HOW SNOWFLAKE DATA CLEAN ROOMS HELP PUBLISHERS AND MARKETERS IMPROVE AD EFFECTIVENESS AND PRESERVE PRIVACY

snowflake

WHITE PAPER

TABLE OF CONTENTS

Why a Cookieless World Requires a New Targeting and Measurement Approach

Consumers have become increasingly concerned about online privacy in recent years. According to a 2023 Pew Research study, roughly four in ten Americans say they are very worried about companies selling their information to others without them knowing (42%) or people stealing their identity or personal information (38%).

Meanwhile, Nielsen's latest Consumer Survey Report, launched in November 2023, revealed that the majority of consumers actively avoid ads across various platforms such as podcasts, streaming services and live TV. This poses a significant challenge for advertisers. Further, the fall 2023 edition of the CMO Survey indicates that inflation and the uncertainties it causes continue to suppress marketing spend.

As marketing leaders remain under fire to prove the ROI of every dollar spent on advertising, they rely on attribution modeling to quantify the impact of each media campaign and ensure spend is optimized. These models have generally relied on common identifiers such as third-party cookies and mobile ad IDs that track people's online activity, as do targeting approaches for delivering personalized ads and content.

However, privacy measures implemented by some of the biggest players in the advertising ecosystem have made such legacy methods of targeting and measurement increasingly precarious. With the general depreciation and increased regulation of common identifiers, data clean rooms have emerged as a strategy for marketers to conduct sophisticated audience segmentation, targeting, attribution and measurement, equipping them to build lookalike models while preserving privacy.

ADDRESSING REGULATORY COMPLIANCE

Media publishers and marketers are facing heightened pressure to meet increasingly stringent data privacy obligations, as they confront both consumer pressure and a surge in regulatory scrutiny. They must comply with strict regulations that affect how to manage and share data, personalize content and target consumers. For example, the European Union is targeting hyperpersonal advertising on social media with the Digital Services Act, which will ban any targeted online ads based on an individual's ethnicity, religion or sexual orientation. Companies that do business in Europe already must comply with the General Data Protection Regulation (GDPR), which imposes requirements on how they collect, store and use personal data. In the U.S., California Privacy Rights Act (CPRA) limits businesses' use of sensitive personal information.

At the same time, company leaders are eager to apply their vast troves of data to improve the customer experience and open new revenue streams but are cautious given industry regulations. Data clean rooms provide a way forward.

THE VALUE OF DATA CLEAN ROOMS

Data clean rooms are secure platforms that allow two or more parties to join and query datasets, without direct access to the other's data.

With data clean rooms, companies can securely share data views and allow data to be joined to other datasets for modeling and analysis without direct access to the other party's data. Data clean rooms can help publishers and marketers improve ad effectiveness and, ultimately, increase ROAS (return on ad spend) without having to directly share the user data and activity data they once made available to marketers.

This white paper explains how Snowflake Data Clean Rooms work and the value they deliver to both marketers and publishers. It also explores how they diverge from other clean room technologies in terms of secure data collaboration, privacy controls, scalability, secure data enrichment and trustworthiness.

HOW SNOWFLAKE DATA CLEAN ROOMS WORK IN 7 STEPS

1. STORE

A marketer and publisher that are existing Snowflake customers decide to launch a data collaboration. They keep their customer data in their own Snowflake accounts, with no need for further ETL or to send copies of that data anywhere else. This helps protect raw data from being viewed or copied by the counterparty without express permission.

2. JOIN

They determine which data and how their data can be joined, taking into account any (individual) data restrictions and each of their goals. The simplest way is with something they both already have, such as lowercased and trimmed email addresses or IP addresses. They can also use advanced forms of identity joining, graph joining, waterfall joining and Boolean expression joining — or they can leverage a third-party identity provider on Snowflake Marketplace, such as Acxiom, LiveRamp or Neustar.

3. CALIBRATE

They figure out how to query the joined data for the intended use case. For example, a retailer may want to query a publisher's anonymized engagement data by content category and find overlaps with its own purchase data. To better understand what types of content its customers consume, the retailer needs to define a query with the appropriate SELECT, GROUP BY, JOIN and WHERE clauses.

4. APPROVE

The requested query then has to be approved by the counterparty, which automatically ensures the query meets the counterparty's rules for joint usage of its data. For example, one such rule could be to disallow any analyses or to suppress any output rows that aggregate fewer than 75 distinct people or devices, to prevent unintended user re-identification.

5. RUN

The requesting party can now run the approved query across both its data and the other party's data, if it respects the rules applied by the counterparty in step 4. Snowflake handles the computation of data between the two (or more) parties.

6. ACTIVATE

Aggregate-level results inform how audiences are segmented and targeted in upcoming campaigns. For example, the retailer in the above example might learn that 7% of people who consume cooking content on the publisher's site bought sneakers from the retailer in the past. This insight will help the retailer better target that segment with ads for the latest sneaker models in the future.

7. MEASURE

After activation, a party can perform another JOIN analysis to measure various campaign performance metrics.

HOW MARKETERS AND PUBLISHERS USE SNOWFLAKE DATA CLEAN ROOMS

If you're a marketer or publisher, you are likely focused on strengthening direct consumer data relationships now that the deprecation of common identifiers has begun. A core pillar of such a strategy is handling data collected directly from users and customers (known as first-party data) in a privacy-preserving manner.

Snowflake Data Clean Rooms allow for sensitive data derived from unique identifiers (such as emails, hashed emails, names, device IDs and IP addresses) to be leveraged for joint analytics while preserving consumer privacy. From there, you can segment and target existing customers by finding overlaps with a publisher's audience, without moving or copying data.

A common approach for an enterprise's data clean room journey is a "crawl, walk, run" framework, through which the number of use cases and collaborators or data sources matures over time. With this framework, you can select one or more collaborators to join the clean room, securely overlap your data on a mutual identity key, and process the overlap of the datasets within minutes.

Crawl: audience sizing and activation

This phase allows those of you just starting your data clean room journey, to choose a simple use case, explore possible solutions and take the first steps toward achieving the benefits of a data clean room. In this sample case, two collaborators move their first-party data into the clean room; any personally identifiable information (PII) is anonymized to allow privacy-enhanced collaboration. You and your collaborators match data and identify audiences based on various factors (such as demographics, behavior or interests), segment the audiences and then develop tailored content to reach out to a selected segment.

Walk: advanced targeting and lookalike modeling

Once you've mastered the basics of using a data clean room, you can start to refine your approach, and create more precise and effective campaigns. After collecting, processing and analyzing data from two to three collaborators in your clean room, you can launch deeper analytics to better target customers. For example, an AI-enabled clean room can use advanced data or lookalike modeling to create accurate lookalike audiences for prospecting opportunities. Ultimately, this approach can enable you to deliver more personalized and relevant content to your target audience, and improve the response rate to your advertisements.

Run: measurement and attribution

In this advanced phase, you can accelerate your approach to understand the effectiveness of your marketing efforts and adjust your advertising strategy accordingly. You'll collate first-party customer data, exposure data and conversion data from multiple collaborators or sources to create a comprehensive view of the customer journey. You can match ad impression data to transactions to calculate the conversion rate. This is particularly helpful for multitouch attribution, sales lift analysis, reach and frequency.

CASE STUDY: HOW THE NBCUNIVERSAL ONE PLATFORM IS ENABLING CROSS-CLOUD DATA COLLABORATION WITH SNOWFLAKE DATA CLEAN ROOMS

NBCUniversal's NBCU Audience Insights Hub is built on a cross-cloud data clean room environment powered by Snowflake, which unlocks data interoperability between NBCUniversal and its advertising ecosystem partners. The new solution, called One Platform, enables NBCUniversal to feed its first-party audience data to advertising partners, who can safely and securely join it with their own datasets without moving, copying or exposing any underlying PII.

Snowflake's framework lets NBCUniversal and its partners govern what data is housed in the clean room, how data can be joined, what types of analyses each party can perform on the data, and what data — if any — can leave the clean room. The approach is distinct from other privacy sandboxes in that it lets participants design the level of protection and transparency that is appropriate for building their audiences, activating campaigns or conducting measurement.

The One Platform encompasses the following capabilities:

- **Digital audience exploration:** In NBCUniversal's centralized clean room environment, partners can explore how audiences and customers overlap, gleaning valuable aggregate insights without exposing any underlying data from either party. This allows brand and agency partners to better find and understand their customers to drive results while keeping viewer information safe.
- **Cross-platform planning:** NBCUniversal will combine the new clean room environment with its proprietary Linear TV APIs, which for the first time will give partners self-service access to NBCUniversal's aggregate linear and digital data — all the necessary ingredients needed for cross-platform media planning — in one integration.
- **Reach and frequency measurement:** The new hub incorporates certified reach measurement models, enabling partners to utilize ad exposure data to conduct their own analyses, and deduplicate campaign reach and frequency for more efficient media planning and measurement.
- **Cross-platform attribution:** NBCUniversal continues to extend its interoperable measurement capabilities, enabling partners to conduct their own self-service multiplatform attribution. In the long term, this will unlock always-on closed-loop attribution capabilities for every campaign.

HOW SNOWFLAKE DATA CLEAN ROOMS ARE DIFFERENT FROM OTHER PRIVACY SANDBOXES

User privacy versus powerful targeting and measurement: These have historically been viewed as competing imperatives. In reality, behavioral tracking has prompted strategies and tactics predicated on a limited number of individual consumer journeys and effectively tiny samples of data. With Snowflake Data Clean Rooms, marketers and publishers can improve the statistical rigor of their media planning and buying by joining and analyzing huge datasets in an anonymized, privacy-preserving manner.

Here are Snowflake's five key differentiators:

1. Secure data collaboration

Marketers and publishers can securely share live datasets across clouds and regions without the other party copying the data, and retain confidence that data will stay fresh and up to date since it exists only in one place instead of being spread out across silos. Data is stored — encrypted by Snowflake — only in a data owner's Snowflake account.

This is significantly different from how other large platforms' walled-garden clean-room environments operate. In those cases, marketers have to effectively hand over their first-party data to join it with aggregate-level data from the other platforms, and then model the combined dataset for targeting and measurement.

2. Privacy controls

Snowflake has powerful multiparty computation capabilities, enabling a protected analysis of joined datasets without revealing a company's data to counterparties.

Yao's Millionaires' problem provides a useful framework for understanding how this works. A well-known cryptography problem, it centers on how two millionaires, Alice and Bob, can determine whose net worth is higher without telling each other how much money they each have. Potential solutions to the problem have historically taken the form of complex and expensive computations, but Snowflake has a simple way to solve the problem, and it's a solution that works across multiple parties and trillions of rows of data, as well as for more complex calculations than just "is number A bigger than number B."

Here's an easy way to think about how Alice and Bob can securely compare their net worth using Snowflake Data Clean Rooms:

1. Alice creates a table called "My Wealth" and enters her wealth — $1.2 million — into it. She creates a secure view of that table in SQL, and she applies clean-room protections to that view. She grants Bob access to that view, but he can use it only for allowed types of analyses, and he cannot simply run a SELECT * statement from the view.
2. Bob does the same thing, but enters $1.1 million into his table.
3. Bob runs a query in Snowflake to determine which number is higher. Alice's clean-room rules allow this query because it is an approved type of joint analysis from Alice's perspective (it follows a preapproved query pattern, or follows rules Alice has set for what Bob can do with his and her data together).
4. Bob learns that Alice is richer than him, but neither learns the exact numbers that underlie that answer.
5. Alice learns that Bob has run this allowed query against their joined data, but she doesn't herself learn the result. If Alice also needs to know the result, the same query above could also be run by Alice, assuming Bob has set up his clean-room configuration to allow that as well.

This concept directly maps to how marketers and publishers use Snowflake Data Clean Rooms to find overlapping segments in their audience without directly comparing shared identifiers such as email addresses and IP addresses. They simply apply the same approach but to millions, billions or even trillions of rows instead of just to two numbers. They also can perform more complex analyses beyond "is a>b?" and even analyze potentially more than two parties' data simultaneously. The scalability of these calculations is the same as if all the data were held by one of the parties, but now the parties can't see each other's data and they can run only certain allowed types of analyses but not others.

The underlying technology can also be leveraged for much more complex problems, such as segmenting and then activating audiences based on a variety of parameters, and training a machine-learning model. The method uses a fraction of the computing resources required by previous solutions to secure multiparty computation problems. This helps companies rein in costs including energy consumption, and makes possible improved marketing and content distribution strategies that may have been unfeasibly slow, insecure or expensive before.

Importantly, access rights for the other party are fully revocable. From a security perspective, this means they can be removed at any time, resulting in the immediate loss of access for the counterparty. From a data freshness perspective, this is an enormous improvement over the legacy approach of sharing copies of data, since a copy can't be kept up to date without incurring ongoing complexity and cost.

3. Per-second pricing with rapid elasticity and linear scaling

With Snowflake, companies pay only for the resources they use — and they can scale up and down as needed instead of paying for a fixed amount of compute resources they may not use. Moreover, a company's costs won't increase if it allows a partner to join a view of its data or to query its data. The party running the query pays the Snowflake compute cost to run the query, even though it may be based all or in part on the other party's data.

4. Secure data enrichment

There will invariably be cases where two parties lack a common identifier, which can make joining the data and finding overlaps on a single key (such as email address) harder to match between the parties. In extreme cases, the marketer might have only email addresses, while the publisher has only IP addresses, but they still want to join their data for allowed analyses.

When Snowflake customers find themselves in this position, they can leverage identity-enrichment providers on Snowflake Marketplace such as Acxiom, Epsilon and Neustar for data enrichment and identity joining. These providers have vast amounts of third-party customer data that marketers and publishers can use to refine and activate their audiences, and boost their join and match rates. The data enrichment step occurs only in the data clean room; third-party data does not get populated into either party's Snowflake account.

5. Data Cloud effect

Since thousands of companies worldwide already trust Snowflake to securely store their sensitive data, the platform is an easy way to share data externally and set up a clean room with a counterparty.

Snowflake runs across the three major cloud providers (AWS, Microsoft Azure and Google Cloud Platform) and across regions, so different parties don't have to move data out of their Snowflake accounts to join it. Meanwhile, **a unified global management structure** and **built-in, consistent data governance** allow for data to be seamlessly joined and analyzed between geographic and cloud provider–specific regions without sacrificing privacy or compliance.

With the **acquisition of Samooha**, Snowflake also empowers customers to create and manage data clean rooms by leveraging Snowflake Native Apps, a suite of applications that run natively on Snowflake. These apps provide capabilities for data ingestion, transformation, governance and consumption, while providing access to the Snowflake Data Cloud. By using Snowflake Data Clean Rooms and **Snowflake Native Apps**, businesses can collaborate more quickly and easily while maintaining control and trust over their data assets.

THE SNOWFLAKE DIFFERENCE: IMPROVE AD PERFORMANCE AND ROI WHILE PRIORITIZING USER PRIVACY

By embracing data clean rooms, organizations can identify audience overlaps and then harness advanced machine learning techniques inside those clean rooms to improve data modeling. In this way, marketers and advertisers can be more rigorous about how they segment and activate audiences. This is an alternative to the current ubiquitous practice of over-indexing on tiny samples of data captured through behavioral tracking to extrapolate customer journeys for huge audiences.

To summarize, Snowflake Data Clean Rooms are effective through these five key actions they unlock:

- **Secure data collaboration** of live, ready-to-query datasets without ETL.
- **Enable privacy controls** to produce a protected analysis across trillions of rows of joined datasets across two or more parties, without revealing a company's data to counterparties.
- **Derive value** from the ability to scale up and down as needed, instead of paying for a fixed amount of compute power.
- **Unlock data enrichment** by leveraging third-party identity providers to help with matching when needed.
- **Enact the Data Cloud effect** through the scale of Snowflake's thousands of customers, built-in governance capabilities, and ability to run across various clouds and geographic regions.

CONCLUSION

With Snowflake Data Clean Rooms, every stakeholder in the value chain wins. Marketers can get a higher return on ad spend thanks to more precisely targeted ads and powerful lookalike modeling. Publishers can improve their performance and provide direct ROAS metrics through one-to-one connections between ad exposures to marketers' purchase data. And consumers win by having their online privacy preferences honored and receiving more relevant and personalized offers.

Want to learn more about Snowflake Data Clean Rooms? Watch our: Data Clean Room Demo.

ABOUT SNOWFLAKE

Organizations use Snowflake's Data Cloud to unite siloed data, discover and securely share data, power data applications and execute diverse AI/ML and analytic workloads across multiple clouds and geographies. Organizations, including 691 of the 2023 Forbes Global 2000 as of January 31, 2024, use the Snowflake Data Cloud to power their businesses.

Learn more at **snowflake.com**