HOW SNOWFLAKE'S DATA CLEAN ROOMS HELP PUBLISHERS AND MARKETERS IMPROVE AD EFFECTIVENESS

snowflake

WHITE PAPER

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 2020 KPMG study, 54% of people don't trust companies to use their personal data in an ethical way, and 53% don't trust companies to ethically collect personal data.[1]

Meanwhile, according to research from executive search firm Spencer Stuart, CMO tenures continue to shrink,[2] and senior marketing leaders are 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 that spend is optimized. These models have generally relied on third-party cookies 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 legacy methods of targeting and measurement increasingly precarious. Specifically, with Apple now requiring users to opt in for their activity to be tracked across apps[3] and the coming deprecation of third-party cookies on Chrome in 2023,[4] data clean rooms have emerged as a strategy for marketers to do sophisticated audience segmentation, targeting, attribution, measurement, and machine learning–boosted lookalike analysis while preserving privacy.

Data clean rooms are secure platforms that allow two or more parties to join and query data sets, without each party being able to see the other's data.

With Snowflake data clean rooms, companies can securely share data views and allow data to be joined to other data sets for modeling and analysis without the other party being able to see any sensitive user information. This also can benefit publishers, which are under enormous pressure to protect user privacy. Snowflake data clean rooms can help them improve ad effectiveness and, ultimately, lower CPMs (cost per thousand impressions) and increase ROAS (return on ad spend) without having to share the user data and activity data they once made available to marketers.

WHY MARKETERS AND PUBLISHERS USE SNOWFLAKE DATA CLEAN ROOMS

With the deprecation of third-party cookies looming over the ad industry,[5] marketers and publishers alike are focused on strengthening their direct consumer data relationships. A core pillar of that strategy is handling the data they collect 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 while preserving privacy. From there, marketers can segment and target existing customers by finding overlaps with a publisher's audience without ETL (that is, without having to move or copy the data).

Targeting in the clean room environment offers a sharp contrast to programmatic ad buying, which often lacks transparency. When marketers buy from a demand-side platform (DSP) or an exchange, there's usually no way for them to tell if they've reached actual customers or whether the reached people have actually converted to buy something, especially when the conversion happens in a non-linear manner. For example, if a consumer sees an ad online or sees an email, but then buys a product in a physical store, these are flows that cookie-based approaches generally couldn't handle to begin with.

With Snowflake data clean rooms, targeting and activation can be refined based on a wide variety of parameters, such as content consumption on the publisher side and purchase history or demographic

data on the marketer side. (For example, existing customer segments that read a lot of content about running can be targeted by retailers through ads for running apparel on a publisher's site.)

Marketers can also leverage machine-learning-driven lookalike modeling in the clean room environment to find cohorts within the publisher's audience who have similar attributes to existing customers or even high-value customers.

This all occurs without the advertiser being able to see the publisher's data, without the publisher being able to see the marketer's data, and without any other third party being able to see either party's data or even knowing that they're performing this kind of analysis between them.

Clean rooms are also helping ROI-driven marketers close the loop. They can simply run a query in the clean room environment to see which of their customer segments who purchased a specific product were exposed to ad impressions on the publisher side to find out conversion information for their campaign—even across campaigns that have run across multiple media channels. This eliminates the need to collect and analyze behavioral data via third-party cookies in order to understand successful customer journeys.

A win for marketers is also a win for the publisher, since more powerful targeting capabilities and accurate attribution can help justify higher ad spend, while sophisticated lookalike modeling of their audiences help publishers to sell more ads.

How NBCUniversal Is Enabling the Future of Cross-Cloud Data Collaboration with Snowflake Data Clean Rooms

In 2021, NBCUniversal announced plans to launch the NBCU Audience Insights Hub, built on a cross-cloud data clean room environment powered by Snowflake that will unlock data interoperability between NBCUniversal and its advertising ecosystem partners. The new solution will enable NBCUniversal to feed its first-party audience data, which advertising partners will then be able to safely and securely join with their own respective data sets, without moving, copying, or exposing any underlying personally identifiable information (PII).

Snowflake's framework will let 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 the participants design the level of protection and transparency that is appropriate for building audiences, activating campaigns, or measurement.

The new capabilitie of the NBCU Audience Insights Hub include:

- Digital audience exploration: In NBCUniversal's new centralized clean room environment, partners will be able to explore how audiences and customers overlap, providing valuable aggregate insights without any underlying data exposed from either party. This will enable 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, giving partners self-service access, for the first time ever, 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 will incorporate certified reach measurement models, which will enable partners to utilize ad exposure data and conduct their own analyses, providing the ability to deduplicate campaign reach and frequency for more efficient media planning and measurement.
- Cross-platform attribution: NBCUniversal will continue to extend its interoperable measurement capabilities, enabling partners to conduct their own self-service multi-platform attribution. In the long-term, this will unlock always-on closed-loop attribution capabilities for every campaign.

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 sending copies of that data anywhere else. This makes it impossible for raw data to be 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 (personal) 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 Data 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, if a retailer wants to query a publisher's anonymized engagement data by content category and find overlaps with its own purchase data, then to better understand what types of content its customers consume, it 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, automatically ensuring that 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 in order 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 the previous step. Snowflake handles the computation of data between 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 publisher in the past. That segment can be targeted with ads for the latest models of sneakers in the future.

7. MEASURE

After activation and the campaign has run (or while it is running), a party can perform another JOIN analysis to measure various campaign performance metrics.

HOW SNOWFLAKE DATA CLEAN ROOMS ARE DIFFERENT FROM OTHER PRIVACY SANDBOXES

User privacy and powerful targeting and measurement have historically been viewed as competing imperatives. In reality, behavioral tracking resulted in strategies and tactics that were predicated on a handful of individual consumer journeys or 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 data sets (all the data) in an anonymized, privacy-safe manner.

Here are five key differentiators:

1. Secure Data Sharing

Marketers and publishers can securely share live data sets across clouds and regions without the other party copying the data, and 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, where marketers have to effectively hand over their first-party data in order to join it with aggregate-level data from those platforms and then model the combined data set for the purposes of targeting and measurement.

2. Privacy Controls

Snowflake has powerful multiparty computation capabilities, enabling a protected analysis of joined data sets 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 been 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. In addition, they 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. It uses a fraction of the computing resources required by previous solutions to secure multiparty computation problems, which helps companies rein in their costs and energy consumption, and it makes things possible 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 millisecond, 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 scale down as needed instead of paying for a fixed amount of compute they may not use. Moreover, a company's costs won't increase if it allows a partner to join a view of its data to theirs or 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 less able to match all the possible data 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 together for allowed analyses. When Snowflake customers find themselves in this position, they can leverage identity enrichment providers on Snowflake Data Marketplace such as Acxiom, Epsilon, and Neustar for data enrichment and identity joining. These providers have vast troves 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 and 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, Snowflake 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 two parties never have to move data out of their Snowflake accounts to join it. Meanwhile, a unified global management structure and an access-governance structure allow for data to be seamlessly joined and a nalyzed between geographic and cloud provider–specific regions.

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

The impending deprecation of third-party cookies should be viewed as an opportunity for marketers and publishers rather than as a looming catastrophe. By embracing data clean rooms to find audience overlaps and harnessing advanced machine-learning techniques inside those clean rooms to improve data modeling, they can be more rigorous about how they segment and activate audiences. This is an alternative to the ubiquitous current 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 in five key ways:

- **Secure data sharing of live,** ready-to-query data sets without ETL
- **Privacy controls** to enable a protected analysis across trillions of rows of joined data sets across two or more parties without revealing a company's data to counterparties
- **Value** from the ability to scale up and down as needed instead of paying for a fixed amount of compute
- **Secure data enrichment** leveraging third-party identity providers to help with matching, when needed
- **Data Cloud** effect from the scale of Snowflake's thousands of customers and ability to run across various clouds and geographic regions

With Snowflake data clean rooms, every stakeholder in the value chain wins. Marketers can get a higher return on ad spend from more precisely targeted ads and powerful lookalike modeling. Publishers can lower CPMs by improving their inventory's performance and proving impact via direct (and anonymous) comparisons of ad exposures to marketers' purchase data. Finally, consumers win by having their online privacy honored and receiving more relevant and personalized offers.

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**.

ENDNOTES

[1] bit.ly/3x2icxl [2] bit.ly/3x4DGZy [3] bit.ly/361OFrL [4] bit.ly/2Trdj26 [5] cnet.co/3A5pUsl [6] bit.ly/367KBpO