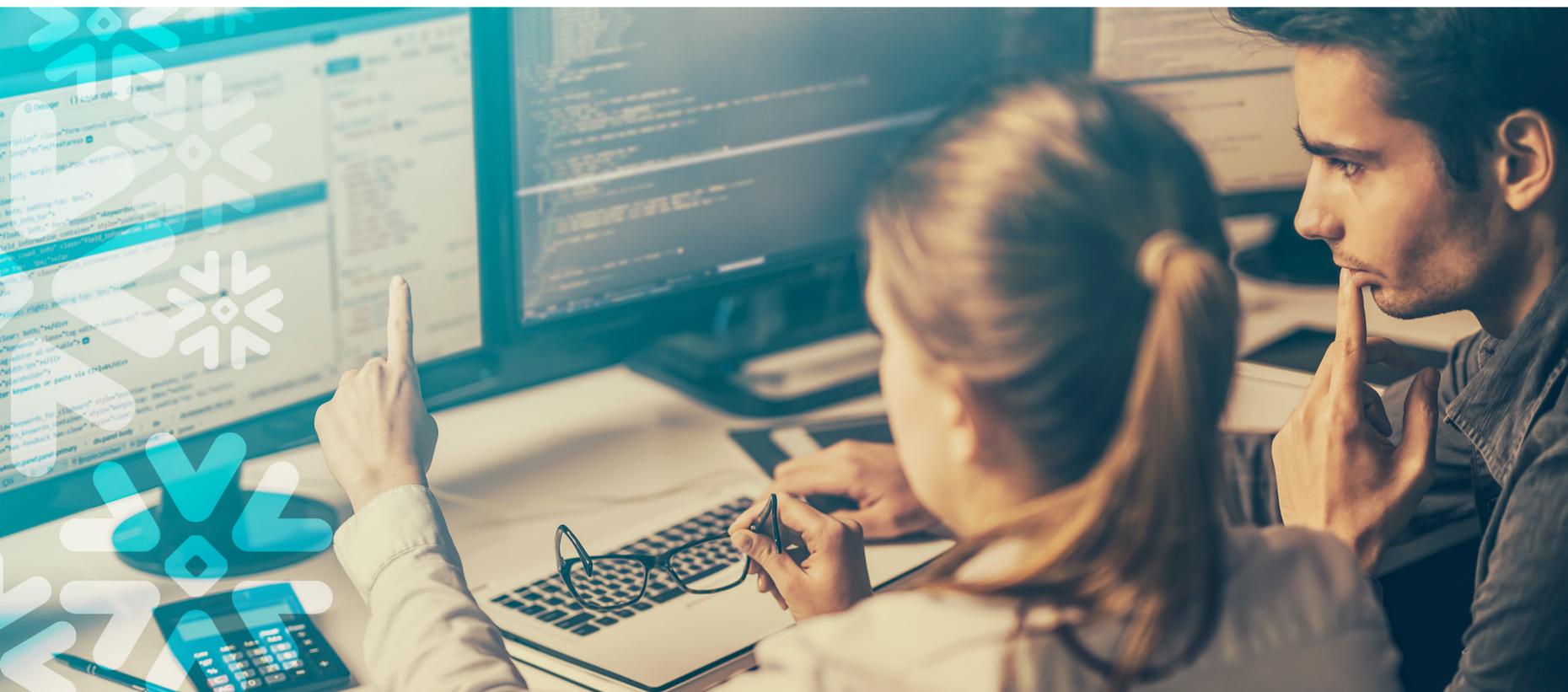# From Data Warehouse to Data Sharehouse™

**TRANSFORMING HOW ORGANIZATIONS CREATE
BUSINESS ASSETS FROM DATA**

snowflake

CHAMPION
GUIDES

What's inside

snowflake

Today's economy is not data-driven, not yet

Why insight derived from data inside the enterprise is no longer enough.

Data has always been at the heart of the modern enterprise. For decades, though, most IT efforts have focused on joining data created from enterprise applications and stored inside an organization— enterprise resource planning (ERP), customer resource management (CRM) applications, point of sale (POS)— and brought together in an enterprise data warehouse to support internal reporting and analysis.

In recent decades, data collected and stored within the enterprise expanded to website data, device data and log data. To store and process these new types of data, organizations have built increasingly complex solutions that enable them to explore and refine that data to enhance their internal reporting and analytics.

Today, it's clear that focusing on internal data for business insight has become dangerously inadequate. A significant share of the value derived from data exists outside the enterprise. External data combined with an organization's internal data can reveal insights never before thought possible. Here are just a few of the new business models available through modern data sharing:

- **Data sharing as a product.** Share live data as a service so data consumers can enrich their own existing data.

- **Data sharing for business efficiencies.** Share live data with business partners to optimize costs, streamline operations and better serve customers.

- **Data sharing to unite data silos.** Acquire a single source of truth by sharing among thousands of data silos across hundreds of business units within a single enterprise.

- **Data sharing as a product differentiator.** SaaS providers can offer the terabytes of data generated from their B2B subscribers' activity back to those subscribers as a full-blown analytics service.

Only modern data sharing would unlock an entire new world of economic potential. Nearly any company could easily create new business opportunities from external data assets. Empowered, these companies could address new markets, offer new and enhanced products and services and drive new revenue and business efficiencies never before possible. They could achieve all this just by leveraging external data stores and tapping the existing economic power of data inside their organizations in new ways.

Out with the old. In with the new.

## Why traditional data sharing methods inhibit innovation

To achieve well beyond current boundaries, organizations must quickly and painlessly share data. However, traditional approaches to sharing data between companies require stitching together a patchwork of disparate methods, since no true platform for sharing data was available.

The most common methods include email, FTP and file sharing. These methods move only low-level bits of raw data between data providers and the consumers of that data. That leaves both providers and consumers with huge effort, time and cost to make shared data usable. Providers spend time and effort extracting data from their own systems, staging that data and publishing information about the content of that data. Consumers must build the infrastructure to store that data and create a process to retrieve the data and updates to that data. They must also configure and maintain a system, such as a data warehouse, that can load that data and make it accessible.

More sophisticated tools like electronic data interchange (EDI) and application programming interfaces (APIs) still focus on moving data, leaving a significant amount of work for both data providers and consumers.

This reality, while considered acceptable for a long time, creates three formidable roadblocks to innovation:

- Moving data is slow, expensive and risky.
- The moment the data moves it becomes static and therefore stale.
- The significant cost and effort required make it impractical to share data for all but a few critical scenarios.
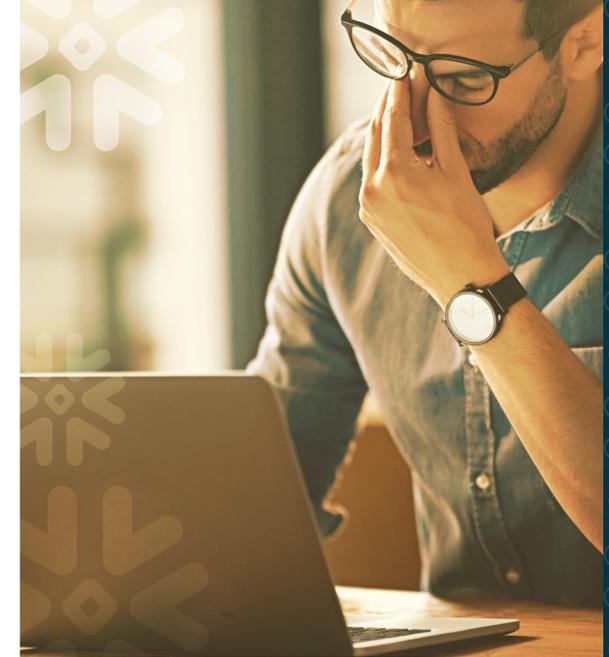
THREE FORMIDABLE ROADBLOCKS TO INNOVATION

» Moving data is slow, expensive and risky.

» The moment the data moves it becomes static and therefore stale.

» The significant cost and effort required make it impractical to share data for all but a few critical scenarios.

Disrupting data sharing

The data sharehouse has arrived

Modern data sharing: Extending the data warehouse to the data sharehouse.

What if data sharing itself were disrupted? Imagine the possibilities if companies could have on-demand access to ready-to-use, live data, and could make immediate use of that data. Data would no longer have to be deconstructed by the data provider, moved to the data consumer and reconstructed by the data consumer. It would be instantly accessible and ready to use inside a secure, governed solution.

A solution that delivers these benefits is now a reality. It extends the architecture and functionality of the modern cloud data warehouse as a platform for data sharing, without friction, transforming the data warehouse into a data sharehouse. Enterprises of any size can share read-only versions of their live, ready-to-use structured and semi-structured data in a secure and governed environment, consuming the same types of data from other organizations to augment their own data analytics.

The data sharehouse is possible only with a data warehouse built for the cloud. The scalability, elasticity and flexibility of the cloud are necessary in order to make it possible to store data from diverse sources and share that data among a large number of data consumers without contention and competition for computing resources.

A data sharehouse solves the challenges of data sharing with a new approach:

- **Eliminate movement and copying of data.** A data sharehouse offers direct access to live data in a controlled, secure solution.

- **Provide data in a ready-to-use environment.** A data sharehouse provides data consumers the full capabilities of a data warehouse, allowing them to use shared data within minutes and combine it with their own data. Security,
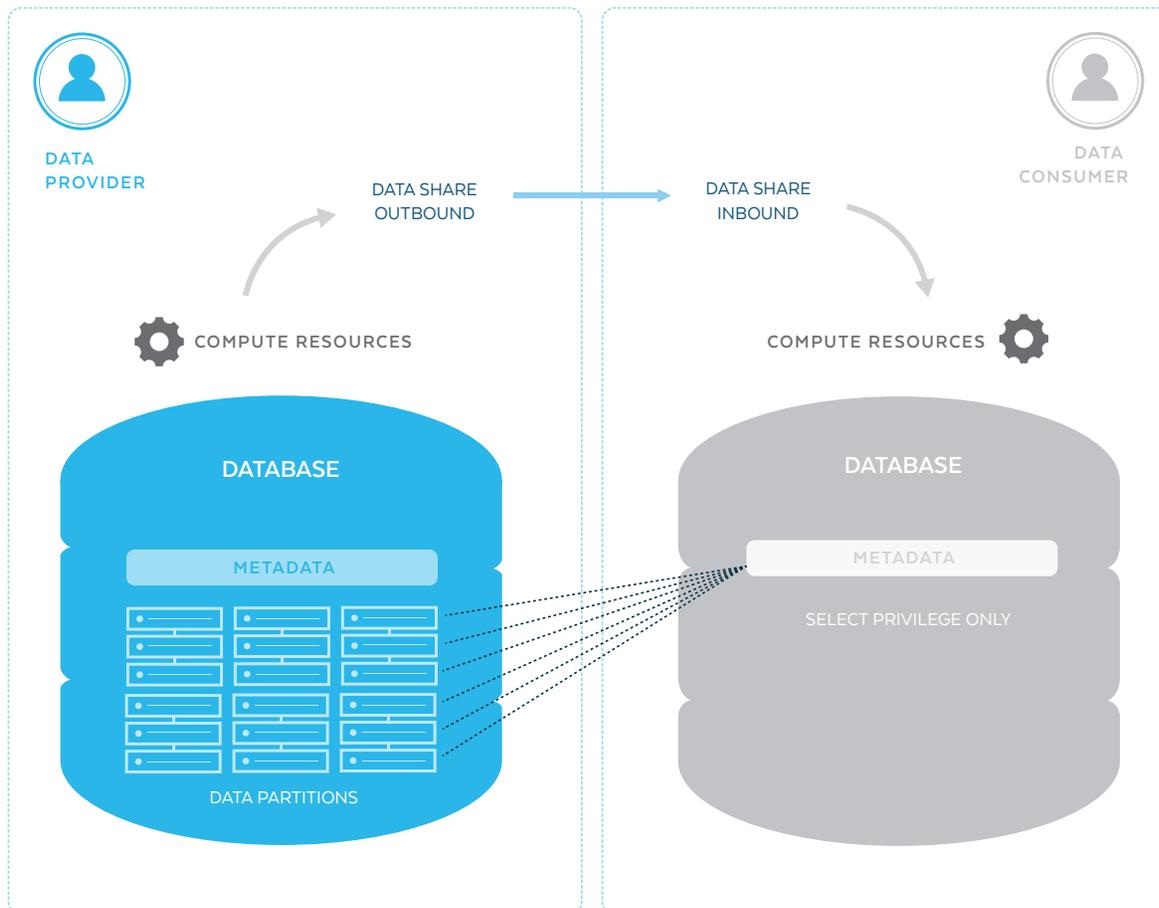
governance, data models and metadata are provided within the data sharehouse.

- **Share data without added costs.** A data sharehouse eliminates the duplicative costs of building the infrastructure needed to store shared data.

- **Share data with unlimited data providers and consumers.** Unlike data consortiums, which serve a limited number of companies, the data sharehouse is a market open to all organizations.

A data sharehouse eliminates the delays, cost and friction of existing methods, which provide only primitive mechanisms for data publishing, access and control. With modern data sharing, enterprises gain access to limitless business opportunities that never could be pursued with just their own data.

The technology behind the data sharehouse

How the data sharehouse works

DATA
PROVIDER

DATA SHARE
OUTBOUND

DATA SHARE
INBOUND

DATA
CONSUMER

COMPUTE RESOURCES

COMPUTE RESOURCES

DATABASE

METADATA

DATA PARTITIONS

DATABASE

METADATA

SELECT PRIVILEGE ONLY

MODERN DATA SHARING SHOULD BE FAST, EASY AND SECURE

Access to live data:
The data sharehouse provides direct access to live data. The data does not move. And unlike with the "black box" of a data drop, a data provider gains insight into how the data is being used through the shared metadata.

Data granularity:
Modern data sharing users gain access not merely to aggregated data but to data at fine granularity. Data consumers can combine and enrich shared data with their own data, without having to duplicate shared data.

Sharing options:
Data providers can maintain a master database and easily share specific objects from it with one or many consumers. Sharing custom views with many users is also easy.

Pay only for the resources used:
Data consumers pay only for the compute they need. They don't pay to store the shared data. Likewise, as noted above, data providers pay only for the storage and compute resources they use.

Data sharing spotlight: Localytics

Speeding direct access to data collected from billions of devices

Localytics is a Boston-based company that provides a mobile engagement platform used in over 37,000 apps on more than 2.7 billion devices worldwide. Localytics gives hundreds of the world's top brands insights about their mobile users and the tools to engage with those users.

Localytics Direct Access, powered by modern data sharing, gives customers access to Localytics' data without exporting that data, solving one of the biggest data challenges marketers face. Previously, users had to connect different sources of customer data from CRM systems, business intelligence (BI) tools, mobile analytics and other sources. Data was often exported and copied into other platforms, manipulated and further analyzed. The process produced multiple copies of the data living in many places, increasing costs and complexity, and producing inconsistent results.

ETL ELIMINATED

Direct Access removes the burden of cumbersome extract, transform and load (ETL) efforts and makes data directly accessible through modern data sharing, creating a much more efficient and reliable way to manage and understand customer data. Specifically, Direct Access employs secure, permissions-based access to enable customers to work with session, event and profile data from Localytics and run their own queries and custom reports against that data. Customers can also connect BI tools, such as Tableau and Looker, directly to their data.

DATA LATENCY REDUCED FROM THREE HOURS TO THREE MINUTES

The Direct Access service stores all its data in a modern cloud data warehouse, augmented with modern data sharing to become a data sharehouse. Instant sharing of live data with Localytics' customers eliminated a previous data latency of three hours. With modern data sharing, fresh data is ready to query in about three minutes. Customers don't need to expend any effort to use the data, which is appropriately filtered through permissions-based sharing.

Localytics

"Customers immediately saw that modern data sharing could potentially eliminate their ETL processes. That's extremely important because they have over-burdened data teams. Our customers are very happy because our solution saves them a lot of time and effort."

MICHAEL KLOS
DIRECTOR OF ENGINEERING, LOCALYTICS

A single solution for a multitude of data types

Extract more value from more information

The rising tide of data has been well-documented, with industry estimates attributing up to 90 percent of the world's data as having been produced in the last two years.[1] From these vast global stores, only 0.5 percent of all available data is effectively analyzed and used.[2]

The data quantity phenomenon is magnified by the rapid proliferation of newer data types: cloud, mobile, Internet of Things (IoT), data from wearables and mushrooming amounts of data from commercial and government sources. Together, these types are a jumble of structured and semi-structured data, requiring significant preparation if an enterprise has to store that data before analyzing it.

MODERN DATA SHAREHOUSING
DEMANDS THE MODERN DATA WAREHOUSE

Enterprises can land all structured and semi-structured data of any size into one place, in a data warehouse built for the cloud. Once there, semi-structured data can be easily joined to structured data using SQL extensions. Users can query all of this data immediately using SQL, the standard language of data exploration and analysis.
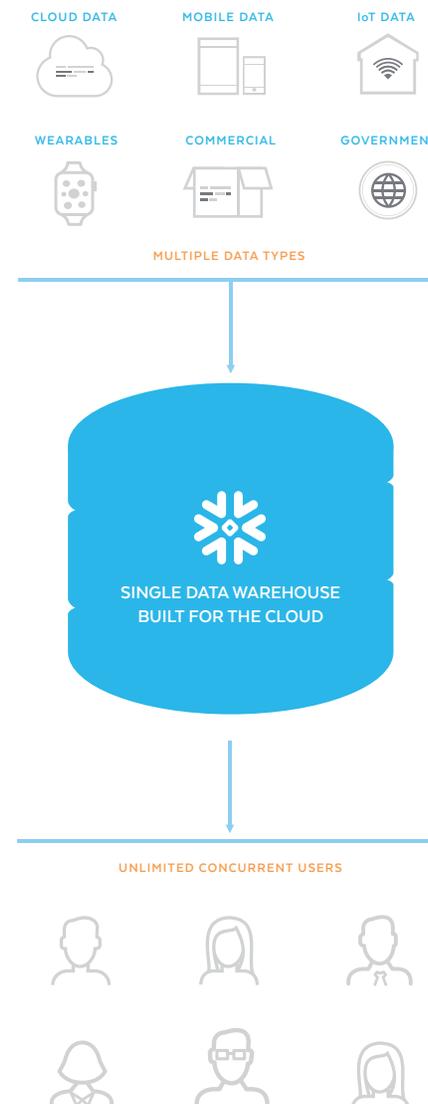
Modern data sharing applies a multiplier effect to the modern data warehouse built for the cloud. Enterprises can share any part of their data warehouse with other

enterprises with simple, read-only, permission-based access. Data sets no longer need to be deconstructed, moved and reconstructed. Instead, they can be instantly shared.

Modern data sharing is possible only if the cloud data warehouse is architected in the following ways:

- **Compute and storage are separate from each other.** Data consumers can apply their own compute power to access, integrate and analyze read-only copies of another enterprise's data in the warehouse.

- **Scaling up, down and out (concurrency) are limitless.** Unlimited numbers of users can access unlimited compute and data resources without impacting the performance of other, existing operations.

- **Separate services layer for shared meta data.** A global metadata map enables large numbers of data sharing consumers to easily find subsets of data without having to search the entire data warehouse.

The business impact of modern data sharing is real and immediate. But the only way to make it easy and effective for data providers and consumers is with a built-for-the cloud data warehouse designed with data sharing in mind.

CLOUD DATA   MOBILE DATA   IoT DATA

WEARABLES   COMMERCIAL   GOVERNMENT

MULTIPLE DATA TYPES

SINGLE DATA WAREHOUSE
BUILT FOR THE CLOUD

UNLIMITED CONCURRENT USERS

[1] "What is big data?" IBM, December 2016. Retrieved from https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html
[2] "Only 0.5% of All Data is Currently Analyzed," A.R. Guess, June 10, 2015. Retrieved from http://www.dataversity.net/only-0-5-of-all-data-is-currently-analyzed/

Instant access to live data

Share live data quickly, securely and without any data movement

DATA PROVIDER

PROVIDES INSTANT PERMISSIONS-BASED-ACCESS

SINGLE DATA WAREHOUSE BUILT FOR THE CLOUD

UNLIMTED CONCURRENT USERS

An enterprise should be able to share data, regardless of data consumers, in way that's fast, secure and extremely simple. The data sharing process should eliminate the complexity of traditional data sharing methods—the alphabet soup of FTP, EDI and API.

SECURE BY DESIGN

Permissions-based access is a central tenet of the modern data warehouse and, therefore, modern data sharehousing. With just a few simple steps, the data provider can provide instant data sharing access to other data warehouse users. And, thanks to individualized access rights, each user gets filtered, personalized views of the same live data without the headache of creating different versions of a table. In this way, the data provider can provide any number of data consumers with direct, frictionless access to live data, as business happens.

DATA SHARING ON DEMAND

As part of a modern cloud data warehouse, data sharehousing is delivered as a service, a fully managed offering. All cloud data customers should be free

from managing any aspect of the data warehouse infrastructure. With compute separate from storage, compute clusters can be dynamically or automatically spun up or down and allocated to specific queries, users or workgroups. And a modern cloud data warehouse enables an infinite level of concurrent processing against the same data without impacting performance.

Modern data sharehousing extends this concurrency to individuals or workgroups external to the enterprise. When a data provider issues permission to the data consumer, read-only data can be accessed instantly, on demand.

CONSTANTLY UPDATED FOR A SINGLE SOURCE OF TRUTH

Sharing live data requires up-to-the-minute updates from the data provider and incorporated into shared data regardless of the number of data consumers accessing the data. With modern data sharing, no updating effort is required by either the data provider or consumer, helping to ensure that decisions are made on the most recent data.

Exploring the economic impact

Three of many ways to share and monetize data assets

Because traditional barriers of on-premises and cloudified data warehouses are removed, modern data sharing enables data providers to quickly monetize their data. And, for data consumers, live, secure data sharing reduces time-to-value. As data is shared, value is unlocked. In this way, modern data sharing can introduce at least three new economic opportunities. All of these usage scenarios enable enterprises to share data as packaged and monetized assets, quickly and securely, powering a true data-driven economy:

DIRECT REVENUE

Many companies produce and sell data, some of which started nearly 100 years ago. Now, any company can turn their data into business assets by granting access to slices of their data warehouse.

This low-cost, low-friction solution enables data companies to immediately meet customers' urgent demands for fresh data with up-to-the-minute accuracy inside a modern cloud data warehouse.

DATA SHARING WITH BUSINESS PARTNERS

Sharing data directly with business partners is not new. But effortless sharing of live data is. Data sharehousing enables:

- Instant data sharing
- Partner queries against live data
- A higher velocity of downstream practices executing against the data

These benefits come without the need to move data using complex data pipelines and at no cost for sharing data.

BREAKING DOWN ENTERPRISE AND ACQUISITION SILOS

Modern data sharing sidesteps the data silo problem by allowing structured or semi-structured data to be stored in a modern cloud data warehouse. Data can be shared seamlessly, without downloading or replication. Systems that were previously siloed can now be tightly integrated, without manual integration or data pipelines.

And data sharing should only exist as the technology enabler. All business arrangements between data providers and consumers are their own.

Data sharing spotlight: Snagajob

Accelerating data delivery to keep customers engaged

Snagajob, America's leading marketplace for hourly workers, connects more than 75 million registered workers with jobs at 300,000 employer locations. Its mobile sourcing and workforce management tools drive results for businesses ranging from single-location operators to Fortune 500 giants.

ENGAGING AT-RISK CUSTOMERS AT PRECISELY THE RIGHT MOMENT

Keeping both job seekers and employers engaged with the site is critical. Snagajob partners with an external marketing analytics firm for email outreach to employers using the site. Specifically, the external partner firm identifies at-risk business customers in need of re-engagement. Modern data sharing plays a critical role in optimizing the matching process, enabling the partner firm to run queries on live Snagajob data. In doing so, the firm can identify and reach out

to at-risk business customers at the right moment with the right incentive to renew their relationship with Snagajob.

BOOSTING REVENUE WHILE REDUCING COSTS

Modern data sharing helps Snagajob achieve two critical goals: revenue stability, via relationship continuity, and cost reduction. Both Snagajob and their marketing analytics partner realize significant cost savings through modern data sharing, since it eliminates the need to move data using complex data pipelines. The external firm also benefits as a data consumer, using data sharing capabilities at no cost.

snagajob

"For Snagajob, modern data sharing really is the Holy Grail. Ever since our implementation went live, I haven't touched it. Conversely, if you do data extracts and physically ship the parts, there's always something breaking. With modern data sharing, I haven't had to worry about this project at all. It just runs. Plus, the entire implementation was extremely fast."

ROBERT FEHRMANN
PRINCIPAL ARCHITECT, SNAGAJOB

Why build when you can buy?
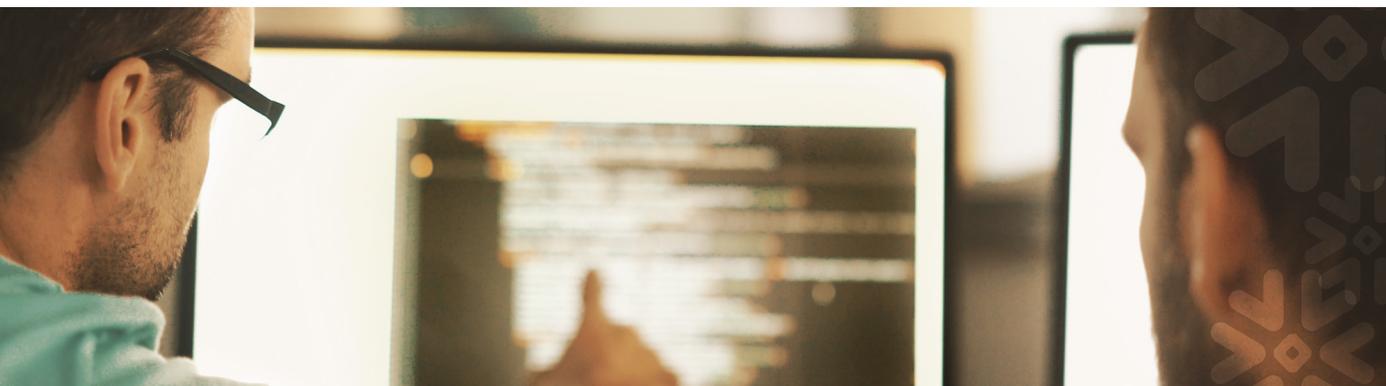
Modern data sharing delivers turnkey value

Build versus buy is a long-running enterprise technology debate, sparked anew in the cloud data sharing arena. Modern data sharing offers turnkey value with compelling capabilities that can't be easily built with commercial cloud resources.

- **Convert data into business assets.** Enterprises can easily share data, unlocking new data monetization opportunities. For the first time, they can drive growth by deriving fast time-to-value from data assets in the cloud.

- **Store once, share with many.** Modern data sharing stores data once and shares a data warehouse with any number of business partners or clients, without any data movement. It significantly simplifies data sharing by eliminating the pains of current data sharing approaches such as FTP, API, EDI or email, saving time, money and effort.

- **Direct access to a live database.** Business partners and clients can run analytics on always-live data. Modern data sharing ensures that partner analytics are fresh, thereby enabling data consumers to take advantage of opportunities as business happens.

- **Individual, secure views to data.** Modern data sharing has the flexibility to provide customizable access based on business needs, by partner, even with overlapping data, creating better business experiences for data providers and consumers.

- **Data sharehouse users only pay for what they use.** Data providers share read-only views of their data so they don't pay for duplicate storage. Data consumers pay only for the compute they use on shared data

Find out more

Transform how your organization turns data into business assets

Snowflake is the only data warehouse built for the cloud. Snowflake delivers the performance, concurrency and simplicity needed to store and analyze all data available to an organization in one location. Snowflake's technology combines the power of data warehousing, the flexibility of big data platforms, the elasticity of the cloud, and live data sharing at a fraction of the cost of traditional solutions. Snowflake: Your data, no limits. Find out more at **www.snowflake.net**

❄️ snowflake