Habu

Data Clean Rooms for Financial Services

How confidential computing unlocks the value of third-party data while preserving privacy

Habu

The emergence of the data clean room

The financial services industry is undergoing profound changes to its data ecosystem. We now have unprecedented opportunities to understand market volatility, reduce risk in credit markets, prevent money laundering, and much more. In the past, these advances haven't been possible because vast amounts of data were siloed, hard to access, and impossible to secure. These limitations have made it difficult for organizations to:

1. **Improve decision-making at scale**
   Financial services organizations need to enable collaboration that opens up more data in a way that's scalable, simple, and secure.

2. **Utilize more decentralized data from a range of platforms**
   When data is siloed in a variety of formats, applications, and languages, organizations need a way to customize or integrate those data locations to get the full benefit from any data source. That can be time consuming.

3. **Democratize data access and use**
   When there are barriers to data access, collaboration is more difficult. But when more people have secure access, collaboration increases and accelerates solutions for a greater number of use cases.

4. **Protect their most-sensitive data while at rest, in transit, and in use**
   This is possibly the most challenging barrier for financial services collaboration. Traditional data-security measures protect data in transit and data in storage. The question is, how do you increase data confidentiality and privacy while collaborating with other organizations?

Now, data clean rooms that utilize confidential computing are changing all this. Using Microsoft Azure confidential computing and AMD EPYC™ processors with Infinity Guard featuring Secure Encrypted Virtualization-Secure Nested Paging (SEV-SNP) technology, a Habu data clean room helps eliminate these barriers to unlock collaborative intelligence with the most advanced data collaboration software available. Using this new technology, financial institutions are able to collaborate with partners without compromising their most sensitive data.

In this guide, we'll show you what data clean rooms are, how they work, and how they provide a single solution that addresses each of these needs and challenges for financial services organizations.

Habu

What is a data clean room?

Let's start with the basics. A data clean room is a relatively new, privacy-first, closed-loop environment that enables more-secure access to decentralized data so it can be safely matched with data in other formats from other sources. This capability offers financial services organizations and their partners a solution that vastly expands data access for collaboration that meets today's strict data governance and privacy regulations.

This approach is having a profound effect on bottom lines in many industries. Analyst firm Gartner recently identified a three-times-greater economic benefit for firms that share data externally,[1] and IDC predicts that by 2024, "65% of Global 2000 enterprises will form data-sharing partnerships with external stakeholders via data clean rooms."[2] Across industries, data-driven enterprises understand the implications: Nearly 80% of organizations want to collaborate with other businesses to share data in the next 12 months, and almost 70% want to expand their current data collaborations.[3]

3x the economic benefit of sharing data externally

65% of enterprises will form data-sharing partnerships

80% of enterprises want to collaborate in the next 12 months

1. Laurence Goasduff. "Data Sharing Is a Business Necessity to Accelerate Digital Business." Gartner. May 20, 2021.
2. Dan Vesset, Marci Maddox, Carl W. Olofson, Stewart Bond, Lynne Schneider, Amy Machado, Dr. Chris Marshall, John Rydning, Philip Carnelley, Takashi Manabe, Enrique Phun, Ali Zaidi, Ray Huo, Steve Charbonnier. "Worldwide Data and Content Technologies 2022 Predictions." IDC FutureScape. October, 2021.
3. State of Data Collaboration Report. Habu, 2022.

Habu

Five high-level benefits of data clean rooms

1. **Unrivaled access to data and intellectual property while maintaining data privacy**
   Legacy tools from some data warehouses are technical and can be too complex for business users and others lacking data engineering skills. A modern data clean room utilizing Azure confidential computing and AMD EPYC processors with Infinity Guard featuring SEV-SNP technology delivers more-secure direct access and highly automated data collaboration on a volume and variety of sensitive data that is simply unavailable by other means.

2. **Expanded access to partners while maintaining data sovereignty**
   To drive essential growth, organizations with large and valuable datasets need to collaborate rapidly and at unrestricted scale. With a data clean room backed by confidential computing, an enterprise can expand the universe of data available for collaboration while all partners can maintain data sovereignty, control, and privacy while providing simple, consistent access to the unique datasets of multiple partners.

3. **Use-case enablement and automation**
   The real value of a data clean room lies in the volume of recurring questions and their increasing scale as teams engage directly with new data. This is what makes them such a powerful innovation. A modern clean room provides not only enablement of use cases from risk reduction to fraud prevention via analytic templates, but the technical automation capabilities that enable organizations to run and scale those use cases faster.

4. **Efficiency and productivity gains at scale**
   The business advantages of data clean rooms increase dramatically when the number of data sources and quantity of data points scale. Leveraging additional data collaboration partners by adding clean rooms dramatically increases the speed and quality of data analytics while reducing risk. When anyone in your organization can quickly and securely query data in plain language, you reduce time-to-insight; and there's no longer a need for specialists to code (and recode) complex queries or interpret results.

5. **Protection for existing investments**
   For data owners, a data clean room is the only means of simply and effectively leveraging your own distributed intellectual property—your machine learning models, containerized microservices, and other proprietary code—in the context of data collaboration. As your data analyses grow in detail and sophistication, this capability becomes more and more important.

Habu

Financial services use cases and the six essential capabilities

Whether your financial services include real estate, consumer finance, banking, insurance, or others, a data clean room can support a number of use cases such as:

- Improved fraud detection and anti-money laundering tactics
- Deeper insights into portfolio risk for mitigation measures
- Refining insurance rate calculations for more accurate pricing
- More tailored recommendations of adjacent products for individual clients

Let's take a look at an example of confidential computing in action in fraud detection.

Habu

Data sample: Detecting fraud without exposing data

To effectively expose and mitigate financial fraud, institutions must analyze a variety of data to identify patterns, anomalies, and suspicious activities. When more data is made available for analysis, prevention becomes significantly faster and more accurate.

Data clean rooms enable partners to securely contribute valuable data without risk of exposing that data. Depending on the context or industry, that data can vary, but here are some common types of data that are often used:

- **Transaction data:** The history of purchases, withdrawals, and payments, plus transaction amounts, dates, times, and involved parties, provides the most meaningful clues.
- **User behavior data:** Sign-in times, device types, IP addresses, browsing history, and interactions with online platforms can add context to other data sources.
- **Account data:** Details about the overall account, including creation date, type, balance fluctuations, and other activity can identify potential for near-term fraud.
- **Location data:** The physical location of a transaction or event can help reveal anomalies in unexpected or distant locations.
- **Machine and sensor data:** Sensors, IoT devices, and machines can provide streaming data for insights into abnormal activities.
- **Communication data:** Email messages, chat logs, and call records can be mined for suspicious interactions or attempts at fraud.
- **Social network data:** User posts and connections within online communities can be used to detect potential collusion or fraud rings.
- **Other external data sources:** Credit bureaus, government databases, and watchlists can be integrated to enhance fraud-detection accuracy.

Six essential capabilities for financial data clean rooms

To achieve these and other goals, data clean rooms for financial services applications should offer these six basic capabilities:

1 Full interoperability with no copying of data

Clean rooms need to be able to orchestrate workflows by securely connecting to data, models, and code across a plethora of cloud platforms—wherever data lives and without copying any of that data. Beyond being a frequent requirement, this reduces latency and the risk of potential data leakage.

2 Security for data at rest, in transit, and in use

Strong encryption methods such as AES and RSA have long protected data at rest. Transport Layer Security (TLS), which is the protocol used across the Internet, protects data in transit. And now, by leveraging new confidential computing technologies, your organization can unlock the value of your most-sensitive data while in use. With confidential computing, data can be stored in any data warehouse. Compute happens inside a Trusted Execution Environment (TEE) backed by Azure confidential computing running on AMD EPYC processors with Infinity Guard featuring SEV-SNP technology to protect data. Because the data is only decrypted within that enclave on that CPU, the attack surface is greatly reduced.

3 Future-proof for data science

To be ready for future use cases such as machine-learning tasks or training a model on a combination of first-, second-, and third-party data, a clean room should provide a secure environment to run any containerized code—SQL, Python, R, Spark, and other data science tools and libraries—without your partners being able to access your proprietary model or see your underlying data.

Habu

4 Enterprise-grade privacy and governance

Full control over how your data, models, and code are used is essential for everyone. Each partner should have a positive opt-in to approve the use case and data assigned to a query. In addition, you should look for all the enterprise privacy features you need to confirm privacy preservation while sharing data and to adjust to evolving regulations.

5 User experience for any use case

As your organization's data needs grow, your clean room will need to accommodate different types of users and their unique needs without significant overhead or complexity. Ensure you can accelerate time-to-insight with a clean room that offers prewritten analytics for common business use cases, APIs, and multiple coding languages to develop advanced analytics and integrations.

6 Multi-party collaboration

Look for a clean room that supports templatized analytics to allow you to work efficiently with data and service-provider partners across multiple clean rooms. This time savings becomes even more powerful with a natural language framework for analytics, which makes it easy to quickly find queries and customize them as needed.

How data clean rooms work

Let's dig into the nuts and bolts of data clean rooms: how they're built, how they connect to and access data, and how they enforce privacy and security.

There are many kinds of niche data clean rooms—some for deriving insights, for training and inference on machine-learning tasks, for enrichment of data, and for creating lists and pursuing activations. However, we're going to focus on the modern, hybrid data clean room, designed for use with multiple platforms, that empowers data collaborators to perform all of these use cases and with any partner—which is particularly valuable in the fast-evolving data-collaboration space.

To create a data clean room, you need four things:

1 Data connections

As the clean room owner, you'll first need to establish connections to the cloud accounts of all data partners. These are simply pointers to where the data, models, or code reside, so nothing needs to be moved or copied. A modern data clean room should enable you to connect to wherever data lives without requiring a fixed schema—so your connections are not limited by data type.

2 Datasets

Datasets are contributed by the clean room owner and collaboration partners (depending on the use case) and are only accessed by the clean room at runtime. This way, data can live in any cloud data warehouse, and only the rows and columns specified in the clean room use-case policies are read into the trusted execution environment enabled by Azure confidential computing on AMD EPYC processors with SEV-SNP technology. And, with a Habu data clean room, processed data is immediately deleted upon completion of the associated data join or query.

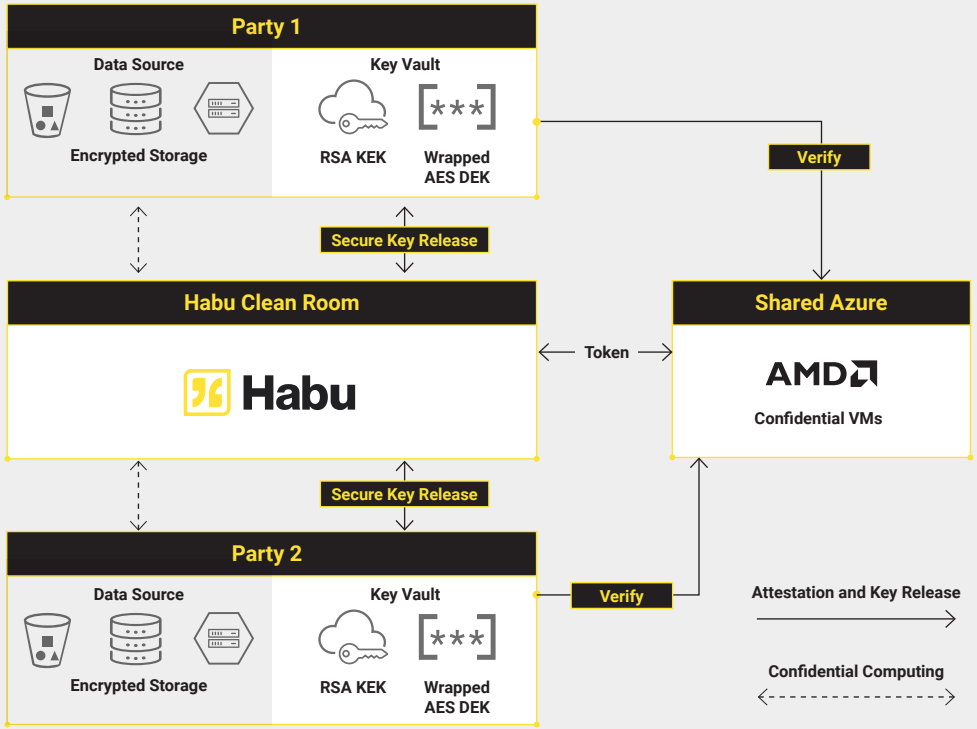
Habu

3 Use cases

For each data collaboration use case defined in the clean room, the clean room owner and all partners must explicitly approve the queries to be run, and each data contributor must also opt in to allowing their data to be used as specified. At runtime, only approved use cases can execute, only approved data can be accessed, and no party can see the underlying assets (data, models, and code). Note that the best clean room vendors—like Habu—offer extreme use-case flexibility. You should be free to run virtually any machine-learning use case with privacy and governance as long as it's technically possible to do with the data. This broad clean room capability with machine-learning use cases in particular is a significant advantage for data-science teams.

4 Roles and permissions

From managing datasets and viewing reports to building and proposing questions and scheduling runs, the clean room owner decides the scope of each partner's available actions during onboarding, and these permissions are easily tracked and updated.



Habu

Bringing the data clean room to life

Solving for interoperability

Privacy-compliant interoperability is a crucial data clean room capability. It enables complete freedom in your choice of collaboration partners.

Many other types of data platforms offer the capability to freely share data across clouds. But that data sharing often does not include privacy controls. A data clean room leverages platform data sharing protocols but adds a crucial layer of policy-driven privacy control that protects against data leakage as you collaborate with diverse partners.

In a data clean room, collaboration is enhanced but neither the data nor the model are shared

Ideally, with a modern data clean room, any number of collaborators—working with any kind of data, code, or models on any number of environments—can come together to share data. Everything is interoperable and privacy preserving. In a world in which everything is also multicloud, that's a crucial capability.

Delivering on privacy

To see how clean rooms enable privacy-preserving data collaboration, we have to dive deep into the analytics process. To begin, clean room owners and collaborators agree to an analysis, assign data, and run one or more queries. For each query, the clean room may offer different privacy controls that the clean room owner can specify while exercising governance over where and how data and code may be exported from the clean room.

Only once a query executes does the clean room solution open connections to each partner's data, and only to the necessary data elements to run the specified analysis. Within the clean room, a policy is created that says the current data connections can only be used for this specific query, with this specific data, at this specific moment. The differential privacy capabilities of a Habu data clean room, plus the security capabilities of Azure confidential computing and AMD EPYC processors with SEV-SNP technology, ensure data is secure throughout the process.

Habu

Note that the data accessed for processing purposes is not shared; instead, it's read into a trusted execution environment, which none of the collaborating parties has access to. After the report is run, the processing environment and associated data is deleted, and the output is sent to an agreed location from which users can view results via embedded analytics in the clean room software or by piping the data to their own BI tools, such as Microsoft Power BI.

Dramatically streamlining analysis with automation

When a query requires a specific set of inputs—data, models, and/or code—without clean room automation, your technical teams would have to manually implement all the necessary primitives to prepare for analysis. Often they need to do that over dozens, hundreds, or thousands of queries. End-to-end automation dynamically implements all the necessary primitives, slashing the time required to prepare and execute queries. That's a huge time-saver.

Caveat: Data collaboration still requires ownership and trust

A data clean room is a powerful tool in the era of decentralized data, but it's important to mention two ways clean rooms assume a level of ownership and trust from their users.

First, keep in mind that most data clean rooms don't provide privacy-enhancing technology (PET). The best data clean rooms—such as Habu—provide access to a broad suite of PETs that data collaborators deploy in various combinations to ensure that each use case is completely compliant from a privacy and governance perspective. Among those PETs are:

- Encryption at rest, in transit, and in use
- Obfuscation
- Data minimization
- Differential privacy
- Injection of random noise

Habu

The Habu data clean room protects your most-sensitive data at rest, in transit, and while in use in memory. This can only be achieved through the unique partnership between Habu, Azure confidential computing, and AMD EPYC CPUs with Infinity Guard featuring SEV-SNP technology.

So, although a clean-room platform cannot ensure you are privacy compliant, the power of Habu, Microsoft, and AMD protects your organization with strong privacy and security tools, making it the most-advanced data collaboration software out there.

Second, data collaboration via a data clean room assumes a base level of trust between collaboration partners. Clean rooms enable a great deal of privacy: partners cannot access the data, code, or models of their counterparties. But all participants in the clean room must still trust that their partners' representations about the efficacy of code, or the quality of data, or the outputs of a model, are what they say they are. The clean room itself cannot vouch for these things.

Habu

Bringing it all together

Habu has partnered with Microsoft and AMD to create the clean room, the cloud environment, and the technology to realize the promise of data clean rooms.

The Habu data clean room platform unlocks collaboration intelligence to maximize the value of data without the risk of sharing sensitive information. It enables teams across your organization to deliver more insights with greater security than ever, protecting data in use with Azure confidential computing utilizing AMD EPYC processors with Infinity Guard featuring SEV-SNP technology.

- **Unlock more data value at scale**
  Seamlessly enable improved insights and decision-making with data collaboration that's smart, scalable, simple, and secure.
- **Collaborate on decentralized data from any platform**
  Collaborate with data from any platform without copying data anywhere. Habu data clean rooms are 100% interoperable with any solution, so there's no need to develop customized solutions or relearn anything to get the full benefit of any data source.
- **Cut time-to-value with flexible, secure data democratization**
  When you can bring more brainpower to every challenge, you accelerate time-to-value with a highly intuitive no-code/low-code user experience that democratizes access to collaborative data for everyone in your organization and for any use case.
- **Protect your most-sensitive data while at rest, in transit, and in use**
  Increase data confidentiality and privacy while collaborating with encryption in memory powered by Azure confidential computing and featuring AMD EPYC processors with Infinity Guard featuring SEV-SNP technology.

Habu

The data is right there. Just reach out.

No matter where your data resides, how decentralized, how seemingly inaccessible, your next step is easy. Find out how the Habu data clean room platform increases data value through simple, accessible, and highly secure collaboration at any scale.

The first step to securely democratizing data access and insight across your organization for reduced time-to-value and improved decision-making is to explore what the innovations of Habu and the security and power of Azure confidential computing on AMD EPYC processors with Infinity Guard featuring SEV-SNP technology can do for you.

STEP INTO DATA CLEAN ROOMS

Habu

Habu provides the world's most advanced data collaboration software to unlock decentralized data in a smart, safe, scalable, and simple way. The Habu data clean room platform helps you harness the power of your most-sensitive data to improve your business outcomes with seamless, secure access to actionable insights.

SEE A DEMO >

Microsoft Azure

The Azure cloud platform includes more than 200 products and cloud services designed to help you bring new solutions to life—to solve today's challenges and create the future. Build, run, and manage applications across multiple clouds, on-premises, and at the edge with the tools and frameworks of your choice.

START FREE >

AMD

AMD pushes the limits of innovation to solve the world's most challenging problems. As the high performance and adaptive computing leader, AMD powers the products and services that help tackle the world's most important challenges. Our technologies advance the future of the datacenter, embedded, gaming, and PC markets.

LEARN MORE >