

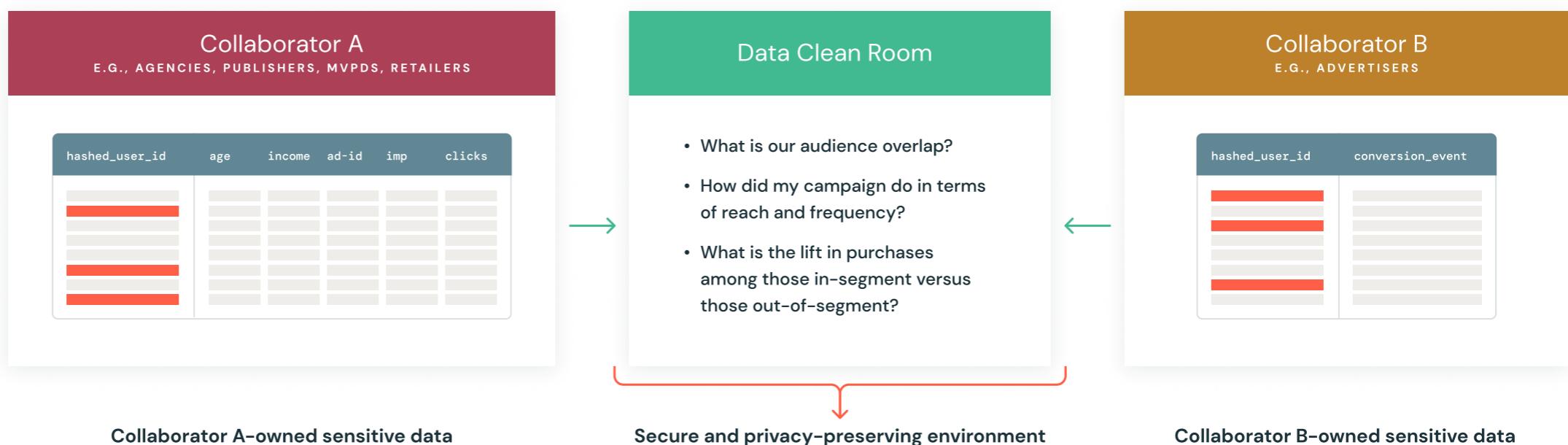
# Chapter 6

## Privacy-Enhanced Sharing With Databricks Clean Rooms

While the demand for external data to make data-driven innovations is greater than ever, there is growing concern among organizations around data privacy. The need for organizations to share data and collaborate with their partners and customers in a secure, governed and privacy-centric way is driving the concept of “data clean rooms.”

### What is a data clean room?

A data clean room provides a secure, governed and privacy-enhanced environment where participants can bring their sensitive data and perform joint analysis on that private data. Participants have full control of the data and can decide which participants can perform what analysis without exposing the underlying sensitive data.



**Figure 9:**  
Data clean room  
diagram example  
for audience  
overlap analysis in  
advertising

Data clean rooms have gained renewed attention as organizations seek privacy-compliant ways to collaborate on data. This trend is driven by stringent data privacy regulations like GDPR and CCPA, which have reshaped how data is collected, used and shared.

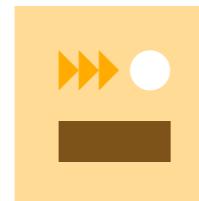
Despite the growing interest in data clean rooms, enterprises are still in the early stages of their data collaboration journey. According to the [2024 IDC External Data Sourcing and Collaboration Survey](#), only one-third of enterprises have started using data clean rooms, and those that have are only working with one or two data collaboration partners. Many organizations face significant challenges related to technology and data management when implementing data clean rooms. Nearly 56% of enterprises have concerns about privacy and/or consent regarding the treatment of data being shared.

## Common data clean room uses cases



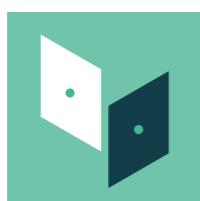
### Campaign optimization for media and entertainment

By creating a clean room environment, media companies can securely unlock the value of their audience data by combining it with their advertiser's first-party data. This allows them to perform in-depth analysis and identify shared audience segments and post-campaign measurements without directly accessing or exposing individual user information.



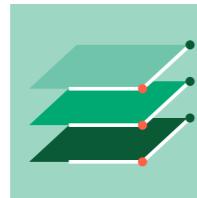
### Supply chain optimization for retail and consumer goods

Clean rooms enable real-time collaboration between retailers and suppliers, ensuring secure information exchange for demand forecasting, inventory planning and supply chain optimization. This improves product availability, reduces costs and streamlines operations for both parties.



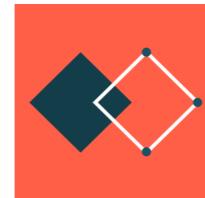
### Real-world evidence (RWE) for healthcare

Clean rooms provide secure access to healthcare datasets, allowing collaborators to connect and query multiple sources of data without compromising data privacy. This supports RWE use cases such as regulatory decisions, safety, clinical trial design and observational research.



## Know Your Customer (KYC) in banking

KYC standards are designed to combat financial fraud, money laundering and terrorism financing. Clean rooms can be used within a given jurisdiction to allow financial services companies to collaborate and run shared analytics to build a holistic view of a transaction for investigations.



## Personalization with expanded interests for retailers

Clean rooms enable retailers to augment their knowledge of consumers to suggest new products and services that are relevant to the individual but haven't yet been purchased.

## Shortcomings of existing data clean rooms

Organizations exploring clean room options are finding some glaring shortcomings in the existing solutions that limit the full potential of the “clean rooms” concept.

First, many existing data clean room vendors require data to be on the same cloud, same region and/or same data platform. Participants then have to move data into proprietary platforms, which results in lock-in and additional data storage costs.

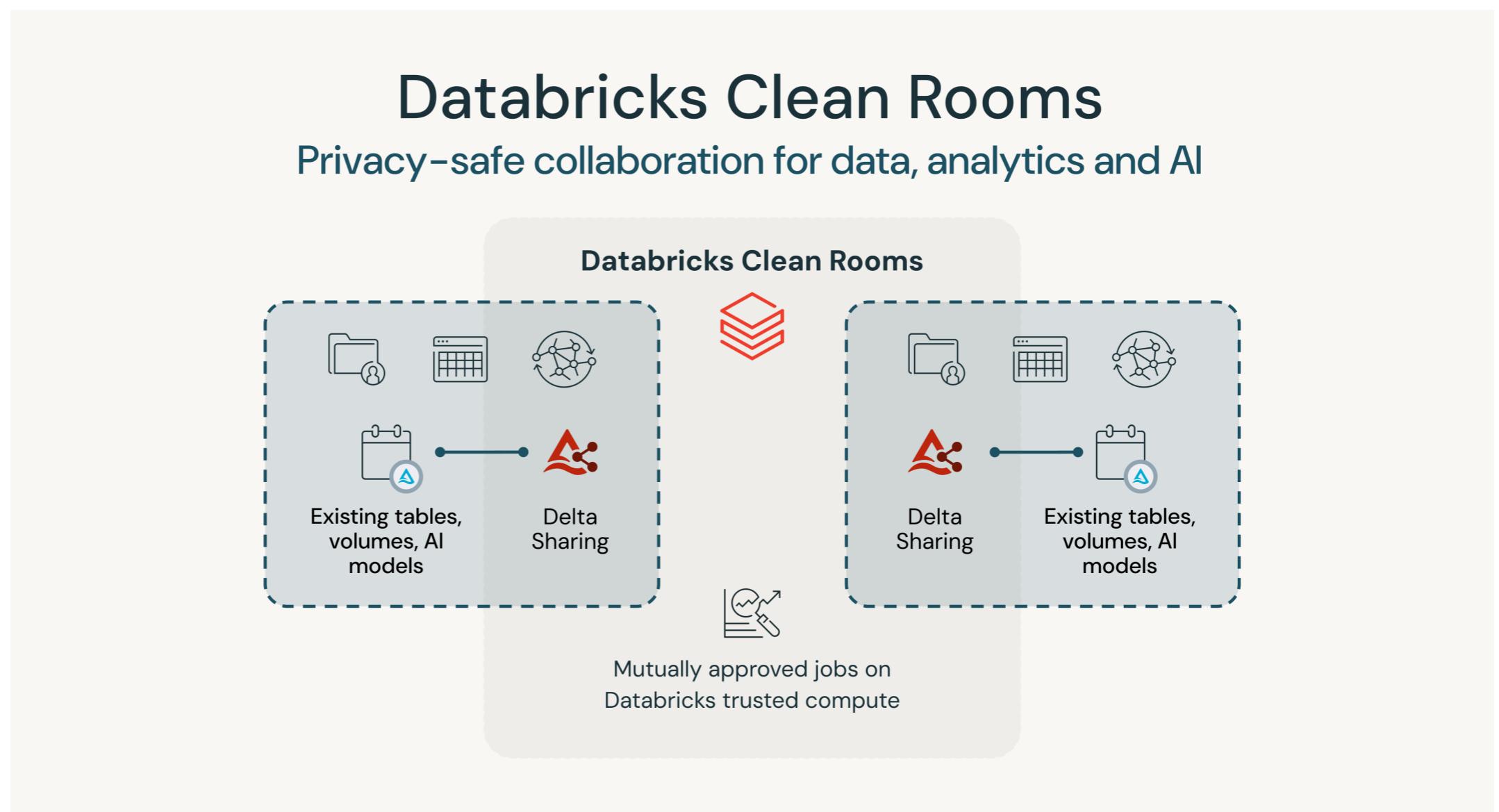
Second, existing clean rooms often lack the flexibility needed to support complex analyses, and they’re pretty much restricted to SQL. While SQL is great and absolutely needed for clean rooms, it can be limiting and prevent organizations from fully leveraging their data for advanced insights and innovation. The inability to run diverse workloads or integrate with AI/ML frameworks limits the potential applications of data clean rooms in unlocking valuable insights.

Finally, setting up data clean rooms can be complex due to a lack of automation. This manual setup process leads to longer ramp-up times and increased total cost of ownership (TCO), making it challenging for organizations to deploy and scale these solutions quickly. The high costs associated with implementation and maintenance can be prohibitive, especially for smaller enterprises.



## Privacy-safe collaboration with Databricks Clean Rooms

Databricks Clean Rooms is powered by Delta Sharing and allows businesses to easily collaborate with their customers and partners on any cloud without compromising privacy or sharing sensitive data. When collaborating in a clean room, your data stays in place and you're always in control of where and how the data is being used.



## Any cloud, any platform

Databricks Clean Rooms is built for collaboration across clouds and across platforms. You can choose any cloud and region to start your clean rooms. Data stays where it is, and collaboration happens without the need to copy data. You collaborate on not only Databricks data, but also source data from outside of Databricks, thanks to Sharing for Lakehouse Federation. Databricks Clean Rooms truly supports you and meets your collaborators' needs wherever their data resides.

## Any scale, any trust level

Databricks Clean Rooms is scalable and supports automating your privacy-safe workload with APIs, SQL commands and built-in workflow orchestration. You can also easily access your Databricks Clean Rooms outputs in notebooks or in your Unity Catalog so you can use them for other workloads. Databricks Clean Rooms also supports collaboration between multiple parties at different trust levels using different approval modes.

## Any language, any workload

When collaborating using data clean rooms, the need for different programming languages becomes essential due to the diverse nature of tasks that participants from multiple organizations may need to perform. For instance, SQL is excellent for data querying and manipulation, while Python is preferred for machine learning and statistical analysis. Scala and Java are often used for building scalable data processing applications. Multilanguage support in Databricks Clean Rooms enables users to choose the best language for their specific workload, whether it's simple data joins or complex ML/AI computations. Databricks Clean Room supports collaboration on any format of data and AI models while protecting the privacy of the raw content. Leveraging the full power of Databricks Notebooks, you can run SQL or Python for complex compute and ML/AI workloads. And more language support is on the way.

“Mastercard is leading with new insights that solve our customers’ needs and real problems. These insights are founded in data that can be a company’s biggest asset” said Andrew Reiskind, Mastercard’s chief data officer. “Accordingly, we’re always looking at how we protect the confidentiality, privacy and security of that data when we use it. Databricks Clean Rooms is a solution that allows us to protect the information aligned to our Data and Tech Responsibility Principles, while giving visibility into trends. Databricks Clean Rooms offers new innovative opportunities for our customers to drive insights and value-added services. Partnering with Databricks, we’ve piloted both frameworks and developed a set of integrated PET capabilities to offer flexibility, scale and transparency.”