

The True Cost Of Building Your Own Data Collaboration Offering

Introduction

In launching a data collaboration initiative, tech-savvy organizations often assume they can simply build their own data clean rooms. But modern clean rooms hide sophisticated technology that can frustrate even experienced teams. We look at the advantages of investing in – and building on top of – a modern data clean room such as Habu.

When launching a data collaboration initiative, an early decision enterprises must contend with is whether to invest in a modern data clean room solution or try to build a solution of their own – the DIY route.

Some organizations look at their cloud data warehouse provider and conclude that building a data collaboration solution on top of their provider's cloud tools could reasonably satisfy their needs. But, if we unpack all the considerations around successful data collaboration, two things become quite clear: the relative cost and complexity of a secure, interoperable, and multi-party data clean room solution capable of orchestrating your workflows – and the strategic advantage to be gained by leveraging the technology of an existing data clean room vendor like Habu.

This whitepaper will explore:

- Why your cloud platform primitives are not sufficient for scalable data collaboration
- How your partners' technical capabilities limit the speed and success of collaboration
- Key elements that result in a build process that is time consuming and very complex
- How orchestration is critical in data collaboration – and difficult to develop
- Why DIY economic costs are hard to justify vs a best in-class data clean room solution
- A case study of how a major media company struggled with the DIY route
- The key capabilities to look for in a modern data clean room solution

1. Your Cloud Platform Primitives Are Not Sufficient For Scalable Data Collaboration

The objective of data collaboration via data clean rooms is to securely leverage your first-party data to connect to large amounts of valuable, second-party data that you would not otherwise have access to, and gain insights from that data to drive the performance of your campaigns and ultimately your bottom line. Some data owners and aspiring data collaborators with experienced technical teams may be tempted to try to build their own solutions with the data clean room features of the cloud data platform they already use.

But, building your own solution on a single cloud platform in this way raises three immediate challenges:

- 1. Your partnerships may be limited.** Obtaining the variety of data you need typically requires multiple data collaboration partners, who likely work across different cloud platforms (e.g., AWS, Google Cloud, Snowflake). Native cloud data sharing tools only operate on their own cloud, and there's no way you'll be able to convince collaboration partners to copy their valuable data to your cloud in order to operate on their data. So, unless you're willing to arbitrarily limit your potential partnerships by locking yourself into a single platform, you will need an interoperable solution capable of connecting directly to all the major cloud providers in order to access data where it lives.
- 2. The technical demands are steep.** The clean room offerings of the major cloud platforms are highly technical environments designed for IT staff and data engineers. They are not the plug-and-play services that enable your business teams to accelerate their data analytics. Building on a cloud platform will require engineering resources capable of using a basic set of building blocks ("primitives") to construct all the elements of your data collaboration solution. These technical demands apply both to a data owner and each partner, so as your collaborations grow, your technical demands also increase.
- 3. You need data collaboration now.** Leveraging data for competitive advantage isn't something that can wait. But building a data clean room takes time, and, as we'll see, can significantly delay your data collaboration agenda.

These challenges pertain just to you. Once you begin to think about requirements beyond your own organization, the challenges multiply.

```
var self = getObject(data, id),
    parents = [];
if(self){
    parents.push(self);
    if(self.pId){//If pId is not 0.
        parents = parents.concat
            (getSelfAndParents(data,
            self.pId));
    }
}
return parents;

function getSelfAndParentsId(data, id) {
    self = getObject(data, id),
    ids = [];
    if{//In the presence of self
        ids.push(self
```



2. You're Only As Capable As Your Partners

A data clean room is by design a multi-party solution. That means that even if you are incredibly confident in your team and your technical resources, your partners will also need to have significant technical resources. Unless you are building for them, they must be able to align their infrastructure, Infosec requirements, and workflows with yours. And, during data collaboration, they must also be able to keep up by taking appropriate steps on their side, in sync and at speed.

So, even if you build a tool that enables you to execute your own tasks seamlessly, your data collaboration will fall down unless that is true for every one of your partners as well. The chances of your partners meeting that requirement are quite low, which means that if you're going the DIY route, in order to ensure the success of your data collaborations you will need to build not just your own internal capability, but also develop self-service workflow tools for your partners.

In other words, if you're not covering your partners with an interoperable, orchestrated data clean room, you're limiting yourself to whatever their technical ability is. And if you are building to cover them, your workload just grew by 2x, 3x, 4x, or more.

3. The Build Is Time-Consuming And Complex

Let's look more closely at the challenges of building data clean room infrastructure. A modern data clean room includes a good deal of complex code under the hood, and brands must consider whether it's worth building all of that themselves, or simply investing in a field-proven data clean room solution where the vendor has experienced staff and has already done that work and much more.

Not all of your partners will have compatible infrastructure for data collaboration

As you move into data collaboration, your partners may show up with a technology stack that's compatible with yours, or they may not. If they don't, you'll need to write additional code in order to work with them, and, since they're not on your chosen cloud, there will be a technical obstacle as your own team comes up to speed on their cloud. If you're using Habu data clean room software though, Habu includes managed infrastructure along with pre-built connectors to all of the major cloud platforms (AWS, Google Cloud, Microsoft Azure, Snowflake, and Databricks) so your partner becomes instantly compatible and interoperable. Leveraging Habu also gives you access to integrations with the native data clean rooms of all the major walled gardens (Amazon Marketing Cloud, Google Ads Data Hub, and Facebook Advanced Analytics) and cloud platforms. That means your sphere of partners increases immediately, because you don't have to limit yourself to compatible infrastructure for scalable data collaboration.



You'll have to develop additional APIs

If you decide to build on top of a major cloud platform (e.g., leveraging Databricks' Delta Sharing), you'll be wrapping platform services with APIs and writing code to address data on the platform. However:

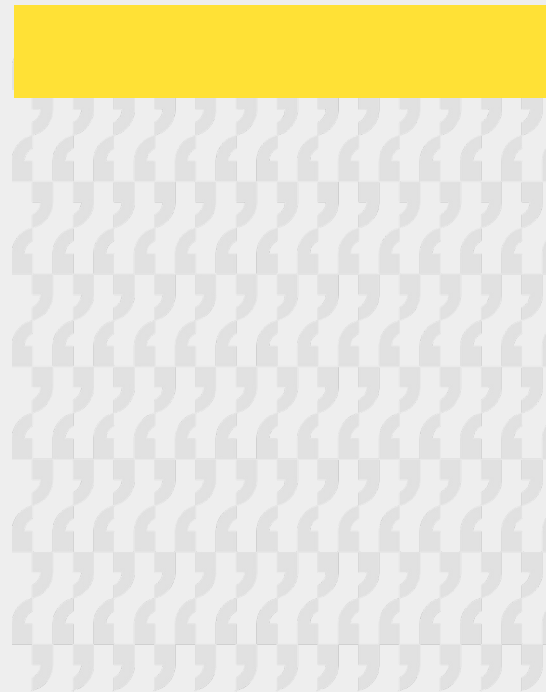
- You probably won't go to the additional effort of extending those APIs to other cloud platforms. But if you don't, you're locked into that particular cloud platform.
- The native data clean room services on the major cloud platforms are nascent, and updated about every six months. Writing your own APIs also means supporting them through constant system updates and maintenance. Plus, the cloud providers' API support itself may be limited as their offerings mature.

Habu has already built all the APIs you'll need, and maintains them through any and all platform iterations. By using Habu, you get the broadest interoperability — and maximum addressable market. However, Habu is a fully-featured data clean room software platform and beyond full interoperability, it includes all the necessary components for modern data collaboration.

Security is essential, but complex to build

Data clean rooms exist to provide a secure environment for data collaboration. If you're even considering a data clean room, it's partly because you must have those security benefits. But architecting all of the many redundant steps that provide data security — checks, validations, handshakes, attestations — is complex, and more so across multiple parties. If you go DIY and don't implement security protocols correctly, you jeopardize the entire security profile of your solution. And, you risk exposing your data or the partners' data to other parties or malicious actors. If that happened, at best they won't be your partner any longer, and, at worst, you could cause significant financial harm to your business due to lack of consumer trust.

Habu provides seamless orchestration of security for you and your data partners. This is even more critical when your data collaboration includes advanced workloads like ML model training and confidential computing. Habu has passed detailed InfoSec scrutiny by some of the most risk-averse and security-conscious businesses around the world, and is constantly innovating to make sure that the latest security protocols and privacy-enhancing technologies are implemented.



4. Orchestration Is Critical – And Difficult



Because data collaboration unavoidably involves multiple discrete parties, smooth orchestration of the data collaboration process is essential to scale. Executing a single query with a partner isn't very hard; you and your partners could run a few queries asynchronously and coordinate over email. But proceeding manually like this is cumbersome and doesn't scale.

If you want to integrate your data clean room into your business processes and really begin to drive value, you need infrastructure that is always-on and automated, and you need a means of synchronizing your data collaboration activities with your partners. If you don't have this kind of orchestration, you're left with slow, manual processes, and/or you're writing code to orchestrate workflow among participants – a task that goes far beyond configuring a database and writing SQL code.

With a modern data clean room such as Habu on the other hand, you can focus on the insights and not the infrastructure. Habu seamlessly handles orchestration of your multi-party data collaborations. With a common UI, you have complete flexibility in establishing data collaborations, generating reports and other analytics, and digging into cross-clean room insights. With Habu, you can quickly move to running your analytics and leveraging new insights without ever needing to manually code for data connections and orchestration.

5. DIY Doesn't Make Sense From An Economic Cost Perspective

As mentioned, companies pursuing the DIY route to data collaboration will need to staff their initiative with capable engineers, including experts in each of the major clouds, security experts, analytic teams to build queries, and more.

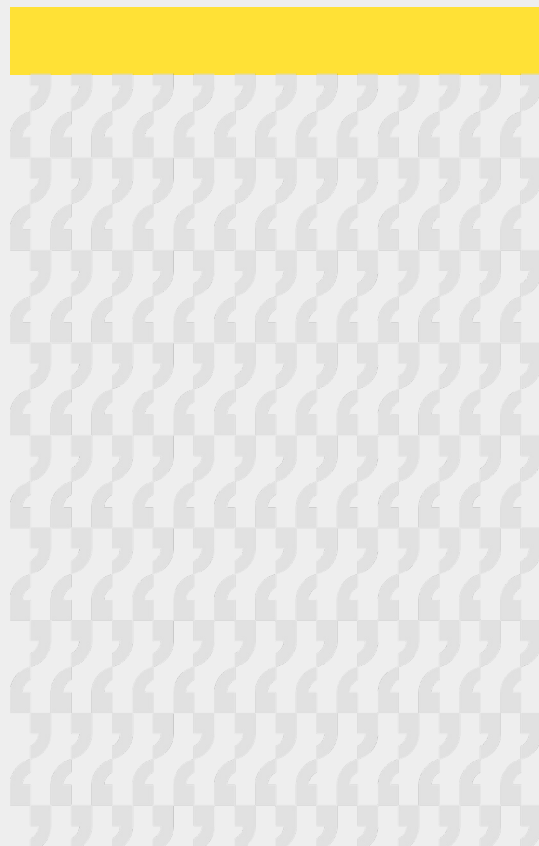
When you compare the annual recurring cost of such a team to the licensing cost for a modern data clean room solution, the conclusion is obvious: just because you have the technical know-how to build doesn't mean that you should – especially given the other high-value projects that your engineering resources could be working on instead.

6. Even A Major Media Company Struggled With The DIY Route

If all of the preceding doesn't convince you of the benefits of investing in a data clean room, consider the experience of a major media company with a huge trove of data and a full complement of data engineers and data scientists. They worked for a year to build a functional data clean room on top of a leading data cloud – and eventually abandoned the effort in favor of investing in Habu.

This media company knew they had a valuable asset in their data, and they wanted a data clean room solution that would both put them on a par with the major walled gardens and deliver far more value to their advertising partners. They particularly wanted a software layer that would simplify and accelerate the deployment of clean rooms for their partners. At first they tried to work within the native clean room environment of their cloud platform, but the complexities of interoperability and multi-party orchestration resulted in slow progress.

The company eventually turned to Habu for help. With full interoperability and high levels of automation, Habu data clean room software was able to jumpstart the company's data collaboration efforts and accelerate the market entry of their data-as-a-service offering by 12-18 months. Instead of painstakingly onboarding partners one by one, on multiple data clouds, the company now brings on new partners in hours rather than weeks, and it enjoys full orchestration among data partnerships.



7. Key Capabilities Of Modern Clean Rooms

As you've seen, going the DIY route is fraught with both technical challenges and questionable economics. But, if you're sure you want to give it a try, here's a list of design principles you'll need to follow:

- ✓ **Interoperability:** Connect to all of the major cloud data warehouse platforms and walled garden clean rooms that you and your partners require.
- ✓ **Data Minimization:** Avoid unnecessary movement and/or duplication of data. Utilize only the minimum required dataset to complete the analysis.
- ✓ **Purpose Specification:** Guarantee that only compliant and approved analysis can run, either through explicit workflow approvals or technical usage constraints.
- ✓ **End-to-End Security:** Protect data throughout its entire lifecycle of usage, leveraging approaches such as encryption and hardware isolation.
- ✓ **Data Privacy:** Ensure that no personal information is exposed with techniques appropriate to the use case such as differential privacy, k-anonymization, and various methods of de-identification.
- ✓ **Data Governance:** Enforce access and usage controls, with granular, role-based permissions.
- ✓ **Transparency:** Maintain a clear audit trail that logs all actions and can be used to validate authorized usage.
- ✓ **Multiparty Computation:** Adapt to support a broad spectrum of workload types, including data and AI/ML models.
- ✓ **Democratization:** Empower technical users with a low-code solution and business users with no-code UX to access insights and outcomes.
- ✓ **Generative AI:** Simplify and scale data clean room initiatives with enhanced productivity and surface new types of analysis for better business results.

Each of these areas will contain a substantial list of core capabilities that need to be built, tested, and maintained for a minimum viable product (MVP). The list of capabilities to approach a market-leading data collaboration solution such as Habu (with vast engineering resources innovating for several years) will be significantly longer.

8. The Path To Data Clean Room Success

At Habu, we've found the most successful data collaboration initiatives come from companies who don't build a data clean room – they skip ahead to building products and services on top of their data clean room. These companies understand that they can move faster toward their data collaboration goals by leveraging Habu's APIs, full interoperability, orchestration capabilities, and built-in security instead of trying to construct them from the primitives of a cloud platform. Take advantage of Habu's experts and APIs to accelerate your data collaboration on top of Habu, instead of on your cloud platform.



Habu enables organizations to master decentralized data, uncovering insights that fuel better business decisions via data collaboration that's smart, safe, scalable, and simple. With flexible, multi-cloud deployment, high levels of automation, and an intuitive interface, a Habu Data Clean Room empowers data teams. Easily connect your data from any source and start accelerating growth today.

[HABU.COM](https://habu.com)