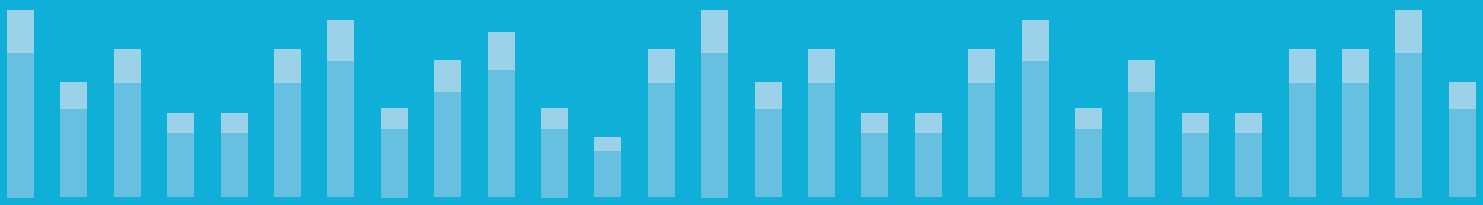
WHITEPAPER

Build vs. Buy Gets Easier with APIs

A CTO's Guide to Getting Data Strategy Right



Keen IO

Contents

Why Now: The Urgency to Build Intelligent Apps

"During the next 10 years, virtually every application and service will incorporate some level of AI in much the same way as consumer appliances have incorporated microprocessors. Some of these apps will be obvious intelligent apps that could not exist without AI and machine learning. Others will be unobtrusive users of AI and machine learning that provide intelligence behind the scenes."[1]

Gartner, October 2016

After years of big data collection and processing sitting squarely in the domain of giants like Google and Facebook, a new crop of technologies on the market today let any size organization use data to compete. The infrastructure and tools companies once found prohibitively expensive and difficult to manage are now more affordable to build or buy as services.

This shift has lead to far more companies using intelligent analytics in their applications, using data to transform user experiences, predict behavior, and even anticipate customer needs before they arise. These capabilities are propelling them ahead of their counterparts—Mckinsey reports a 126% profit improvement for those using making extensive use of customers analytics over those who don't.[2] Other research attributes 50% higher growth rates to the use of big data in organizations.[3]

Others are following suit. Over half of senior executives and decision makers planning to increase investments in big data over the next three years in response to stiff competition from data-enabled startups.[4] Analysts like Gartner urge decision makers to emulate the tech-driven leaders, moving away from IT-centric analytics deployments in favor of more democratized, free-form exploration of data. As Gartner points out, "Every business is an analytics business, every business process is an analytics process and every person is an analytics user."[5]

In spite of this trend, many companies still struggle to find the value in their data. They may have limited resources, making it hard to recruit top engineering talent or build the infrastructure to store and query large datasets. They also may have trouble navigating a market flush with analytics tools that aren't quite right for their use case. However, the need for better analytics is clear—only one-third of decision makers from organizations around the world say they trust the analytics they generate from their business operations.[6]

If you're ready to make intelligent analytics a priority, your biggest decision is the first one you'll make: "Do I build, or do I buy?". Your options are three: Start from scratch and build it yourself, buy an analytics product off the shelf, or use an API platform like Keen to get the benefits of both.

Option 1: Buying an Off-the-Shelf Solution

Engineers inherently like to build. Robust analytics isn't easy, and when it's a core component of your product, your engineers may not feel comfortable relegating it to a third-party vendor. But there are instances when an off-the-shelf solution works just fine.

You need analytics fast, and it doesn't have to be perfect

When your biggest goal is speed, buying a solution like Mixpanel or Looker can get you close enough to where you want to be. Your engineering team will be free to do other work, you'll get basic reporting and some integrations, and you may even be able to build custom integrations or extensions down the line.

Analytics is a nice-to-have, not a core function of your business.

If your company is only using analytics to boost efficiency, you can probably get away with outsourcing it. Your engineering team may want to focus on their core product and not worry about maintaining a custom analytics tool. However, if analytics is crucial to the success of your business and a major influence on customer satisfaction, you'll need more control than an out-of-the-box platform can offer.

You can't afford to build it yourself.

Depending on your in-house capabilities and budget, building an analytics product from scratch might be out of the question. While off-the-shelf options may seem enticing, be sure to do the math to consider the opportunity cost of buying something that only fits 70-80% of your needs. If you think a more customized solution might open up new sources of revenue or lower operating costs over time, you find your answer in Option #3.

Just like big companies trust PayPal or Stripe for payments or AWS for hosting, many organizations will outsource their analytics to save money and time. Forrester analysts encourage CIOs to do this, based on a survey result that "only 29% of firms are good at connecting analytics results to business outcomes".[8] But an abundance of choice and many unknowns makes it hard to choose from the slew of new tools on the market, and companies are left to consider what they can get by doing some of the work in-house: more control over future iterations, fewer disparate systems, and greater customization for their teams and customers.

Pros and Cons of Buying Off the Shelf

Pros:

» Predictability in terms of Total Cost of Ownership (TCO)

» No burden on engineering resources

» Standard data analytics to quickly answer a set of predetermined questions

» Fast implementation

Cons:

» Overwhelming choices with limited functionality

» Limited extensibility and integration with other solutions

» Opinionated platforms that limit the questions you can answer with your data

» Vendor lock-in

Option 2: Building Custom Analytics In House

Building a custom solution is a big undertaking, so it's important to estimate the true cost before deciding on this option. If you have unique analytics needs that can only be solved with a in-house solution, here are some things to consider:

Do you have the engineering resources to commit to it?

The total cost of building a solution largely comes from support, maintenance, and improvements. Try to anticipate feature requests, customer requirements, and how growth will affect your analytics needs for an accurate estimate of cost over the next several years. If you have to add headcount to get it done, factor that in as well—good data scientists and engineers are some of the most in-demand and highest-paid people on the market.[9]

How easy is it to build?

Complicated analytics capabilities can take months to ship. Decide if it's worth waiting or if getting to market faster worth losing some customization. You should also consider the importance of other projects—are you willing to take engineers off of other core features to work on analytics?

Will your analytics be customer-facing?

Some companies need analytics to make internal decisions, while others use them to surface important data to customers. If analytics is a core part of your product and will add value for your customers, a custom product may be better than build. However, there's still a third option to consider, which we'll cover later.

One of the biggest risks in building your own analytics platform is that you may not end up with a successful product. The scope might be too big, updated architecture might call for refactoring, or changing priorities put a the whole operation on the back burner. If you plan to build from scratch, do it with extensive knowledge of your roadmap, your customers, and your engineers' ability to deliver.

A final consideration with a custom build is scale. Storage and traffic requirements are extremely difficult to predict in a growing company, so many internal teams see downtime and performance issues when a glitch, traffic spike, or even a DDoS attack throws a wrench in things.

Read more about Keen's own experience of creating an extensible data platform in this blog post.[12]

According to Gartner, more than half of all analytics projects fail because they can't deliver the features and benefits that are optimistically agreed on at their outset, or because they aren't completed within budget or on schedule.[10]

Pros:

» Control over the infrastructure

» Flexibility in querying and displaying data

» Full ownership of data analytics: important if it's core to the business and may be valuable for IP

» Great for the team's resumes and engineering experience

Cons:

» Big risks of making fatal mistakes or failing to complete the project

» Expensive and slow: lots of time and resources required

» The hardest talent to hire and keep is permanently required for development and maintenance

» TCO is difficult to estimate and scalability is a huge challenge

Calculating Total Cost of Ownership: Can Your Bottom Line Take a Custom Build?

1. Cost of storage & infrastructure	Per month:	Depending on the project scale
2. Cost of people who build/manage the service	± $50-60K per month	Minimum in the team: 2 data engineers (± $300-500K per year for 2 salaries with insurance and other costs) + project manager (add salary) + CTO/supervisor hours. *+ Don't underestimate morale/burnout factor of being on-call and fixing problems in the middle of the night!*
3. Cost of documenting & training people to use the platform	$5000 (one-time cost)	If you're building your own data stack, you'll need to document how everyone in your company will interface with their data: from how they securely connect to how they query, extract, and build reports on that data. At the very least, you'll need to carefully document your solution, so that new engineers joining the team will be able to learn your stack.
4. Maintenance costs	Technical debt and stack creep. Equivalent to #1 + #2.	Consider the impact to your technical debt when you add another 5-10 data queuing, storing, and caching technologies to your stack.
5. Opportunity cost	Fatigue and other roadmap items that didn't get built. Other initiatives commentary that didn't get the needed support.	Make a list of potential missed opportunities and resulting costs: » What's the opportunity cost of spending engineering time on data architecture vs. other key initiatives? » What is the impact to your on-call team when v1 of your architecture inevitably hits unforeseen scale limits, either losing data or sending query response times to a crawl? » What other customer requests or features will be put on hold to address common data scalability challenges? » If you didn't hire two data engineers to build this data platform from scratch, who would you have hired and what would you build?

Be mindful of the time requirements: getting the data analytics infrastructure right takes a small team of 2-3 engineers from a few months to a year. Once the infrastructure grows, you might think that you can easily run five more servers in AWS. In reality, at some point you are probably going to max out your cluster and need your team to deal with it in the middle of the night while the system permanently loses (or even corrupts!) data. So, don't consider just the time and resource you need right now. **Imagine potential future requirements in 12, 24 and 36 months.**

While going over budget and over time are almost guaranteed in software, both of these issues scale alongside the project. Gartner interviewed 199 members of its Research Circle and found out that "while nearly three quarters of respondents said their organization has invested or is planning to invest in big data, a wide majority of these projects are still stuck at the pilot stage." Only 15 percent of businesses had their big data projects in production explaining that "too many pilots and experiments are being built with ad-hoc technologies and infrastructure that are not created with production-level reliability in mind."[11]

Option 3: Building a Data Platform with Keen

"Customer-facing metrics help SaaS companies retain customers by demonstrating value, attract new customers with data-rich UX, and encourage desirable usage patterns. Many of our customers have been with us for over half a decade. We've helped them grow, and we've scaled with them. What a cool experience to see your customers succeed and play a role in their success. All in all, our customers trust Keen to scan over one trillion rows of data each day, which speaks for itself when it comes to our strength and scale."

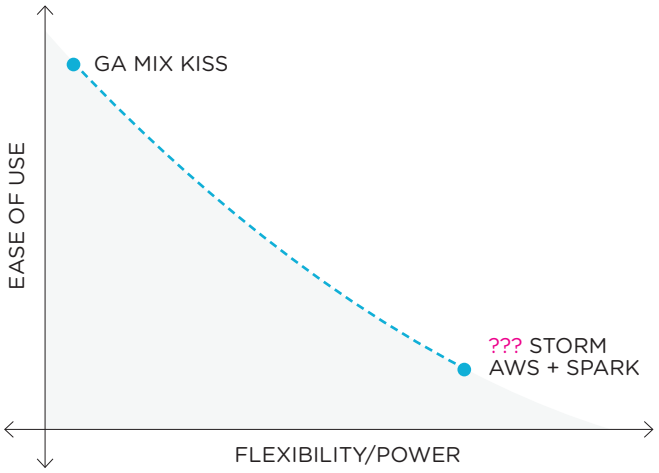
Ben Kuhn, CEO, Keen

If you need customized analytics, you'll ultimately need to build something. It only depends on how much time you have—and how much of the data platform you want to build from scratch. Previously, the only two options to choose from were:

1) Try to match your needs with off-the-shelf analytics products

2) Build an in-house platform, taking on all of the challenges that come with it

Now there's a third option: developers can use APIs to build a custom analytics solution without having to manage infrastructure and other backend components like data queuing, enrichment, storage, compute, and cache functionality.



APIs are getting more attention in recent years as a way to connect technologies and extend a product's functionality. TechCrunch notes that the rise of APIs allows developers and companies "to focus on their core capabilities, more efficiently bring applications to market and develop differentiated functionality, their 'secret sauce,' at higher velocity",[13] while API-focused companies like Apigee are getting scooped up by data giants like Google.

Companies using APIs to build custom analytics products are getting to market faster, with less risk and total cost.

Product owners and technical teams are starting to explore this third dimension in the "Build vs. Buy" debate. Those using Keen's intelligence APIs see a notable acceleration of time to market while significantly reducing delivery risk and TCO. The results are in line with the adoption of other cloud services, where most companies have decided, for example, that they are happy to pay AWS to avoid managing expensive sysadmin and hardware resources. Similarly, an external team of experts can manage the complexities of data storage, recovery, consistency, recoverability and other demanding tasks to take the load off of your team.

Quickly Build Intelligence into Apps and Workflows

With Keen's APIs, you can instantly begin collecting and programmatically querying large volumes of data on a highly available, production-ready platform. Retain the flexibility of a custom buildout while leaving the scalable backend and infrastructure to Keen.

How does it work?

For example, Keen enables Pixlee to provide completely white-labeled in-app analytics dashboards to clients like Levi's and Kimpton Hotels with real-time, beautiful representations of end user engagement and conversion. "Keen saved us months of work building real-time analytics into our product, and our clients love seeing actionable insights as a native part of the user experience," said Jeff Chen, Co-Founder and Director of Engineering.

For Mic the biggest thing was flexibility and the ability to collect data from any interaction on any media property. Anthony Sessa, VP of Product and Engineering, was really impressed with Keen's ability to power live data and well-designed dashboards that have "an amazing emotional effect" on the entire organization. "With Keen IO, there's no black box around our data: it enables us to rely upon our own view of what's important to us," said Anthony.

Plan for the Future with Scalable Architecture

Even if you've perfectly architected something that meets your requirements today, what about the requirements of tomorrow? Will your team be ready to upgrade SDKs when there's a new release of iOS? Or build new integrations when new platforms emerge? By using a service like Keen, you not only relieve yourself of ongoing upgrades and maintenance, but also reap the benefits of a 50,000+ strong developer community that maintains and creates new open source libraries as technology evolves. Keen's API platform is built for maximum extensibility and designed to be future-proof. It's built using world class frameworks that engineers already respect (more in our blog post[12]), and will continue to evolve to use the latest innovations in data. Top this solid foundation of readily available APIs with code of your own, and you can build amazing data products for your teams and your customers.

Here's what you get with Keen:

Software: A data platform with APIs that let you build intelligence into anything connected to the internet. With a production-grade uptime and enterprise SLAs.

Services: Data experts and platform engineers that are available to partner with your team to review your data strategy and recommend solutions based on their experiences serving thousands of customers. A proven onboarding methodology that ensures you will architect things correctly the first time and go-live on schedule

Savings: Lower TCO and decreased delivery risk. You don't have to recruit and hire a data engineer (or a few) to build and manage your data platform. You also don't need to train a new one if someone from your team leaves the job.

Business value: You can quickly become a more intelligent and data-informed organization while keeping full control of your data—same as when you operate your own data platform. Start building and selling new data products and features for your teams and customers. Build automation on your data and become a more efficient business. Take advantage of Keen's limitless customization and extensibility, and stream data to other systems.

Pros:

» Gain speed to market and build intelligent products and applications faster

» Cost reduction: use fewer resources and time to build your customized platform

» Limitless extensibility and low risk of failure; easy integration with existing platforms and applications

» Supported by a team of engineers that dedicate their careers to data engineering: can help you future-proof your strategy, respond to redundant server failure in minutes and lead adoption of key technologies, i.e. Apache Flink

Cons:

» Not a turnkey solution (won't work for companies that don't want to build at all)

» Not applicable if your strategy is focused on centralized analytics deployment

» Not a necessity if you have the time and resources to experiment and make mistakes

A CTO's View on the Build vs. Buy Debate

"If you are a startup like us, you want to build only the things that are at the core of your company's current market offering, add value to your IP and direct value for your customers. As a CTO, I want to build as little as possible because I have to move quickly. I simply wouldn't understand if my team would want to build their own analytics platform from scratch any more than if they wanted to build their own cloud service system from scratch. You don't want to hire staff to support the servers, queuing systems and databases, especially if you're only going to have a couple of events in your system. It's a different story if you are dealing with more than two trillion events per day. Then you might consider building your own system, because one of the things you'd need is hyperscale that's dedicated."

Mike Greer, Co-Founder and CTO of TAPP, Ex-CTO of the Onion

Checklist of questions to future-proof your choice

- ✓ **Is data analytics of core value to my business?**
  Note: If gathering and storing data, as well as running queries against that data is not a core differentiating element to what you do, then use Google Analytics or an off-the-shelf tool. If it is a competency you want to build, consider building something real on Keen.
- ✓ **Is there any urgency? If yes, which solution will help me justify ROI and get value faster?**
  Note: Consider both time to value, as well as future speed for iterations.
- ✓ **What is the TCO including all of the human resources, maintenance and infrastructure? Now, and in 12, 24 and 36 months from now.**
- ✓ **Flexibility: am I evaluating the needs of today or building a future-proof system that allows for scale?**
- ✓ **Control: what happens if the person who builds and maintains this system leaves the job?**
- ✓ **How will we collect data from all the sources that are generating information from events?**
- ✓ **What kinds of questions do we need to answer with the help of this data analytics platform?**
- ✓ **If we go for an off-the-shelf analytics tool, how opinionated is it going to be?**
- ✓ **How will I be able to give secure access to other people, so they don't always have to run through me or my team for data?**
- ✓ **Will it be easy enough to train people across the organization to generate the needed reports and ask questions they want?**
- ✓ **Can I build on top of it to deliver a high performance data product for a broad audience?**
- ✓ **How will it link to other systems or the existing platform?**
- ✓ **Who will own the operational burden—am I going to need another team running on pager?**

Talk to experts at Keen IO:

Contact sales@keen.io to speak with Keen experts about your upcoming data analytics projects and future-proof your data strategy. Leverage Keen's intelligence APIs and expert team to accelerate deployment of real-time, intelligent data applications and quickly embed intelligence throughout your business. Keen's data platform is already used by more than 3,500 customers and 50,000 developers to embed intelligence throughout their businesses.

Sources

1 Gartner, "Top 10 strategic technology trends for 2017," October 2016

2 McKinsey, "Five facts: how customer analytics boosts corporate performance," July 2014

3 Dell, "Global technology adoption index 2015," October 2015

4 Capgemini, "Big & fast data: the rise of insight-driven business," March 2015

5 Gartner, "Gartner says worldwide business intelligence and analytics market to reach $16.9 billion in 2016," February 2016

6 KPMG, "Building trust in analytics," October 2016

7 Forrester, "Forrester's 2016 Predictions: Turn Data Into Insight And Action," November 2015

8 Computer Weekly, "Forrester urges CIOs to outsource analytics to data scientists," June 2016

9 CareerCast, "10 best IT & engineering jobs 2016," October 2016

10 Gartner, "Predicts 2015: Big Data Challenges Move From Technology to the Organization

11 Gartner, "Gartner survey reveals investment in big data is up but fewer organizations plan to invest," October 2016

12 Keen IO Blog, "So you've decided to build analytics in-house," July 2015

13 TechCrunch, "The rise of APIs," May 2016