HEVO

Data Pipelines

The Dilemma of
Build vs Buy

Table of Contents

Build Vs Buy - The Hard Call

Building a great data-driven business involves more than just investing in the right reporting tools. Leading business understand that to be able to make the most of existing data (mostly present in different systems in various formats), they will also have to smartly invest in the right infrastructure to clean, organize, and make the data available in real-time for reporting.

Naturally, building a robust data pipeline to extract, transform, and load the data into a central storage (mostly a data warehouse) is the first inevitable step to building a great data infrastructure. One of the greatest dilemma in the minds of the decision makers under these circumstances is:

Should we build a custom ETL/Data Pipeline solution in-house or simple buy a Data Pipeline Platform?

We have tried to decode this very problem in this ebook!

There can broadly 6 factors in consideration to decide whether to go with a ready-to-use Data Pipeline Platform such as Hevo or build your own.

Let us analyze these factors:

1 Will it be cost effective to build a solution in-house?

Well, there is no free lunch. Whether you build or buy, some costs are apparent while others are not.

A. Engineering Team Bandwidth

Engineering bandwidth to build and maintain the product would direct costs that you can take into account. You would additionally have to invest in the technology infrastructure needed to build the solution. You might also want to consider the opportunity cost at which you are trying to build a custom data integration solution for your team. This likely means that they are not working on other projects that could have a larger and direct impact on your business.

B. Pricing/Cost Consideration

Often, the cost of using a ready solution to build pipelines is only a small fraction of the cost involved in building and maintaining the Data Pipeline infrastructure in-house. Think of this example similar to using Gmail vs hosting and managing mail server in-house for communication. Time to deployment is also lesser compared to building a custom solution in house.

2 Will an external Data Pipeline solution meet all my business requirements?

This is a possibility. A Data Pipeline solution may not be able to provide you an exact use case match. The solutions you are considering may solve parts of your Data Integration problems and may not be able to handle all your use cases. This could be a show stopper.

However, that is not a common case. Usually, the requirement for bringing data from multiple sources such as MySQL, MongoDB, HubSpot, Clevertap, Mixpanel, Google Ads, Facebook Ads, etc is similar across most companies. More often than not, the solutions are designed to specifically solve this problem and have all the use cases covered.

In fact, often third party solutions have much deeper capabilities that you would only start realizing as you move a step closer to using the product for various use cases. Hence, you definitely might want to consider a platform like Hevo before shunning the option away.

3 Will the third party solution scale as my volume grows?

As per your current requirement, you might need to build connectors to just few OLTP systems like MySQL or PostgreSQL. Building an in-house solution would look optimal if your needs do not change very often. But, seldom is this the case.

Slowly, marketing teams that use various tools like MailChimp for emails, Google Analytics for attribution, Facebook for advertising, etc. will want their data integrated. Requests from other business teams will follow. Building connectors for all these systems would need you to put in an ongoing effort on the solution. Additionally, the source schema/API would keep changing from time to time and will regularly have to be updated.

A third party data integration platform like Hevo, will have an ever-growing wide coverage of various sources and destinations. New features keep getting added and you can always request for the addition of new sources as per your need.

To your surprise, many businesses that are 10-100x your scale might already be using a solution like this to power their data infrastructure. This ensures that they won't have to deal with scaling challenges as they grow.

4 System Performance

When dealing with data, it is an absolute necessity to build a foolproof system. Data inconsistency can become an everyday phenomenon if all exceptions are not handled.

In a homegrown system, you would need to invest heavily in engineering bandwidth and DevOps to build such a system. You would additionally have to build instrumentation and monitoring systems to ensure all errors are caught. It would be comparatively easy to handle errors in a home built system as someone from engineering team would know and understand the system's whereabouts and can fix errors fast.

Generally, solutions like Hevo are designed grounds-up to handle all kinds of exceptions that you deal with while working with various data sources.They ensure zero data loss and real-time availability of data, no matter what the scale is.

Tools like Hevo are slightly more robust than homegrown solutions as they additionally implement all kinds of instrumentation, monitoring, and alerting into the system. Even for the days when errors go unhandled, star customer success teams are deployed to be available at the beck and call of the user to sort out all the issues.

5 Dependencies

More often than not, teams change, individuals move (to other projects or out of the organization) and projects taken up outside the scope of direct business metrics take a toll. Given a small team of engineers would be deployed to build a custom data integration solution, it is important for homegrown systems to eliminate all dependencies through an elaborate documentation and proper knowledge transfer. In absence of this, the small team that is maintaining this platform becomes overwhelmed and eventually becomes the bottleneck.

A self-service platform such as Hevo allows various teams to create and configure their own data pipelines specific to their needs. All the necessary documentation is made available at the user's fingertips. This would eliminate major dependencies on engineering. Additionally, there is a dedicated team at Hevo that is responsible for building new functionalities and enhancing the performance of the platform.

6 | Security Concerns

You can argue that building a solution in-house gives complete control and visibility on your data. We understand that this is important. Platforms like Hevo allow you get the best of both worlds. A fully managed, completely secure solution that can be setup and run in your virtual private cloud behind your own firewall. This addresses your data security concerns while providing the same powerful data integration features.

Conclusion

When it comes to Data Pipelines, building, and buying have their own share of pros and cons. The answer to this question depends completely on the unique needs and requirements of your growing business. In a nutshell, if your data integration requirements are very unique, building might be your only option. However, be sure to set clear expectations before the process begins, and design it to handle scale.

On the other hand, a fully managed third party solution like Hevo can constantly serve you alongside your evolving business needs. Hevo integrates with a variety of data sources ranging from SQL, NoSQL, SaaS, File Storage Base, Webhooks, etc. with the click of a button. With a wide array of features Hevo can help in your journey as you leverage data to build a better business. Hevo provides live support on slack and email ensuring that the team is available at your beck and call 24*7.

Check out Hevo Data Pipelines here.