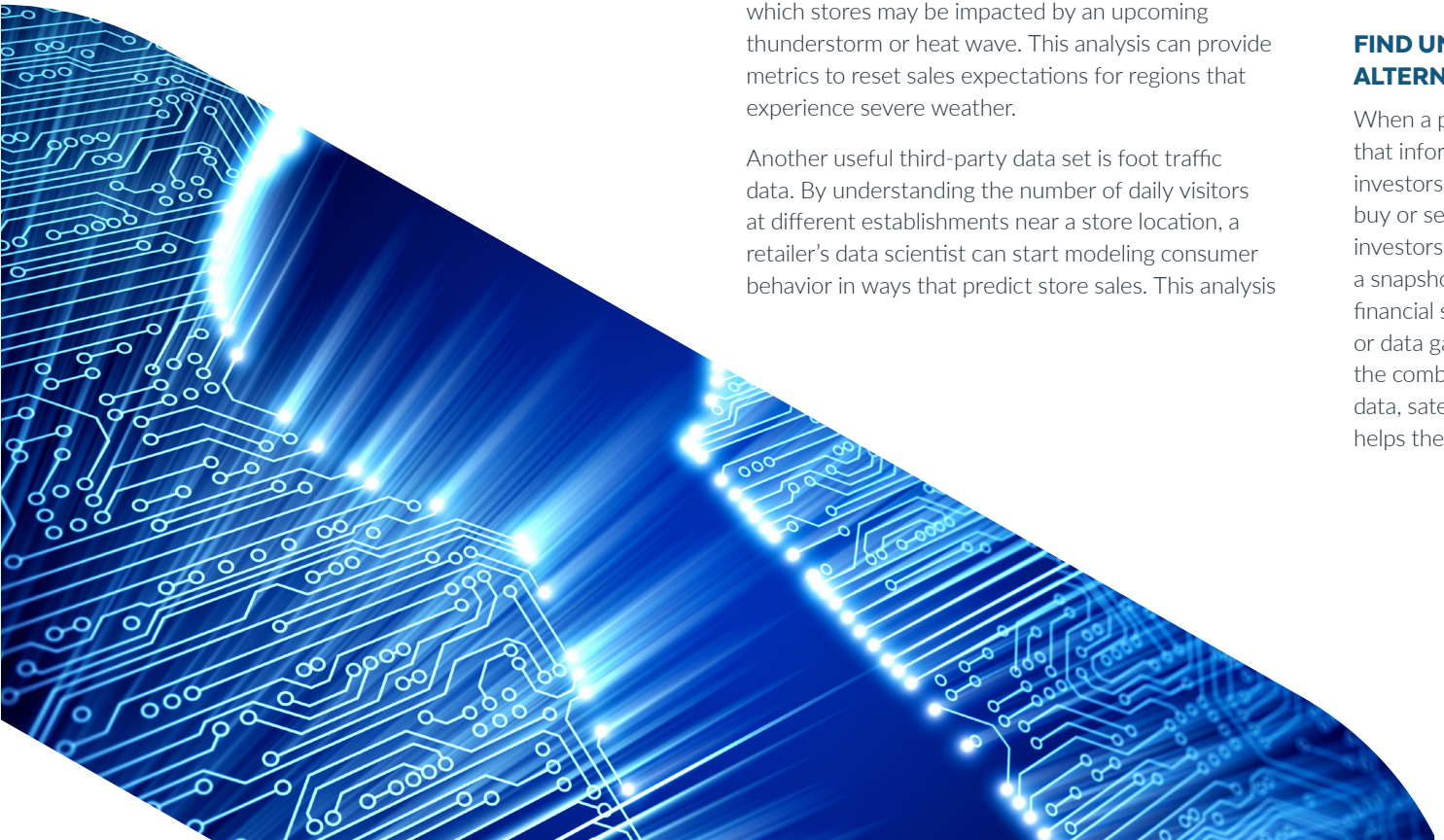
With new sources of relevant data, organizations reap benefits from improved data analytics, stronger data modeling, and decreased time to insight. Third-party data also helps validate hypotheses that are otherwise challenging to prove without external validation.

[1] bit.ly/2Fh8WzF
[2] gtnr.it/3nBrc8j

THREE WAYS TO UNLOCK THE HIDDEN VALUE IN YOUR DATA

While the benefits of using external data from a data vendor are clear, organizations extract even more value from third-party data when they integrate it with internal data and analytics.

To illustrate, here are three examples that demonstrate how external data can shed light on consumer behavior, financial analysis, and the impact of global events, regardless of industry.



ENRICH INTERNAL DATASETS FOR BETTER BUSINESS DECISIONS

Organizations that provide consumer goods or services often use internal data to understand store trends. For example, retailers analyze store sales numbers to compare store performance to sales forecasting. Now consider what happens when a retailer combines historical weather data with internal data to identify if weather patterns affect its store sales. In addition to looking at past trends, the retailer can use real-time weather forecast data to predict which stores may be impacted by an upcoming thunderstorm or heat wave. This analysis can provide metrics to reset sales expectations for regions that experience severe weather.

Another useful third-party data set is foot traffic data. By understanding the number of daily visitors at different establishments near a store location, a retailer's data scientist can start modeling consumer behavior in ways that predict store sales. This analysis can help the company better allocate inventory to accommodate demand patterns at individual stores.

Marketing teams have been using third-party data to spot consumer trends for a long time. By enriching customer data from CRM systems with demographics data, purchasing data, and online behaviors, marketers better understand their customers' interests and how to best incentivize them to do more business with the company. Sales teams can also benefit from this enriched data to better target their sales efforts and accelerate sales cycles.

FIND UNIQUE TRENDS WITH ALTERNATIVE DATA

When a public company issues quarterly reports that inform the public about its financial situation, investors use this information to decide whether to buy or sell shares in that company. However, smart investors recognize that a balance sheet provides only a snapshot in time. That's why hedge funds and other financial services organizations use alternative data, or data gathered from nontraditional sources. With the combination of social media streams, employment data, satellite imagery, and more, alternative data helps these financial organizations predict whether a

listed company will meet its earnings forecast before traditional financial numbers are announced publicly.

For example, financial analysis is bolstered by foot traffic data, which demonstrates whether consumers are heading in droves to a company's stores or if there's a downward trend. This information can be overlaid with demographic data to determine if in-store consumers match this company's target audience. By looking at anonymized credit card transaction data, product returns, and sales at different locations, analysts can obtain a richer view of the company and whether its numbers are likely to go up or down next quarter.

LEVERAGE EVENTS DATA TO UNDERSTAND BUSINESS IMPLICATIONS

To understand the impact of global events such as the COVID-19 pandemic, organizations across industries are using third-party data to make decisions about operations, employees, and customers. For example, a construction company uses COVID-19 data to make decisions about site closures by pinpointing infection incidences and cross-referencing those locations with its construction sites and how many employees work in the area.

A real estate management company is combining COVID-19 data with its own building occupancy data to share insights with landlords on how the spread of the disease may impact rent collection. Businesses with physical locations are sourcing health data and business recovery data (such as number of stores opened in a region or sales per store in a region) and overlaying these data points with business locations to make decisions about reopening offices and stores.

If external data can be so beneficial, what is preventing companies from using more of it? Traditionally, using external data has created challenges.

SOURCING THIRD-PARTY DATA CAN BE CHALLENGING

While the reasons to source external data are sound, the practice of doing so is inefficient. Two major obstacles stand in the way for many organizations, namely:

- Traditional data marketplaces are confusing
- 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 external data. It comes down to a problem of scalability. There's no efficient process for contacting each vendor, evaluating its data, and acquiring it.

Organizations spend enormous amounts of money staffing teams to conduct 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 copying files and intensive engineering work to extract, transform, and load data (ETL). Even APIs are a burden, as 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. Plus, APIs are inefficient for receiving large volumes of data.

But the biggest challenge with these data sharing methods is that they always result in stale data copies that are expensive to acquire. Manual transfer efforts are prone to human error, and your organization faces potential security and compliance issues. Worst of all, you're left with questionable data that results in poor analytics.

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

THE DATA CLOUD SOLVES TRADITIONAL DATA SHARING ISSUES

Sourcing and accessing external data shouldn't take longer than data analysis. The numerous barriers presented by traditional data marketplaces and old 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 all 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 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 enables secure data sharing, which means accessing live data shared from its original location. Data copies are never created, and data transmissions are never made. Data doesn't move.

Instead, a single, current, and centralized version of the data exists. Anyone granted access simply references the data in a controlled and secure manner, without gaining physical custody of the data. 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, without any latency or contention due to concurrent users.

Secure data sharing also means 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 become a problem of the past.

And, with its cloud-agnostic architecture, Snowflake enables organizations to have seamless and immediate access to all shared data, regardless of cloud infrastructure, geographic location, or cloud provider (AWS, Azure, or Google Cloud Platform). Secure data sharing even lets organizations share data with companies that don't have a Snowflake account, which makes it a global and inclusive feature.

SNOWFLAKE DATA MARKETPLACE: POWERED BY SECURE DATA SHARING

Snowflake Secure Data Sharing is the technology foundation for a new and modern data marketplace.

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

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 copy 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, which 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 no data movement (only data access), you also don't pay storage costs for third-party data.
- **Personalization:** Request personalized, secure data feeds that are customized to your specific data needs.
- **Enrich internal data:** Use enrichment services to improve the data quality for your own data by securing sharing slices of your data with providers.
- **All-access:** Receive fast access to third-party data on any major cloud provider.

BECOME AN EMPOWERED DATA CONSUMER TODAY

With Snowflake Data Marketplace, the ability to securely access and quickly combine data from third-party sources ensures the delivery of richer analytics, better insights, and more informed decision making.

Discover for yourself the difference Snowflake Data Marketplace makes. Go to snowflake.com/data-marketplace to sign up for free trial. You can discover data providers in the marketplace and start querying data immediately.

There's only one question remaining: Are you ready to mobilize your 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**.