Confidential Clean Room

Use Cases for Multi-party Data Analytics

SafeLiShare

Contents

3 The State of Cloud

5 Confidential Clean Room

6 What are the Data and Model Protections

8 Clear and Present Danger

9 Industry Use Cases: Public Sector, Financial, and Healthcare

11 Considerations when Building a Confidential Clean Room

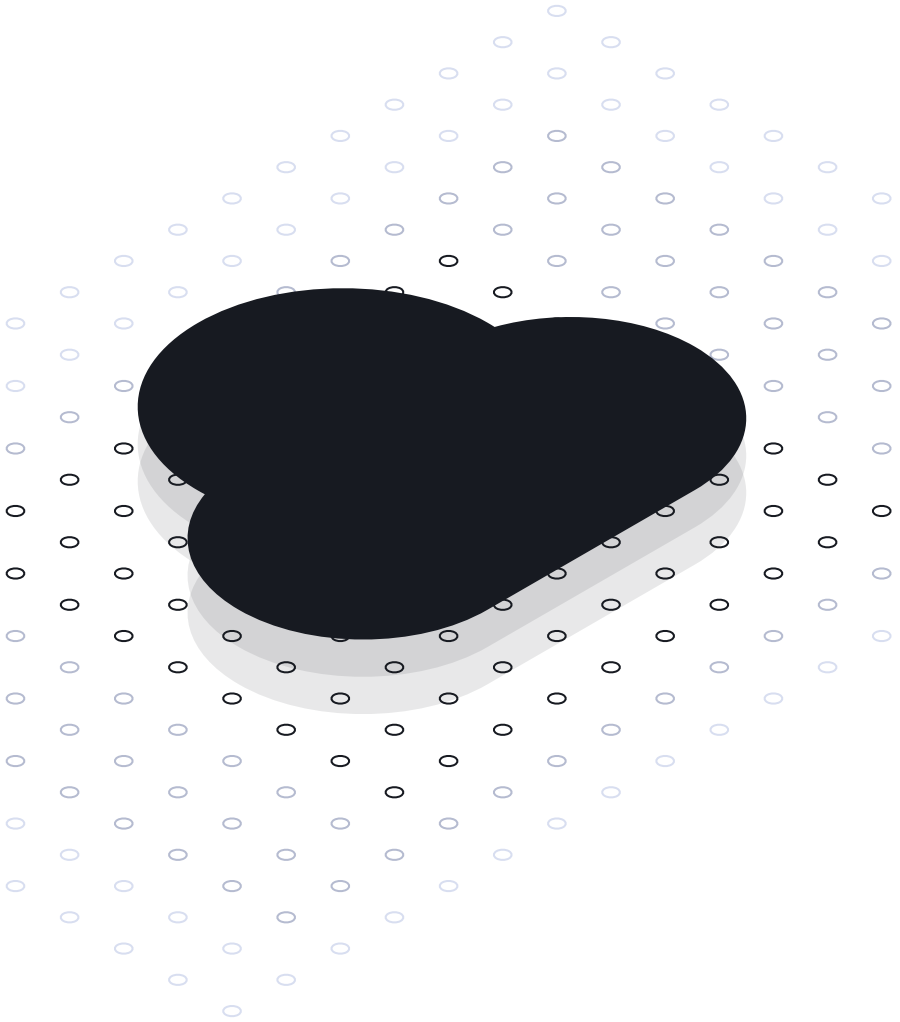
11 Summary

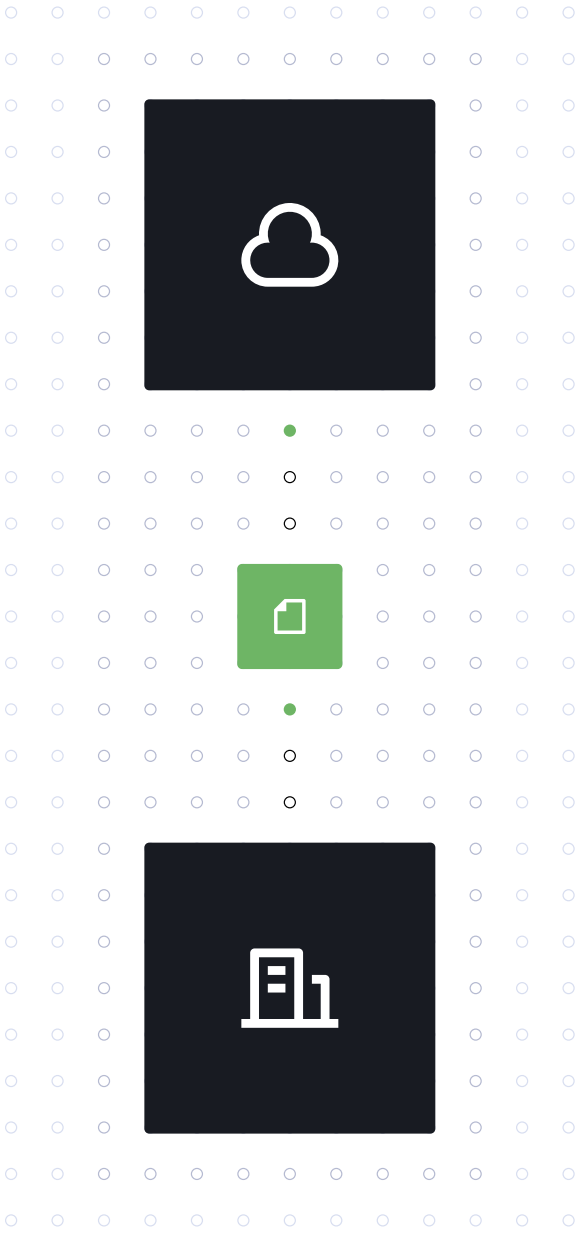
The State of Cloud

Multicloud is the new reality in enterprise technology with increasing priority

As organizations faced new challenges such as increased levels of remote work and collaboration with new business partners and suppliers, they adopted a multicloud strategy to gain the flexibility and scalability they needed for this new reality.

Multicloud strategies give enterprises more control over where and how their data is stored and used, , while also ensuring businesses can control the costs of their cloud operations by adjusting which services they use from different providers.





Growing adoption and initiatives in 2023:

98% of Enterprises Using Public Cloud have adopted a Multicloud Infrastructure Provider Strategy (451 Research)

64% Report showing security concerns in cloud computing (CSA)

41% Data sovereignty is the top driver of multicloud strategies restricted by GDPR

54% Data redundancy is the most anticipated future use case

87% Organizations embrace multi-cloud (The State of the Cloud 2023)

42% The cost of a breach in the healthcare industry went up since 2020 (IBM DBIR)

Protecting the Frontier

Confidential Clean Room and Multi-party Data Analytics

SafeLiShare provides a foundation for solutions that enable multiple parties to collaborate on data. There are various approaches to solutions, researchers, data scientists and data providers to collaborate on data while preserving privacy. This overview covers some of the approaches and existing use cases that can be used with SafeLiShare's Confidential Clean Room.

In a multicloud environment, organizations may use multiple cloud providers with Microsoft Azure, AWS Nitro, GCP, or IBM to host their applications and data. Confidential computing can be used to protect sensitive data and workloads in a multicloud environment. This allows organizations to maintain control and security over their data, even when it is being processed in a third-party cloud environment.

The term " clean room" originally referred to a controlled environment used in the semiconductor industry to prevent contamination during the manufacture of computer chips. In the context of data, a clean room is a secure environment where sensitive or confidential data is processed or analyzed without risking exposure to unauthorized individuals.

The concept of a clean room has been used in various industries for many years, such as in pharmaceuticals, where a clean room may be used to prevent contamination during drug manufacturing. In the field of data, the clean room concept is applied to prevent unauthorized access or use of sensitive data, such as personal or financial data, during analysis or processing.

Therefore, while the concept of a clean room in the context of data is not new, it has become increasingly relevant as organizations seek to protect sensitive data and comply with privacy regulations.

Prevent data from top cloud security threats and data breaches with SafeLiShare's Confidential Clean Room:

- Algorithmic complexity attacks
- System and application vulnerabilities
- Man-in-the-middle attacks (MITM)
- In-memory attacks
- Malicious insiders
- Ransomware

Data and Model Protection

Multi-party data analytics and AI ML Modeling

When analyzing data from multiple parties, each party contributes data to the analysis without revealing their sensitive data to the other parties. This approach enables organizations to collaborate on data analytics projects while maintaining data privacy and security. It becomes increasingly important In the context of AI and ML. Multiparty data analytics and AI/ML can enable organizations to collaborate on data-driven projects, such as disease prediction, fraud detection, and supply chain optimization, while protecting the privacy and security of sensitive data. However, implementing these techniques can be challenging, and organizations must carefully consider the trade-offs between data privacy, security, and performance when designing their multiparty analytics systems.

Simplify All Cryptographic Operations

SafeLiShare's unified data security platform provides a rich set of integration points and simplifies all your cloud provisioning and cryptographic operations with an intuitive management console, command line toolkits, or RESTful API so you can deploy the Confidential Clean Room during development or production phases when your developers and security teams need to control and audit data security across any cloud environment.

Confidential clean room solutions allow data providers to combine data for processing using agreed upon code, queries, or models. The data is often considered sensitive and not directly shared with other participants. Confidential Computing is cloud computing with privacy enforcement that's widely available in AWS, Azure and GCP. Confidential computing can be used to ensure security and privacy of the data and models during processing. This verifies that participants do not have access to the data or models.

A Clear and Present Danger

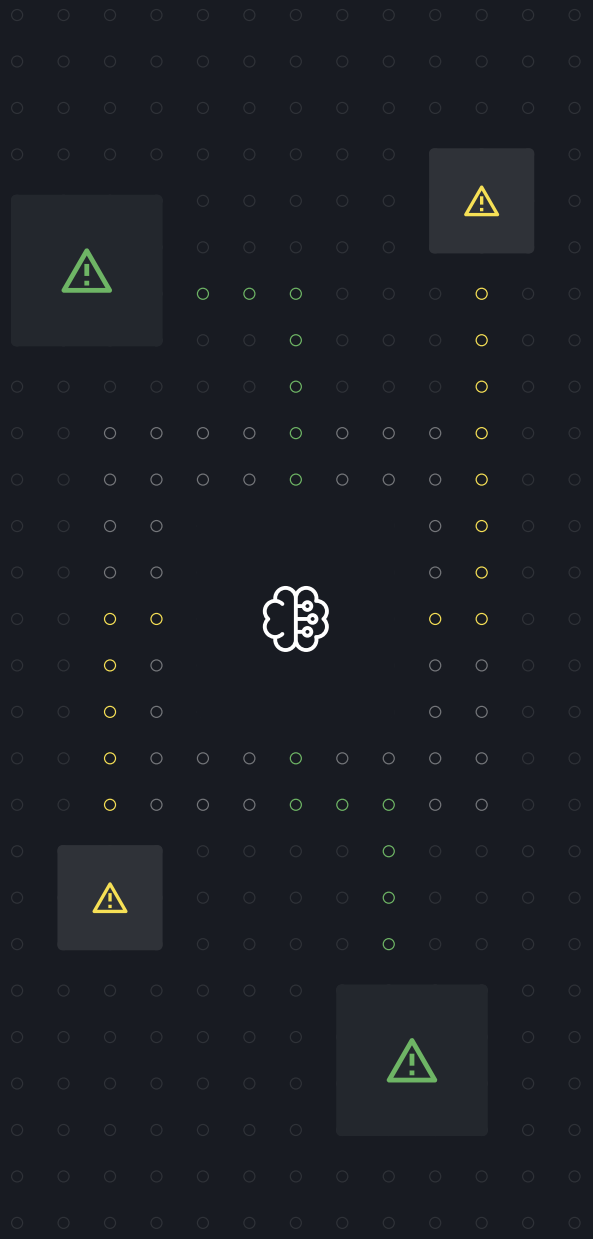
Data leakage nightmare in AI ML Model Training

AI algorithms can be used to analyze large amounts of data from multiple sources in order to identify patterns and insights that would not be possible to detect through traditional analysis methods. Multiparty data analytics involves analyzing data from multiple parties, such as different organizations or individuals, in order to gain a more comprehensive understanding of a particular topic or issue. AI can be used to analyze this data in a way that is more efficient and accurate than traditional methods, enabling organizations to make more informed decisions based on the insights gleaned from the data.

Additionally, AI can be used to protect the privacy and security of the data being analyzed, which is essential in multiparty data analytics where sensitive information may be involved.

Artificial intelligence is viewed as a problem-solving solution and time-saving tool for humanity. However, it's crucial to acknowledge the potential issues that may arise and their impact on humans and the environment.

One of the significant concerns with artificial intelligence (AI) is the problem known as "data leakage." This issue pertains to a machine learning problem where the data employed to train the model may contain unforeseen information, leading to an overestimation of the model's effectiveness when run with actual data.



Confidential Clean Room Use Cases

Banking

By combining merchant data with bank data, personalized offers can be created for customers based on their spending habits and preferences. This can be achieved using confidential computing virtual machines (VMs) and SQL in secure enclaves. Confidential computing VMs allow for secure processing of sensitive data by encrypting it while it is being processed. SQL can be used to query and analyze the combined merchant and bank data, allowing for the creation of personalized offers based on the customer's spending behavior. Secure enclaves ensure that the data is protected from unauthorized access, ensuring customer privacy and security. This solution can improve customer satisfaction and loyalty by providing personalized offers that meet their individual needs and preferences.

Public Sector

AI (Artificial Intelligence) can be used to analyze cross-bank money flows to identify potential instances of money laundering. By utilizing confidential computing, the AI can flag any suspicious activity that may be linked to human trafficking. This method allows for the detection of illegal activity without compromising the privacy and security of confidential financial information. The use of AI in this manner has the potential to significantly reduce instances of money laundering and human trafficking.

Confidential Clean Room Use Cases

Healthcare

Clinical trials for rare diseases can be challenging due to the limited number of patients available for participation. One approach to address this challenge is to use confidential computing to identify potential candidates for clinical trials. Confidential computing allows for the secure and private analysis of sensitive medical data, such as electronic health records, genetic data, and clinical trial data. By using this approach, researchers can identify patients who meet the specific criteria for a clinical trial without compromising their privacy. This can help accelerate the development of treatments for rare diseases by enabling the identification of suitable candidates for clinical trials.

Payment Services

Connecting data across banks for fraud and anomaly detection involves the use of a confidential clean room, which is a secure environment that enables the sharing of sensitive data without compromising privacy. In this context, banks can share transactional data with each other to identify patterns and anomalies that may indicate fraudulent activity.

The clean room ensures that the data is anonymized and encrypted, and that there is no access to personally identifiable information. This means that the banks can collaborate on fraud detection without violating privacy regulations or compromising customer trust.

By connecting data across banks, it is possible to identify suspicious transactions that may involve multiple accounts or institutions. This allows for more effective fraud detection and prevention, as well as the ability to respond quickly to potential threats.

Using confidential clean rooms is a powerful tool for improving the security and integrity of financial systems. It helps to protect customers and institutions from fraudulent activity, while maintaining the highest standards of privacy and data protection.

Data Clean Rooms

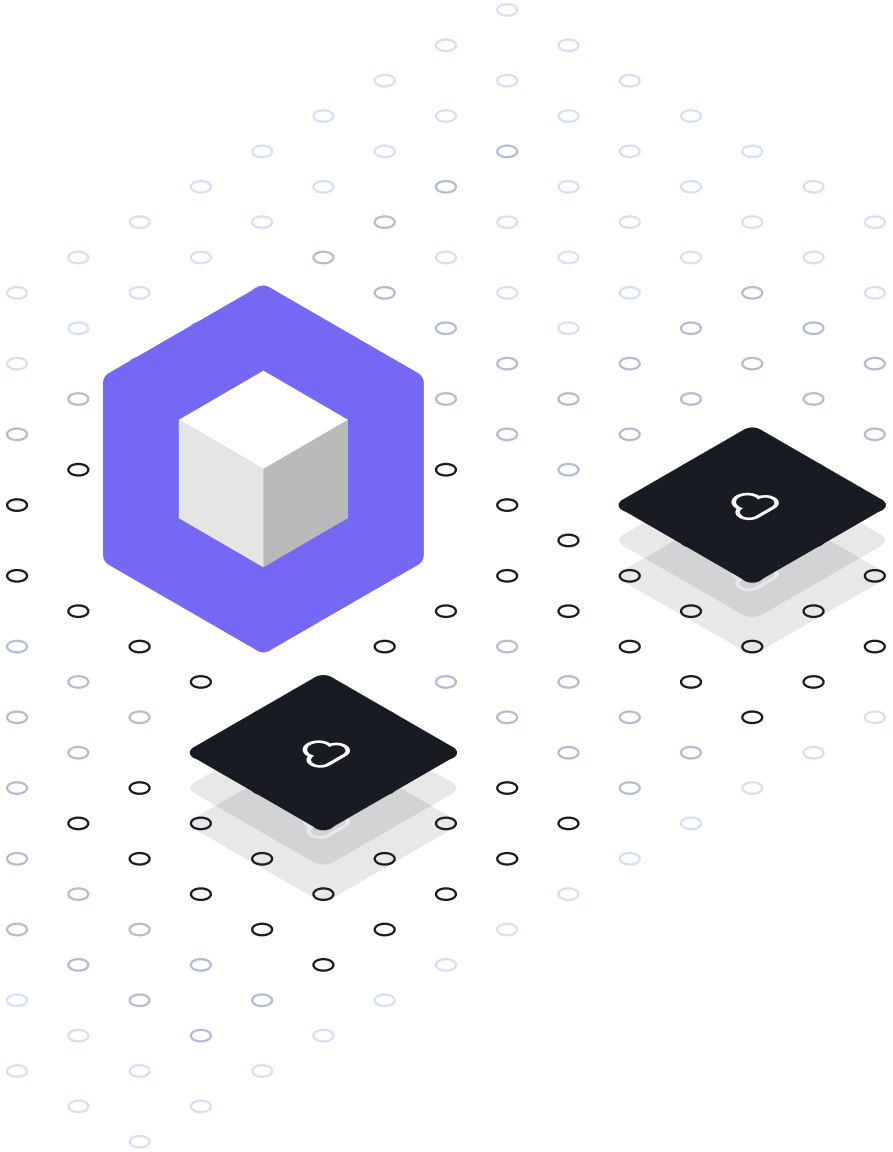
Summary

Data clean rooms have been around for a while, but Confidential Clean Rooms are not. The recent advancements in confidential computing have opened up more possibilities for utilizing cloud scale with larger datasets, protecting the intellectual property of AI models, and complying with data privacy regulations. Previously, certain data may have been inaccessible due to various reasons.

There may be challenges in sharing data across industry companies due to competitive disadvantages or regulations.

- Anonymization can potentially impact the quality of data insights or require significant resources in terms of time and cost
- Data is restricted from cloud processing and must remain in specific locations due to security concerns.
- Legal processes can be expensive and time-consuming in cases where liability is involved due to data exposure or misuse.

With SafeLiShare Confidential Clean Room solution, it provides a turnkey secure enclave to solve above challenges with encryption in use and reinvent PKI technology to deliver complete or effective datasets that result in bigger insights, or less time needed in training and using AI models.



When designing or using a confidential clean room solution, consider the size of datasets and speed of insights. “Offline” data can be loaded into a verified and secured computing environment for batch analytics. This allows for evaluation of large datasets with models and algorithms that don’t need immediate results. Batch analytics are useful for ML inferencing across millions of health records. Real-time insights are necessary for identifying fraud in near real-time transactions between multiple entities.

In confidential clean rooms, zero-trust participation is a significant factor. It entails the ability to safeguard data and model IP from all parties involved, including data providers, code and model developers, solution providers, and infrastructure operator admins. When creating a solution or onboarding, it’s crucial to consider what needs to be protected and from whom, including the code, models, and data.

Federated learning is a solution where models process data in the data owner’s tenant and insights are aggregated in a central tenant. Federated learning often iterates on data multiple times to improve the model’s parameters after insights are aggregated.

Customers have data stored in various clouds and on-premises. Collaboration involves data and models from different sources. Confidential Clean Room solutions can help bring data and models from these other locations to the designated cloud platform. If data can’t move to the cloud from an on-premises data store, some clean room solutions can run on site where the data is stored.

SafeLiShare’s Confidential Clean Room enables the creation of solutions that function across multiple organizations. Consensus among participants is required before adding code logic and analytic rules. Updates to the code are recorded for auditing using tamper-proof logging, which is available through hardware-enforced ledgers in confidential computing.

SafeLiShare

More information

SafeLiShare provides policy-driven encrypted data clean rooms where access to data is auditable, trackable, and visible, while keeping data protected during multi-party data sharing.

Visit **SafeLiShare** for more information or contact us for a demo.